International Journal of Innovations in Engineering and Science, www.ijies.net

College Enquiry Chatbot

Shriya Kawle, Chrisbin Thomas, Joe Mathew, Jayshri Harde

Department of Computer Science and Engineering (Data Science)
St. Vincent Pallotti college of engineering and technology ,Nagpur, India

shriyakawle.21@stvincentngp.edu.in

Abstract – The primary focus on the use of chatbots in the education industry, specifically refers to it as an inquiry chatbot. It covers the advantages of implementing a chatbot system in colleges, such as providing 24/7 access to information and reducing waiting time. The technical concepts of creating a chatbot, such as natural language processing, WordNet, and GUI are also discussed. Additionally, the paper highlights the evolution of chatbot technology and its future potential in the education industry. Overall, this paper provides a valuable overview of the use of chatbots as inquiry systems in education, highlighting their potential to improve customer service and streamline information access for students.

Keywords- NLP (Natural Language Processing), Machine Learning, Deep Learning, Artificial Intelligence, LSTM (Long Short-Term Memory), NLTK (Natural Language Toolkit), WordNet, retrieval-based and rule-based chatbot

I. INTRODUCTION

A dialog system that helps users get instant responses to their queries throughout the day is referred to as artificially intelligent software. It impersonates the process of how a human would reply to a conversation. The increase in the amount of data all over the world led to the increase in questions surrounding these data. A dialog system, commonly referred to as a chatbot, is used in various sectors like medical, e-commerce, education, and many more. Some of the most used examples of chatbots include "ChatGPT", "Snapchat My AI" and many more, all having different functionalities. The applications of this software may vary according to the requirements, but the purpose remains the same and that is to provide a satisfactory answer to the user.

The most commonly used technology in building chatbots is NLP (Natural Language Processing), which helps the system understand what the user is trying to convey. The other additional features of chatbot may include voice recognition, image recognition, etc. The main job of a chatbot is to make intelligent guesses like a human and also to save time utilized to read through whole available data by providing quick responses. These dialog systems also reduce an individual's workload by automating the process of doubt-solving and acting as a help desk.

The most effective use of chatbots is in the education sector. Some chatbots may be used to provide answers to a particular theoretical or practical question and others to get details about a particular institution or a university. The latter type of chatbot is mostly used by students or parents for getting information about a university, its fee structure, eligibility criteria, number of seats, courses it offers, and many more common doubts. Users can directly input their queries on the user-friendly GUI (Graphical User Interface) of the chatbot from where the query is mapped with the keyword which then connects the chatbot to the knowledge base to get the appropriate answer and then the final response is provided to the users. This GUI is directly integrated with the official university or college website to make it easy for users to access the information in more detail and can be accessed from anywhere. Direct integration helps easy access to components like department information, faculties, their experience years, events, and extra-curricular and co-curricular activity information even before entering the campus of the university.

II. LITERATURE REVIEW

Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) technologies have led to the development of chatbot systems that utilize Natural Language Processing (NLP) and deep learning models to achieve high accuracy in responding to user queries. In the education sector, college chatbots are being designed to enhance student support and communication, providing personalized assistance and complete information on various college-related topics.

International Journal of Innovations in Engineering and Science, www.ijies.net

Contrary to the traditional conversation systems like ELIZA, which are largely built with hand-written rules, that may or may not provide an appropriate response researchers have recently started to develop principled and data-driven approaches to build open-domain conversational systems due to the advantages of the large-scale social conversation data that is publicly available and the rapid advancement of deep learning approaches.[1] The authors also focus on the use of LSTM (Long Short-Term Memory) to keep a record of the ongoing conversations to provide correct responses according to the context. It refers to the chatbot having a memory to keep track of the queries that can be related and giving the responses based on the prior conversations held.

Despite being an intelligent system that provides quick responses to queries, most chatbots are rule-based chatbots, where the chatbot is connected to a knowledge base of tags, intents or keywords, and responses which limits the capability of a chatbot to the responses only present in the knowledge base. Following this limitation, research is being done in making the regular rule-based chatbots to be informative, responsive, and complete the correspondence in a conversational human language. This requires the incorporation of Natural Language Processing (NLP) and Machine Learning (ML) technologies into the college chatbot system, ensuring smooth interaction of the system's performance with AI. Educational institutions can develop an interactive online environment to facilitate smoother communication.[2] There are different techniques to implement a chatbot which differs with requirements for the organization being the biggest contributing factor. The chatbot building mainly depends upon the purpose it serves, the users it has, the programming language used for its building and many others. Therefore, selecting an optimal chatbot framework is the most basic and important step to create a proper functional, and efficient chatbot that serves the purpose.

Chatbots are modeled on various strategies such as knowledge base, and machine learning-based. Machine learning primarily based on chatbots yields greater practical results. Chatbot which gives responses primarily based on the context of conversation tends to be extra consumer-friendly. [3] In this literature survey, the author mainly emphasizes the use of TensorFlow as it has powerful tools to teach computers to understand human language and NLP is used to keep the context of the conversation. Chatbots of this kind can mostly be used in small businesses for cost saving and reducing the dependency on humans for minor issues. The authors also focus on the existing architecture of a chatbot and the implementation of chatbots in various domains.

In the literature survey, the author primarily focuses on the implementation of the existing architecture of

a chatbot and providing appropriate responses to the user's queries, may it be a student or a parent. Its primary objective is to reduce the manual efforts of the management and to provide the required information to the user in an automated manner. [4]

In the literature review, the primary focus is on "What is NLP? ". The term NLP refers to the machine learning technique that is used to make the computer understand the human spoken language. It performs tasks like speech recognition, name entity recognition, and many more. NLP also uses NLTK (Natural Language Toolkit) which includes libraries to perform NLP tasks efficiently. [5]

In the literature review, WordNet is a database of English words that groups them into synonyms called "Synsets" and records semantic relationships between them. It's used in natural language processing to retrieve information and mine text. [6]

III. PROBLEM STATEMENT

Consider a scenario where a student has freshly passed his/her 12th standard and is searching for a good university that provides the course he/she may be interested in. The most basic approach they undertake is to visit the official college website.

Many college websites can be confusing as it has a lot of information and majorly would not have the immediate answers for the student queries and at the end, students will have to visit the college campus in person. This can be a tiring job for both the college staff and the students and also proves to be a less efficient way of getting the answer to your questions. One of the major setbacks for this procedure will be the limitation of time in which they can get answers to their queries as the college operates only during working hours. Even if the student reaches college, sometimes due to their hesitating nature, he/she gets answers to every question they wish to ask and thus majorly relies on the wrong information.

The main goal of developing this intelligent dialog system is to help the users, in this case students and parents to get an accurate and appropriate response to their query without any human intervention. Students can ask as many doubts as they want irrespective of the time, and place and get reliable information directly from the college website.

IV. BASIC SYSTEM ARCHITECTURE

The existing architecture uses simple keyword matching for the input query to the knowledge base and gives the appropriate response according to the query. The basic architecture used is referred to as a "Rule-based" chatbot. This type of chatbot utilizes a knowledge base of pre-defined responses and rules. This type of chatbot is used to

International Journal of Innovations in Engineering and Science, www.ijies.net

respond to the keywords that are defined during the training phase of the model and in case the keyword is not found, it gives the message to contact the administrative college for further updating of the knowledge base.

It follows the following steps: –

1. Input the query on the GUI (Graphical User Interface)

2. Parse the input query for a keyword.

3. Use WordNet to find the synonyms of the keyword

4. Match the keyword with the knowledge base

5. When the keyword is matched with the query present in the knowledge base, give back the response to the user

6. When the keyword is not found, the user does not get the requested response and thus gets the default response to contact the administrative department of the institution
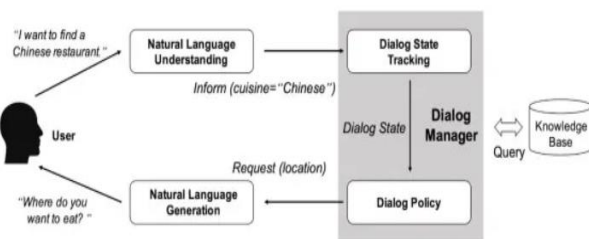

Fig 1. The architecture of rule-based chatbot. [7]

The above-mentioned architecture was used in the creation of the first-ever human-made chatbot ELIZA created by Joseph Weizenbaum at MIT Laboratories and was the first chatbot that attempted to beat the Turing test which tested the computer against a human based on their conversational abilities. ELIZA is a simple dialog system that uses pattern recognition as a way to understand the intent of the user query and give back a response.

Though being the first-ever human-created intelligent system, ELIZA did not always give the correct responses to the asked questions.

Fig 2. Conversation with ELIZA. [8]

While ELIZA failed to perform a normal conversation, DR. Wallace started working on the ALICE ((Artificial Linguistic Internet Computer Entity), a chatbot similar to ELIZA but used AIML (Artificial Intelligence Markup Language) to shape responses for the input questions. The essential structure highlight of AIML is moderation and from all the talk of robot programming languages, AIML is maybe the least complex.

Though the upgrade of ELIZA to ALICE was done after innovating ELIZA on various grounds, which made it win the Turing test and the Loebner prize, it failed to address the following issues -

• Unable to successfully perform NLP as it worked on the pattern recognition and pre-fed data during training which make it difficult to be used and practically implemented in daily life.

• Due to its restrictions on giving responses based on pre-determined templates, it was not able to provide efficient and useful responses for larger and more complex queries.

• The use of pre-fed training data for the rule-based chatbot made scalability a big hurdle in improving the working since the maintenance cost for the pre-fed instructions and response data was high.

V. ANALYSIS OF IMPLEMENTED TECHNIQUES

To overcome the limitation of the basic architecture of chatbots mentioned above, the most implemented architecture for a college inquiry chatbot is "Retrieval-based". This type of chatbot is used for inquiry purposes as the conversation between users and the system is straightforward and majorly does not require the generation of new answers. The working of a retrieval-based chatbot is explained in the following steps:

1. Take the input query via a GUI

International Journal of Innovations in Engineering and Science, www.ijies.net

2.   Parse the input query and find the keyword

3.   Once the keyword is found, check for the appropriate response for the keyword from the knowledge base.

4.   If the keyword is not found in the knowledge base, give back the feedback to the administrator of not getting the correct response for the given input query and give a default response to the user.

5.   Store the query to which the response was not generated into the knowledge base, which can later be updated to have an appropriate response either by the administrator of the dialog system itself.
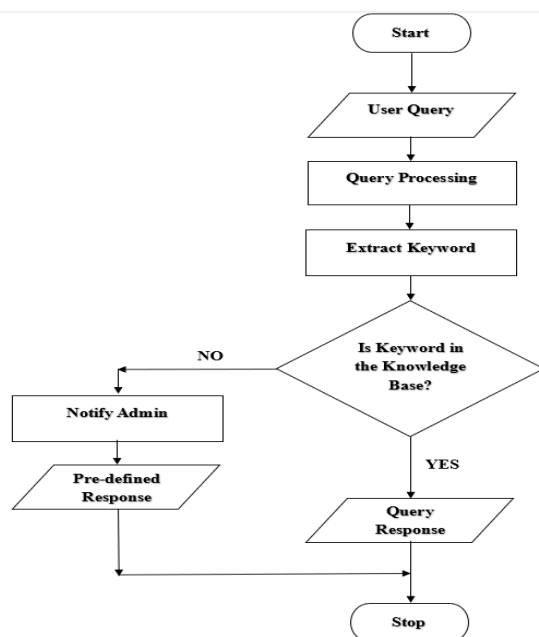

Fig. 3 Architecture of working of a retrieval-based chatbot.

Even though the "retrieval-based" models are widely used and implemented, they cannot overcome the limitation of generating an appropriate answer based on the type of query inputted. To overcome this limitation, a new type of chatbot was introduced which generated responses according to the requirements of the user. These types of chatbots are popularly known as "generative chatbots" which do not fully depend on pre-defined queries and their responses from the knowledge base but generate a response by understanding the intent of the user input query. They are mostly built with the help of Deep Learning concepts like Neural Networks, and LSTM (Long Short Term Memory) which not only make intelligent responses but also keep track of the ongoing conversation between the chatbot system and the user.

VI.   BASIC WORKFLOW FOR A WORKING SYSTEM

1.   Data collection –

   The basic step of the creation of any working model is the collection of appropriate data that is relevant to the chosen domain. The collected data should be in a proper format, suitable to train the model, and easier to find errors and correct them. Datasets can be downloaded from open-source platforms like "Kaggle" according to the preferences and requirements.

2.   Data pre-processing –

   Once the data is collected and in proper formats like Excel, CSV, JSON, and many others, it should be pre-processed to make it usable and free from any missing and noisy values. The second step of pre-processing the data is important to obtain appropriate and accurate results from the developed system.

   The data should contain three basic entities of query, tag, and response which are important to get responses for the user-inputted questions. The declaration of tags, commonly known as keywords should be done properly concerning the intended response. The pre-processed data sets are then used as a 'knowledge base" for the system.

3.   Build a GUI –

   The most basic way for a chatbot to interact with the user is via a user-friendly GUI (Graphical User Interface). So, the third step in the system flow is the creation of UI to facilitate smooth communication of queries to the system while keeping it user-friendly.

4.   User interaction and input processing –

   Once the UI is created, it is then used by the user to put in their queries and wait for the response. The input is then pre-processed which may involve tokenization, which refers to breaking down a piece of text or data into smaller parts called tokens. Each word in the input sentence is broken down into small tokens which reduces sentence complexity and then parsed for keywords.

5.   Keyword identification and pattern matching –

   Once the query is inputted, the system then identifies the keyword from it and the selected keyword is then used to perform pattern matching via which the keyword is matched to the closest matching question or query from the knowledge base where pre-defined queries and responses are stored. This also takes the help of a lexical database referred to as WordNet, which helps in

International Journal of Innovations in Engineering and Science, www.ijies.net

understanding the meaning of the keywords and keeping data about synonyms of various keywords which helps in their pattern matching.

6. If the pattern matched and the response was found

7. Retrieve the responses and deliver them–

Once the step of pattern matching is completed and the appropriate response is detected, the system then retrieves it from the knowledge base and provides it to the user as the final answer for their (user) question or query.

8. ELSE

9. Take a feedback

If the keyword is not matched to any of the queries available in the knowledge base, they are sent a default response and have the option to send feedback to the administrator in the absence of an appropriate response. The query for which the response was not available is stored in the knowledge base for future reference. The feedback sent by the user can be utilized to improve the system.

10. Improving the system quality –

While maintaining the quality of the currently working system to keep it functioning smoothly and user-friendly, upgrading the system according to the obtained feedback helps in improving the response generation over time. This also includes periodically updating the knowledge base with new entries to keep us with the growing demands of the user.

The above-mentioned system works on the methodology of a retrieval-based chatbot with a feedback system which helps the administrator understand the needs of the users, their differing preferences and requirements and to keep up with most of them.

The working of this type of chatbot is similar to the 'rule-based' chatbot which uses pattern recognition to give back responses to user queries, with the biggest difference being the ability of the 'retrieval-based' chatbot to self-learn and improve over time.

While following the basic architecture till keyword identification and pattern matching, generative chatbots unlike the rule or retrieval-based chatbots can generate a new response based on the large amount of training data it was trained on. The basic working of this type of chatbot refers to the mapping of a user-given input to the response generated by the computer. This saves time for both the admin and the user to reduce giving feedback and waiting to make changes in the knowledge bases to get the desired output. But sometimes even the responses generated by computers become repetitive due to their supervised learning nature and

thus reinforcement learning is applied in the generative chatbots to make the conversation sound more human and informative, popularly termed "generative AI".

VII. CONCLUSION

Chatbot as an intelligent dialog system was created to majorly address student college inquiry queries instantly, saving the efforts for both the administration department and the student, making it time-effective and cost-effective at the same time. Even when the user does not get the requested response accurately, they can share the feedback with the admin to make the required changes in the chatbot and keep on updating the system to meet the requirements of the user.

The knowledge base used for the system should also be continuously updated with new queries and their respective responses which were left unanswered by the system before the feedback from the user.

The proposed model also highlights the use of tokenization and WordNet which helped in improving the process of pattern matching for the intents. Also, the use of NLP in helping the system understand what a particular user is trying to say in a human language efficiently is one of the most contributing factors in the proper functioning of the chatbot.

Retrieval-based chatbots have proven to be highly effective in addressing common inquiries, such as admission procedures, course details, campus facilities, and event schedules. This is particularly true when responses can be accurately predicted based on pre-existing queries in the database.

Retrieval-based chatbots, despite their limitations in managing open-ended conversations and producing innovative responses, are still highly favored for tasks that involve information retrieval and simple interactions.

VIII. FUTURE SCOPE

- This retrieval-based chatbot can be modified and innovated to become a generative type of chatbot that does not necessarily require a knowledge base and gives the output based on its understanding of the user input query.

- Additional features like speech and image recognition can be added to improve the functionalities of the chatbot system and make it more automated and user-friendly.

- The chatbot can also be customized for the students already a part of the institution to help them manage their academic formalities like getting academic calendars, department-specific elective

International Journal of Innovations in Engineering and Science, www.ijies.net

lists, event updates, information about extra-curricular and co-curricular club activities, and many more to help them stay updated.

- The chatbot can also be customized to support more than one language other than English.

- Additional features like LSTM (Long Short-Term Memory) which keeps the track record of the previously held conversations to help improve response delivery and ease-of-use of the chatbot.

REFERENCES

[1] M. Suruthi, Mr. P.Anbumani," Smart college inquiry chatbot using Deep Learning algorithm", IJRPR, 2023.

[2] Hrushikesh Koundinya K, Ajay Krishna Palakurthi, Vaishnavi Putnala, Dr. Ashok Kumar K, "Smart College Chabot Using ML and Python," IEEE, 2020.

[3] Ajinkya Huddar, Chaitanya Bysani, Chintan Suchak, Uttam D Kolekar, Kaushiki Upadhyaya," Dexter, the college FAQ chatbot", IEEE, 2020.

[4] Kumar Shivam, Mrs. Sheetal S. Patil," Chatbot for college website", IJCAT, 2022.

[5] The working of NLP and its tasks are mentioned at https://www.ibm.com/topics/natural-language-processing#:~:text=the%20next%20step-,What%20is%20NLP%3F,and%20generate%20text%20and%20speech.

[6] Working of WordNet is available at https://wordnet.princeton.edu/

[7]The evolution of chatbots from a rule-based system to ChatGPT is mentioned at https://www.analyticsvidhya.com/blog/2023/05/chatbot-evolution-chatgpt-vs-rule-based/

[8] Siddhartha B. S, Divya Khyani," An Interpretation of Lemmatization and Stemming in Natural Language Processing", Journal of University of Shanghai for Science and Technology, 2020