



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Special Issue 2, April 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)



# AI BASED CHATBOT TO ANSWER FAQS

Kuldeep L , Priyadharshini P , Priyadharshini M , Gladwin I , Karthikeyan K , Kirupalakshmi G

U.G. Students, Department of Computer Engineering, Velammal Engineering College, Chennai, Tamil Nadu, India

**ABSTRACT:** Artificial intelligence (AI) chatbots are gaining popularity across industries such as education, healthcare, finance, and customer service. Chatbots are software programs that interact with users in a text-based manner, providing quick and automated responses without the need for human intervention. Chatbots use natural language processing (NLP) and machine learning (ML) algorithms to understand user queries and provide accurate and relevant responses. Using semantic search and contextual processing, chatbots can provide personalized recommendations, fix issues, and guide users in finding information. Startups are creating thousands of chatbots to give users better guidance and engage them in friendly, lighthearted conversations. Chatbots have the potential to reduce costs, increase efficiency, and improve customer satisfaction. One of the key capabilities of chatbots is to use semantic similarity to understand user queries and provide relevant responses. In addition to text-based queries, chatbots allow users to interact with voice commands, making them accessible to visually impaired users. It has also been trained to understand queries in multiple languages, making it easier to use in different regions. Question datasets can be self-learned using AI. Chatbots allow users to enjoy a simpler and more convenient experience when interacting with businesses and organizations.

**KEYWORDS :** Chatbot, Artificial intelligence (AI), Natural language processing (NLP), Machine learning (ML), Semantic search, Multilinguistic, Dataset, Transformers, Algorithms, API , SBERT, .

## I. INTRODUCTION

A chatbot which operates by receiving input from users and processing it through natural language processing (NLP) techniques to understand the user's query. The chatbot then searches for the relevant information by utilizing a knowledge base or corpus of data. The chatbot is designed to allow users to raise their queries through popular messaging platforms such as Telegram and Whatsapp. It supports multiple languages. For example, users can interact in different languages such as Tenglish, and the chatbot can respond in the appropriate language such as Tamil. The chatbot employs semantic similarity to find the most appropriate response to the user's query. This means that the chatbot is able to analyze the user's question and identify the underlying meaning behind it. It then searches for similar questions in its corpus or knowledge base to find the most appropriate response. One of the key features of the chatbot is the ability to self-train the question datasets using artificial intelligence (AI) leading to higher accuracy and greater efficiency. By leveraging AI and machine learning technologies, the chatbot can identify patterns and trends in user behavior, and use this information to optimize its performance and provide more relevant and personalized responses. The chatbot is well-equipped to handle a wide range of user queries and provide valuable insights and assistance across a variety of domains.

## II. RELATED WORKS

Chatbots were first introduced in the 1960s with ELIZA, which simulated a psychotherapist. Since then, advances in artificial intelligence, machine learning, and natural language processing have improved chatbot technology. Chatbots provide personalized support and scalability, handling multiple exchanges simultaneously and available 24/7. However, chatbot development requires a multidisciplinary approach involving experts in computer science, natural language processing, and psychology. Machine learning algorithms train chatbots on vast datasets, while natural language processing analyzes user input and generates appropriate responses. Chatbots raise ethical and privacy concerns, requiring unbiased and inclusive design.

**2.1 An Evaluation of the Performance and User Experience of the chatbot discussed in this paper with ChatGpt(text-davinci-003 ) and Meena Chatbot**

When comparing it with chatGPT and its API. we can specify like generative model has no fact checking and it is pre trained of some other dataset but our can be trained when new data arrives via the admin panel

Meena is a Google-developed open-domain chatbot that focuses on generating engaging responses across various topics and queries. It is not focused on the users dataset whereas the FAQ bot should be focused.

**III. METHODOLOGY**

Developing a chatbot requires a systematic approach consisting of various essential steps. The initial phase involves identifying the primary purpose and scope of the chatbot, which is to provide customers with a conversational experience that feels natural and allows them to have their queries resolved efficiently. The subsequent step involves understanding the target audience, their preferences, and requirements, which requires the developers to empathize with the end-users and deliver a user-friendly interface that aligns with their needs. The subsequent step is crucial as it requires the developers to possess knowledge of various prerequisites such as chatbot platforms, frameworks, design, and programming skills that cater to the specific requirements. Additionally, creating decision trees that help guide the bot's responses to user inputs is essential, as is training the bot using Natural Language Processing (NLP) techniques to enable it to understand and respond to user inputs accurately.

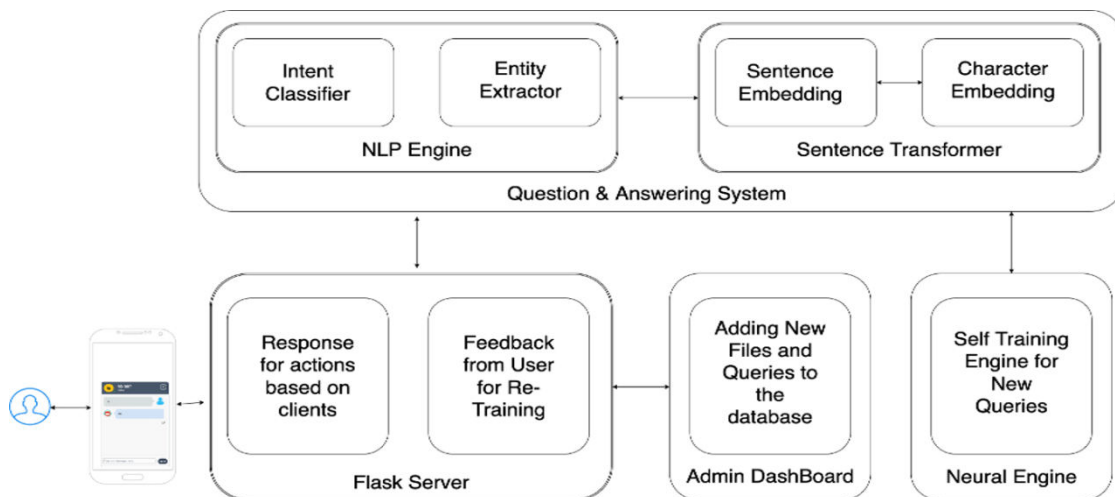


Fig.1 Architecture of the chatbot

The development stage follows next, where the developers need to gather the corpus text data that includes questions and responses that the bot is likely to encounter. Preprocessing the data by removing irrelevant and redundant information, including stop words, stemming, and lemmatization, and tokenization of the preprocessed data into phrases or words are vital steps. The tokens are then converted into vectors using advanced embedding techniques that map each word or phrase to a numerical vector representation. A semantic similarity model, all-miniLM-L6-v2 is trained using the preprocessed data, allowing the model to calculate the similarity between two vectors by comparing their cosine values and euclidean distance. Fine-tuning the model by adjusting hyperparameters such as the learning rate, the number of layers, and batch size improves its performance and ensures optimal functionality.

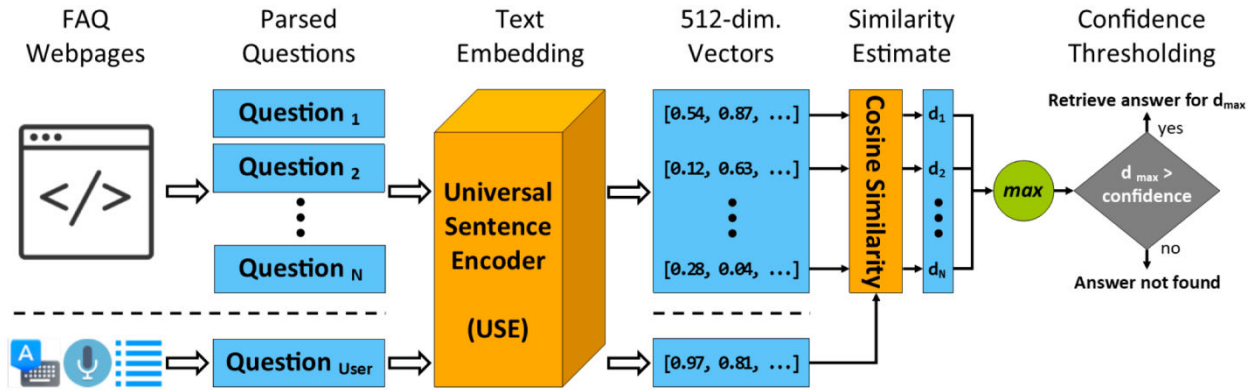


Fig.2 Working of semantic similarity model

### 3.1 TRANSFORMERS

The Transformer architecture used in chatbots consists of an encoder and a decoder. An encoder takes user input and transforms it into a sequence of vectors. Each vector represents a specific word in the input. This process involves multiple layers of intra-attentional and feed-forward neural networks that allow the model to capture relationships between different words and learn contextual representations of each word. A decoder, on the other hand, takes an encoded input sequence and produces a response based on it. It also contains multiple layers of intra-attentional and feed-forward neural networks, yet uses yet another attentional mechanism to selectively focus on relevant parts of the encoded input sequence to generate responses. During the training process, the model is trained on pairs of input and output sequences, with the goal of minimizing the discrepancy between predicted and actual outputs. Backpropagation is used to adjust the model's internal parameters based on the error between the predicted output and the actual output. Our chatbot leverages a pre-trained Transformer model to generate appropriate responses based on user input and uses attention mechanisms to focus on the most relevant parts of the input. Contextual understanding obtained through the intra-attention and attention mechanisms of encoders and decoders enables chatbots to provide accurate and relevant responses to user queries.

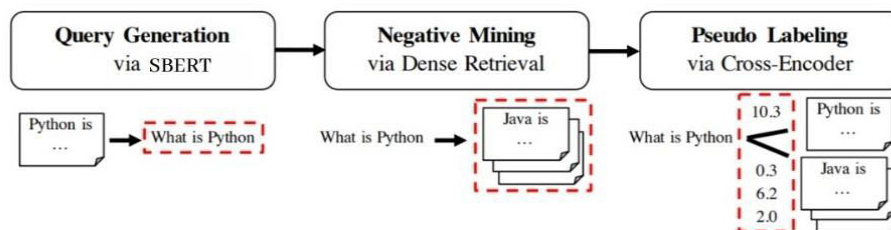


Fig.3 GPL Architecture

In the next phase, we have extended the multilingualization of our chatbot to improve the personalized customer experience and increase customer satisfaction. It does this by leveraging the power of the Google Translate API, which uses advanced machine learning techniques to automatically translate text from one language to another. First, the Google Translate API receives input provided by the chatbot. The Translation API uses language detection algorithms to automatically identify the language of the input text. The entered text is passed to the neural translation system.



Neural translation systems translate from a source language to a target language by dividing text into small units called tokens and matching them with corresponding words or phrases in the target language. After the translation is complete, the API performs post-processing to adjust the output and improve the quality of the translation. This includes removing unnecessary words, correcting grammatical and syntactic errors, and ensuring that the translation accurately reflects the intended meaning of the original text.

To further simplify user accessibility, chatbots will be accessible via popular instant messaging platforms such as Telegram and WhatsApp. To enable access via Telegram, the developer must create a Telegram bot using Telegram Bot API. This will generate an API token for authentication. Following this, we set up a server that receives messages from Telegram and forwards them to our chatbot, and forwards replies back to Telegram. The Twilio cloud communications platform is used to enable access via WhatsApp. This requires creating a Twilio account and WhatsApp activation. Creating his Twilio sandbox for WhatsApp allows your chatbot to send and receive messages from his WhatsApp number without purchasing a dedicated WhatsApp number. You can communicate with Twilio by setting up a server and deploying this bot to your server. Connect your bot to Twilio so it can receive and reply to messages from WhatsApp. Sending your bot a message from your WhatsApp number and testing it ensures that your users get a real-time response from your chatbot.

## V. WORKFLOW

### 5.1 Data Collection

Data is collected with important technical keywords and their respective answers with the correct meaning to the domain-based. The dataset would be of various domains covering all customer queries and different information to answer the question to the user in an accurate manner.

### 5.2 SBERT Encoding

This approach uses the BERT model to encode the question  $q_i$  and the deduced conversation context  $c$  into a contextualized SSE text representation. BERT is a cutting-edge, pre-trained language model for NLU that uses transformers to learn deep bidirectional representations. In a training example  $(q_i, c)$ , you pair questions and conversation contexts into a single sequence. The contextual information provides the BERT model with the functionality necessary to understand the contextual problem. ConvQA's BERT model is limited to a text of 512 tokens per sequence. It makes it suitable for processing long sequence data from multi-turn interactions related to each conversation. If the sequence data exceeds 512 tokens per sequence, understanding the data becomes a challenge.

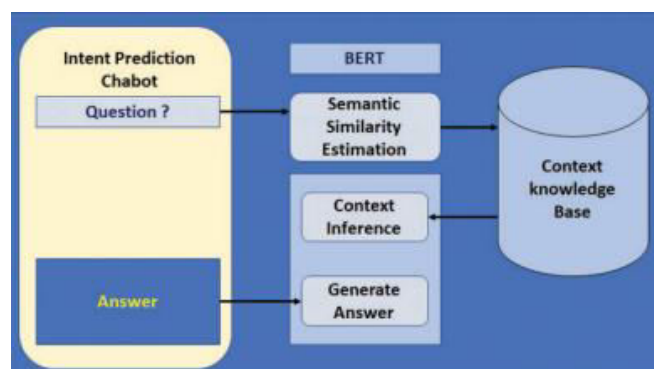


Fig.6 Working of BERT in Chatbot

### 5.3 Semantic Similarity

Given the user's question, the model performs Semantic Similarity Estimation by comparing the similarity of the question to the contextual representations of each previous CS in the knowledge base. By converting text (question and context) to term frequency inverse document frequency (TF-IDF) vectors, we calculate the cosine similarity between



the pronunciation of questions and the contextual conversation data in the knowledge base. The questions are represented as vectors, and the context data as matrices

$$\text{Similarity} = \cos^{-1} \theta = \frac{q \cdot c}{\|q \cdot c\|} = \frac{\sum_{i=1}^n q_i c_i}{\sqrt{\sum_{i=1}^n q_i^2} \sqrt{\sum_{i=1}^n c_i^2}} \quad (1)$$

## VI. IMPLEMENTATION OF QUESTION AND ANSWERING

The key to finding answers to questions is to determine the most relevant conversational context for those questions and to determine user intent using the context. If the word and the hidden vector are provided, the probability that the word is the START/END of the answer range is calculated as

$$P_i^S = \frac{\text{Exp}^{S.T_i}}{\sum_j \text{Exp}^{S.T_j}}, \quad P_i^E = \frac{\text{Exp}^{E.T_i}}{\sum_j \text{Exp}^{E.T_j}} \quad (2)$$

The task is to predict the answer using the inferred context. In response prediction tasks, the model represents input questions and implicit contexts as a single-pair sequence of context-contextual questions using embedded questions.

$$P_i^S = \frac{\text{Exp}^{S.T_i}}{\sum_j \text{Exp}^{S.T_j}} \quad (3)$$

The same formula is used for computing the end of the answer text span,

$$P_i^E = \frac{\text{Exp}^{E.T_i}}{\sum_j \text{Exp}^{E.T_j}} \quad (4)$$

The score of a candidate span from position 'i' to position 'j' is defined as, Eq. (5)  $S.T_i + E.T_j$

VII. RESULTS

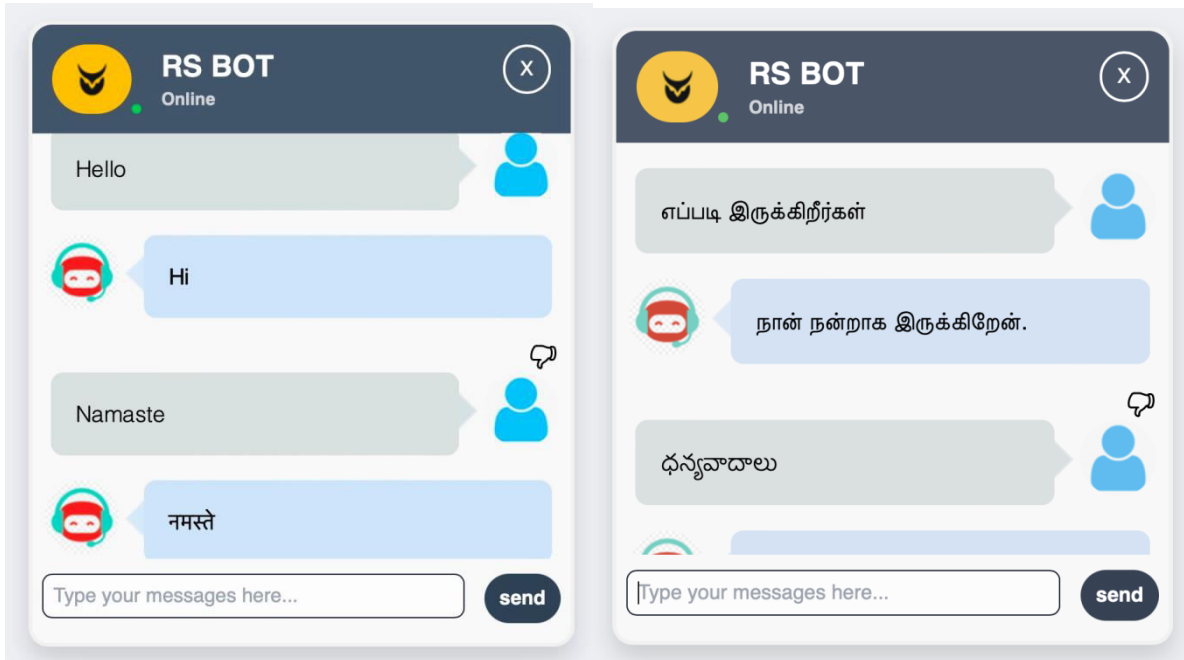


Fig.8 Multi Linguistic Nature of Chatbots.

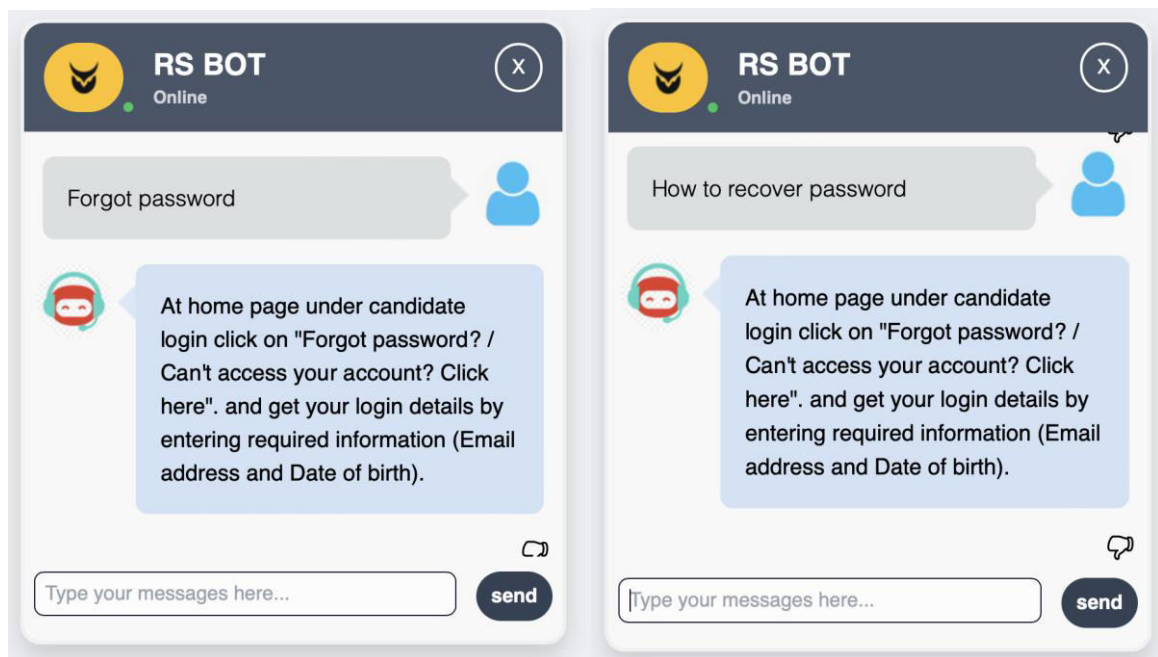


Fig.9 Implementation of semantic similarity in chatbot

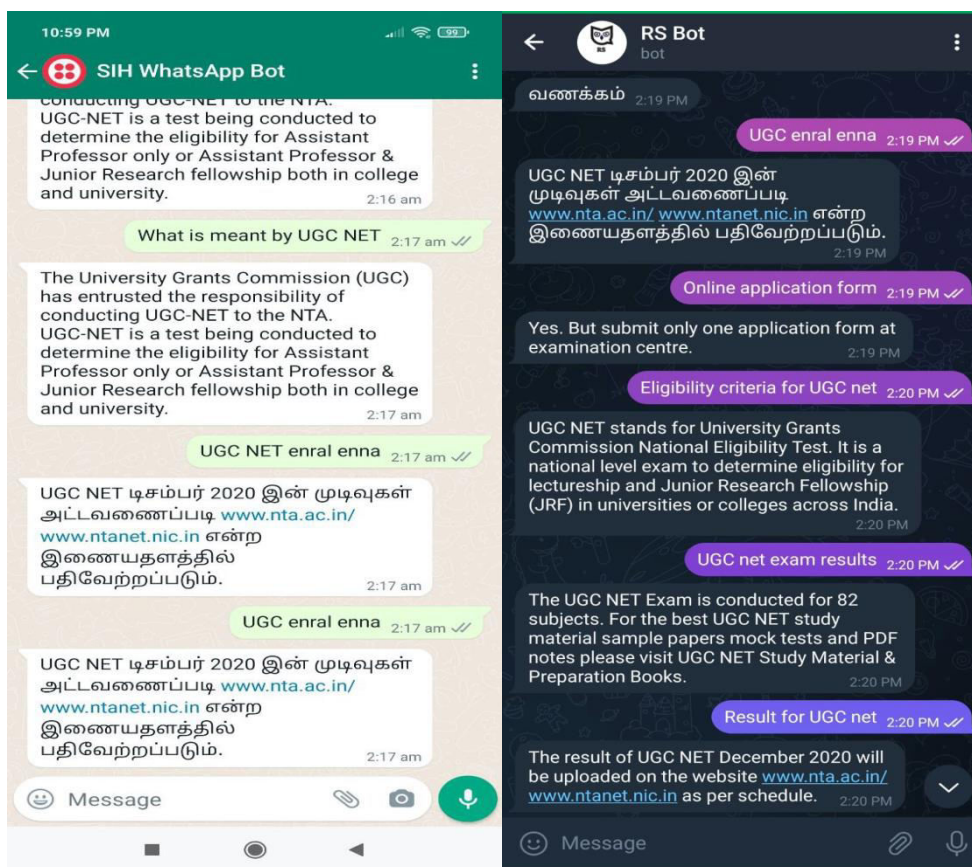


Fig.10 Working model of chatbot accessible from whatsapp and telegram.

The study presents the successful development of a chatbot system that utilizes advanced AI-based training methods and sophisticated NLP techniques to efficiently address user inquiries in multiple languages. The chatbot's language prediction abilities and learning from failures resulted in a 99% accuracy rate. It is easy to integrate, customizable, and offers an admin panel and a separate dataset to meet client-specific needs.

### 7.1 Comparison with Human

Although AI-powered chatbots can handle numerous queries promptly and efficiently, they cannot replace human communication as they lack emotional intelligence and empathy. Human communication is more adaptable to complex scenarios and can interpret subtleties in language, culture, and environment. However, chatbots offer benefits such as managing a vast amount of inquiries simultaneously, 24/7, without fatigue or errors. They can also be integrated into messaging platforms like WhatsApp and Telegram, providing a more accessible avenue for users to seek information and assistance. This research presents a sophisticated chatbot system utilizing advanced NLP techniques, language prediction capabilities, and efficient training methods for improved customer support and engagement.

## VIII. ETHICAL AND PRIVACY CONSIDERATION

The proliferation of chatbots in recent times has led to their increased use in collecting stoner data, including particular information similar as names, dispatch addresses, and phone figures. Still, it's critical to consider the ethical and sequestration counter accusations of data collection by chatbots.





In conclusion, the development of chatbots must be approached with a keen mindfulness of the ethical and sequestration considerations girding stoner data collection. translucency, security, delicacy, availability, and ethical use are essential aspects to consider, creating a positive stoner experience and fostering stoner trust.

## IX. CONCLUSION

It can be concluded that the use of natural terminology processing and simple machine learning algorithms has proven to be a work approach for developing conversational agents that can sympathize and react to homo language. AI Chatbot are not going to be a single standalone model it will be compressed with various models with each has the ability to transfer learning Factors so much as the accuracy of natural nomenclature processing, the complexity of the informal flow, and the level of customization and personalization put up a great touch down on the user's perception and satisfaction with the chatbot.

Future search in the domain of chatbots should focalize on improving the natural language processing in multi language and machine eruditeness algorithms to better sympathize the nuances of man language, as well as enhancing unfold domain noesis base of chatbots to be correlated with exploiter preferences.

## REFERENCES

1. "A Study on Chatbot using Natural Language Processing and Artificial Intelligence Techniques" by S. S. Sankpal and S. S. Sankpal, International Journal of Computer Science and Mobile Computing, Volume 9, Issue 9, September 2020 - <https://www.ijcsmc.com/docs/papers/September2020/V9I9202001.pdf>
2. "Design and Development of a Chatbot for a Customer Service Application using Artificial Intelligence" by S. S. Alam, S. Hossain, and A. Paul, International Journal of Advanced Computer Science and Applications, Volume 11, Issue 1, January 2020. - <http://dx.doi.org/10.14569/IJACSA.2020.0110156>
3. "Integrating Chatbots in Online Courses: Using Telegram to Improve Learning Outcomes" by D. D. Díaz, D. R. Amado-Salvatierra, and N. Muñoz, International Journal of Educational Technology in Higher Education, Volume 15, Issue 1, February 2018. - <https://doi.org/10.1186/s41239-018-0096-2>
4. "A Multilingual Chatbot for User Engagement in E-commerce Websites" by M. M. Rahman and N. N. Singh, International Journal of E-Services and Mobile Applications, Volume 11, Issue 3, July 2019. - <https://doi.org/10.4018/IJESMA.2019070101>
5. "A Multilingual Chatbot for Mobile Devices" by A. Hossain, S. Chatterjee, and S. Bhattacharya, IEEE Transactions on Consumer Electronics, Volume 63, Issue 2, May 2017. - <https://ieeexplore.ieee.org/document/7898942>
6. "Design and Development of a Chatbot for Visually Impaired People" by N. K. Roy and R. K. Chaurasiya, International Journal of Intelligent Systems and Applications, Volume 10, Issue 2, February 2018. - <https://doi.org/10.5815/ijisa.2018.02.03>
7. "Design and Development of a Multilingual Chatbot System" by V. Prakash, M. G. Gowda, and M. K. Bhaskar, International Journal of Computer Applications, Volume 170, Issue 3, July 2017. - <https://www.ijcaonline.org/archives/volume170/number3/28012-2017915026>
8. "Building a Customer Service Chatbot Using Semantic Similarity and Fuzzy Logic" by D. Kumar and S. B. Patil, International Journal of Advanced Research in Computer Science, Volume 8, Issue 2, February 2017. - <https://www.ijarcs.info/index.php/Ijarcs/article/view/2565>



**SJIF: 8.423**



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INTERNATIONAL CENTRE



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details