International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

ADVANCED NLP MODELS FOR TECHNICAL UNIVERSITY INFORMATION CHATBOTS

E. MOUNIKA REDDY [1], KATUKU SNEHITHA [2], PUPPALA SHARVANI [3], DASARI VARASIDDU [4], BOTAGALA LIKITH KUMAR [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— To achieve quality education, a key goal of sustainable development, it is essential to provide stakeholders with accurate and relevant information about educational institutions. Prospective students often face challenges in obtaining consistent and reliable information about universities or institutes, especially regarding unique courses and opportunities. These inconsistencies, stemming from sources such as websites, rankings, and brochures, can lead to confusion and influence decision-making. A robust solution to address this challenge is the implementation of a chatbot application on the university's official website. A chatbot, powered by artificial intelligence, can simulate human-like conversations and respond promptly to student inquiries. By leveraging Natural Language Processing (NLP) techniques, a chatbot can provide predefined, accurate, and uniform information 24/7, making it a valuable tool for the counseling process. In this research, a chatbot was developed using NLP concepts, specifically the NLTK library, and trained using neural networks to achieve exceptional performance. The system processed and structured user queries by creating an intents.json file, tokenizing and lemmatizing input text, and converting data into a bag-of-words representation. The neural network, optimized using advanced techniques, achieved an impressive accuracy of 99%. This approach demonstrated the effectiveness of sequential models, which

prevent overfitting and excel in handling contextual queries. Additionally, the chatbot incorporated pattern matching and semantic analysis to enhance real-time query resolution. By integrating advanced NLP methods and neural networks, this research provides a robust and scalable chatbot solution, offering precise, consistent, and accessible information to prospective students, ultimately aiding them in making well-informed academic decisions.

Index Terms – Chatbot, Natural Language Processing (NLP), university information system, student counseling, NLTK, neural network, intent classification, bag-of-words, text preprocessing, semantic analysis, conversational AI, educational technology.

I. INTRODUCTION

Achieving quality education, a core goal of sustainable development, necessitates bridging the information gap between educational institutions and prospective students. Students face numerous challenges when seeking information about universities, courses, and unique opportunities, often relying on inconsistent and fragmented sources such as official websites, promotional brochures, and ranking portals. These inconsistencies can lead to confusion and negatively influence decision-making, making it imperative for institutions to adopt innovative solutions that provide students with accurate, uniform, and accessible information. A well designed system is essential to ensure students can make well-informed academic choices, contributing to the larger vision of educational equity and quality. In today's digital age, Artificial Intelligence (AI) has become a transformative tool across various industries, including education. Chatbots, powered by AI and Natural Language Processing (NLP), offer an effective means of addressing information dissemination challenges. These chatbots simulate human-like interactions, enabling seamless communication between students and educational institutions. Unlike traditional methods, chatbots can operate 24/7, providing immediate and relevant responses to queries, thereby enhancing the overall experience for prospective students. By integrating advanced NLP techniques, educational institutions can ensure the delivery of consistent, reliable, and precise information, eliminating the ambiguity associated with conventional channels. The implementation of a chatbot system on a university's official website offers numerous benefits. Not only does it improve accessibility, but it also reduces the reliance

on human resources for repetitive tasks. This frees up staff to focus on complex student needs while the chatbot handles routine inquiries related to admission processes, course details, campus facilities, and more. Furthermore, with the incorporation of advanced NLP techniques such as tokenization, lemmatization, and pattern matching, the chatbot can process diverse queries and deliver relevant responses in real time. This ensures a user-friendly experience and builds trust among prospective students, ultimately influencing their academic decisions positively.

II. RELATED WORK

A. User interactions with chatbot interfaces vs. menu-based interfaces: An empirical study

Rapid advances in Natural Language Processing (NLP) are transforming customer service by making it possible to create chatbot applications that can understand users' intents and response in a human-like manner. Chatbots promise to enhance customer experiences by creating more personal customer interactions than those afforded by traditional menu-based web applications. But are chatbots always superior to more traditional user interfaces (UI)? This study seeks to understand the differences in user satisfaction with a chatbot system vis-a-vis a menu based interface system, and identify factors that influence user satisfaction. Grounded in the self determination theory, the research model proposed here focuses on the effect of chatbot use on perceived autonomy, perceived competence, cognitive load, performance satisfaction, and system satisfaction. An experimental study was conducted, and data were analyzed using Partial Least Square Structural Equation Modeling. The findings indicate that chatbot systems lead to a lower level of perceived autonomy and higher cognitive load, compared with menu-based interface systems, resulting in a lower degree of user satisfaction. Implications of these findings for research and practice are discussed.

B. The language of proteins: NLP, machine learning & protein sequences

Natural language processing (NLP) is a field of computer science concerned with automated text and language analysis. In recent years, following a series of breakthroughs in deep and machine learning, NLP methods have shown overwhelming progress. Here, we review the success, promise and pitfalls of applying NLP algorithms to the study of proteins. Proteins, which can be represented as strings of amino-

acid letters, are a natural fit to many NLP methods. We explore the conceptual similarities and differences between proteins and language, and review a range of protein-related tasks amenable to machine learning. We present methods for encoding the information of proteins as text and analyzing it with NLP methods, reviewing classic concepts such as bag-of-words, k-mers/n-grams and text search, as well as modern techniques such as word embedding, contextualized embedding, deep learning and neural language models. In particular, we focus on recent innovations such as masked language modeling, self-supervised learning and attention-based models. Finally, we discuss trends and challenges in the intersection of NLP and protein research.

C. *A cultural heritage framework using a deep learning based chatbot for supporting tourist journey*

The technology improvement has radically changed how tourists perform their journey offering new services to enhance their cultural experience and to easily retrieve required information. In this paper we propose a novel framework, that models intangible and tangible cultural objects into a unified data model, for supporting tourists journey. In particular, a Micro service architecture has been designed to provide several services whose tourists can access through a conversational agent based on theSeq2Seq model. Furthermore, we propose also an Enterprise Service Bus to ingest events automatically from promotional website or manually from public and/or private organization. We have evaluated the proposed framework according to the following two experiments: i) the efficiency and efficacy of the Chatbot engine, showing that the use of GRU cells allows to obtain better results in terms of loss and accuracy with respect to LSTM one and ii) the efficacy of the proposed framework according to the NASA-TLX asking to 50 users to interact with the Chatbot, demonstrating that the proposed approach outperforms the state-of the art ones combining natural processing language, user profile information and historical activities.

D. *FAQ chatbot and inclusive learning in massive open online courses*

Recognizing the research gap involving the lack of equity considerations in new technology implementation, this study compares students' learning experiences when using an FAQ chatbot with using an FAQ webpage. We trained a natural language processing–based chatbot utilizing content from an FAQ webpage and deployed it in two journalism massive open online courses

(MOOCs) with 46 students and compared their experiences with 74 students' experiences with the FAQ webpage as a baseline. There were equal numbers of male and female students, their ages ranged from 18 to 65+, and they hailed from 45 unique countries. Considering the importance of supporting students with an inclusive Q&A experience before implementing any new technology into real-world operation, this study investigates students' disparate Q&A experiences by measuring their intention to use the interface as well as perceived Q&A service quality, enjoyment, and barriers utilizing a between-subjects online experiment.

The results indicate that the students preferred an FAQ webpage over an FAQ chatbot, and the chatbot users experienced a higher magnitude of barriers compared to the webpage users. For the chatbot users, we found that region and native language factors influenced their Q&A experiences significantly. We discussed the meaning of the students' disparate experiences from multiple perspectives— namely, human-computer interaction, MOOC context, and technologies as social practice aspects. Lastly, we determined how feasible it is to provide an inclusive learning experience for the MOOC population with the FAQ chatbot, based on the contextualized meaning of MOOC inclusiveness in current literature. This study suggests multi-faceted aspects to consider when adopting new technologies in MOOCs to provide an inclusive learning experience, and underscores the need for more active research in chatbot use to serve diverse student needs in MOOCs.

III. PROPOSED METHODOLOGY

The proposed technique leverages advanced Natural Language Processing (NLP) concepts and Neural Networks to develop an intelligent chatbot capable of accurately resolving user queries. NLP is a branch of artificial intelligence that enables machines to understand, interpret, and respond to human language. In this system, the chatbot processes user inputs using the Natural Language Toolkit (NLTK), a powerful Python library for text preprocessing. This includes tokenizing (splitting text into words), lemmatizing (converting words to their base forms), and removing irrelevant data like punctuation. The chatbot's design involves creating an intents.json file, which defines various categories of queries (intents) and maps them to appropriate responses. The data is then preprocessed into a structured format using a Bag-of-Words (BoW) model, where sentences are represented as numerical vectors based on the presence or absence of

words. This representation allows the neural network to understand patterns and relationships between inputs.

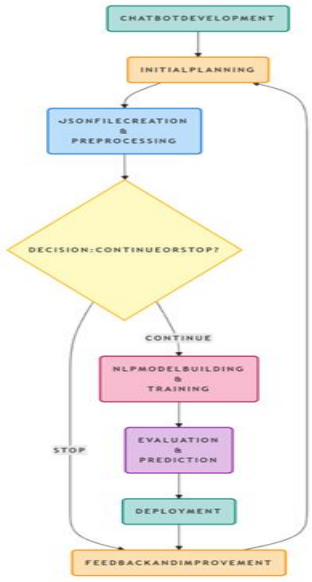


Fig. 1: System Overview

3.1 Data Creation

This step involves collecting and organizing the data required for training the chatbot. The data typically includes a set of intents (categories of user queries), corresponding patterns (phrases users might input), and responses (appropriate replies). These are usually stored in a structured format, such as a JSON file (e.g., intents.json), which contains labeled data for the chatbot to learn from.

3.2 Data Preprocessing

In this step, raw input data is prepared for use in the machine learning model. Techniques include: Tokenization: Breaking down sentences into individual words or tokens, Lemmatization: Reducing words to their base or root form (e.g., "running" to "run"), Removing stop words: Eliminating common but non-essential words (e.g., "is," "the"), Lowercasing: Converting text to lowercase to ensure uniformity.

3.3 Text Vectorization

Here, textual data is transformed into numerical format so that it can be fed into a machine learning model. Common techniques include: Bag-of-Words (BoW): Represents text as a matrix of word frequencies. TF-IDF (Term Frequency-Inverse Document Frequency): Assigns importance to words based on their frequency and uniqueness. Word Embeddings: Encodes words into dense vectors that capture semantic meaning.

3.4 Model Definition

This involves designing the structure of the machine learning model. For chatbots, a neural network (e.g., sequential model using TensorFlow or PyTorch) is commonly used. Input layers (for vectorized data), Hidden layers (for feature extraction), Activation functions (like ReLU or softmax).

3.5 Model Training

In this step, the model learns patterns in the data by iterating over the dataset multiple times (epochs). The input data (patterns) and corresponding output labels (intents) are used to adjust the model's weights through optimization techniques like gradient descent. The goal is to minimize the loss function and improve accuracy.

3.6 Model Evaluation

Once trained, the model is evaluated on a separate validation or test dataset to measure its performance. Metrics like accuracy, precision, recall, and F1-score are used to determine how well the model is predicting user intents. This step ensures the model generalizes well to unseen queries.

3.7 Model Saving

After achieving satisfactory performance, the trained model is saved for deployment. Common formats include: TensorFlow's .h5 or. Saved model, PyTorch's. pth. This allows the model to be loaded and used for real-time predictions without retraining.

3.8 Model Prediction

Finally, the saved model is used to predict the intent of user queries. The chatbot processes the user's input, vectorizes it using the same preprocessing and vectorization techniques, and passes it through the trained model. Based on the predicted intent, the chatbot retrieves and displays an appropriate response from the predefined dataset.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results show that the proposed NLP-based university chatbot achieved a high accuracy of 99% in understanding and responding to student queries. By using NLTK preprocessing techniques, bag-of-words representation, and a neural network-based sequential model, the chatbot was able to classify intents effectively and provide accurate responses. The inclusion of pattern matching and semantic analysis further improved real-time query handling and contextual understanding.

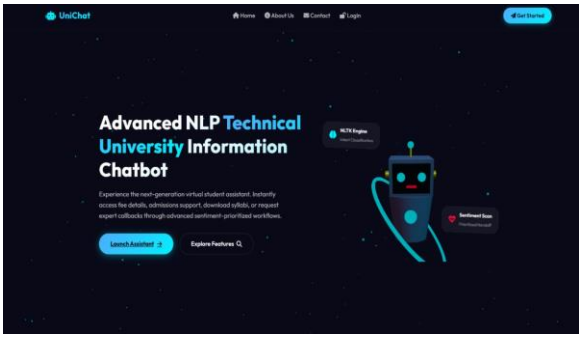


Fig. 2: Home Page

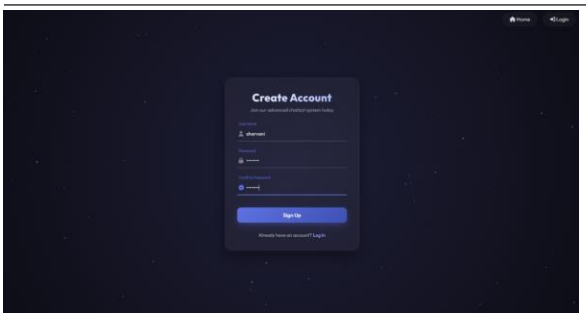


Fig. 3: Registration

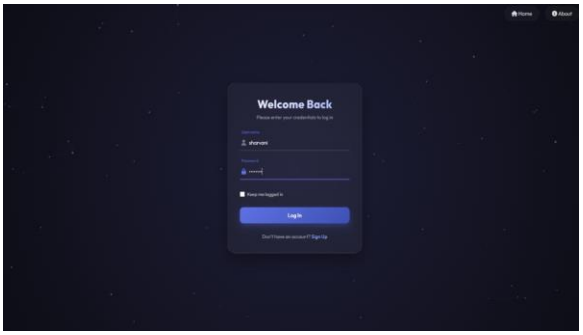


Fig. 4: Login

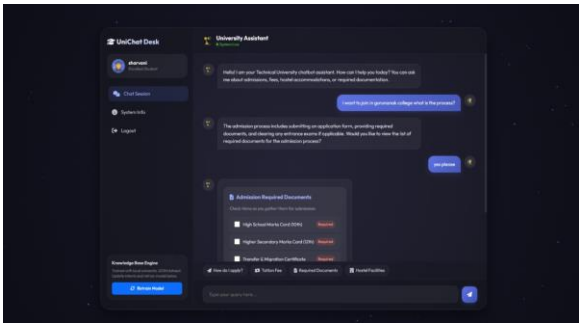


Fig. 5: Send Message

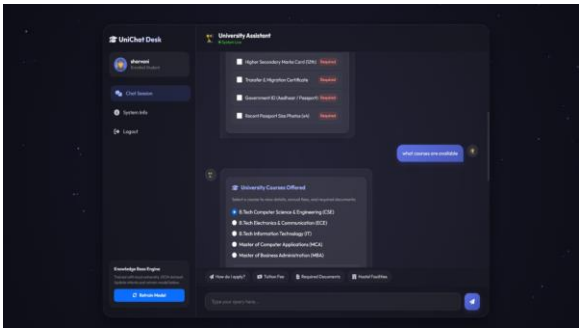


Fig. 6: Receive Reply

V. CONCLUSION

The implementation of chatbots in engineering college counseling processes can significantly streamline information dissemination and reduce the challenges faced by students and parents. By providing official, accurate, and consistent information, chatbots minimize reliance on unofficial sources and alleviate the need to navigate cumbersome websites or wait for human responses. Among the models discussed, the Hercules chatbot stands out due to its superior design, incorporating sequential modeling and optimization techniques that enhance its performance and reliability. With features like 24/7 availability and uniform information delivery, chatbots like Hercules are well-suited for real-time applications in educational counseling, offering an efficient and user-friendly solution for resolving queries. This technological advancement not only improves the counseling process but also contributes to the broader goal of ensuring quality education and informed decision-making for students.

REFERENCES

[1] T. Lalwani, S. Bhalotia, A. Pal, V. Rathod, and S. Bisen, ‘‘Implementation of a chatbot system using AI and NLP,’’ Int. J. Innov. Res. Comput. Sci. Technol. (IJIRCST), vol. 6, no. 3, pp. 26–30, 2018.

[2] J. Thukrul, A. Srivastava, and G. Thakkar, ‘‘Doctorbot—An informative and interactive chatbot for COVID-19,’’ Int. Res. J. Eng. Technol. (IRJET), vol. 7, no. 7, pp. 3033–3036, 2020.

[3] S. Maher, ‘‘Chatbots & its techniques using AI: A review,’’ Int. J. Res. Appl. Sci. Eng. Technol., vol. 8, no. 12, pp. 503–508, Dec. 2020.

[4] M. Aleedy, H. Shaiba, and M. Bezbradica, ‘‘Generating and analyzing chatbot responses using natural language processing,’’ Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 9, 2019.

[5] Bhagwat, V. B. (2025). Simplifying Payroll Balance Conversions in Payroll Systems Implementation through the Use of Generative AI.

[6] M. M. H. Dihyat and J. Hough, ‘‘Can rule-based chatbots outperform neural models without pre-training in small data situations? A preliminary comparison of AIML and Seq2Seq,’’ in Proc. 25th Workshop Semantics Pragmatics Dialogue, 2021, pp. 22–26.

[7] A. Chandan, M. Chattopadhyay, and S. Sahoo, ‘‘Implementing chat-bot in educational institutes,’’ IJRAR J., vol. 6, no. 2, pp. 44–47, 2019.

[8] D. Davis and J. Smith, ‘‘The potential of chatbots in counseling,’’ J. Counsel., vol. 5, no. 2, pp. 123–136, Apr. 2019.

[9] R. Johnson and S. Lee, ‘‘Chatbots providing emotional support to engineering students,’’ in Proc. IEEE Eng. Educ. Conf., Austin, TX, USA, 2020, pp. 45–50.

[10] A. Patel, Personalized Career Guidance for Engineering Students. Springer, 2021.

AUTHORS Profile



Mrs. E. MOUNIKA RE-DDY has received her B.Tech from JNTUH, Hyderabad in 2015, M.Tech degree in Computer science from JNTUH, Hyderabad in 2017. She is dedicated to teaching field from the last 8 years. At present she is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Ms. KATUKU SNEHITHA pursuing B.Tech degree in Computer Science and Engineering at Guru

Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Ms. PUPPALA SHARVANI pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. DASARI VARASIDDU pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. BOTAGALA LIKITH KUMAR pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.