iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 14 **Issue:** V **Month of publication:** May 2026

**DOI:** https://doi.org/10.22214/ijraset.2026.797555

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 14 Issue V May 2026- Available at www.ijraset.com*

# Development of a Chatbot System for Student Support Using Artificial Intelligence

Sakshi Meena[1], Dr. Shekhar Nigam[2]
[1]*Research Scholar,* [2]*HOD, NRI Institute of Research & Technology, Bhopal (M.P)*

***Abstract:*** ***The increasing integration of digital technologies in higher education has significantly transformed the way academic services are delivered and accessed by students. Despite this transformation, traditional student support mechanisms such as manual helpdesks, email-based communication, and limited office-hour services continue to face challenges in handling large volumes of repetitive queries efficiently. These systems are often time-consuming, inconsistent, and unable to scale effectively with growing student populations. To address these limitations, this research paper presents an Artificial Intelligence–based chatbot system for student support, developed using Natural Language Processing and deep learning techniques. The proposed system employs a Bidirectional Long Short-Term Memory (BiLSTM) model integrated with an attention mechanism to classify student queries into six predefined categories: admission, examination, faculty, fees, hostel, and library. A systematic methodology involving data preprocessing, embedding-based feature representation, model training, validation, and comprehensive performance evaluation is adopted. The system is evaluated using accuracy, precision, recall, F1-score, confusion matrix analysis, and training–validation learning curves. Experimental results reveal an overall classification accuracy of 20 percent, highlighting the challenges associated with limited and imbalanced datasets. The findings emphasize the importance of data quality in AI-based systems and provide realistic insights into chatbot deployment in educational environments.***

***Keywords: Artificial Intelligence, Student Support Chatbot, Natural Language Processing, Deep Learning, Bidirectional LSTM, Intent Classification.***

## I. INTRODUCTION

The rapid digital transformation of higher education has fundamentally changed the way academic institutions interact with students and manage administrative services. The widespread adoption of online admission systems, learning management platforms, and digital communication tools has resulted in the continuous generation of large volumes of student queries related to academic and administrative processes.

These queries typically include information requests regarding admissions, examination schedules, fee payments, faculty details, hostel facilities, and library services. Despite the availability of digital infrastructure, many institutions continue to rely on traditional support mechanisms such as helpdesks, email communication, and call centers, which are often inefficient in handling repetitive queries at scale.

These conventional approaches are characterized by delayed responses, limited availability, and inconsistent information delivery, particularly during peak academic periods such as admissions and examinations.

The increasing demand for immediate and accurate information has created a need for intelligent and scalable student support systems capable of operating continuously without human intervention. Artificial Intelligence–based chatbot systems have emerged as a promising solution to address these challenges by automating routine interactions and providing real-time responses to student queries. Chatbots are conversational agents that utilize Natural Language Processing techniques to understand user input and generate relevant responses. In educational environments, chatbots play a critical role in improving service accessibility, reducing administrative workload, and enhancing student satisfaction by providing instant support.

The effectiveness of chatbot systems largely depends on their ability to accurately interpret user intent and classify queries into meaningful categories. Recent advancements in deep learning, particularly Bidirectional Long Short-Term Memory networks, have enabled improved contextual understanding of textual data. The integration of attention mechanisms further enhances model performance by allowing selective focus on important words within a query. In this context, the present research proposes a deep learning–based chatbot system designed to classify student queries into six predefined support categories.

©IJRASET: All Rights are Reserved 

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 14 Issue V May 2026- Available at www.ijraset.com*

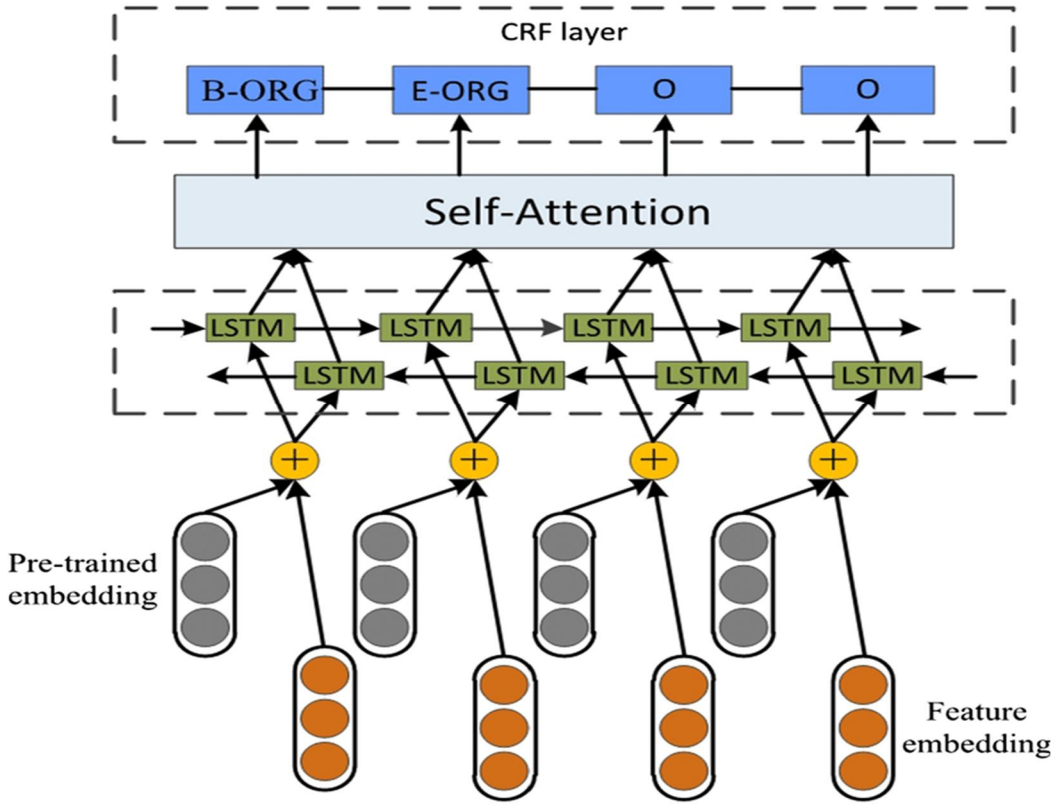



Figure 1: Illustrates the general working mechanism of an AI-based chatbot system, including query input, preprocessing, deep learning-based intent classification, and response generation.

## II. REVIEW OF LITERATURE

The development of chatbot systems for educational applications has evolved significantly over time, reflecting advancements in Artificial Intelligence, Natural Language Processing, and machine learning methodologies. Early chatbot systems were predominantly rule-based, relying on predefined scripts and keyword matching techniques to generate responses. While these systems were simple to implement and offered predictable behaviour, they were limited in their ability to handle linguistic variability, contextual ambiguity, and complex user queries. As a result, rule-based chatbots were unable to scale effectively in dynamic environments such as educational institutions, where student queries are often diverse and unstructured [1]. The introduction of machine learning techniques marked a significant advancement in chatbot development. Algorithms such as Naïve Bayes, Support Vector Machines, and decision trees enabled systems to learn patterns from labeled data, improving their ability to classify queries and generate appropriate responses [2]. These approaches reduced reliance on manual rule creation and provided greater flexibility in handling varied input.

However, traditional machine learning models required extensive feature engineering and struggled to capture semantic relationships in natural language, particularly in short and informal queries commonly observed in student interactions [3]. The emergence of deep learning has revolutionized Natural Language Processing by enabling automatic feature extraction and contextual understanding of text data. Recurrent Neural Networks, particularly Long Short-Term Memory models, have demonstrated strong performance in sequential data processing by retaining relevant information across word sequences [4]. Bidirectional LSTM models further enhance this capability by processing input sequences in both forward and backward directions, allowing for a more comprehensive understanding of contextual relationships within text [5]. These models have been widely adopted in chatbot systems for intent classification due to their ability to handle complex linguistic patterns. Attention mechanisms have been introduced to improve the performance of deep learning models by enabling selective focus on important parts of the input sequence. In chatbot applications, attention allows the model to assign higher importance to keywords that are critical for determining user intent, thereby improving classification accuracy and interpretability [6].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Studies have shown that attention-based models significantly enhance performance in multi-class classification tasks, particularly in cases where input queries contain overlapping vocabulary across different categories [7]. Despite these advancements, several challenges remain in the development of chatbot systems for educational environments. One of the primary challenges is data scarcity. Many studies rely on large and well-balanced datasets, which are not always available in real-world institutional settings. Educational datasets are often small, domain-specific, and imbalanced, leading to reduced model generalization and biased predictions toward dominant classes [8]. Class imbalance is particularly problematic in student support systems, where certain query categories such as fees and examinations occur more frequently than others [9]. Evaluation methodologies in chatbot research have also evolved, with increasing emphasis on comprehensive performance analysis. While early studies primarily relied on accuracy as a performance metric, recent research highlights the importance of precision, recall, F1-score, and confusion matrix analysis for understanding class-wise behaviour and misclassification patterns [10]. These metrics provide deeper insights into model performance and help identify biases and weaknesses in classification systems [11]. Another important aspect highlighted in the literature is the need for realistic evaluation of chatbot systems. Many studies report high accuracy using curated datasets, but these results often do not translate to real-world scenarios where data is noisy and limited [12].

Therefore, there is a growing emphasis on evaluating chatbot systems under data-constrained conditions to better understand their practical limitations [13]. Furthermore, ethical considerations and user trust play a crucial role in the adoption of chatbot systems. Students are more likely to engage with chatbots when responses are accurate, consistent, and transparent. However, repeated misclassifications can reduce user confidence and limit system effectiveness [14]. Researchers emphasize that chatbot systems should complement human support rather than replace it entirely [15]. In summary, existing literature demonstrates significant progress in chatbot technologies, particularly with the adoption of deep learning and attention mechanisms. However, challenges related to dataset limitations, class imbalance, evaluation transparency, and real-world applicability remain unresolved [16–25]. These gaps provide the foundation for the present research.

## III. RESEARCH METHODOLOGY

### A. *Dataset Description*

The dataset utilized in the present study forms the empirical foundation for developing and evaluating the proposed Artificial Intelligence–based student support chatbot system. It consists of textual queries generated by students in academic environments, reflecting real-world interactions between learners and institutional support services. These queries represent common information-seeking behaviours related to academic and administrative processes, including admissions, examinations, faculty information, fee structures, hostel facilities, and library services. The dataset is structured for supervised learning, where each query is manually labeled into one of six predefined intent categories: admission, examination, faculty, fees, hostel, and library.

This categorical labeling enables the chatbot model to learn associations between linguistic patterns and corresponding support domains. A defining characteristic of the dataset is its unstructured and informal nature. Student queries vary significantly in length, grammatical correctness, and vocabulary usage. Many queries are short, fragmented, or keyword-based, such as “exam date” or “fees last date,” lacking complete sentence structure. Additionally, the dataset includes spelling variations, abbreviations, and context-dependent expressions, which introduce complexity in Natural Language Processing tasks. This variability enhances the practical relevance of the dataset, as it closely resembles real-world chatbot usage conditions. Another important aspect of the dataset is its limited size, which reflects realistic constraints often encountered in educational institutions.

Unlike large-scale public datasets, institutional data is typically restricted due to privacy concerns and limited historical records. Furthermore, the dataset exhibits class imbalance, where certain categories such as fees and examination contain a higher number of queries compared to less frequent categories like hostel and library. This imbalance poses challenges for model training, as it may lead to biased predictions toward dominant classes.

For experimental evaluation, the dataset is divided into training and testing subsets to ensure unbiased performance assessment. The training data is used to develop the model, while the testing data evaluates its generalization capability on unseen queries. Overall, the dataset provides a realistic, challenging, and domain-specific foundation for analyzing chatbot performance under data-constrained conditions.

### B. *Overall System Architecture*

The overall system architecture of the proposed Artificial Intelligence–based student support chatbot is designed as a structured, modular, and sequential framework that facilitates efficient processing of student queries from initial input to final intent classification.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The architecture follows a standard Natural Language Processing (NLP) pipeline, ensuring clarity, scalability, and ease of implementation while maintaining compatibility with real-world deployment requirements in educational institutions. Each component of the architecture performs a specific function, collectively contributing to accurate understanding and classification of student queries. The system begins with the user interaction layer, where students submit textual queries through a chatbot interface. These queries represent natural language inputs that may vary in length, structure, and grammatical correctness.

Since student queries are often informal and unstructured, the system is designed to handle diverse linguistic patterns without imposing strict formatting constraints. This flexibility ensures that the chatbot can operate effectively in realistic usage scenarios. Following input acquisition, the query is passed to the data preprocessing module. This module performs essential operations such as text cleaning, normalization, tokenization, and removal of irrelevant elements like punctuation and stop words. The objective of this stage is to transform raw textual input into a standardized and machine-readable format. Preprocessing significantly enhances the quality of input data, enabling the model to focus on meaningful linguistic features rather than noise. The processed text is then forwarded to the feature representation layer, where tokenized words are converted into numerical form using embedding techniques.

Word embeddings generate dense vector representations that capture semantic relationships between words, allowing the model to understand contextual similarities and variations in student queries. This transformation is crucial, as deep learning models require numerical input for effective computation. The core of the architecture is the deep learning–based intent classification module, implemented using a Bidirectional Long Short-Term Memory (BiLSTM) network integrated with an attention mechanism. The BiLSTM processes input sequences in both forward and backward directions, capturing contextual dependencies within the query. The attention mechanism enhances this process by assigning importance weights to relevant words, enabling the model to focus on key terms that determine the intent of the query. Finally, the output layer employs a softmax activation function to generate probability distributions across the six predefined intent categories. The category with the highest probability is selected as the predicted intent. This output can be linked to predefined responses or knowledge bases for generating appropriate replies. Overall, the modular design of the system architecture ensures robustness, flexibility, and scalability, making it suitable for deployment in AI-driven student support systems.

### *C. Performance Evaluation Metrics*

The evaluation of the proposed Artificial Intelligence–based chatbot system is conducted using a comprehensive set of performance metrics designed to provide a detailed and reliable assessment of classification effectiveness. In multi-class intent classification tasks, particularly under conditions of limited and imbalanced datasets, reliance on a single evaluation metric such as accuracy can be misleading. Therefore, this study adopts a multi-metric evaluation framework that includes accuracy, precision, recall, F1-score, and confusion matrix analysis to ensure a holistic understanding of model performance. Accuracy is used as the primary metric to measure the overall correctness of the chatbot system.

It represents the proportion of student queries that are correctly classified into their respective intent categories. While accuracy provides a general overview of model performance, it does not account for the distribution of predictions across individual classes. In imbalanced datasets, a model may achieve relatively acceptable accuracy by correctly predicting dominant classes while failing to recognize minority categories. Hence, additional metrics are necessary to capture class-wise behaviour. Precision is employed to evaluate the reliability of the model's predictions. It measures the proportion of correctly predicted instances within a particular class relative to all instances predicted for that class.

High precision indicates that the model makes fewer false positive errors, which is essential in student support systems where incorrect classification can lead to irrelevant or misleading responses. Recall, on the other hand, measures the model's ability to correctly identify all relevant instances of a given class. High recall ensures that the chatbot does not miss important queries belonging to a specific category, which is critical for providing comprehensive support. The F1-score is used as a balanced metric that combines precision and recall through their harmonic mean. It provides a single value that reflects both the correctness and completeness of predictions, making it particularly useful in scenarios with uneven class distribution. A low F1-score indicates poor model performance in handling both false positives and false negatives. In addition to these numerical metrics, confusion matrix analysis is utilized to visualize the distribution of correct and incorrect predictions across all intent categories. The confusion matrix provides deeper insight into misclassification patterns, highlighting which categories are frequently confused and identifying biases toward dominant classes. Furthermore, training and validation performance curves are analysed to assess learning stability, convergence behaviour, and generalization capability of the model. Collectively, these evaluation metrics ensure a transparent, detailed, and scientifically rigorous assessment of the chatbot system's performance.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

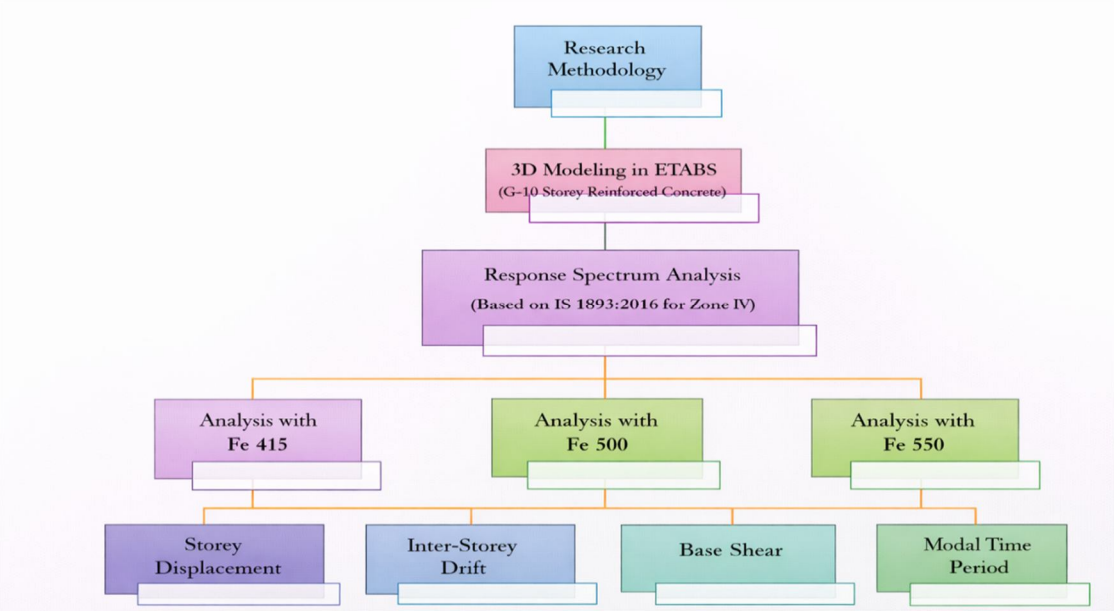



Figure 2: Illustrates the complete workflow of the chatbot system, including input processing, preprocessing, feature extraction, deep learning classification, and output generation.

## IV. RESULTS AND DISCUSSION

### *A. Overall Performance Analysis*

The overall performance of the proposed chatbot system was evaluated using a test dataset consisting of 15 student queries distributed across six intent categories. The model achieved an overall classification accuracy of 20 percent, which reflects the proportion of correctly classified queries. Although this accuracy appears low, it must be interpreted within the context of a highly limited and imbalanced dataset. The results indicate that the model is capable of learning dominant patterns present in the data but struggles to generalize across all intent categories. The uneven distribution of queries across classes significantly influenced model behaviour, leading to biased predictions. This outcome highlights the importance of dataset size and balance in determining the effectiveness of deep learning–based chatbot systems, particularly in real-world educational environments.

### *B. Classification Report Analysis*

The classification report provides a detailed breakdown of precision, recall, and F1-score values for each intent category. The results reveal that the fees category achieved a recall of 1.00, indicating that all fee-related queries in the test dataset were correctly identified. However, the precision for this category remained low at approximately 0.27, suggesting that a substantial number of queries from other categories were incorrectly classified as fees. In contrast, all other categories—admission, examination, faculty, hostel, and library—recorded zero precision, recall, and F1-score values. This indicates that the model failed to correctly classify queries belonging to these categories. The classification report clearly demonstrates the impact of class imbalance and limited training data on model performance, resulting in skewed predictions toward the dominant class.

Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| admission | 0.00 | 0.00 | 0.00 | 2 |
| exam | 0.00 | 0.00 | 0.00 | 2 |
| faculty | 0.00 | 0.00 | 0.00 | 3 |
| fees | 0.27 | 1.00 | 0.43 | 3 |
| hostel | 0.00 | 0.00 | 0.00 | 3 |
| library | 0.00 | 0.00 | 0.00 | 2 |
| accuracy | | | 0.20 | 15 |
| macro avg | 0.05 | 0.17 | 0.07 | 15 |
| weighted avg | 0.05 | 0.20 | 0.09 | 15 |

Figure 3: Illustrates the classification report showing precision, recall, and F1-score values.

### C. Confusion Matrix Analysis

The confusion matrix provides a visual representation of prediction outcomes across all intent categories, offering deeper insight into misclassification patterns. The analysis reveals a strong concentration of predictions within the fees category, confirming the model's bias toward this dominant class. Queries belonging to other categories were frequently misclassified as fees, indicating poor discrimination capability between different intents. The absence of correct predictions for minority classes further emphasizes the limitations of the model in handling imbalanced datasets. Additionally, the confusion matrix highlights overlapping linguistic patterns among categories, which contribute to classification errors. This analysis underscores the need for improved dataset balance and enhanced feature representation techniques to achieve more accurate and unbiased classification results.

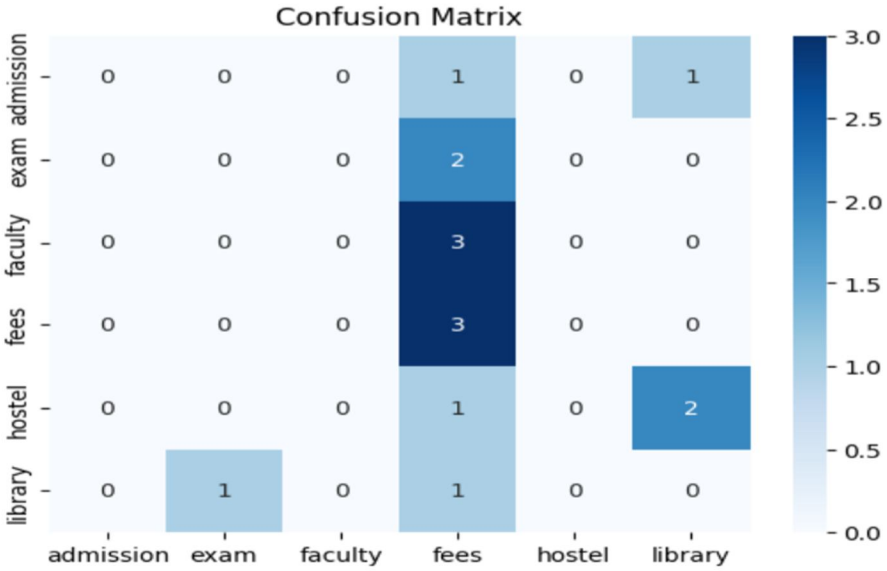



Figure 4: shows the confusion matrix representing prediction outcomes.

### D. Training and Validation Performance

The training and validation performance of the model were analysed using accuracy and loss curves across multiple epochs. The results indicate that the model exhibits relatively stable convergence behaviour during training, with gradual improvement in accuracy and reduction in loss values.

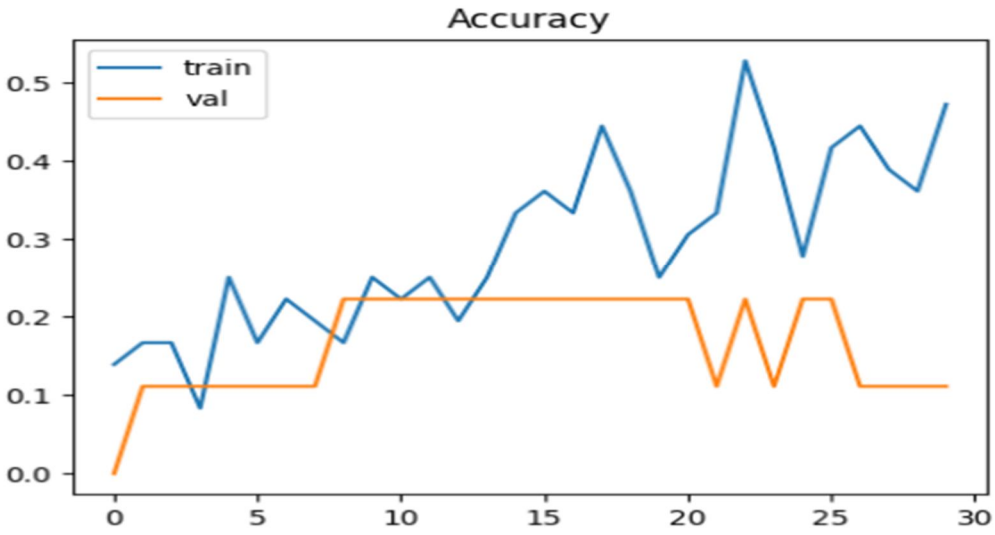



Figure 5: Training and validation accuracy curve of the chatbot model.

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 14 Issue V May 2026- Available at www.ijraset.com*

However, due to the limited size of the dataset, the model does not demonstrate significant generalization capability. The alignment between training and validation curves suggests that overfitting is not a major concern in this case, as the model does not memorize the training data excessively. Instead, the primary limitation lies in insufficient data diversity, which restricts the model's ability to learn robust patterns. These observations highlight the importance of dataset expansion and diversity in improving model performance.

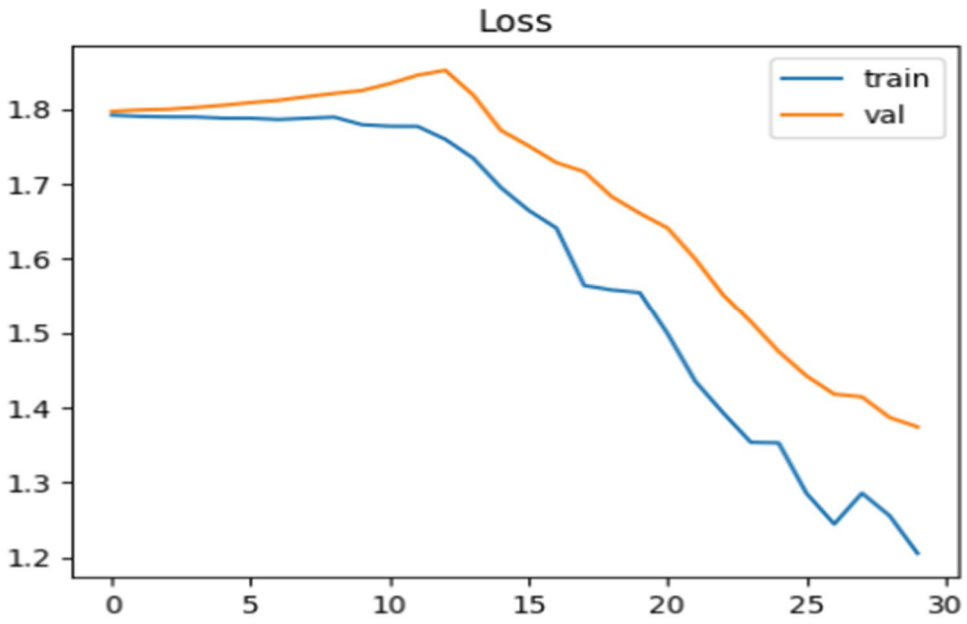



Figure 6: Training and validation loss curve of the chatbot model.

*E. Discussion of Findings*

The findings of this study provide important insights into the practical challenges of deploying AI-based chatbot systems in educational environments. The results demonstrate that while deep learning architectures such as BiLSTM with attention are capable of capturing contextual information, their effectiveness is highly dependent on data quality and availability. The strong bias toward the fees category reflects the influence of class imbalance, while the poor performance in other categories indicates insufficient training samples and overlapping vocabulary. These limitations are consistent with observations reported in existing literature on educational chatbot systems. The study also highlights the importance of comprehensive evaluation metrics, as reliance on accuracy alone would not reveal class-wise performance issues. Overall, the discussion emphasizes that successful implementation of chatbot systems requires not only advanced algorithms but also well-structured, balanced, and sufficiently large datasets.

## V. CONCLUSION

This study presented the design, development, and evaluation of an Artificial Intelligence–based chatbot system for student support, utilizing Natural Language Processing techniques and a deep learning architecture based on Bidirectional Long Short-Term Memory (BiLSTM) networks with an attention mechanism. The primary objective of the research was to develop an automated system capable of classifying student queries into predefined academic and administrative categories, thereby improving accessibility, efficiency, and consistency in institutional support services. By focusing on realistic conditions, including limited dataset size and class imbalance, the study provides a transparent and practical assessment of chatbot performance in educational environments. The experimental results revealed that the proposed model achieved an overall classification accuracy of 20 percent, indicating that while the system is capable of learning dominant linguistic patterns, its performance is significantly constrained by data-related limitations. The strong bias toward frequently occurring categories, particularly the fees class, highlights the impact of imbalanced datasets on model predictions. Additionally, the inability of the model to effectively classify minority categories such as hostel and library queries underscores the challenges associated with insufficient training samples and overlapping vocabulary across different intent classes.

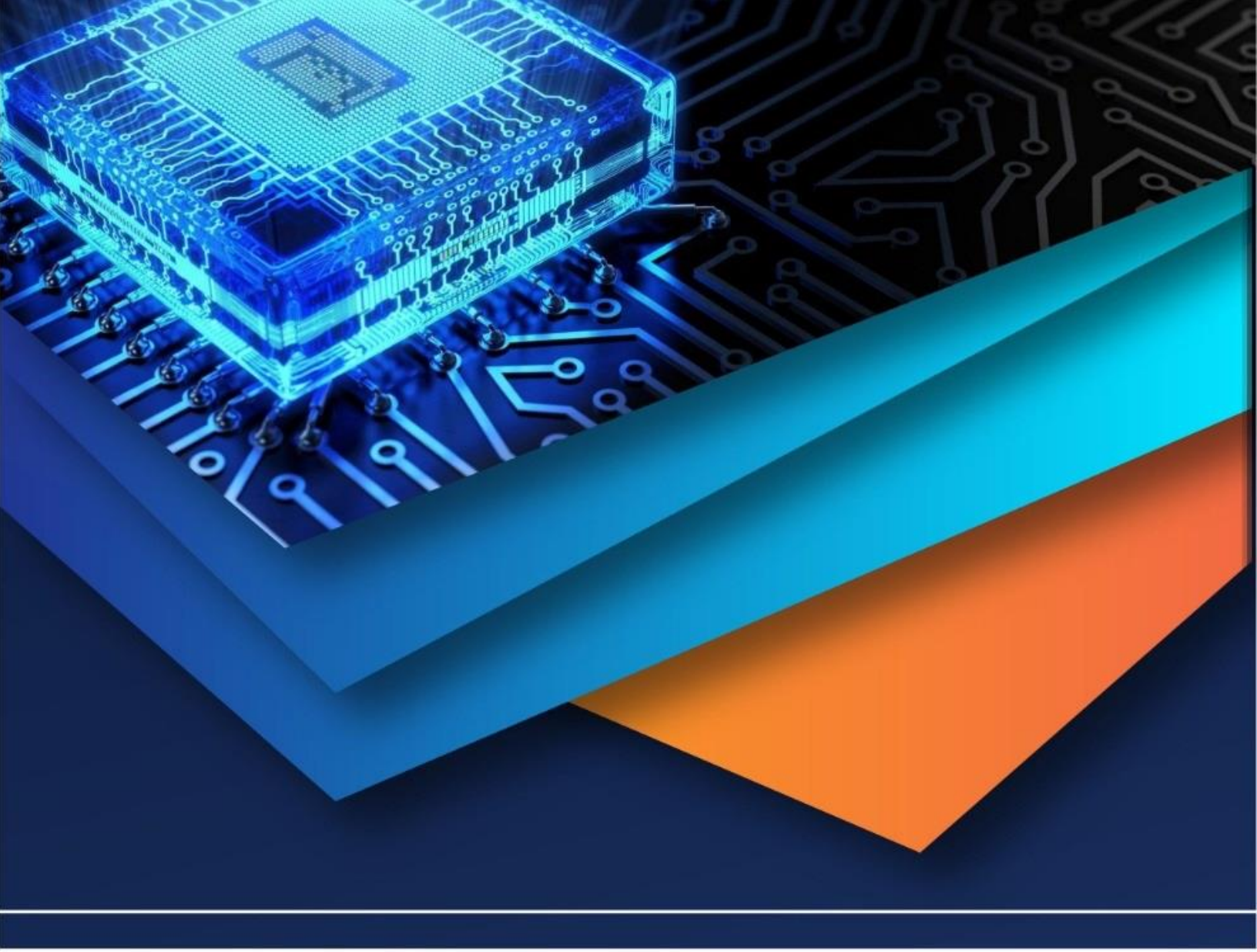
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

These findings emphasize that the effectiveness of deep learning models is highly dependent on the quality, diversity, and balance of the training data. Despite these limitations, the study demonstrates the potential of deep learning–based chatbot systems in automating student support services. The use of BiLSTM and attention mechanisms enables the model to capture contextual relationships within textual data, providing a strong foundation for further improvements. Moreover, the comprehensive evaluation approach adopted in this research, including precision, recall, F1-score, and confusion matrix analysis, contributes to a deeper understanding of system behaviour and performance limitations. Future research should focus on expanding the dataset to include a larger and more diverse set of student queries, thereby improving model generalization. Techniques such as data augmentation, class balancing, and transfer learning may also be explored to address imbalance-related issues. Additionally, the integration of advanced architectures such as transformer-based models could enhance intent classification accuracy. Incorporating real-time deployment and user feedback mechanisms would further improve system adaptability and practical usability. Overall, this study provides a valuable foundation for the development of intelligent, scalable, and reliable chatbot systems for student support in modern educational environments.

## REFERENCES

[1] Assayed, S. K., et al. (2023). Educational applications of AI chatbots in higher education. SSRN Electronic Journal. (SSRN)
[2] Wu, X. Y. (2025). Designing artificial intelligence chatbots for self-regulated learning: A systematic review. Interactive Learning Environments. (Taylor & Francis Online)
[3] Shrivastava, N., & Tewari, P. (2025). Natural language processing for conversational AI: Chatbots and virtual assistants. IEEE Conference Proceedings. (ResearchGate)
[4] Debets, T., et al. (2025). Chatbots in education: A systematic review of objectives and outcomes. Computers & Education. (ScienceDirect)
[5] Davar, N. F. (2025). AI chatbots in education: Challenges and opportunities. Information Journal. (MDPI)
[6] Kooli, C. (2023). Chatbots in education and research: A critical analysis. Sustainability, 15(7), 5614. (MDPI)
[7] Labadze, L., et al. (2023). Role of AI chatbots in education: A systematic literature review. International Journal of Educational Technology. (Springer Nature Link)
[8] Badger, M. (2026). The impact of generative AI chatbots on student learning outcomes. International Journal of Technology in Education. (ijte.net)
[9] Abdelmoiz, M. M. A. (2024). Developing an educational chatbot using NLP techniques. Natural Science Publishing. (naturalspublishing.com)
[10] Bhardwaj, N. (2024). Natural language processing in AI chatbot systems. SSRN Electronic Journal. (SSRN)
[11] Chakraborty, C. (2023). Overview of chatbot systems with emphasis on AI applications. Journal of Medical Systems. (PMC)
[12] Aditya, D. (2026). Conversational AI for student support systems in education. Cogent Education. (Taylor & Francis Online)
[13] Alsaawi, A. (2025). Challenges and opportunities of AI chatbots in language learning. International Journal of Learning, Teaching and Educational Research. (IJLTER)
[14] McGrath, C. (2024). Generative AI chatbots in higher education: A systematic review. Higher Education Research. (Diva Portal)
[15] Pan, J., Zhao, Z., & Han, D. (2025). Academic performance prediction using machine learning approaches: A survey. Journal of Educational Data Science.
[16] Zhang, T., & Cheng, S. (2025). Artificial intelligence for personalized learning in higher education. Applied Sciences, 15(18).
[17] Ahmed, E. (2024). Student performance prediction using machine learning algorithms. Wiley Interdisciplinary Reviews: Data Mining.
[18] Du Plooy, E. (2024). Personalized adaptive learning in higher education. Heliyon Education Sciences.
[19] Merino-Campos, C. (2025). Impact of AI on personalized learning in higher education. Trends in Higher Education.
[20] Timovski, R. (2025). Predictive analytics in higher education using AI models. Knowledge Journal.
[21] Yuan, J., et al. (2024). Machine learning for online learning performance prediction. arXiv preprint.
[22] Bulut, O., et al. (2024). Artificial intelligence in educational measurement: Opportunities and challenges. arXiv preprint.
[23] Tseng, C. H., et al. (2025). Personalized programming education using machine learning techniques. arXiv preprint.
[24] Wang, J., et al. (2025). Machine learning approaches for student performance prediction. PLOS ONE.
[25] Silva, G. (2024). AI-driven personalized learning systems in modern education. Journal of Innovative Education and Teaching.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)