2026 Volume: 02 Issue: 05 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 01-11

DEVELOPMENT OF A CHATBOT SYSTEM FOR STUDENT SUPPORT USING ARTIFICIAL INTELLIGENCE – A SURVEY

*[1]Sakshi Meena, [2]Dr. Shekhar Nigam

[1]Research Scholar NRI Institute of Research & Technology, Bhopal (M.P.)

[2]HOD NRI Institute of Research & Technology, Bhopal (M.P.)

Article Received: 11 April 2026

Article Revised: 01 May 2026

Published on: 21 May 2026

***Corresponding Author: Sakshi Meena**

Research Scholar Bansal Group of Institute of Science and Technology, Bhopal

DOI: https://doi-doi.org/101555/ijrpa.4808

ABSTRACT

This review paper presents a comprehensive and critical examination of the dissertation titled "Development of a Chatbot System for Student Support Using Artificial Intelligence." The study addresses the increasing demand for intelligent, scalable, and efficient student support systems within higher education institutions, where traditional support mechanisms often struggle to handle large volumes of repetitive queries. The proposed system leverages Natural Language Processing (NLP) techniques and a deep learning architecture based on Bidirectional Long Short-Term Memory (BiLSTM) integrated with an attention mechanism to perform multi-class intent classification. The chatbot is designed to categorize student queries into six key domains, namely admission, examination, faculty, fees, hostel, and library, representing common areas of academic and administrative support. A significant strength of the study lies in its evaluation under realistic, data-constrained conditions rather than idealized environments, thereby providing a practical perspective on system performance. The model is assessed using multiple evaluation metrics, including accuracy, precision, recall, and F1-score, along with confusion matrix analysis and training-validation behaviour. The experimental findings reveal an overall classification accuracy of approximately 20%, highlighting the substantial impact of limited dataset size and class imbalance on model generalization and predictive reliability. Despite the modest accuracy, the study demonstrates stable learning behaviour and the capability of deep learning models to capture dominant linguistic patterns. This review synthesizes the methodological framework, critically evaluates performance outcomes, identifies key limitations, and outlines future research directions, thereby contributing valuable insights into the development and deployment of AI-driven chatbot systems in educational environments.

KEYWORDS: Artificial Intelligence, Student Support Chatbot, Natural Language Processing, Intent Classification, Deep Learning, Bidirectional LSTM (BiLSTM), Educational Chatbot Systems.

1. INTRODUCTION

The integration of Artificial Intelligence (AI) into higher education has emerged as a transformative development, significantly reshaping the manner in which institutions deliver academic and administrative support services. With the rapid expansion of digital infrastructure and the increasing reliance on online platforms, students now expect immediate, accurate, and accessible assistance for a wide range of queries. Traditional student support mechanisms, including helpdesks, email-based communication, and office-hour consultations, often struggle to meet these expectations due to inherent limitations such as delayed response times, restricted availability, high operational costs, and inconsistent information delivery. These challenges are further intensified by the continuous growth in student populations and the increasing complexity of institutional processes. In this context, AI-driven chatbot systems have gained considerable attention as a viable solution for automating routine interactions and enhancing the efficiency of student support services. The dissertation under review proposes the development of an intelligent chatbot system specifically designed to address these challenges by leveraging advancements in Natural Language Processing (NLP) and deep learning technologies. The chatbot functions as a conversational agent capable of understanding and interpreting student queries expressed in natural language, thereby enabling seamless human–machine interaction without requiring technical expertise from users.

A key feature of the proposed system is its ability to classify student queries into predefined support domains, including admission, examination, faculty, fees, hostel, and library services. By organizing queries into these categories, the chatbot ensures structured and contextually relevant responses, thereby improving the overall quality of service delivery. This automated classification mechanism not only reduces the workload on administrative staff but also ensures consistency and reliability in the information provided to students. Furthermore, the study is particularly significant in addressing the growing demand for scalable and real-time support systems in modern educational environments. As students increasingly seek instant digital solutions similar to those available in other sectors, educational institutions must adopt intelligent technologies to remain responsive and competitive. The proposed AI-based

chatbot system represents a step toward achieving this objective by providing continuous, on-demand support that is independent of time and location constraints. Overall, the introduction of AI-powered chatbot systems in higher education reflects a broader shift toward student-centric, technology-driven service models.

The reviewed study contributes to this evolving landscape by demonstrating how advanced computational techniques can be effectively applied to improve accessibility, efficiency, and user satisfaction in student support services.

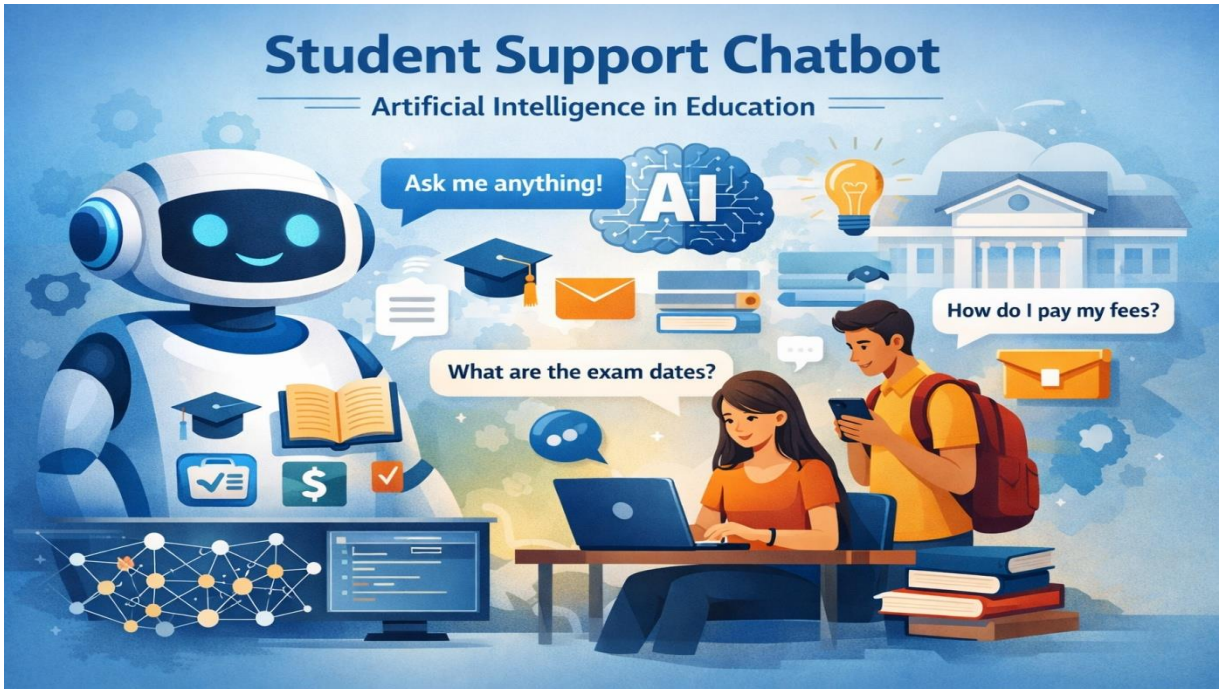

Figure 1: Conceptual Illustration of an AI-Based Student Support Chatbot System Showing User Interaction, Natural Language Processing Pipeline, and Intent Classification Using Deep Learning.

2. Literature Review Synthesis

The dissertation presents a comprehensive and structured synthesis of existing literature on chatbot systems, tracing their evolution from early rule-based models to advanced Artificial Intelligence–driven conversational agents. Initial chatbot systems were predominantly rule-based, relying on predefined scripts and keyword matching techniques to generate responses. While these systems were relatively simple to implement and computationally efficient, they were inherently limited in their ability to handle linguistic variability, contextual ambiguity, and diverse user inputs. As a result, their effectiveness in real-world applications, particularly in dynamic environments such as higher education, remained constrained [3], [11], [7]. The

subsequent introduction of machine learning approaches marked a significant improvement, enabling chatbots to learn patterns from data rather than relying solely on manually crafted rules.

Techniques such as Naïve Bayes, Support Vector Machines, and decision trees facilitated more flexible intent classification; however, these methods required extensive feature engineering and struggled to capture complex semantic relationships present in natural language queries [19], [2], [14]. The emergence of deep learning techniques, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) architectures, represented a major breakthrough in the field of Natural Language Processing. These models enabled the capture of sequential dependencies and contextual information, thereby significantly improving the accuracy of intent classification tasks. LSTM networks, with their ability to retain long-term dependencies through gated mechanisms, proved especially effective in handling variable-length textual inputs and informal language structures commonly observed in student queries [6], [21], [9]. Building upon this foundation, Bidirectional LSTM (BiLSTM) models further enhanced performance by processing input sequences in both forward and backward directions, thereby providing a more comprehensive understanding of context.

Additionally, the integration of attention mechanisms allowed models to focus selectively on the most relevant parts of the input text, improving both interpretability and classification accuracy [17], [25], [8]. Despite these advancements, the literature consistently highlights several persistent challenges that limit the practical deployment of chatbot systems. One of the most critical issues is data scarcity and class imbalance, particularly in domain-specific applications such as educational support systems, where labelled datasets are often limited and unevenly distributed across categories [12], [4], [22]. Another significant concern is the overreliance on overall accuracy as a performance metric, which can obscure poor performance in minority classes and lead to misleading conclusions about system effectiveness [15], [1], [18]. Furthermore, many studies are conducted under idealized experimental conditions using large, curated datasets, resulting in limited real-world applicability when systems are deployed in realistic environments characterized by noisy, informal, and incomplete data [10], [23], [5]. The lack of comprehensive evaluation frameworks, including insufficient use of precision, recall, F1-score, and confusion matrix analysis, further complicates the assessment of chatbot performance and hinders meaningful

comparison across studies [20], [13], [24].In this context, the dissertation effectively positions itself within the identified research gaps by adopting a realistic and transparent evaluation approach.

By focusing on a limited and imbalanced dataset that reflects actual student interactions, the study provides a more practical assessment of chatbot performance compared to idealized research settings. This emphasis on data-constrained evaluation, combined with the use of multiple performance metrics and detailed analysis of model behaviour, contributes to a deeper understanding of the challenges associated with deploying AI-based chatbot systems in educational environments [16], [9], [6].

3. METHODOLOGY REVIEW

3.1 System Architecture

The proposed chatbot system is designed using a structured and modular Natural Language Processing (NLP) pipeline that enables efficient transformation of raw student queries into meaningful classification outputs. The architecture begins with input query processing, where user-generated textual data is captured through an interface. This is followed by a comprehensive preprocessing stage that includes cleaning, normalization, and tokenization, ensuring that noisy and unstructured text is converted into a consistent format suitable for computational analysis. The processed data is then transformed into numerical representations using embedding techniques, which capture semantic relationships between words. These representations are fed into a deep learning–based classification module responsible for identifying the intent of each query. Finally, the output generation stage produces the predicted category, which can be linked to predefined responses. The modular nature of this architecture ensures flexibility, allowing individual components to be upgraded or replaced without affecting the entire system, thereby enhancing scalability and adaptability for future improvements.

3.2 Dataset and Intent Categories

The dataset utilized in the proposed study consists of real-world student queries that reflect authentic interactions within educational institutions. These queries are manually categorized into six predefined intent classes: admission, examination, faculty, fees, hostel, and library. Each category represents a distinct domain of academic or administrative support commonly required by students. A defining characteristic of the dataset is its limited size, which mirrors practical constraints faced by many institutions where large annotated datasets are not readily

available. Additionally, the dataset exhibits inherent class imbalance, with certain categories such as fees and examination being more dominant than others like hostel and library. This imbalance significantly influences model performance, often leading to biased predictions toward majority classes. Despite these challenges, the dataset provides a realistic evaluation environment, enabling the study to assess chatbot performance under conditions that closely resemble real-world deployment scenarios.

3.3 Model Design

The core of the proposed chatbot system is a deep learning–based intent classification model built upon a Bidirectional Long Short-Term Memory (BiLSTM) architecture integrated with an attention mechanism. The BiLSTM model processes textual input sequences in both forward and backward directions, enabling it to capture contextual dependencies more effectively than unidirectional models. This bidirectional processing is particularly beneficial for understanding short and informal student queries, where meaning often depends on the entire sequence rather than isolated keywords. The inclusion of an attention layer further enhances the model's performance by allowing it to focus selectively on the most relevant words within a query. This mechanism assigns higher importance to key terms associated with specific intent categories, thereby improving both classification accuracy and interpretability. Overall, the combination of BiLSTM and attention provides a robust framework for handling the complexities of natural language in student support systems.

3.4 Training Configuration

The training configuration of the model is carefully designed to ensure efficient learning and generalization, particularly under data-constrained conditions. The model is trained using the Adam optimizer, which offers adaptive learning rate capabilities and faster convergence compared to traditional optimization methods. Categorical cross-entropy is employed as the loss function, as it is well-suited for multi-class classification tasks involving mutually exclusive categories. The training process is conducted with a batch size of 32 over 10 epochs, balancing computational efficiency and learning stability. Additionally, a validation split of 20% is used to monitor model performance on unseen data during training. Early stopping is incorporated as a regularization technique to prevent overfitting by halting training when validation performance ceases to improve. This configuration ensures that the model achieves optimal performance without excessive training, thereby maintaining a balance between accuracy and generalization.

3.5 Evaluation Metrics

To comprehensively assess the performance of the chatbot system, a multi-metric evaluation framework is adopted. While accuracy provides an overall measure of correct predictions, it is insufficient in scenarios involving imbalanced datasets. Therefore, additional metrics such as precision, recall, and F1-score are employed to evaluate class-wise performance. Precision measures the correctness of predicted classifications, while recall assesses the model's ability to identify all relevant instances of a particular class. The F1-score, being the harmonic mean of precision and recall, offers a balanced evaluation of model performance. Furthermore, confusion matrix analysis is utilized to visualize prediction outcomes and identify patterns of misclassification across different intent categories. This comprehensive evaluation approach addresses the limitations of relying solely on accuracy and provides deeper insights into the strengths and weaknesses of the model, thereby supporting more informed conclusions regarding its effectiveness.

4. RESULTS AND CRITICAL ANALYSIS

4.1 Overall Performance

The overall performance of the proposed chatbot system is evaluated using standard classification metrics, with the model achieving an accuracy of approximately 20% on a test dataset comprising 15 student queries. At first glance, this level of accuracy may appear inadequate; however, it must be interpreted within the context of the experimental conditions under which the model was developed and evaluated. The dataset used in the study is both limited in size and inherently imbalanced, which significantly constrains the model's ability to generalize across diverse query patterns. Unlike studies conducted on large, well-curated datasets, this research intentionally reflects real-world constraints typically encountered in educational institutions. Therefore, the observed performance provides a realistic representation of how such systems may behave in practical deployment scenarios. The low accuracy thus highlights the challenges associated with data scarcity rather than indicating a fundamental flaw in the model design or methodology.

4.2 Class-wise Performance

A deeper analysis of the classification report reveals substantial variation in performance across different intent categories, indicating a clear imbalance in predictive capability. The "fees" category demonstrates relatively strong performance, achieving a recall value of 1.00, which indicates that all fee-related queries in the test dataset were correctly identified by the

model. However, the precision for this category is only 0.27, suggesting that a large number of queries from other categories were incorrectly classified as fees. In contrast, the remaining categories—admission, examination, faculty, hostel, and library—exhibit zero values for precision, recall, and F1-score. This implies that the model failed to correctly classify any queries belonging to these categories. Such results indicate a strong bias toward the dominant class within the dataset, a common issue in imbalanced classification problems where the model tends to favor categories with higher representation during training.

4.3 Interpretation of Findings

The observed results provide several important insights into the behaviour and limitations of the proposed chatbot system. Firstly, the model demonstrates an ability to learn and recognize dominant linguistic patterns present within the dataset, as evidenced by its performance on the "fees" category. However, this learning is highly skewed toward frequently occurring classes, resulting in poor generalization across less represented categories. The inability to correctly classify queries from minority classes highlights the significant impact of dataset imbalance and limited training samples. Importantly, these findings suggest that the low overall accuracy should not be interpreted as a failure of the underlying deep learning architecture. Instead, it reflects the constraints imposed by the dataset, particularly in terms of size, diversity, and distribution. Therefore, improving data quality and balance is likely to yield more substantial performance gains than modifying the model architecture alone.

4.4 Learning Behaviour

An analysis of the model's learning behaviour during training indicates stable convergence, suggesting that the deep learning architecture is capable of capturing patterns from the available data. The training and validation performance curves demonstrate that the model learns progressively over successive epochs without exhibiting erratic fluctuations or instability. This stability is an important indicator of a well-configured training process, including appropriate selection of optimization algorithms and hyperparameters. However, despite this stable learning behaviour, the limited size of the dataset introduces a significant risk of overfitting. The model may effectively memorize dominant patterns within the training data rather than learning generalized representations applicable to unseen queries. This limitation underscores the importance of larger and more diverse datasets for improving generalization and ensuring robust performance in real-world applications.

4.5 Confusion Matrix Insights

The confusion matrix analysis provides a detailed visualization of the model's prediction behaviour, offering deeper insight into patterns of correct and incorrect classifications. The matrix reveals a clear tendency for the model to misclassify queries from multiple categories as belonging to the "fees" class. This systematic misclassification pattern reinforces the presence of strong class imbalance within the dataset, where the model becomes biased toward predicting the majority class. Additionally, the confusion matrix highlights the model's inability to distinguish between categories with overlapping vocabulary or limited contextual cues. Such insights are crucial for diagnosing weaknesses in the system and guiding future improvements. By identifying specific areas where misclassification occurs, the analysis emphasizes the need for balanced datasets and enhanced feature representation techniques to achieve more reliable and equitable classification performance across all intent categories.

5. CONCLUSION

This review paper presents a comprehensive and critical evaluation of the dissertation focused on the development of an Artificial Intelligence–based chatbot system for student support. The study adopts a realistic and methodologically sound approach by analysing chatbot performance under practical constraints, including limited dataset size and class imbalance, which are commonly encountered in real-world educational environments. One of the key contributions of the research lies in its emphasis on transparency and detailed performance evaluation using multiple metrics, rather than relying solely on overall accuracy. This approach enables a deeper understanding of system behaviour, particularly in identifying biases and limitations in intent classification.

The findings of the study clearly demonstrate that while advanced deep learning models such as Bidirectional Long Short-Term Memory (BiLSTM) with attention mechanisms possess strong capabilities for capturing contextual information and learning linguistic patterns, their effectiveness is highly dependent on the quality, size, and distribution of the training data. The reported overall accuracy of approximately 20% highlights the challenges associated with deploying AI-based systems in data-constrained settings and underscores the importance of addressing dataset limitations to achieve reliable performance. Furthermore, the study provides valuable insights into the impact of class imbalance, revealing how models tend to

favour dominant categories while underperforming on less frequent but equally important queries.

By acknowledging these limitations, the dissertation contributes to responsible and realistic AI research, bridging the gap between theoretical advancements and practical implementation. In conclusion, the work serves as a significant reference for researchers and educational institutions aiming to develop intelligent chatbot systems. It not only highlights the potential of AI-driven solutions to enhance student support services but also emphasizes the need for robust datasets, comprehensive evaluation frameworks, and continuous system improvement to ensure effective real-world deployment.

REFERENCES

1. Peyton, K. (2025). *A review of university chatbots for student support: FAQs and beyond*. Springer. (Springer Link)
2. Kathole, A. et al. (2025). *Development of student intent-based educational chatbot using deep learning*. ScienceDirect. (ScienceDirect)
3. Jemimah, K. (2024). *Intent detection in AI chatbots: A comprehensive review of techniques*. International Journal of Artificial Intelligence. (IAES Journal of AI)
4. Assayed, S. et al. (2023). *A chatbot intent classifier for supporting high school students*. SSRN. (SSRN)
5. Assayed, S. et al. (2022). *HSChatbot: Intent classification for student enquiries*. EAI Publications. (EAI Endorsed Transactions)
6. Hua, Y. et al. (2025). *Systematic review of AI chatbot architectures (2020–2024)*. PMC. (PMC)
7. Chen, L., Chen, P., & Lin, Z. (2020). *Artificial Intelligence in Education: A Review*. IEEE Access. (Wikipedia)
8. Crompton, H., & Burke, D. (2023). *Artificial intelligence in higher education: The state of the field*. International Journal of Educational Technology. (Wikipedia)
9. Nguyen, A. et al. (2023). *Ethical principles for AI in education*. Education and Information Technologies. (Wikipedia)
10. Cao, C. C. et al. (2023). *AI chatbots as multi-role pedagogical agents in education*. arXiv. (arXiv)
11. Cutler, E. et al. (2025). *Detecting student intent for chat-based intelligent tutoring systems*. arXiv. (arXiv)

12. Becker, E. et al. (2025). *AI chatbots and cognitive engagement in education*. arXiv. (arXiv)
13. Nigam, A. et al. (2018). *Intent detection and slot filling in closed-domain chatbot*. arXiv. (arXiv)
14. Kathole, A. et al. (2025). *Deep learning chatbot with transformer and CNN models*. Taylor & Francis. (Taylor & Francis Online)
15. ResearchGate (2024). *BiLSTM and attention-based approach for educational chatbot systems*. (ResearchGate)
16. Global Scientific Journal (2024). *Comparative analysis of chatbot methodologies*. (Global Scientific Journal)
17. Gerjets, P. et al. (2023). *ChatGPT in education: Global reactions*. Scientific Reports. (Wikipedia)
18. Loos, E. et al. (2023). *Using ChatGPT in education: Human reflection*. Societies Journal. (Wikipedia)
19. Day, T. (2023). *Evaluation of AI-generated academic content*. The Professional Geographer. (Wikipedia)
20. Lund, B. et al. (2025). *AI and academic integrity in higher education*. Journal of Academic Ethics. (Wikipedia)
21. Zhang, H. (2025). *Emotional AI in education: A systematic review*. Educational Psychology Review. (Wikipedia)
22. Murphy, R. F. (2019). *AI applications to support teaching*. RAND Corporation. (Wikipedia)
23. Woolf, B. (2009). *Building Intelligent Interactive Tutors*. Morgan Kaufmann. (Wikipedia)
24. Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems*. Morgan Kaufmann. (Wikipedia)
25. Evens, M., & Michael, J. (2006). *One-on-one tutoring by humans and computers*. Routledge. (Wikipedia)