IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153148

College Enquiry Chatbot Using Machine Learning: An Intelligent Conversational System for Academic Information Retrieval

Sonu Yadav[1], Vivek Chavan[2], Sufyan Hawaldar[3], Samruddhi Chinchavale[4], Prof. Ashwini Chavan[5]

Department of Computer Engineering, Bharat College of Engineering,

Bharat College of Engineering, Badlapur, Maharashtra, India[1-4]

Professor, Department of Computer Science and Engineering,

Bharat College of Engineering, Badlapur, Maharashtra, India[5]

Abstract: The **College Enquiry Chatbot** is an intelligent conversational system designed to automate the handling of student and visitor enquiries related to college admissions, courses, fees, faculty, facilities, and academic schedules. The system leverages **Machine Learning (ML), Natural Language Processing (NLP), and deep learning-based intent classification** to understand and respond to user queries in real time. Built using Python, NLTK, TensorFlow/Keras, and a Flask-based web interface, the chatbot delivers accurate, context-aware responses without human intervention.

Unlike traditional FAQ pages or static information portals, this system understands the intent behind user queries, handles variations in phrasing, and provides personalized responses. The system is evaluated against benchmark metrics including accuracy, F1 score, and response latency. Experimental results demonstrate a classification accuracy of over 92% on a domain-specific college enquiry dataset. Future work includes multilingual support, voice integration, and CRM system connectivity.

Keywords: Artificial Intelligence, Chatbot, Deep Learning, Intent Classification, Machine Learning, Natural Language Processing, Neural Network.

I. INTRODUCTION

With the growing volume of student enquiries received by educational institutions every year, it has become increasingly challenging for administrative staff to respond promptly and accurately to every query. Students and prospective applicants seek information on topics ranging from **admission procedures, eligibility criteria, fee structures, hostel availability, examination schedules, and placement records.** Delays in responses often lead to student dissatisfaction and missed opportunities.

To address this challenge, the College Enquiry Chatbot is developed as an AI-driven conversational agent capable of automatically answering common college-related questions. The system uses Machine Learning for intent recognition and NLP for language understanding, delivering instant and accurate responses through a user-friendly web interface. Unlike rule-based bots that rely on keyword matching, this chatbot understands the semantic meaning of queries, making it more adaptable and reliable for real-world deployment.

This project represents a significant step in the digital transformation of college administration, reducing the burden on staff and providing 24/7 availability of information to students, parents, and visitors.

II. BACKGROUND

The concept of chatbots originated in the 1960s with ELIZA, a rule-based system that simulated conversation using pattern matching. Over the following decades, chatbot technology evolved from simple command-response systems to sophisticated AI-powered agents capable of multi-turn conversations and contextual understanding.

DOI: 10.17148/IJARCCE.2026.153148

In the educational domain, AI-driven virtual assistants have been explored for student support, learning management, and administrative automation. However, most existing solutions are either domain-agnostic (and therefore inaccurate for specialized college queries) or rely on expensive cloud-based proprietary platforms that are difficult to customize.

The availability of open-source NLP libraries such as NLTK, spaCy, and deep learning frameworks like TensorFlow and PyTorch has enabled the development of highly accurate, domain-specific chatbots. Our College Enquiry Chatbot builds upon these tools, combining intent classification using neural networks with a structured knowledge base of college-specific information.

III. PROBLEM STATEMENT

Educational institutions handle thousands of enquiries annually through phone calls, emails, and in-person visits. The administrative staff is often overwhelmed by repetitive questions that follow predictable patterns — questions about fee payment deadlines, admission dates, available branches, scholarship criteria, and campus facilities.

Current solutions such as static FAQ pages, email response systems, and human helpdesks suffer from several limitations:

- Static pages are not conversational and require users to manually search for answers.
- Email-based responses involve delays of hours or even days.
- Human helpdesks are available only during working hours and are prone to human error.
- Existing general-purpose chatbots lack the domain-specific knowledge needed to answer college-specific queries accurately.

The College Enquiry Chatbot addresses these gaps by providing an always-available, intelligent, and conversational interface specifically trained on college domain data.

IV. OBJECTIVES

The primary objective of this project is to develop an intelligent, ML-powered chatbot capable of accurately understanding and responding to college-related enquiries in natural language. Specific objectives include:

- To build an intent classification model using neural networks that correctly identifies the purpose behind a user query.
- To create a comprehensive, updatable college-specific knowledge base covering admissions, fees, courses, faculty, and facilities.
- To develop a user-friendly web interface using Flask accessible from any device.
- To achieve a minimum intent classification accuracy of 90% on a held-out test set.
- To design the system for easy scalability, allowing new intents and responses to be added without full retraining.

V. CONTRIBUTION

This project makes the following key contributions to the field of educational AI and conversational systems:

- A domain-specific NLP pipeline tailored for college enquiry scenarios, handling informal language, abbreviations, and multi-phrasing of the same intent.
- A labeled college enquiry dataset containing over 500 unique query samples across 25+ intent categories, which can serve as a benchmark for future research.
- A lightweight, deployable chatbot architecture based on a feedforward neural network that achieves high accuracy without requiring GPU resources.
- An end-to-end Flask web application with a real-time chat interface, readily deployable in any college's existing IT infrastructure.
- A modular JSON-based knowledge base allowing non-technical staff to update responses without retraining.

VI. RELATED WORKS

Several AI-powered virtual assistants, such as **Google Assistant, Siri, and Alexa,** have been developed to assist users with voice-based commands and task automation. These systems leverage **Natural Language Processing (NLP)** to understand user queries but lack advanced document comprehension and domain-specific reasoning.

Traditional Optical Character Recognition (OCR) models and rule-based chatbot frameworks such as **AIML (Artificial Intelligence Markup Language)** have been widely used for FAQ automation. However, their accuracy varies with paraphrase variations and they cannot handle unseen query forms.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153148

Recent advancements in multi-modal AI models, such as **BERT and transformer-based architectures,** have improved intent understanding significantly. However, these models require substantial computational resources. Our College Enquiry Chatbot builds upon these foundations using **NLTK for text preprocessing, TensorFlow/Keras for intent classification, and Flask for deployment,** providing a lightweight yet accurate solution for educational domain automation.

VII. METHODOLOGY

The **College Enquiry Chatbot** follows a structured, multi-stage pipeline to ensure efficient processing of user queries. The methodology includes **data collection, NLP preprocessing, model training, and intelligent response generation.**
Step 1 – Knowledge Base Creation: A comprehensive set of 25+ intent categories is defined (admissions, fees, hostel, placement, etc.). For each intent, 15–25 sample utterances are written to capture natural language variation, and responses are mapped to each intent in a structured intents.json file.
Step 2 – Text Preprocessing: User input is cleaned using tokenization, lowercasing, stopword removal, and stemming via NLTK. A bag-of-words (BoW) feature vector is generated for each input sentence.
Step 3 – Intent Classification Model: A feedforward neural network is built using TensorFlow/Keras with two hidden layers (128 and 64 neurons, ReLU activation), dropout regularization (0.5), and softmax output for multi-class classification. Trained using Adam optimizer and categorical cross-entropy loss for 200 epochs.
Step 4 – Response Generation: Based on predicted intent, the system retrieves a response from the knowledge base. A confidence threshold of 70% is applied; inputs below this trigger a fallback prompting the user to rephrase.
Step 5 – Deployment: The chatbot is integrated into a Flask web application with a real-time AJAX-powered chat UI, accessible from any browser on desktop or mobile.

VIII. SYSTEM ARCHITECTURE

The system architecture of the College Enquiry Chatbot follows a three-layered modular design, ensuring efficient handling of user queries and providing accurate, context-aware AI responses. The architecture consists of the following key components:

1. Input Layer
User Interface (Flask Web App) – Accepts text input from students/users via a browser-based chat window.
Input Preprocessing Module – Tokenizes, lowercases, removes stopwords, applies NLTK stemming, and converts input to a BoW feature vector.
2. Core Processing Layer
Intent Classification Engine – Trained feedforward neural network predicts the intent category of the user input.
Confidence Threshold Filter – Validates model output confidence before response selection (threshold: 70%).
Knowledge Base (intents.json) – Structured JSON file containing all intent categories, sample utterances, and mapped responses.
Response Selector – Randomly selects one of the pre-defined responses for the predicted intent to add variety to replies.
3. Output Layer
Response Delivery – Returns selected response through the chat interface in real time.
Session Context Manager – Maintains conversation context within an active user session.
Fallback Handler – Provides a default response when model confidence falls below the threshold.

IX. IMPLEMENTATION FLOW

The implementation of the College Enquiry Chatbot follows a structured workflow optimized for ML-based conversational AI processing.
Chatbot Training – The intents.json knowledge base is parsed and all sample utterances are converted to BoW vectors. The neural network is trained on this labeled dataset for 200 epochs with batch size 5. Model weights and class labels are serialized as pickle files for runtime inference.
Query Processing – User input is received via the Flask interface, preprocessed through the NLP pipeline, converted to a BoW vector, and fed to the trained model to predict intent.
Response Retrieval – The predicted intent label is looked up in the knowledge base and an appropriate response is returned to the user.
User Interface (UI) – The Flask-based UI provides a real-time chat experience with dedicated input and response panels, supporting both desktop and mobile browsers.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153148

X. EXPERIMENTAL EVALUATION

The College Enquiry Chatbot is evaluated through controlled experiments across various real-world application scenarios. The system undergoes benchmark testing for intent classification accuracy, response correctness, and system response time.
User trials assess the assistant's ability to handle diverse query phrasings, spelling variations, abbreviated inputs, and multi-topic questions. Performance metrics include intent classification accuracy, F1 score, response latency (milliseconds per query), and user satisfaction. Comparisons with a rule-based baseline highlight improvements in conversational intelligence and domain coverage.

A. Datasets Used

To train and evaluate the College Enquiry Chatbot, the following datasets are utilized:

- Custom College Enquiry Dataset – 525 labeled utterances across 25 intent categories, manually created to represent real student queries.
- Real-World Trial Data – Queries collected from 40 students during a one-week live trial, used for user acceptance testing.
- Augmented Paraphrase Set – Additional paraphrased variants generated using back-translation to improve model robustness.

These datasets provide a robust foundation for model training, ensuring high accuracy and adaptability to real student language patterns.

B. Evaluation Criteria

The College Enquiry Chatbot is evaluated based on:

- Intent Classification Accuracy – Percentage of correctly identified intents on the held-out test set.
- Macro-Averaged F1 Score – Per-intent precision and recall averaged across all 25 categories.
- Response Latency – Average time in milliseconds from query submission to response delivery.
- User Satisfaction – Rated 1–5 through real-world usability testing and structured surveys with students.

These criteria ensure that the assistant meets high standards of efficiency, accuracy, and real-world usability.

XI. RESULTS

Table 1: Performance Evaluation Results

Metric	Proposed System	Rule-Based Baseline
Intent Classification Accuracy	92.4%	74.1%
Macro-Averaged F1 Score	0.91	0.70
Avg. Response Latency	210 ms	85 ms
User Satisfaction (1–5)	4.3 / 5	3.1 / 5
Fallback Rate	7.6%	25.9%

The proposed system achieves an intent classification accuracy of 92.4%, significantly outperforming the rule-based baseline (74.1%). The macro-averaged F1 score of 0.91 confirms consistent performance across all 25 intent categories. User satisfaction scores from 40 student participants averaged 4.3 out of 5 during the trial period. The fallback rate of 7.6% indicates the system correctly identifies uncertain inputs. Response latency of 210 ms is well within acceptable limits for real-time conversation.

XII. STRENGTHS

✓ High Intent Classification Accuracy – Achieves 92.4% accuracy using a lightweight neural network without requiring GPU resources.
✓ Domain-Specific Knowledge Base – Covers all major college enquiry categories trained specifically on educational domain data.
✓ 24/7 Availability – Provides instant responses at any time, reducing dependency on human administrative staff.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153148

✓ Easy to Update – JSON-based knowledge base allows non-technical staff to add or modify responses without retraining.
✓ Lightweight and Deployable – Runs efficiently on standard college server hardware without specialized infrastructure.
These strengths position the College Enquiry Chatbot as a leading tool for intelligent educational domain automation.

XIII. LIMITATIONS

⚠ Limited Context Memory – The system does not maintain context across multiple conversational turns, making multi-step queries challenging.
⚠ Training Data Dependence – Model accuracy depends on dataset quality; rare or unseen intents may not be handled correctly.
⚠ No Multilingual Support – The current version supports only English; regional language queries are not handled.
⚠ Static Knowledge Base – Responses must be manually updated; the system cannot learn from new queries automatically without retraining.
Future iterations will focus on reducing these limitations for better scalability and real-world adaptability.

XIV. INPUT SENSITIVITY

The chatbot's performance is influenced by the phrasing of user input. Highly abbreviated or grammatically incorrect queries may fall below the confidence threshold and trigger the fallback mechanism. To mitigate this, the BoW model is trained with a diverse set of utterance variations per intent, and the NLTK stemmer reduces inflectional variants to a common base form.

Additionally, spelling correction using the TextBlob library is integrated as a preprocessing step, improving the system's ability to handle common typographical errors. Future iterations will explore the use of word embeddings and transformer-based encoders to further improve robustness against input variation.

XV. CONCLUSION

The College Enquiry Chatbot presented in this research demonstrates a powerful integration of Machine Learning, NLP-based intent classification, and a domain-specific knowledge base into a single deployable conversational system. By leveraging technologies like NLTK, TensorFlow/Keras, and Flask, the system efficiently processes and responds to student enquiries with high accuracy and real-time performance.

Experimental evaluations highlight its 92.4% intent classification accuracy, strong F1 scores, and high user satisfaction, making it a robust AI-driven solution for college administrations seeking to automate student interaction.

While the chatbot achieves notable improvements, challenges remain in multi-turn context management, multilingual support, and automatic learning from new data. Future enhancements will focus on voice integration, ERP connectivity, and extended language support to make the system even more adaptable and widely applicable.

In conclusion, the College Enquiry Chatbot sets a new benchmark for intelligent educational chatbot systems. By continuously improving AI-driven automation and domain personalization, it has the potential to revolutionize how students and institutions interact with administrative services.

XVI. FUTURE WORK

Context-Aware Multi-Turn Dialogue – Implementing memory across conversation turns using RNNs or transformer-based models such as BERT for more natural conversation flow.

Multilingual Support – Extending the chatbot to support regional Indian languages (Hindi, Marathi, Gujarati) using multilingual NLP models.

Voice Interface Integration – Adding speech-to-text and text-to-speech modules for voice-based student interaction.
Automatic Learning from Interactions – Implementing online learning to update the model based on new queries and administrator-confirmed responses.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.153148

Integration with College ERP/SIS – Connecting the chatbot to live college databases to provide real-time information on results, fee payment status, and timetables.

REFERENCES

[1]. Ranoliya, B. R., Raghuwanshi, N., & Singh, S. (2017). "Chatbot for university related FAQs." 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE.

[2]. Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). "Delivering cognitive behavior therapy using a fully automated conversational agent (Woebot)." JMIR Mental Health, 4(2).

[3]. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media. (NLTK Library)

[4]. Abadi, M., et al. (2016). "TensorFlow: A system for large-scale machine learning." 12th USENIX Symposium on Operating Systems Design and Implementation.

[5]. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of deep bidirectional transformers for language understanding." NAACL-HLT 2019.

[6]. Chollet, F. (2017). Deep Learning with Python. Manning Publications. (Keras Framework)

[7]. Grinberg, M. (2018). Flask Web Development: Developing Web Applications with Python (2nd ed.). O'Reilly Media.

[8]. Benedetto, L., et al. (2021). "A comparative study of intent detection methods for task-oriented dialogue systems." ACL Anthology.