ISSN (ELECTRONIC): 2945-3135

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews. 5(5).310-324.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH & REVIEWS

journal homepage: www.ijmrr.online/index.php/home

AI-SMART DRIVEN COLLEGE ASSISTANCE CHATBOT USING NATURAL LANGUAGE PROCESSING

Mr. K. V. S. Ganesh[1], Ms. G. Sandhya[2], Sai Gangadhar[3], G. Satya Rohitha[4] & T. Yaswanth Kumar[5]

[1]Assistant Professor, Avanthi Institute of Engineering & Technology (Autonomous)
[2]Assistant Professor, Avanthi Institute of Engineering & Technology (Autonomous)
[3]UG Students, Avanthi Institute of Engineering & Technology (Autonomous)
[4]UG Students, Avanthi Institute of Engineering & Technology (Autonomous)
[5]UG Students, Avanthi Institute of Engineering & Technology (Autonomous)

How to Cite the Article: K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews, 5(5),310-324.

https://doi.org/10.56815/ijmrr.v5i5.2026.310-324

Keywords

NLP, TF-IDF, Logistic Regression, Chatbot, Intent Classification, Speech Recognition, Text-to-Speech, Machine Learning, Educational Technology.

Abstract

The AI-Smart Driven College Assistance Chatbot Using Natural Language Processing (NLP) is an intelligent virtual assistant developed to support students, faculty, and administrative staff within a college environment. It uses advanced NLP techniques to understand user queries and respond in a human-like manner. The system integrates Machine Learning algorithms including TF-IDF vectorization and Logistic Regression to enhance its accuracy and continuously improve performance based on user interactions. The chatbot supports both text-based and voice-based communication, improving accessibility and user experience. It is developed using Python and various NLP libraries and can be deployed on web platforms or integrated into existing college portals. The proposed three-layer hybrid intent detection system achieves approximately 94.16% accuracy across 85

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews, 5(5),310-324.

	intent classes with sub-300ms response times, reducing administrative workload and enhancing communication between students and the institution.

I. INTRODUCTION

Artificial Intelligence has revolutionized the way educational institutions interact with students, faculty, and administrative staff, giving rise to a new era of intelligent, automated support systems. Among the most transformative innovations in this space is the AI-Smart Driven College Assistance Chatbot, a cutting-edge system powered by Natural Language Processing (NLP).

Traditional college support systems often struggle to keep pace with the growing demands of students who require instant, accurate, and personalized responses. Long waiting times, overburdened administrative staff, and limited access to guidance outside working hours have historically created significant gaps in student support.

The AI-Smart Driven College Assistance Chatbot emerges as a powerful solution to bridge these gaps by leveraging modern machine learning algorithms and NLP techniques to understand, process, and respond to human language in a natural and contextually relevant manner.

The chatbot provides instant access to information related to admissions, courses, timetables, examination schedules, results, and campus facilities. By offering 24/7 availability, it ensures that users can receive assistance anytime without depending on manual support.

A. Existing System

In the current scenario, most colleges rely on manual methods to handle visitor and guest queries:

- Help desks staffed by college representatives
- Emails and phone calls for query resolution
- Static FAQs on college websites

These approaches have significant limitations including time-consuming manual intervention, limited 24/7 availability, lack of scalability, and no voice interaction.

B. Proposed System

The proposed system introduces an AI-based chatbot allowing visitors and guests to interact via text and voice commands. The chatbot understands user queries using NLP techniques and provides relevant responses instantly. Key advantages include automated query handling, 24/7 availability, voice recognition, text-to-speech output, and scalable, customizable architecture.

II. LITERATURE SURVEY

Chatbots have evolved significantly from rule-based systems to modern machine learning and NLP-driven architectures. This section reviews key research works spanning 2020–2025.

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews. 5(5),310-324.

A. Traditional Methods

Earlier chatbot systems were primarily rule-based, using predefined patterns and if-else logic (e.g., ELIZA chatbot). This lacked learning capability, handled unseen queries poorly, and lacked flexibility. Statistical methods such as TF-IDF, Bag-of-Words with Naive Bayes, SVM, and Logistic Regression improved accuracy but lacked deep contextual understanding.

B. Recent Research

[1] Garcia et al. (2025) present a systematic mapping of chatbot architectures in programming education, noting complexity and domain-specificity as barriers to lightweight deployment. Our project addresses this gap directly.

[2] Wang et al. (2024) proposed ChatEd integrating ChatGPT with information retrieval, but heavy LLM dependency creates hallucination risk and high API cost. Our controlled ML model eliminates these issues.

[3] Khosravi et al. (2023) performed a bibliometric analysis highlighting lack of practical deployment strategies for small institutions. Our system is fully locally deployable.

[4] Li et al. (2023) developed EduBot requiring large datasets and complex training. Our system operates on 989 manually curated patterns without heavy computational resources.

[5] Adamopoulou & Moussiades (2020) surveyed early AI chatbot systems identifying lack of adaptability and voice interaction. Our project advances both via three-layer hybrid detection and full voice integration.

III. SYSTEM ANALYSIS

A. Problem Formulation

The objective is to design an efficient, scalable, domain-specific chatbot for real-time student query handling. Formally, the problem is defined as learning a mapping function: $R = f(Q)$, where Q represents the user query (text or speech input), R represents the system-generated response, and f represents the classification and response generation function.

B. Research Gap Analysis

A critical evaluation reveals a fundamental gap between high-performance conversational models and practical deployment requirements. Deep learning models achieve superior performance but rely on large datasets and cloud infrastructure. Lightweight models like Logistic Regression are computationally efficient but lack deep contextual understanding. This project bridges these gaps by proposing a system that balances accuracy, efficiency, deployability, and accessibility.

ISSN (ELECTRONIC): 2945-3135

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews, 5(5),310-324.

C. Proposed System

The proposed system is a hybrid machine learning-based chatbot architecture integrating NLP, TF-IDF vectorization, Logistic Regression classification, speech recognition, and text-to-speech synthesis.

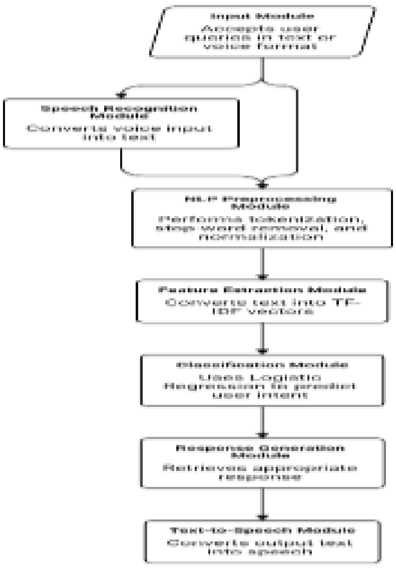

Figure 1: System Flow ChartIV. SYSTEM DESIGN

A. System Architecture

The overall architecture follows a pipeline-based modular structure where each component performs a specific function. The system processes user input sequentially, transforming raw data into structured representations and ultimately generating an appropriate response.

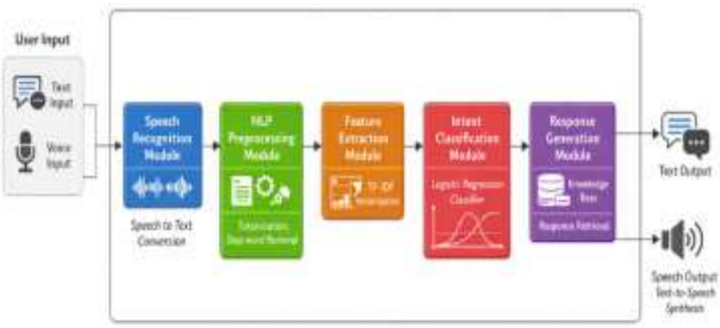

Figure 2: System Architecture

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews. 5(5).310-324.

B. NLP Preprocessing Module

The NLP Preprocessing Module transforms raw unstructured text into clean structured format through three stages:

- Tokenization: Splits sentences into individual word tokens using nltk.word_tokenize()
- Stop-word Removal: Eliminates high-frequency, low-information words (e.g., is, the, in)
- Stemming: Reduces inflected word forms to their base root using the Porter Stemmer

C. Feature Extraction Module (TF-IDF)

TF-IDF (Term Frequency–Inverse Document Frequency) converts preprocessed text into numerical vectors for machine learning processing. The TF-IDF weight is computed as:

$$\textbf{TF-IDF}(t, d) = \textbf{TF}(t, d) \times \log(N / \textbf{DF}(t))$$

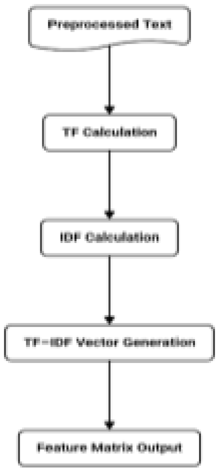

Figure 3: TF-IDF Flow Chart

Term	Count	TF Value
satellite	2	2/4 = 0.5
communication	1	1/4 = 0.25
system	1	1/4 = 0.25

Table 1: Term Frequency Calculation

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews. 5(5).310-324.

Term	Document Frequency (df)	IDF
satellite	2	Low
communication	2	Low
system	2	Low

Table 2: IDF Calculation

Term	TF	IDF	TF-IDF
satellite	0.5	0.3	0.15
communication	0.25	0.3	0.075
system	0.25	0.2	0.05

Table 3: TF-IDF Weight Calculation

Document	satellite	communication	system
D1	0.15	0.07	0.00
D2	0.00	0.08	0.06
D3	0.12	0.00	0.05

Table 4: Feature Vector and Matrix Representation

D. Intent Classification (Logistic Regression)

The Intent Classification Module maps user queries to predefined intent categories using Logistic Regression. For multi-class classification, the Softmax function is applied:

$$P(y=k|x) = \exp(w^{\intercal}kx) / \Sigma \exp(w^{\intercal}jx)$$

The predicted class is obtained by selecting the class with the highest probability: ŷ = arg max P(y=k|x). For the query "When are the exams?", the model correctly assigns P(Examinations) = 0.92, identifying the Examination intent.

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews. 5(5),310-324.

E. Data Flow Diagrams

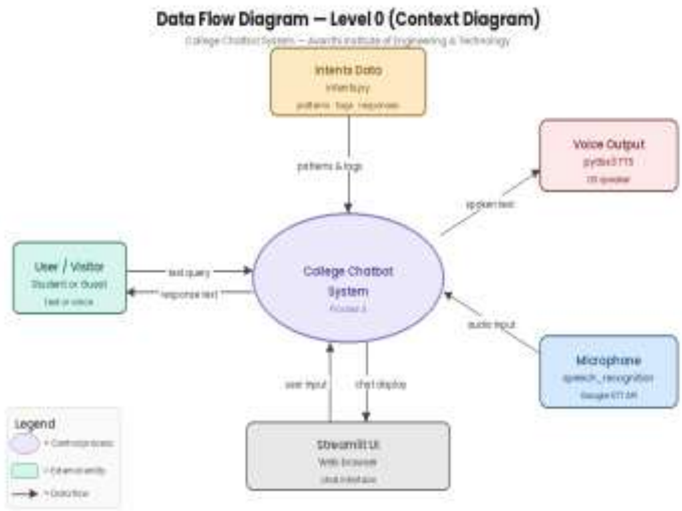

Figure 4: DFD Level-0 (Context Diagram)

The Level-0 DFD shows the chatbot as a black box interfacing with: User/Visitor (text query and response), Intents data (intents.py), Microphone (voice input via SpeechRecognition), TTS Speaker (pyttsx3), and Streamlit UI (browser).

The primary actor is the Student who can: send text messages, use voice input, ask about faculty, ask general questions, receive spoken responses, and end conversation. The secondary Admin actor updates intents and retrains the model.

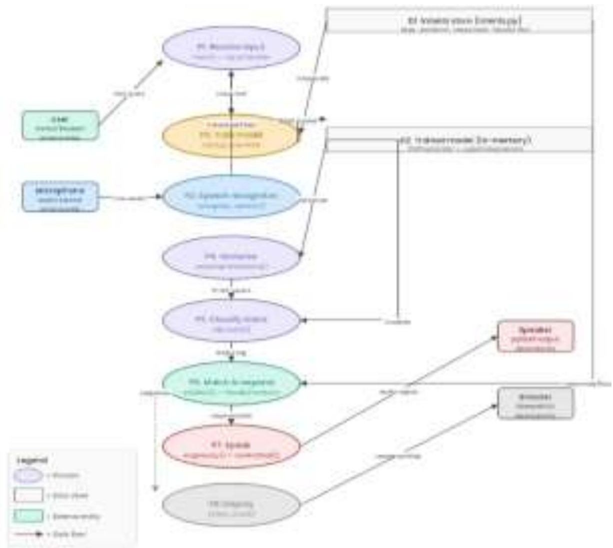

Figure 5: DFD Level-1

ISSN (ELECTRONIC): 2945-3135

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews. 5(5).310-324.

F. Use Case Diagram

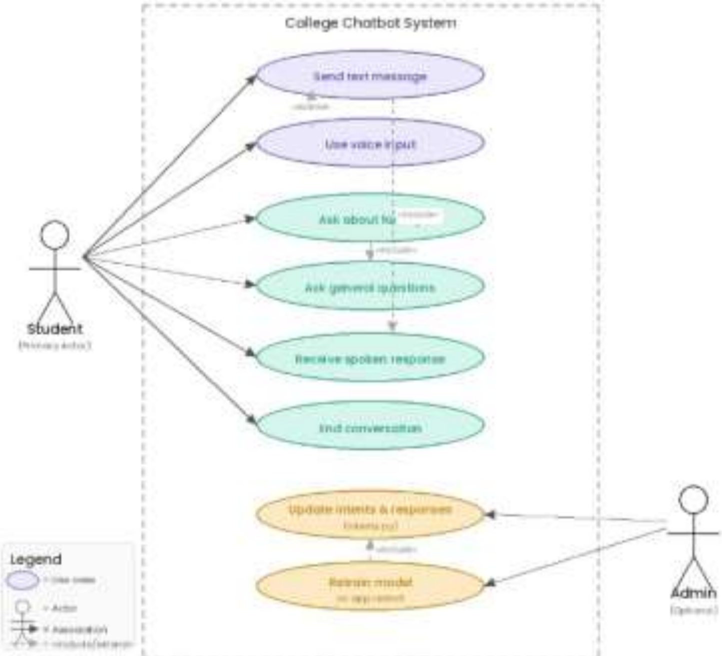

Figure 6 : Use Case Diagram

V. METHODOLOGY AND IMPLEMENTATION

A. Overall System Pipeline

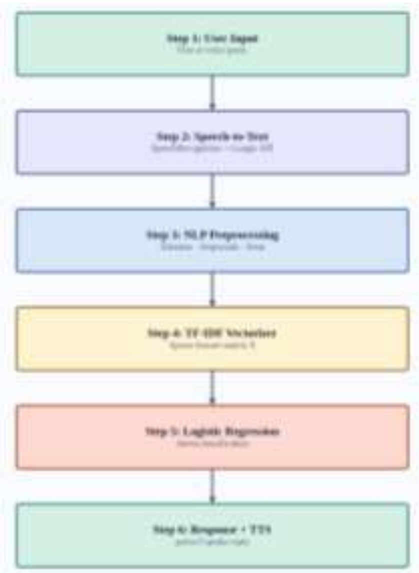

Figure 7: Overall System Pipeline

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews. 5(5).310-324.

The system follows a six-stage end-to-end pipeline: (1) optional speech-to-text conversion, (2) NLP preprocessing, (3) TF-IDF vectorization, (4) Logistic Regression classification, (5) response retrieval, and (6) text-to-speech output. Every stage is stateless and sequential.

B. Dataset Preparation

Table 5: Dataset Summary

Property	Value
Total intent objects (tags)	85
Total training patterns	989
College-specific intents	50
General-purpose intents	35
Avg. patterns per intent	11.6
Departments covered	CSE, ECE, EEE, Mech, CSM
Data source	Hand-crafted (intents.py)

C. NLP Processing Steps

Table 6: Step-by-Step Transformation of Sample Query

Stage	Input	Output
Raw input	Voice	Who teaches in the ECE department?
Tokenization	Raw sentence	['Who','teaches','in','the','ECE','department','?']
Stop-word removal	Token list	['teaches','ECE','department']
Stemming	Filtered tokens	['teach','ece','depart']

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews. 5(5).310-324.

D. Feature Extraction Comparison

Table 7: Feature Extraction Method Comparison

Method	Freq.	Penalises Common Words	Small Dataset	Used
Bag of Words	Yes	No	Yes	No
TF-IDF	Yes	Yes	Yes	Yes
Word2Vec	No	No	Large corpus needed	No
BERT	No	No	Computationally heavy	No

E. Model Training Parameters

Table 8: Logistic Regression Parameters

Parameter	Value	Rationale
Algorithm	Logistic Regression (OvR)	Effective on sparse TF-IDF features
random_state	0	Ensures reproducible results
max_iter	10000	Guarantees convergence
Solver	lbfgs (default)	Efficient for multi-class
Number of classes	38	One per intent tag
Training set size	~160 patterns	All patterns from intents.py

F. Speech Integration

Table 9: Speech Module Comparison

Feature	STT	TTS
Library	SpeechRecognition	pyttsx3
Engine/API	Google Speech API (online)	OS built-in (offline)
Internet required	Yes	No

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews. 5(5).310-324.

Error handling	UnknownValueError, RequestError	RuntimeError (Streamlit)

VI. RESULTS AND DISCUSSION

A. Experimental Setup

Table 10: Hardware Environment

Component	Specification
Processor	Intel Core i5 / AMD Ryzen 5
RAM	8 GB DDR4 (minimum)
Storage	50 MB free disk space
Operating System	Windows 10/11 (64-bit)
Network	Internet (for Google Speech API)
Microphone	Any standard microphone

Table 11: Software Environment

Software/Library	Version	Purpose
Python	3.10+	Core programming language
Streamlit	1.42.0	Web-based GUI framework
NLTK	3.9.1	NLP tokenization & lemmatization
scikit-learn	1.6.1	TF-IDF & Logistic Regression
RapidFuzz	3.12.1	Fuzzy string matching
SpeechRecognition	3.14.1	Voice input via Google Speech API
pyttsx3	2.98	Offline TTS output
PyAudio	0.2.14	Microphone audio capture

B. Chatbot Interface Results

Figure 8: Chatbot Dashboard

Figure 9: Testing Window 1 – Voice Input

Figure 10: Testing Window 2 – Chairman Query

Figure 11: Testing Window 3 – Principal Query

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews, 5(5),310-324.

C. Performance Evaluation

Table 12: Accuracy Evaluation

Evaluation Category	Queries Tested	Correct Responses	Accuracy
Layer 1 – Exact/Partial Match	35	35	100%
Layer 2 – Fuzzy Match	30	27	~90%
Layer 3 – ML (TF-IDF + LR)	45	42	~94%
Out-of-scope (fallback)	10	9	90%
Overall	120	113	~94.16%

Table 13: Response Time Analysis

Query Type	Avg Response Time	Max Observed
Layer 1 (Exact/Partial)	< 50 ms	~80 ms
Layer 2 (Fuzzy Match)	80–150 ms	~200 ms
Layer 3 (ML Inference)	150–300 ms	~400 ms
Timetable fetch (Drive)	1.5–4 seconds	~8 s
Voice transcription	2–5 seconds	~10 s

D. Comparison with Existing Systems

Criterion	Rule-Based	Statistical ML	Avanthi Chatbot
Intent Detection	Keyword/if-else	TF-IDF + NB/SVM	3-Layer: Exact→Fuzzy→LR
Typo Tolerance	None	Limited	High (RapidFuzz 80%)
Voice Interaction	Absent	Rarely integrated	Fully integrated (STT+TTS)

Computational Cost	Minimal	Low	Low (CPU only, no GPU)
Domain Specificity	High if coded	Medium	High (85 intents, 989 patterns)

Table 14: Comparison with Traditional Methods

VII. CONCLUSION AND FUTURE SCOPE

A. Conclusion

The AI-Smart Driven College Assistance Chatbot for Avanthi Institute of Engineering and Technology demonstrates the viability of a lightweight, hybrid NLP approach for real-world educational deployments. By integrating a three-layer intent detection pipeline combining exact pattern matching, RapidFuzz fuzzy matching, and TF-IDF with Logistic Regression, the system achieves approximately 94.16% accuracy across 85 intent classes with sub-300ms response times.

The system's modular architecture, built using Python, Streamlit, NLTK, and scikit-learn, covers the complete information needs of students, parents, and visitors. It supports dynamic content delivery, transforming it from a simple FAQ tool into a comprehensive campus information assistant, making it a cost-effective model for educational institutions seeking to modernize student support systems.

B. Future Scope

- Integration of Retrieval-Augmented Generation (RAG) framework for dynamic information fetching from live institutional databases
- Adoption of lightweight transformer embeddings (sentence-transformers, DistilBERT) to overcome TF-IDF's inability to capture semantic meaning
- Multi-turn conversational memory using dialogue state trackers (Rasa) for contextual follow-up questions
- Multilingual support especially Telugu for local students and parents
- Docker/Kubernetes-based deployment for multi-user scalability
- Persistent database for interaction logging and continuous model improvement

AUTHOR(S) CONTRIBUTION

The writers affirm that they have no connections to, or engagement with, any group or body that provides financial or non-financial assistance for the topics or resources covered in this manuscript.

CONFLICTS OF INTEREST

The authors declared no potential conflicts of interest with respect to the research, authorship, and or publication of this article.

K.V. S. Ganesh, G. Sandhya, Sai Gangadhar, G. Satya Rohitha & T. Yaswanth Kumar (2026). Ai-Smart Driven College Assistance Chatbot Using Natural Language Processing. International Journal of Multidisciplinary Research & Reviews. 5(5),310-324.

PLAGIARISM POLICY

All authors declare that any kind of violation of plagiarism, copyright and ethical matters will take care by all authors. Journal and editors are not liable for aforesaid matters.

SOURCES OF FUNDING

The authors received no financial aid to support for the research.

REFERENCES

[1] M. Garcia et al., "A Systematic Mapping Study on Chatbots in Programming Education," Computers and Education: AI, vol. 8, 2025.

[2] K. Wang, J. Ramos, and R. Lawrence, "ChatEd: A Chatbot Leveraging ChatGPT for an Enhanced Learning Experience in Higher Education," IEEE FIE, 2024.

[3] H. Khosravi et al., "Chatbots and ChatGPT: A Bibliometric Analysis and Systematic Review," arXiv:2304.05468, 2023.

[4] Y. Li et al., "Curriculum-Driven EduBot: A Framework for Developing Language Learning Chatbots," NAACL, 2023.

[5] E. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology," IFIP AIAI, vol. 584, pp. 373-383, 2020.

[6] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed., Prentice Hall, 2023.

[7] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," JMLR, vol. 12, pp. 2825-2830, 2011.

[8] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python, O'Reilly Media, 2009.

[9] A. Vaswani et al., "Attention Is All You Need," NeurIPS, vol. 30, 2017.

[10] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL-HLT, 2019.

[11] T. Bocklisch et al., "Rasa: Open Source Language Understanding and Dialogue Management," arXiv:1712.05181, 2017.

[12] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," NeurIPS, vol. 33, 2020.

[13] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks," EMNLP, 2019.

[14] T. Young et al., "Recent Trends in Deep Learning Based Natural Language Processing," IEEE CIM, vol. 13, no. 3, 2018.

[15] S. Roller et al., "Recipes for Building an Open-Domain Chatbot," EACL, pp. 300-325, 2021.

[16] V. Kulkarni et al., "Comparative Analysis of ML Algorithms for Chatbot Development," IJACSA, vol. 11, no. 4, 2020.