IJARSCT

ISSN: 2581-9429

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Volume 6, Issue 4, June 2026

Impact Factor: 8.2

Collexa : An AI Powered Calling Agent

Mr. Nishant Ethape, Mr. Harsh Solanki, Mr. Parth Deshpande, Mr. Rohit Borde, Prof. Minal Patil

Department of Computer Science

Indira College of Engineering and Management, Pune, Maharashtra, India.

(An Autonomous Institute affiliated to SPPU)

Savitribai Phule Pune University, Pune, Maharashtra, India.

Abstract: *Educational institutions receive thousands of repetitive phone calls daily regarding admissions, fees, scholarships, office timings, and documentation. Manual call handling results in long queue times, inconsistent responses, and a high burden on reception staff. This paper presents Collexa, an AI-powered automated calling agent that integrates cloud telephony (Twilio/Exotel), Speech-to-Text (STT), Natural Language Processing (NLP), Retrieval-Augmented Generation (RAG), and neural Text-to-Speech (TTS) models to autonomously answer incoming calls. The system performs real time intent classification, multilingual FAQ retrieval, and natural voice response generation while ensuring seamless human handoff for unresolved queries. Collexa supports concurrent calls, performs automatic analytics logging, and delivers 24×7 availability. This paper discusses the end-to-end architecture, microservices, model selection strategies, optimization techniques, performance analysis, and limitations, providing a scalable framework for AI-driven telephony systems.*

Keywords: Telephony AI, Voice Automation, NLP, RAG, Twilio, Exotel, Fast API, Neural TTS, Whisper, Google Speech.

I. INTRODUCTION

The reception desk plays a vital role in educational institutions by serving as the primary communication channel between students, parents, and administrative departments. During peak periods such as admissions, examination sched ules, scholarship deadlines, and fee submissions, reception staff receive an overwhelming number of phone calls. These calls often revolve around repetitive questions, leading to long waiting queues and missed calls. Traditional IVR (Interactive Voice Response) systems used by many organizations provide only menu-driven interactions, which are rigid and fail to understand natural human language. With advancements in Automatic Speech Recognition (ASR), Natural Language Understanding (NLU), and cloud telephony APIs, it has become possible to design conversational AI systems capable of autonomous call handling. Collexa is developed to address this gap by offering a natural, conversational, multilingual calling agent that operates seamlessly with existing telephony infrastructure. Unlike classical IVR systems, Collexa interprets free-form speech, retrieves contextually relevant responses grounded in institutional FAQs, and speaks back using near-human neural voices. The system reduces operational costs, improves availability, and ensures consistent communication. This paper presents an in-depth description of the architecture, system design, implementation, optimization strategies, and performance evaluation.

II. PROBLEM DEFINITION AND OBJECTIVES

Problem Definition

Manual call answering at college receptions leads to long wait times, inconsistent communication, human error, and inefficiency. Callers frequently ask routine questions that can be automated using AI-driven conversational systems. The goal is to create an autonomous calling agent capable of identifying caller intent, retrieving accurate information from an institutional knowledge base, and responding through natural speech, while escalating complex queries to humans.

Objectives

• Automate call handling for routine reception queries.
• Implement robust STT and TTS for natural conversations.
• Use NLP and RAG for accurate and grounded responses.
• Provide multilingual support (English and Hindi).
• Support human handoff when confidence is low.
• Maintain logs for analytics and iterative improvements.

III. LITERATURE SURVEY

The development of Collexa draws on a broad set of prior works spanning speech recognition, conversational AI, retrieval-augmented generation, text-to-speech synthesis, and cloud telephony systems. The following subsections summarize the major contributions and how they inform Collexa's design.

Speech Recognition: Advancements from HMM-GMM models to deep neural networks significantly improved accuracy in Automatic Speech Recognition (ASR). Modern systems like Whisper and Google Speech-to-Text provide robust performance across accents and noisy environments, forming the basis for reliable speech-to-text conversion.[1][7]

Natural Language Processing: Conversational understanding has evolved from statistical dialog systems to transformer-based models such as BERT. These models enable effective intent detection and semantic similarity matching, which are essential for mapping user queries to FAQ-based responses.[2]

Text-to-Speech (TTS): Neural TTS models like Tacotron and WaveNet have improved speech naturalness and quality. Cloud-based TTS services (Azure, Google, Amazon Polly) provide low-latency and expressive voice output suitable for real-time applications.[3]

Conversational AI Systems: Modern virtual assistants integrate intent recognition, dialogue management, and response generation. These systems emphasize context handling and user interaction strategies, which are adopted in Collexa for efficient conversation flow.[4]

Telephony Integration: Cloud platforms like Twilio and Exotel enable real-time voice communication through APIs and webhooks. These systems support seamless integration of ASR, NLP, and TTS for deploying voice-based agents.[6]

AI in Education: AI-driven automation in educational institutions improves administrative efficiency and accessibility. Collexa aligns with these goals by providing automated, 24×7 response handling.[8]

IV. SYSTEM ARCHITECTURE

The architecture of Collexa is designed as a modular and scalable voice automation framework that integrates cloud telephony, speech processing, natural language understanding, knowledge retrieval, and speech synthesis. The system enables end-to-end automation of college reception calls while maintaining flexibility for future enhancements.

As illustrated in Figure 1, incoming calls are received through the Exotel cloud telephony platform and forwarded to the FastAPI backend for processing. The backend coordinates multiple AI services, including Whisper for Speech-to-Text conversion, a knowledge base for FAQ retrieval, OpenAI-based response generation, and Text-to-Speech synthesis for voice responses. The generated responses are delivered back to callers through the telephony gateway.

The system maintains a centralized database containing institutional information, FAQs, call logs, and user interaction records. This architecture ensures efficient query handling, real-time response generation, scalability, and continuous monitoring of system performance.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, June 2026
Impact Factor: 8.2

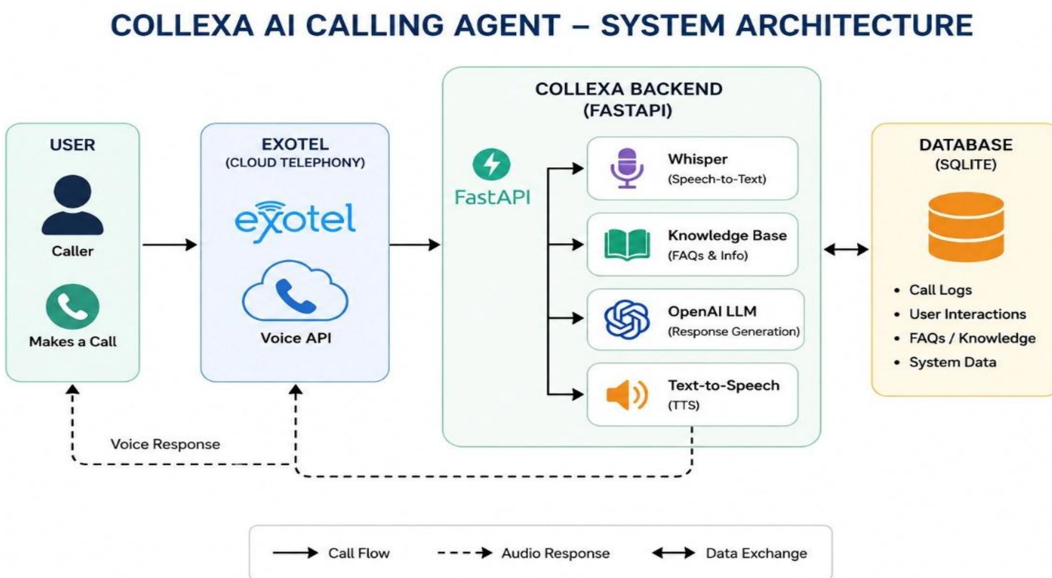

Fig 1. System Architecture

V. METHODOLOGY

A. Call Flow When a caller dials the reception number:

1) Telephony provider (Twilio/Exotel) forwards the call to backend via webhook.
2) The backend initializes a session and streams audio to the STTService.
3) STTService returns transcription; NLPEngine extracts intent and embeddings.
4) RAG retrieves the highest scoring FAQ or generates a grounded response.
5) TTSService synthesizes speech and the telephony provider plays audio to caller.
6) On low confidence, the system performs a human handoff with context.
7) All interactions are logged for analytics and retraining.

B. Technology Stack:

The Collexa system utilizes Exotel for telephony integration, FastAPI for backend services, Whisper for speech recognition, OpenAI LLM for response generation, SQLite for knowledge storage, and Text-to-Speech services for voice synthesis.

C. Multilingual Strategy Language detection is performed using fastText embed dings. FAQ entries exist in English and Hindi; the system selects the appropriate language model for STT/TTS and retrieval.

VI. SYSTEM IMPLEMENTATION

The Collexa system was implemented using a modular architecture integrating cloud telephony, speech processing, natural language understanding, and knowledge retrieval components. Exotel serves as the telephony gateway responsible for receiving incoming calls and forwarding audio streams to the backend service. The backend was developed using FastAPI and manages call sessions, speech processing, response generation, and database interactions.

Speech-to-Text conversion is performed using Whisper, which transcribes caller speech into text. The transcribed query is analyzed using NLP techniques and matched against the institutional knowledge base containing admission details, fee structures, office timings, scholarship information, and academic policies. When a suitable answer is identified, the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, June 2026
Impact Factor: 8.2

response is generated and converted into speech using a Text-to-Speech engine before being delivered back to the caller.

The system maintains conversation logs and query history in a database, enabling future analysis and continuous improvement of response quality.

Fig. 2. Call Processing Workflow

Figure 2 presents the end-to-end call processing workflow. The system receives an incoming call through Exotel, converts speech into text using Whisper, analyzes user intent, retrieves relevant knowledge from the database, generates a response using AI models, converts the response into speech, and delivers it back to the caller before terminating the session.

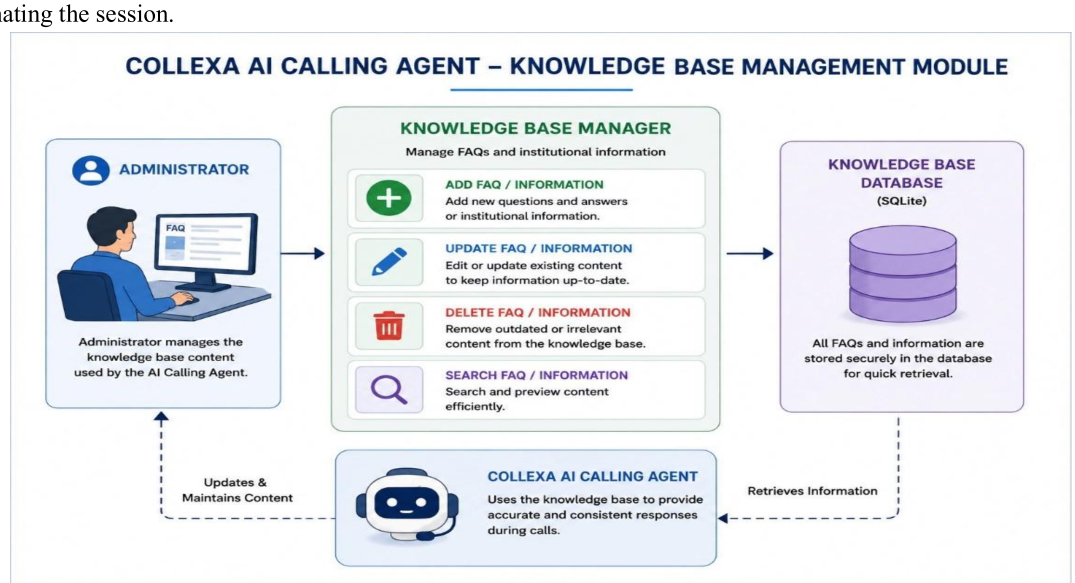

Fig. 3. Knowledge Base Management Module

IJARSCT
ISSN: 2581-9429
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 6, Issue 4, June 2026
Impact Factor: 8.2

Figure 3 illustrates the administrative knowledge management process. Administrators can add, update, delete, and search institutional information stored in the knowledge base. This ensures that the AI calling agent always responds using accurate and up-to-date information.

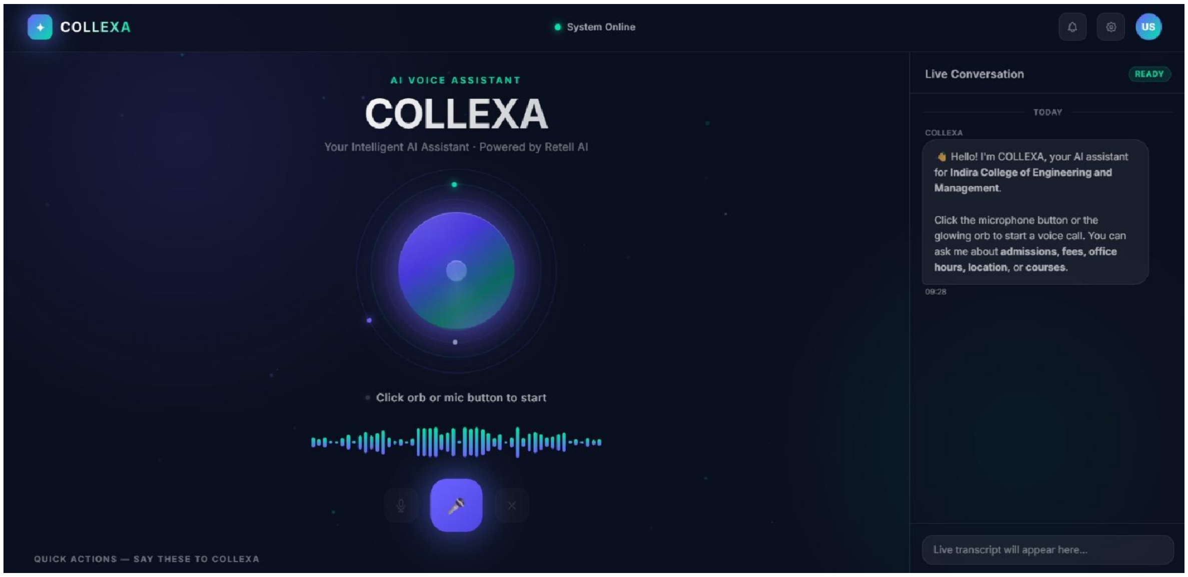

Fig. 4. COLLEXA Voice Assistant Interface

Figure 4 shows the implemented COLLEXA voice assistant interface. The interface provides real-time conversation monitoring, live transcript generation, voice interaction controls, and system status indicators. The conversational panel displays user interactions and generated responses, allowing administrators to monitor system behavior during active sessions.

VII. RESULTS AND PERFORMANCE ANALYSIS

The Collexa system was evaluated using multiple real-world call scenarios covering admission inquiries, fee-related questions, office timings, and document requirements. Manual quality assurance testing was conducted across all major modules, including call handling, speech recognition, response generation, and telephony integration.

Module	Completion
Call Handling	100%
STT Module	100%
TTS Module	100%
NLP/RAG	85%
Multilingual Support	70%
Admin Panel	80%

Results:
Metric Value

Metric	Value
Average API Response Time	< 2 sec
Concurrent Calls Supported	10+

Metric	Value
Target System Availability	99.9%
Manual Test Cases Passed	6/6

The system successfully handled admission inquiries, fee-related questions, office timing requests, and document verification queries during testing.

The results indicate that the system is capable of providing near real-time responses while maintaining reliable performance under concurrent call conditions. The speech-to-text and text-to-speech modules demonstrated stable operation during testing, enabling smooth conversational interaction. The telephony integration successfully handled call routing and response delivery without significant delays.

The multilingual functionality supported communication in both English and Hindi. The system effectively answered frequently asked institutional queries, reducing the dependency on manual reception staff for routine information requests. The experimental evaluation demonstrates that Collexa can automate a significant portion of repetitive reception queries, improving operational efficiency and response consistency. The use of AI-generated responses ensures uniform information delivery irrespective of workload conditions. However, speech recognition performance may vary depending on caller accents and background noise, highlighting the need for further model optimization and expanded multilingual training datasets.

VIII. CONCLUSION

Collexa demonstrates an end-to-end approach for automating college reception calls using modern speech and NLP technologies. By integrating RAG, robust STT/TTS services, and a microservices architecture, Collexa offers scalable, multilingual, and reliable voice automation while retaining human oversight through seamless handoff mechanisms. The system reduces administrative load, improves response consistency, and provides actionable analytics for continuous improvement.

REFERENCES

[1] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., & Kingsbury, B., "Deep Neural Networks for Acoustic Modeling in Speech Recognition," IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, 2012.

[2] Young, S., Gasic, M., Thomson, B., & Williams, J. D., "POMDP-Based Statistical Spoken Dialog Systems: A Review," Proceedings of the IEEE, vol. 101, no. 5, pp. 1160–1179, 2013.

[3] Zen, H., Tokuda, K., & Black, A. W., "Statistical Parametric Speech Synthesis," Speech Communication, vol. 51, no. 11, pp. 1039–1064, 2009.

[4] Hoy, M. B., "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants," Medical Reference Services Quarterly, vol. 37, no. 1, pp. 81–88, 2018.

[5] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., & Riedel, S., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems (NeurIPS), 2020.

[6] Paul, S., & Dutta, A., "Automated Call Handling System Using Cloud Telephony and Artificial Intelligence," International Journal of Advanced Computer Science and Applications, vol. 10, no. 6, pp. 221–226, 2019.

[7] Huang, J.-T., Li, J., Gong, Y., & Yu, D., "Cross-Language Knowledge Transfer for Deep Neural Network Based Speech Recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 1586–1597, 2020.

[8] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F., "Systematic Review of Research on Artificial Intelligence Applications in Higher Education," International Journal of Educational Technology in Higher Education, vol. 16, no. 1, 2019.