

# **Collexa : An AI Powered Calling Agent**

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

Department Of Computer Science

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

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, FastAPI, 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 schedules, 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. MOTIVATION**

Educational institutions face communication challenges such as:

- High call traffic during admissions and deadlines.
- Repetitive inquiries that reduce staff productivity.
- Limited availability of staff outside office hours.
- Instant, consistent responses with 24×7 availability.
- Multilingual support for a broader user base.



### **III. 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.

### **IV. 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 (Hinton et al., 2012) 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.

**Natural Language Processing:** Conversational understanding has evolved from statistical dialog systems (Young et al., 2013) 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.

**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.

**Conversational AI Systems:** Modern virtual assistants integrate intent recognition, dialogue management, and response generation (Hoy, 2018). These systems emphasize context handling and user interaction strategies, which are adopted in Collexa for efficient conversation flow.

**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.

**AI in Education:** AI-driven automation in educational institutions improves administrative efficiency and accessibility. Collexa aligns with these goals by providing automated, 24x7 response handling.

### **V. SYSTEM ARCHITECTURE**

*Figure 1 illustrates the complete system architecture consisting of telephony, backend microservices, NLP/RAG engines, STT/TTS modules, database, and admin dashboard.*



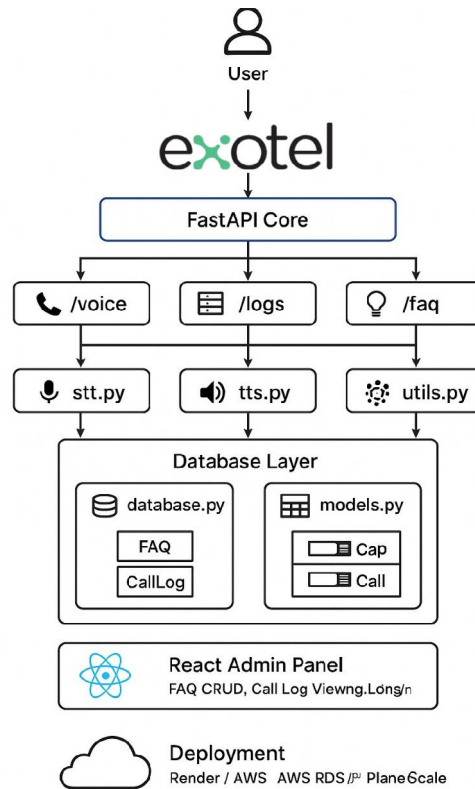


Fig 1. System Architecture

## VI. 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; NLP Engine 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. Model Selection and Fallbacks Primary STT: Whisper-medium (robust to noise) Fallback STT: Google Speech API (low-latency) Primary TTS: Azure Neural TTS / Google WaveNet Fallback TTS: Amazon Polly

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

## VII. 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]. Dhiliban Swaminathan, Dhileep Kumar, M. Kowshik, S. Sundaramoorthi Vinayagam, S. Sundhar, "AI-Powered Voice Assistant for Hospital Reception Automation," 2025.
- [2]. Indukuri Manikanta Sai Varma, Kalidindi Pranith Varma, Karri Gnana Keerthana, "Voice Assistant Using Artificial Intelligence," 2022.
- [3]. George Terzopoulos, Maya Satratzemi, "Voice Assistants and Smart Speakers in Everyday Life and Education," 2020.
- [4]. Marcello M. Mariani, Novin Hashemi, Jochen Wirtz, "Artificial Intelligence Empowered Conversational Agents: A Systematic Literature Review and Research Agenda," Journal of Business Research, 2023.

