International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3377-3382
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0435

AI-Based Voice Assistants for Automated Customer Care Systems

Mr. P. R. Kulkarni[1], Ms. Chanchal Sandip Aher[2], Ms. Akanksha Balasaheb Darade[3], Mr. Parth Gajanan Patil[4], Mr. Roshan Jankiram Pawara[5]
[1]Assistant Professor, Computer Engineering, GCOERC, Nashik
[2,3,4,5]Students, Department of Computer Engineering, GCOERC, Nashik
Emails: piyushrkulkarni@gmail.com[1], aherchanchal07@gmail.com[2], akankshadarade@gmail.com[3], roshanpawara335@gmail.com[4], parthgajananpatil@gmail.com[5]

Abstract
AI-powered voice assistants automate customer support through phone calls using advanced Artificial Intelligence (AI) and Natural Language Processing (NLP). The sys- tem converts spoken customer queries into text using Speech-to-Text (STT), retrieves accurate information using Retrieval-Augmented Generation (RAG), and delivers re- sponses through Text-to-Speech (TTS). The assistant provides context-aware, real-time, and policy-compliant answers, minimizing the need for human agents. It also allows call transfers to human representa- tives when necessary and logs conversations for future anal- ysis. The system ensures 24/7 assistance, reduces opera- tional costs, and enhances customer experience by deliver- ing consistent, efficient, and intelligent support.

Keywords— *AI Voice Assistant, Speech-to-Text (STT), Text-to-Speech (TTS), Retrieval-Augmented Generation (RAG), Large Language Model (LLM), Conversational AI, Customer Care Automation, Voice Interaction.*

1. Introduction

In today's fast-changing business world, customer satisfaction mainly depends on how effectively customer support services operate. Traditional call centers often face problems such as long waiting times, inconsistent communication, and high operational costs. As organizations expand globally, the number of customer interactions continues to grow, making it difficult for human agents alone to consistently maintain high service quality. The proposed system addresses these issues by introducing an intelligent, automated voice assistant that can under- stand, interpret, and respond to customer queries in natural language. Developed using technologies such as Speech Recognition, Retrieval-Augmented Generation (RAG), and Large Language Models (LLMs), the assistant converts spoken input into text, identifies user intent, retrieves relevant organizational information, and delivers accurate verbal responses. Unlike conventional Interactive Voice Response (IVR) systems, the proposed system supports dynamic interaction, manages multiple calls simultaneously, and operates continuously without fatigue. The system's AI engine improves gradually by learning from previous interactions and adapting to organizational needs. Its 24/7 availability and multilingual support make it a suitable solution for industries such as banking, e-commerce, telecommunications, and healthcare.

2. Literature Survey

Recent studies in artificial intelligence and automated customer service highlight the increasing importance of conversational AI and Retrieval-Augmented Generation (RAG) models in improving both response accuracy and system efficiency. Veturi et al. [1] introduced a RAG- based question-answering approach that integrates infor- mation retrieval with Large Language Models (LLMs) to produce reliable and context-aware answers, thereby reduc- ing hallucinations and enhancing factual correctness. Khan and Iqbal [2] evaluated AI-driven customer service solu- tions and reported that although automation improves ef- ficiency and lowers response time, the best outcomes are achieved through hybrid systems that combine automated support with human intervention for emotionally sensitive or complex queries. Shafeeg et al. [3] developed a voice assistant integrated with generative models, incorporating Automatic Speech

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3377-3382
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0435

Recognition (ASR) and Text-to-Speech (TTS) for real-time spoken interaction, while noting that latency and recognition accuracy remain open challenges. Almeida and Xexe´o [4] reviewed word-embedding methods such as Word2Vec, GloVe, and BERT, emphasizing that contextual embeddings significantly enhance semantic understanding in NLP applications. Malkiel et al. [5] proposed GPT-Calls, a framework that generates synthetic customer service dialogues using LLMs to improve call segmentation and intent labeling, thereby reducing man- ual annotation efforts and improving training efficiency. Rau et al. [6] enhanced RAG performance through con- text embeddings that limit document retrieval to highly relevant sources, leading to faster response times and reduced computational overhead in real-time systems. Uzok et al. [7] analyzed AI-powered chatbots and demonstrated their ability to manage large volumes of customer queries with consistency and scalability, suggesting that AI will play a dominant role in future automated support services. A study on Hindi-Marathi code-switching ASR [8] show- cased multilingual recognition capabilities that are critical for inclusive regional AI solutions in India. Furthermore, research on vector databases [9] examined platforms such as FAISS, Pinecone, and Milvus, which enable efficient high-dimensional embedding searches and serve as a core component of RAG-based retrieval systems. Mehta and Wang [10] explored AI-based voice analytics for detecting sentiment and emotion in customer calls using transformer- based deep learning models, allowing more personalized responses and improved user experience. Taken together, these studies demonstrate that the com- bined use of speech recognition, natural language under- standing, and Retrieval-Augmented Generation forms a strong foundation for building intelligent, scalable, and context-aware voice-based customer support systems [11-15].

3. Problem Statement

In many organizations, responding to customer queries through voice calls remains a slow and error-prone task. Human agents often face difficulties in managing large call volumes, understanding different regional languages, and correctly classifying or routing customer requests. These issues commonly lead to longer waiting times, repeated conversations, and lower levels of customer satisfaction. To overcome these challenges, the proposed system introduces an AI-powered voice assistant capable of under- standing, processing, and responding to customer queries in real time. It utilizes Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Text-to-Speech (TTS) technologies to generate context-aware and natural-sounding responses. The system is designed to support human agents by automating routine interactions, improving call handling efficiency, and enabling faster, more accurate, and multilingual customer support.

4. Objective

The primary objectives of the proposed system are outlined as follows:

- Automate Customer Query Handling: To develop an AI-powered voice assistant that can understand and respond to customer queries in real time, reducing the need for manual intervention.
- 2Enhance Communication Efficiency: To minimize customer waiting time and improve response accuracy by automating routine and repetitive support tasks.
- Support Multilingual and Code-Switched Speech: To enable the system to understand and respond in multiple languages, including English, Hindi, and Marathi, as well as mixed-language (code-switched) conversations.
- Integrate AI Components for Real-Time Interaction: To incorporate key modules such as Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Text-to-Speech (TTS) to ensure smooth and continuous two-way communication.
- Improve Customer Satisfaction and Agent Productivity: To support human agents by automatically handling simple and frequently asked queries, allowing them to focus on more complex customer issues.
- Enable Data-Driven Insights: To provide analytical tools and dashboards for monitoring call performance, identifying

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3377-3382
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0435

query trends, and evaluating system accuracy for ongoing improvement.

5. Proposed System

The proposed system is designed to automate customer support using voice interaction. The system allows users to ask questions through speech and receive instant responses without human assistance. It combines speech recognition, language understanding, and automated response generation to provide efficient customer service. Techniques from Natural Language Processing are used to analyze and understand user queries. Working Methodology

The system follows the steps below:

- Voice Input
  The user speaks a query through a microphone or voice-enabled device.
- Speech-to-Text Conversion
  The system converts the spoken voice into text using speech recognition technology.
- Text Processing
  The text is processed to remove unnecessary words and extract important keywords.
- Intent Identification
  The system analyzes the processed text and identifies the user's intent, such as order inquiry, complaint registration, or product information.
- Knowledge Base Search
  Based on the detected intent, the system searches a database containing customer support information.
- Response Generation
  The system generates an appropriate answer related to the user's request.
- Text-to-Speech Conversion
  The generated text response is converted into voice using speech synthesis.
- Voice Response to User
  The final response is delivered to the user through audio output.

The proposed methodology enables the system to respond quickly to customer queries and operate continuously without human intervention.

6. System Architecture

The system architecture of the proposed AI-based voice assistant explains how a user's voice call is received, processed, and answered using Artificial Intelligence components and backend services. The complete workflow brings together speech processing, AI-driven reasoning, and voice synthesis to deliver real-time and natural-sounding customer support.

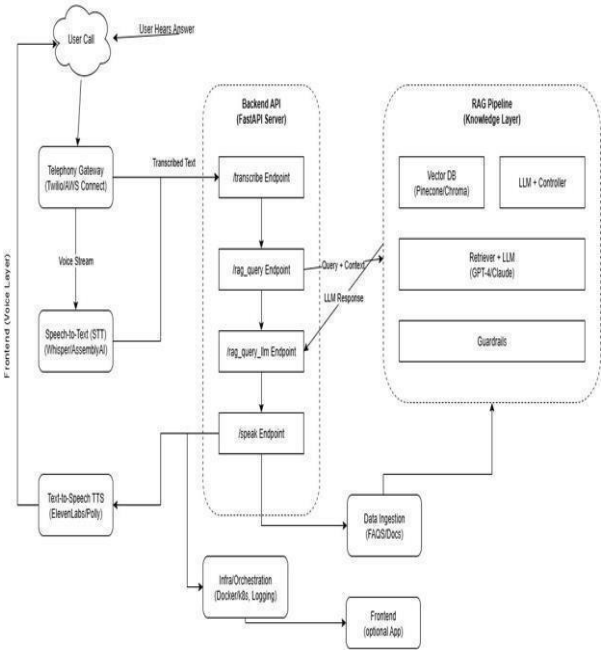

Figure 1 System Architecture

The major components of the system are described be- low:

- User Call (Frontend – Voice Layer): The process starts when a user places a call through telephony gate- ways such as Twilio or AWS Connect, which route the call to the system.
- Speech-to-Text (STT): The user's spoken input is converted into text using tools like Whisper or AssemblyAI. The transcribed text is then forwarded to the backend API for further processing.
- Backend API (FastAPI Server): The backend, built using FastAPI, coordinates communication between different modules through multiple endpoints:
  – transcribe – Receives the transcribed text from the STT module.
  – rag query – Sends the user query along

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3377-3382
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0435

with contextual information to the RAG pipeline.

– rag query llm – Connects with the Large Language Model (LLM) to generate intelligent responses.

– speak – Converts the final AI-generated text out- put back into speech.

- RAG Pipeline (Knowledge Layer): This layer per- forms information retrieval and reasoning to produce accurate and context-aware answers. It includes:

  – Vector Database (Pinecone/Chroma): Stores and retrieves knowledge base content such as FAQs and documents.

  – LLM + Controller: Oversees and manages the operation of the language model (e.g., GPT-4 or Claude).

  – Retriever + LLM: Fetches the most relevant in- formation and generates the final AI-based re- sponse.

  – Guardrails: Ensure that responses remain accurate, safe, and contextually appropriate.
- Data Ingestion: Handles the addition of FAQs, documents, and other knowledge sources into the database to enhance system accuracy and efficiency.
- Text-to-Speech (TTS): After the backend generates the response, tools such as ElevenLabs or Amazon Polly convert the text into audio so the user can hear the reply.
- Infrastructure and Orchestration: System deployment, scaling, and monitoring are handled using Docker, Kubernetes (K8s), and logging tools to ensure reliable and scalable operation.
- Frontend (Optional Application): In addition to voice interaction, the system may also include a web or mobile interface for users who prefer text-based communication.

7. Implementations

The proposed system is implemented using artificial intelligence and speech processing technologies. The system integrates voice recognition, language processing, and automated response generation modules to handle customer queries effectively. Technologies used in the system include:

- Programming Language: Python is used for system development and AI integration.
- Speech Recognition: converts user voice input into text for processing.
- Natural Language Processing: techniques are used to analyze user queries and detect intent.
- Machine Learning models are used to classify customer queries into different categories.
- Text-to-Speech engine converts generated responses into voice output.
- Database stores frequently asked questions, product details, and customer support information.
- Web frameworks: such as Flask or Django are used for building the backend system.

These technologies collectively enable the system to process voice commands, understand user intent, and generate appropriate responses.

8. Results and Discussion

The proposed system was tested using different customer queries to evaluate its performance and response accuracy. The system successfully converted voice input into text and identified the correct intent for most queries shown in Figures 1-4.

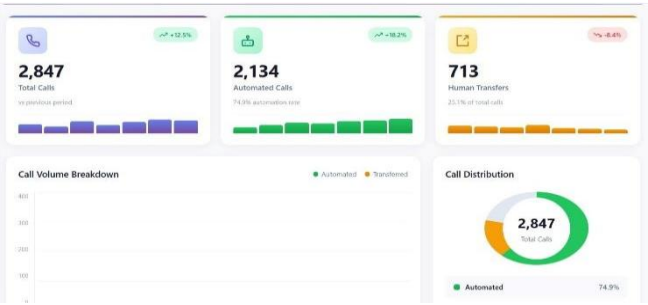

Figure 2 Call Volume Breakdown

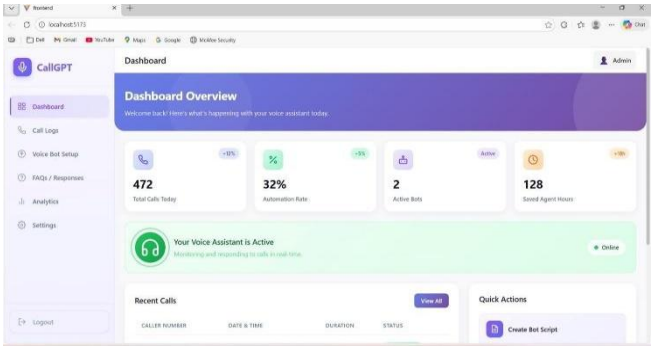

Figure 3 Dashboard Overview

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3377-3382
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0435

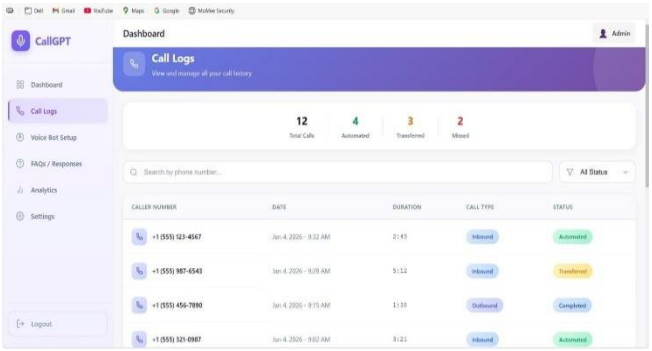

Figure 4 Call Logs

9. Advantages

The proposed voice assistant system offers several advantages.

- It provides 24/7 customer support without human intervention and reduces the workload on human customer care representatives.
- The system ensures faster response time compared to traditional support systems and is capable of handling multiple customer queries simultaneously.
- It improves customer satisfaction by providing quick and efficient assistance. Additionally, the system helps in reducing operational costs for organizations.

10. Applications

The AI-based voice assistant system can be applied in various industries for automated customer service. Major application areas include

- banking and financial customer support
- e-commerce order tracking and product inquiries, telecommunication service assistance
- healthcare appointment and information systems, government helpline services
- technical support centers.

These applications help organizations improve service quality and enhance customer interaction.

11. Future Work

Although the proposed system provides efficient automated customer support, further improvements can enhance its performance.

Possible future enhancements include:

- Support for multiple regional languages.
- Integration with mobile applications and smart devices.
- Emotion detection to understand user sentiment.
- Advanced deep learning models for better conversation understanding.
- Personalized responses based on customer history.
- Integration with real-time databases and cloud services.

Conclusion

The integration of Retrieval-Augmented Generation (RAG) with speech technologies has created new possibilities for building intelligent, voice-based customer care systems. An analysis of existing RAG architectures, vector databases, and multimodal learning approaches shows that combining Large Language Models (LLMs) with real-time voice interaction can greatly improve both user experience and service efficiency. The findings from earlier research provide a strong basis for developing an AI-based voice assistant capable of delivering context-aware, multilingual, and natural conversational support through phone calls. This study emphasizes the importance of dependable Speech-to-Text (STT), contextual information retrieval, and Text-to-Speech (TTS) components working together within a coordinated AI pipeline. By using these advancements, the proposed system can help narrow the gap be- tween automated responses and human-like communication. Future work will focus on improving speech recognition accuracy for regional languages, reducing latency in real- time interactions, and expanding the knowledge base to better address a wide range of customer queries.

References

[1] A Comprehensive Platform for Resume Building,Job Serach, Matching, and skill Enhancement, Narayan Attarde, Piyush Kulkarni,Yash Vaidya, Akshad Shelare, Meet Sali,GRADIVA REVIEW JOURNAL,2024, vol-10 issue-3

[2] S. Veturi, S. Vaichal, R. L. Jagadheesh, N. I. Tripto, and N. Yan, “RAG-based Question Answer- ing for Contextual Response Prediction System,” Proc. 1st Workshop on GenAI and RAG Systems for Enterprise at CIKM, arXiv Preprint, 2024. DOI:

10.48550/arXiv.2409.03708.

[3] J. Antony, M. Trovati, and S. Bolton, "Retrieval- Augmented Generation to Generate Knowledge As- sets and Creation of Action Drivers," Applied Sci- ences (MDPI) – International Journal, 2025.

[4] D. Rau, S. Wang, H. De´jean, and S. Clinchant, "Con- text Embeddings for Efficient Answer Generation in RAG," arXiv Preprint, 2024.

[5] A. Uzok, D. Cadet, and P. Ojukwu, "Leveraging AI- Powered Chatbots to Enhance Customer Service Ef- ficiency and Future Opportunities in Automated Sup- port," Computer Science & IT Research Journal – In- ternational Journal, 2024.

[6] Y. Han, C. Liu, and P. Wang, "A Comprehensive Sur- vey on Vector Databases," arXiv Preprint – Survey Paper, 2024. DOI: 10.48550/arXiv.2310.11703.

[7] I. Malkiel, U. Alon, Y. Yehuda, S. Keren, O. Barkan, R. Ronen, and N. Koenigstein, "GPT-Calls: Enhanc- ing Call Segmentation and Tagging by Generating Synthetic Conversations via Large Language Mod- els," Proc. Int. Conf. on Information and Knowledge Management (CIKM), IEEE Conference, 2023.

[8] A. Shafeeg, I. Shazhaev, D. Mihaylov, A. Tu- larov, and I. Shazhaev, "Voice Assistant Inte- grated with Chat GPT," Indonesian Journal of Com- puter Science – International Journal, 2023. DOI: 10.33022/ijcs.v12i1.3146.

[9] H. Palivela, M. Narvekar, D. Asirvatham, S. Bhushan, V. Rishiwal, and U. Agarwal, "Code-Switching ASR for Low-Resource Indic Languages: A Hindi-Marathi Case Study," IEEE Access, 2023. DOI: 10.1109/AC- CESS.2025.3527745.

[10] S. Khan and M. Iqbal, "AI-Powered Customer Ser- vice: Does it Optimize Customer Experience?," Proc. 8th Int. Conf. on Reliability, Infocom Technologies and Optimization (ICRITO), IEEE Conference, 2020.

[11] F. Almeida and G. Xexe´o, "Word Embeddings: A Survey," arXiv Preprint – Survey Paper, 2019.

[12] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilac- qua, F. Petroni, and P. Liang, "Lost in the Middle: How Language Models Use Long Contexts," arXiv preprint arXiv:2307.03172 [cs.CL], 2023.

[13] Yu. A. Malkov and D. A. Yashunin, "Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs," arXiv preprint arXiv:1603.09320 [cs.DS], 2018 .

[14] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ Questions for Machine Com- prehension of Text," in Proc. 2016 Conf. on Em- pirical Methods in Natural Language Processing (EMNLP), Austin, Texas, pp. 2383–2392, 2016. DOI: 10.18653/v1/D16-1264.

[15] N. Reimers and I. Gurevych, "Sentence-BERT: Sen- tence Embeddings Using Siamese BERTNetworks," arXiv preprint arXiv:1908.10084 [cs.CL], 2019. Google Research,"ScaNN: Efficient Vector Similarity Search," 2020. https://github.com/google research/google/research/tree/master/scan