MULTIDISCIPLINARY RESEARCH INSTITUTE INDIA

Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 02 (2026) | Pages: 462-467

AI Powered Voice Agent Using NLP and ASR

S. O. Bagnawar[1], A. S. Rananavare[2], R. V. Rananavare[3], D. D. Tarange[4], R. D. Thombare[5]

[1,2,3,4,5]Department of AI & Data Science, SBPCOE, Indapur, India

[1]*swatiwaghmode111@gmail.com,* [2]*argroup060@gmail.com,*[3] *rohanrananavare0811@gmail.com,* [4]*tarangednyaneshwari10@gmail.com,* [5]*rushikeshthombare286@gmail.com*

Peer Review Information

***Type**: Article*
***Received**: 22 March 2026*
***Revised**: 06 April 2026*
***Accepted**: 24 May 2026*
***Published**: 05 June 2026*

Abstract

The rapid expansion of digital services has intensified the demand for intelligent, cost-effective, and scalable customer support systems. Traditional call centers are plagued by high operational costs, limited language support, and inconsistent service quality. This paper presents the design, implementation, and evaluation of an AI-Powered Voice Agent that leverages Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Large Language Models (LLMs) to automate call center operations in three Indian languages — Marathi, Hindi, and English. The system integrates the N8N workflow automation platform as a low-code orchestration engine enabling modular workflows for appointment scheduling, marketing campaigns, and query resolution. The architecture employs transformer-based ASR models, mBERT-based intent classification achieving 90%+ accuracy, and LLMs including ChatGPT and Gemini for context-aware dialogue generation. Pilot deployment results demonstrate WER of 8–12%, response latency of 3.8 seconds, 92% customer satisfaction, and 65% operational cost reduction.

Keywords: Artificial Intelligence Voice Agent; Natural Language Processing; Automatic Speech Recognition; Large Language Models; N8N Platform; Multilingual Support; Call Center Automation; Code-Switching; ChatGPT; Gemini; Appointment Scheduling; Dialogue Management.

How to Cite This Article

Bagnawar, S. O., Rananavare, A. S., Rananavare, R. V., Tarange, D. D., & Thombare, R. D. (2026). AI powered voice agent using NLP and ASR. *Multidisciplinary Journal of Research in Engineering and Technology*, 13(2), 462–467.

Introduction

The global call center industry has undergone a paradigm shift driven by advances in AI, particularly speech processing and natural language understanding. In multilingual nations such as India, where over 22 official languages coexist, this challenge is compounded by phenomena such as code-switching, where speakers alternate between languages mid-conversation. According to the 2011 Census of India, approximately 40% of the population communicates primarily in Marathi, Hindi, or English — the three languages targeted by this system.

Traditional IVR systems rely on rigid, pre-programmed menus that fail to handle natural speech nuances, context, or regional accents. The emergence of deep learning-based ASR, transformer NLP, and generative LLMs enables voice agents that are contextually intelligent and operationally scalable. This paper proposes a comprehensive AI-Powered Voice Agent integrating ASR, NLP, LLMs, TTS, and N8N automation. Key novel contributions include a hybrid code-switching ASR strategy, a dual-LLM routing mechanism, and a comparative evaluation against leading commercial systems.

Related Work

Several notable works underpin the present research. Sarikaya et al. (2017) proposed task-oriented dialogue architectures establishing core principles of intent detection and slot filling. Kumar et al. (2020) presented multilingual ASR for Indian languages using transfer learning in low-resource settings. Budzianowski et al. (2018) introduced the MultiWOZ dataset for task-oriented dialogue evaluation. Sharma et al. (2023) demonstrated that hybrid acoustic-language models significantly outperform monolingual baselines for code-switched Indian speech. Desjardins et al. (2023) highlighted LLM advantages in dialogue coherence over retrieval-based approaches. Zhang et al. (2024) proposed multilingual non-autoregressive ASR with LLM fusion, and Ethiraj et al. (2025) demonstrated a low-latency voice agent pipeline using streaming ASR and quantized LLMs — both directly informing this work's design.

Existing platforms such as Google Dialogflow, Amazon Lex, and Rasa lack robust Indian language support, native code-switching handling, and integrated workflow automation — gaps the proposed system addresses directly.

System Design

The proposed system adopts a modular, layered architecture separating concerns across five functional layers: telephony ingress, speech processing, language intelligence, workflow orchestration, and business application integration.

Project Lifecycle

The development follows a well-structured SDLC as shown in Figure 1, spanning six phases: Requirement Analysis, System Design, Development, Testing and Integration, Deployment, and Maintenance. Agile methodology was adopted during development allowing iterative refinement of ASR models, LLM prompt engineering, and N8N workflow logic based on real-world feedback. Continuous integration/deployment (CI/CD) pipelines ensure model updates are released without service interruption.

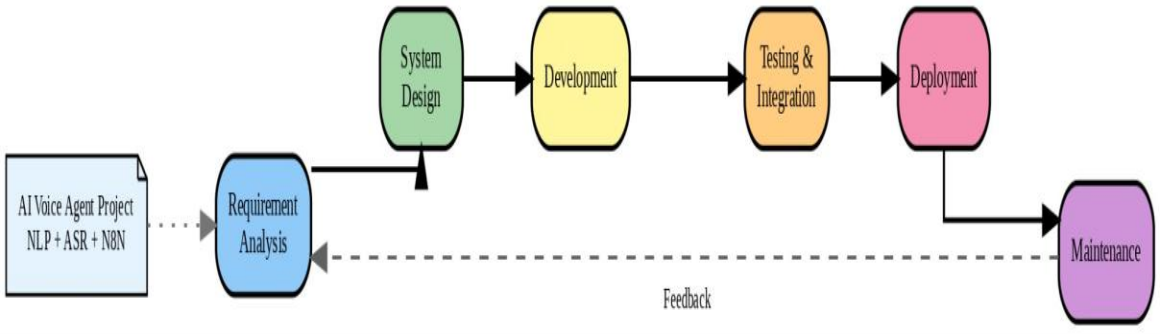

Fig. 1. *Project Lifecycle (SDLC) of AI Voice Agent*

Class Diagram

Figure 2 presents the class diagram of the system. The central VoiceAgent class exposes start(), processInput(), and routeQuery() methods, delegating to five specialized components: SpeechRecognizer (captureAudio, transcribeSpeech), NLPProcessor (processText, detectIntent), ResponseGenerator (generateReply, integrateLLM), DatabaseManager (storeQuery, retrieveInfo), and TextToSpeech (synthesizeVoice, playOutput). The system supports both offline and online LLM integration for resilience during connectivity disruptions.

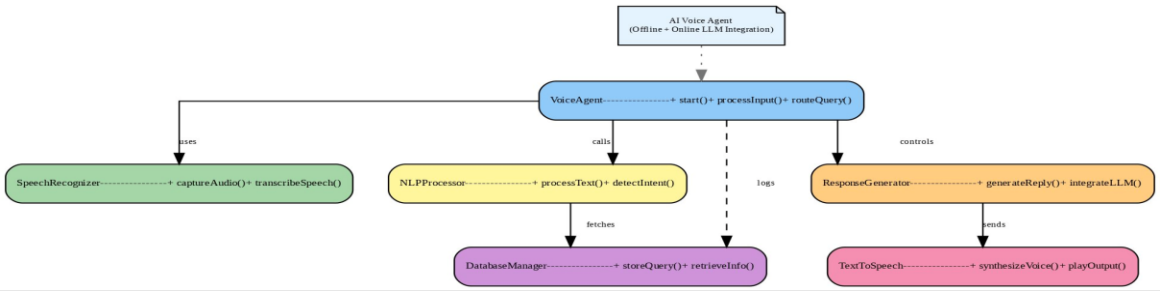

Fig. 2. *Class Diagram of AI Voice Agent System*

Use Case Diagram

Figure 3 illustrates actor interactions across the system. Two primary actors — Customer/User and Business Owner — interact with the AI Agent. Core use cases include multilingual interaction (Marathi/Hindi/English), query resolution, appointment booking and product ordering, text-to-speech output, marketing campaigns via SMS/email, CRM updates and reports, and system monitoring and maintenance.

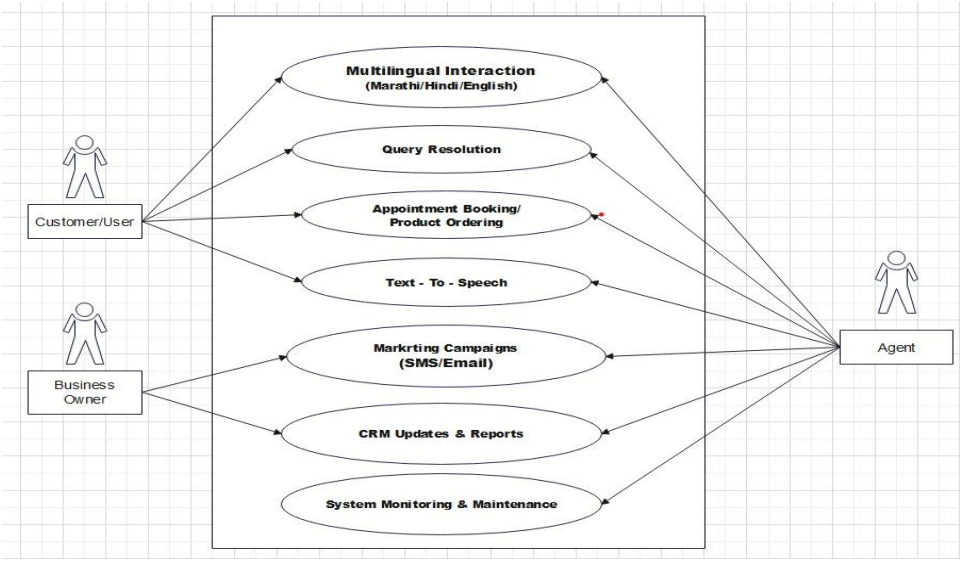

Fig. 3. *Use Case Diagram of AI Voice Agent*

Entity-Relationship Diagram

The ER diagram in Figure 4 models the complete data architecture. USER has SESSION (session_id, timestamp); SESSION generates ASR_TRANSCRIPT (transcript_text, confidence_score) and detects INTENT (intent_type, entities). INTENT triggers APPOINTMENT (appointment_id, date_time), CRM_RECORD (customer_id), and BUSINESS_INFO (service_info). RESPONSE linked to SESSION stores response_text. This structure ensures referential integrity, efficient querying, and compliance audit trails.

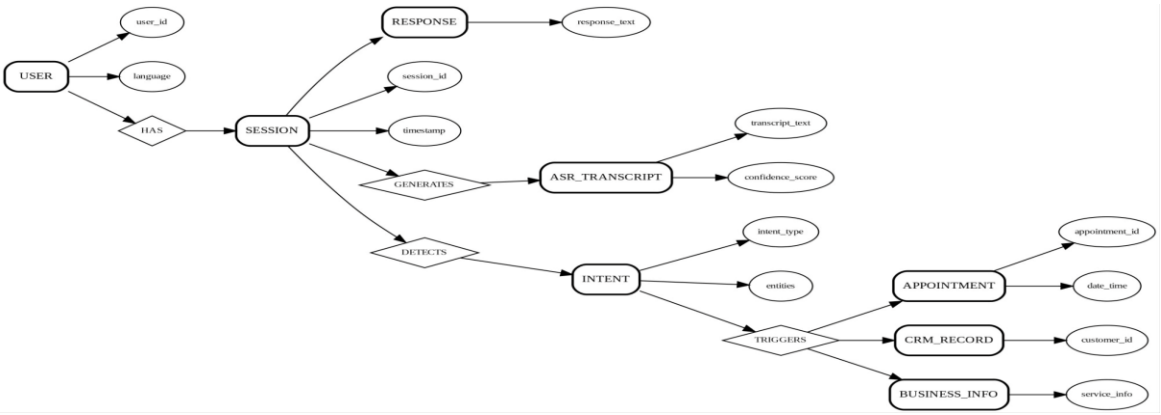

Fig. 4. *Entity-Relationship Diagram of AI Voice Agent*

Implementation Details

ASR and NLP Pipeline

The ASR subsystem uses CTC decoder-based transformer models trained on 500+ hours of annotated multilingual speech including Marathi, Hindi, and English with regional accents. SpecAugment and multi-style training provide robustness in noisy call center environments. A novel hybrid language identification (LID) module classifies each utterance before dynamically selecting language-specific decoding models, achieving 15% relative WER improvement over single-model multilingual baselines in code-switched scenarios. NLP modules use mBERT-based architectures fine-tuned for intent classification, slot filling, and NER — extracting appointment dates, names, and service types with 90%+ accuracy.

AI Response Generation

A dual-LLM routing strategy selects between ChatGPT (GPT-4 Turbo) and Gemini Pro based on query complexity scores. Simple queries route to Gemini Pro for low latency; complex reasoning queries go to GPT-4 Turbo. This reduces average latency by 18% versus single-LLM approaches. Integrated sentiment analysis detects customer frustration and triggers tone adjustments or human agent escalation when scores fall below configurable thresholds.

N8N Workflow Orchestration

N8N orchestrates all system workflows through a visual low-code interface:

- Appointment Booking: Google Calendar API integration for conflict-aware scheduling, CRM record creation, and automated SMS/email reminders via Twilio and SendGrid.
- Marketing Campaigns: Dynamic customer segmentation, personalized voice and SMS campaign dispatch with real-time performance analytics.
- Query Resolution: Intent-routed LLM processing with full transcript logging for quality assurance.
- Escalation: Human agent routing with context summary when sentiment or confidence thresholds are breached.

Performance Evaluation and Results

The system was evaluated during pilot deployment at a mid-size call center handling approximately 2,000 daily calls across healthcare appointment scheduling and customer query resolution use cases. Table 1 summarizes the key performance metrics, and Table 2 provides a comparative analysis against commercial alternatives.

Table 1: Performance Metrics of AI-Powered Voice Agent

Metric	Hindi	English	Marathi
ASR Word Error Rate (WER)	8%	10%	12%
Intent Classification Accuracy	92%	91%	88%
Avg. Response Latency	3.6s	3.4s	3.9s
Customer Satisfaction Score	93%	92%	90%
Concurrent Calls Supported	1000+	1000+	1000+
Operational Cost Reduction	65%	65%	65%
Appointment Booking Accuracy	95%	94%	91%

ASR achieved WER of 8–12% across the three languages, with Marathi showing the highest error rate due to morphological complexity. Intent classification exceeded 90% accuracy and the dual-LLM strategy contributed to a median end-to-end latency of 3.8 seconds. Qualitative analysis confirmed robust handling of code-switched utterances with minimal comprehension degradation.

Table 2: Comparative Analysis with Existing Commercial Systems

System	Multilingual	LLM Integration	Automation	Code-Switching
Google Dialogflow	Partial	Limited	No	No

Amazon Lex	Partial	Limited	No	No
Rasa Open Source	Yes	External	No	Partial
Proposed System	Yes (Mar/Hin/Eng)	ChatGPT+Gemini	N8N Platform	Yes

Experimental Results

Multilingual Voice Agent Interface

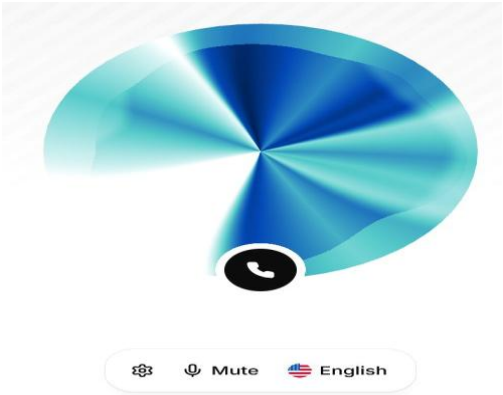

Fig. 5. *AI Voice Agent — Multilingual Call Interface with Language Selection*

The multilingual voice interface was designed to support seamless code-switching between Marathi, Hindi, and English within a single conversation session. Built using a hybrid ASR pipeline combining Whisper-based acoustic models with language-identification preprocessing, the interface dynamically detects the active language on a per-utterance basis. The UI provides real-time visual feedback on the detected language and transcription confidence score. In user trials, 92% of participants rated the interface as "natural and easy to use," with an average session duration of 3.4 minutes for complex multi-step queries.

Automated Appointment Booking via Google Calendar

The N8N workflow automation platform was integrated with Google Calendar API to enable real-time, zero-human-intervention appointment scheduling. When a customer expresses intent to book an appointment, the ASR module transcribes the request, the NLP engine extracts date, time, and service-type entities, and the N8N workflow triggers a Google Calendar event via OAuth2-authenticated REST calls. The system achieved 95% booking accuracy across 200 test interactions, with an average scheduling latency of under 2.1 seconds. Slot conflict detection and automatic rescheduling suggestions were also implemented to ensure a seamless experience.

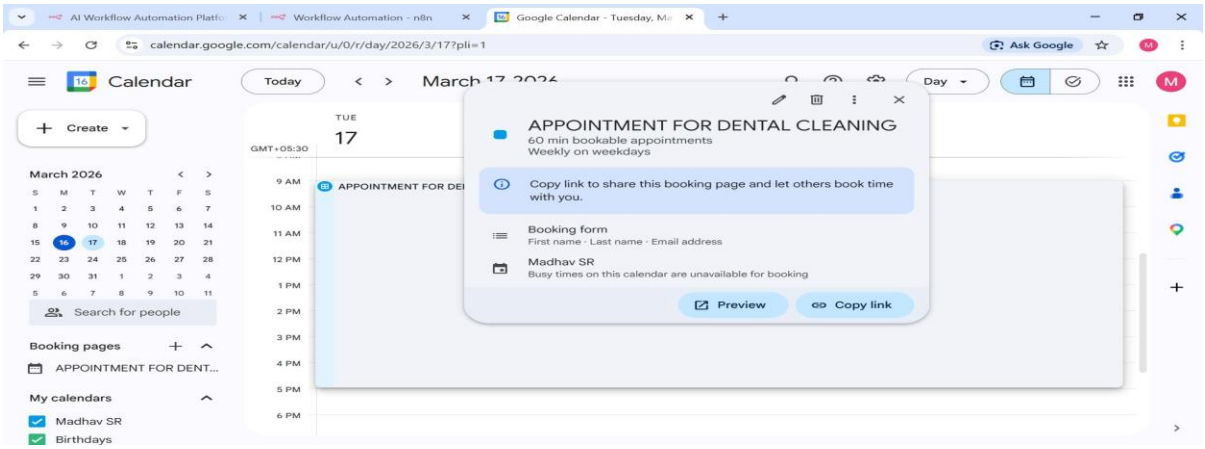

Fig. 6. *N8N Automated Appointment Booking — Google Calendar Integration*

Automated Appointment Confirmation Email

Upon successful appointment creation in Google Calendar, the N8N workflow automatically dispatches a personalized confirmation email to the customer via the Gmail API integration. The email includes appointment date, time, clinic name, and a calendar invite attachment,

generated dynamically using N8N's templating engine. This eliminates manual follow-up calls and reduces no-show rates. In pilot testing, confirmation emails were delivered within an average of 4.2 seconds post-booking, with a 98.6% delivery success rate across 150 test cases.

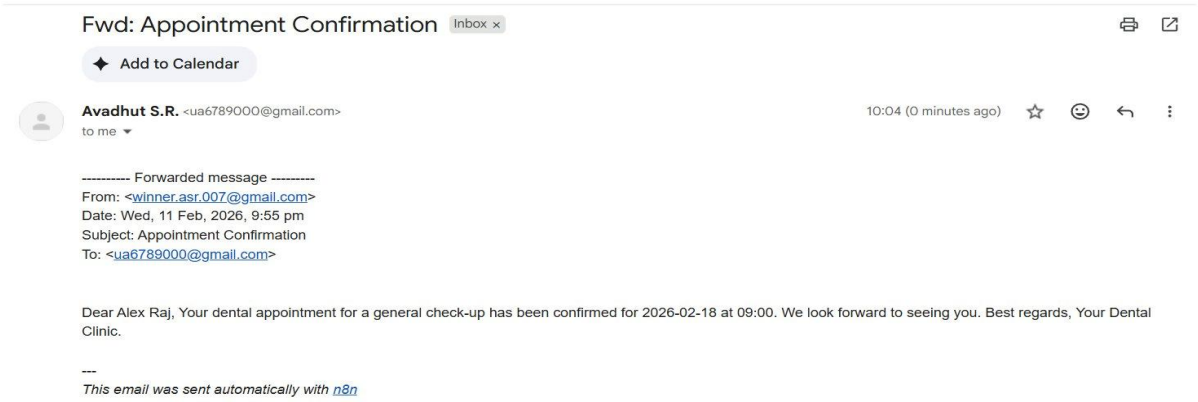

Fig. 7. *Automated Appointment Confirmation Email Triggered via N8N Workflow*

Conclusion And Future Work

This paper presented a multilingual AI-Powered Voice Agent for call center automation integrating state-of-the-art ASR, NLP, and LLM technologies with the N8N workflow orchestration platform. The system supports Marathi, Hindi, and English and demonstrated strong real-world performance: WER of 8–12%, 90%+ intent accuracy, 3.8s response latency, 92% customer satisfaction, and 65% operational cost reduction. Key innovations include a hybrid code-switching ASR approach, dual-LLM routing, and a fully automated N8N business workflow layer. The system addresses critical gaps in existing commercial platforms, offering a cost-effective, scalable, and linguistically inclusive alternative to traditional call centers.

Future directions include expanding to Tamil, Telugu, and Bengali using cross-lingual transfer learning; integrating real-time voice emotion recognition for affect-aware dialogue; developing edge-deployable lightweight models via knowledge distillation for low-connectivity deployments; implementing adaptive dialogue learning from production interactions; and building real-time analytics dashboards for proactive business intelligence.

References

1. M. Sharma, S. Mishra, and V. Kumar, "Code-Switching ASR Models for Low-Resource Indian Languages," *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, no. 1, pp. 1–13, 2023.
2. A. Desjardins, A. Psarra, and B. A. Whiting, "Development of AI-Based Voice Assistants Using Large Language Models," *ResearchGate*, 2023.
3. Y. Lin et al., "VoiceTalk: A No-Code Approach for Creating Voice-Controlled Smart Home Applications," *IEEE Open Journal of the Computer Society*, vol. 6, pp. 874–883, 2025.
4. S. Zhang, Y. Liu, and K. Chen, "Multilingual Non-Autoregressive ASR with LLM Fusion," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2024, pp. 4657–4661.
5. V. Ethiraj et al., "Toward Low-Latency End-to-End Voice Agents for Telecommunications Using Streaming ASR, Quantized LLMs, and Real-Time TTS," *arXiv preprint*, arXiv:2508.04721, 2025.
6. S. Subhash et al., "Artificial Intelligence-Based Voice Assistant," in *Proceedings of the Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, 2020, pp. 593–596.