Tele-AI – Conversational AI Voice Agent for Telecom Customer Support

Project Team

Zaid Basharat 22i-2513
Altaf Ur Rehman 22i-2519
Muhammad Ali 22i-1516

Session 2022-2026

Supervised by

Ms. Mahnoor Tariq

Department of Software Engineering

National University of Computer and Emerging Sciences
Islamabad, Pakistan

June, 2026

Contents

1 Problem Statement **3**

1.1 Introduction . . . 3
1.2 Problem Statement . . . 3
1.3 Motivation . . . 4
1.4 Problem Solution . . . 4
1.5 Stake Holders . . . 5

2 Project Description **7**

2.1 Scope . . . 7
2.2 Modules . . . 8
2.2.1 Speech Processing Module . . . 8
2.2.2 Dialogue Management Module . . . 8
2.2.3 Knowledge Management Module . . . 9
2.2.4 CRM Integration Module . . . 9
2.2.5 Webhooks and Workflow Automation Module . . . 9
2.2.6 Admin Console Module . . . 10
2.2.7 Analytics and Monitoring Module . . . 10
2.3 Tools and Technologies . . . 10
2.4 Work Division . . . 11
2.5 Timeline . . . 12

References **13**

List of Figures

1 Architecture Diagram 1

List of Tables

2.1 Work Division 11
2.2 Project Timeline 12

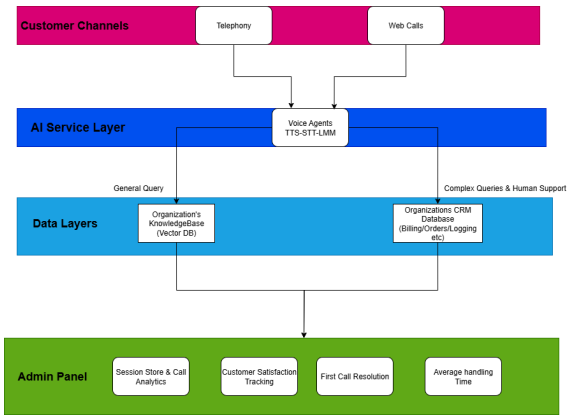

Figure 1: Architecture Diagram

CHAPTER 1

PROBLEM STATEMENT

1.1 Introduction

This proposal specifies the purpose, scope, and design of a conversational AI system tailored for telecom customer support

1.2 Problem Statement

The telecom industry continues to face growing challenges in managing customer support services effectively. Traditional call centers rely heavily on large human teams, which are expensive to maintain and difficult to scale as customer bases grow. Customers often face long waiting times, especially during peak hours, which negatively impacts satisfaction and trust. Human agents, while capable, are subject to fatigue, stress, and inconsistency, leading to variable service quality. Many customers demand 24/7 availability, but maintaining night shifts and holiday staff is costly and impractical for most telecom companies. Complaints such as network issues, billing disputes, and package inquiries are repetitive and consume significant agent time, even though they could be automated. In many cases, first call resolution rates are low because agents must escalate issues to higher levels. This creates frustration for customers and operational inefficiency for organizations. The lack of intelligent automation also limits the ability of companies to serve thousands of concurrent customers at once. Current IVR systems provide only basic menu-driven support and fail to deliver natural, human-like experiences. As a result, many customers abandon calls or churn to competitors offering better support systems. Moreover, telecom companies need consistent feedback on performance metrics like Customer Satisfaction (CSAT), Average Handling Time (AHT), and First Call Resolution (FCR), but existing human-driven processes often lack structured data collection. Without automation, orga-

nizations continue to overspend on human resources while underperforming in customer experience. There is a clear need for an AI-powered conversational system that addresses these challenges. Such a system must be cost-effective, scalable, always available, and able to simulate human-like interactions while providing measurable improvements in KPIs.

1.3 Motivation

The motivation behind this project arises from the critical role that customer service plays in the success of telecom companies. In highly competitive markets, customer loyalty is determined not only by service quality but also by how effectively issues are resolved. Customers expect immediate assistance, but most call centers fail to deliver during peak hours. This gap between expectation and delivery highlights the importance of automated, always-available systems. Advances in artificial intelligence and natural language processing now allow us to simulate near-human conversations, making it feasible to automate Tier-1 customer support. The availability of pre-trained speech-to-text engines and large language models provides a strong foundation for building telecom-specific solutions. By fine-tuning these models on telecom data, accuracy in speech recognition and response generation can be improved significantly. At the same time, retrieval-augmented generation (RAG) allows systems to dynamically fetch information from a telecom knowledge base, ensuring accuracy and adaptability. Automating repetitive tasks such as package inquiries, complaint registration, and service activation reduces pressure on human agents. Human resources can then focus on more complex queries requiring empathy and higher-level problem-solving. The inclusion of analytics also provides organizations with real-time insights into customer satisfaction and operational efficiency. This not only reduces cost but also creates measurable business value. The ability to provide consistent and scalable support directly impacts brand reputation and customer retention. Therefore, developing Tele-AI is strongly motivated by both technical feasibility and industry demand.

1.4 Problem Solution

The proposed solution, Tele-AI, is a managed multi-tenant platform that provides telecom organizations with conversational AI voice agents. The system leverages speech-to-text engines fine-tuned on telecom industry data to ensure accurate transcription of customer queries. Once transcribed, the text is processed by an intelligent language model integrated with a retrieval system that accesses the organization's knowledge base. This enables real-time, contextually accurate responses tailored to telecom operations. The

system will support key functions such as complaint registration, package inquiry, and service requests. To demonstrate CRM readiness, the system will integrate with mock APIs that simulate service activations, cancellations, and billing inquiries. Escalation mechanisms will allow unresolved issues to be forwarded to human agents, ensuring a smooth handoff. Tele-AI will operate as a managed service, with the internal project team configuring prompts, embeddings, and knowledge bases for each organization through an admin console. This reduces complexity for organizations and ensures security, compliance, and quality. The system will also support real-time webhooks and workflow automation to connect with external services such as notification systems, scheduling platforms, and CRM updates. This ensures adaptability and seamless integration into existing telecom infrastructures.

The platform will also track performance metrics such as FCR, CSAT, and AHT, enabling continuous improvement. The design focuses on Tier-1 customer support, which is repetitive and well-suited for automation. Unlike traditional IVR systems, Tele-AI will provide natural, human-like conversations rather than rigid menu navigation. The architecture is extensible, allowing integration with real telecom CRMs in future deployments. The solution balances automation and human intervention to maximize efficiency and customer satisfaction. Overall, Tele-AI aims to deliver a cost-effective, scalable, and future-ready customer support system that addresses the shortcomings of existing processes.

1.5 Stake Holders

The success of the proposed Tele-AI system depends on multiple stakeholders, each playing a distinct role in its development, deployment, and usage. The key stakeholders are:

- **Customers:** End-users who call telecom helplines for inquiries, complaints, or service requests. They directly interact with the AI voice agent and provide feedback that measures satisfaction.
- **Telecom Providers:** Organizations that adopt the Tele-AI platform to improve efficiency, reduce operational costs, and deliver better customer service. They provide requirements, policies, and knowledge bases for system configuration.
- **Customer Support Officers:** Human agents who handle escalated cases beyond the scope of the AI system. They benefit from reduced workload and improved case prioritization.
- **Project Team (Developers):** The students responsible for designing, implementing, testing, and presenting the Tele-AI system. They ensure quality, accuracy, and extensibility.

- **Supervisors and Academic Evaluators:** Faculty members overseeing the project. They guide technical decisions, validate the methodology, and assess project outcomes.
- **Future Integrators:** Potential system administrators or third-party developers who may extend Tele-AI with real CRM systems, multilingual features, or regulatory compliance modules.

CHAPTER 2

PROJECT DESCRIPTION

2.1 Scope

The proposed system, Tele-AI, falls within the domain of intelligent customer service automation for the telecom sector. The scope of the project is limited to Tier-1 customer support, focusing on repetitive and high-volume queries that can be effectively automated. This includes package inquiries, complaint registration, billing questions, SIM activation, and basic troubleshooting guidance. The system will not attempt to replace human agents completely but will instead act as the first line of support, capable of resolving the majority of issues without escalation. For unresolved or complex queries, the system will provide a seamless handoff to a human support officer, ensuring service continuity. The voice interaction pipeline will make use of a fine-tuned speech-to-text model specialized in telecom terminology, combined with a retrieval-based knowledge system for accurate responses. An internal admin dashboard will be developed for the project team to onboard organizations, upload their knowledge bases, configure system prompts, and assign numbers for their AI agents. Direct CRM integrations will be demonstrated with mock APIs, but the architecture will remain extensible for real-world telecom systems. In addition to core conversational features, the system will provide webhook and workflow automation to allow real-time communication with third-party services. This will enable tasks like complaint status updates, automatic callback scheduling, and integration with external notification systems. The system will also provide built-in analytics for performance tracking, focusing on key performance indicators such as First Call Resolution (FCR), Customer Satisfaction (CSAT), and Average Handling Time (AHT). These analytics will allow organizations to evaluate the effectiveness of the AI agent and optimize support strategies. The platform will follow a managed model, where configurations are handled internally by the project team rather than by external organizations, ensuring quality and consistency. Furthermore, the design will emphasize scalability, making it possible to

handle thousands of concurrent calls without degradation in performance. The project will not include advanced multilingual features or field-level technical support, which remain outside the scope of this iteration. However, the architecture will allow for future expansion into these areas if required. The ultimate scope of Tele-AI is to build a reliable, always-available, and cost-effective conversational agent that demonstrates the feasibility of applying AI-driven voice technologies to telecom customer support in a controlled academic setting.

2.2 Modules

The proposed system is divided into six modules, each designed to handle a distinct responsibility within the Tele-AI platform. This modular breakdown ensures clarity, maintainability, and parallel development by different team members.

2.2.1 Speech Processing Module

This module is responsible for handling all voice-related inputs and outputs in real time.

1. Convert customer speech into text using a fine-tuned speech-to-text engine.
2. Handle noise reduction and telecom-specific vocabulary recognition.
3. Generate human-like spoken responses using a text-to-speech system.
4. Ensure smooth conversation flow with barge-in and interruption handling.

2.2.2 Dialogue Management Module

This module governs the flow of conversation and determines how the system interacts with users.

1. Identify user intent and context from transcribed queries.
2. Manage multi-turn dialogues to maintain continuity across a session.
3. Apply predefined business rules and escalation logic.
4. Route unresolved queries to the human agent escalation module.

2.2.3 Knowledge Management Module

This module handles all knowledge retrieval and contextual response generation.

1. Preprocess and ingest organization-specific knowledge bases.
2. Store and index knowledge in a vectorized format for efficient retrieval.
3. Use retrieval-augmented generation (RAG) to ground responses in relevant data.
4. Update knowledge dynamically to reflect new packages or service changes.

2.2.4 CRM Integration Module

This module ensures interaction with organizational systems for service requests.

1. Provide mock APIs to simulate telecom CRM functions (e.g., package activation, cancellation).
2. Register and track customer complaints with unique IDs.
3. Update customer status in CRM when packages are upgraded or cancelled.
4. Maintain extensibility to integrate with real CRMs in production deployments.

2.2.5 Webhooks and Workflow Automation Module

This module enables real-time communication between Tele-AI and external systems through event-driven workflows.

1. Trigger outbound webhooks when specific events occur (e.g., call started, call ended, complaint registered).
2. Receive inbound webhooks from external systems to dynamically inject data (e.g., customer name, package details).
3. Automate workflows such as scheduling callbacks, sending SMS/email confirmations, or updating records.
4. Ensure secure webhook handling with authentication and retry mechanisms.

2.2.6 Admin Console Module

This module provides internal configuration and management tools for the project team.

1. Onboard organizations into the system with unique profiles.
2. Upload and manage knowledge bases on behalf of organizations.
3. Configure system prompts, assign phone numbers, and voice settings.
4. Securely manage credentials and ensure controlled system access.

2.2.7 Analytics and Monitoring Module

This module provides insights into system performance and user experience.

1. Track KPIs such as First Call Resolution (FCR), Customer Satisfaction (CSAT), and Average Handling Time (AHT).
2. Generate reports and dashboards summarizing performance.
3. Log conversations for audit and improvement.
4. Conduct stress testing and monitor concurrent call handling capacity.

2.3 Tools and Technologies

The proposed system will be developed using a combination of artificial intelligence, web technologies, and cloud deployment tools. The tools and technologies are described in general categories to ensure flexibility and extensibility.

- **Speech Processing:** Pre-trained speech-to-text and text-to-speech engines, fine-tuned on telecom-specific data to improve recognition and response accuracy.
- **Natural Language Understanding:** Large language models and dialogue management frameworks to interpret customer intent and generate context-aware responses.
- **Knowledge Retrieval:** Vector-based databases and embedding models (e.g., Hugging Face ecosystem) for storing and retrieving knowledge base documents efficiently.

- **CRM Simulation:** Mock API services representing telecom operations such as package activation, billing inquiries, and complaint logging, designed to be extendable to real CRM systems.

- **Webhooks and Workflows:** Event-driven communication using webhook handlers and workflow automation to support callbacks, notifications, and external integrations.

- **Application Frameworks:** General-purpose backend and frontend frameworks for building scalable APIs and responsive admin dashboards.

- **Database Systems:** Relational databases to store organizational data, customer interactions, and analytics results.

- **Analytics and Monitoring:** Tools for logging, performance measurement, and KPI tracking (FCR, CSAT, AHT).

- **Cloud and Deployment:** Containerization, orchestration, and CI/CD pipelines for reliable deployment and scalability.

- **Security:** Secure credential storage, role-based access control, and data protection mechanisms to ensure privacy and compliance.

2.4 Work Division

Each team member is assigned modules and responsibilities to ensure parallel development and accountability.

Name	Reg. No.	Responsibility
Altaf Rehman	22i-2519	Speech Processing Module, Dialogue Management Module
Muhammad Ali	22i-1516	Knowledge Management Module, CRM Integration Module
Zaid Basharat	22i-2513	Admin Console Module, Analytics & Monitoring Module,
Altaf, Zaid, Ali	22i-2519, 22I-2513, 22i-1516	Webhooks & Workflow Automation Module

Table 2.1: Work Division

2.5 Timeline

The project will be executed in four iterations across eight months. Each iteration builds on the previous one and integrates assigned modules.

Iteration	Time Frame	Tasks/Modules
01	Months 1–2	Backend setup, database schema, basic voice call demo (Speech + Dialogue foundations)
02	Months 3–4	STT/LLM/TTS pipeline, knowledge base ingestion, retrieval system (Speech + Knowledge modules)
03	Months 5–6	Admin dashboard, CRM mock APIs, basic outbound webhooks (Admin + CRM + Webhooks phase 1)
04	Months 7–8	KPI tracking, escalation logic, inbound webhooks, workflow automation (Analytics + Webhooks phase 2)

Table 2.2: Project Timeline

Bibliography

[1] Josh Achiam and Steven Adler. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[2] Alexei Baevski and Yuhao Zhou. wav2vec 2.0: A framework for self-supervised learning of speech representations, 2020.

[3] Sebastian Borgeaud and Arthur Mensch. Improving language models by retrieving from trillions of tokens. *arXiv preprint arXiv:2112.04426*, 2021.

[4] Jacob Devlin and Ming-Wei Chang. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[5] Hugging Face. Hugging face: Open-source machine learning platform. https://huggingface.co/docs, 2023. Accessed: 2025-09-12.

[6] Patrick Lewis and Ethan Perez. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

[7] OpenAI. Introducing chatgpt and whisper apis. https://openai.com/blog/introducing-chatgpt-and-whisper-apis, 2023. Accessed: 2025-09-12.

[8] Alec Radford and Jong Wook Kim. Robust speech recognition via large-scale weak supervision. *arXiv preprint arXiv:2212.04356*, 2022.

[9] Ashish Vaswani and Noam Shazeer. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[10] Thomas Wolf and Lysandre Debut. Huggingface's transformers: State-of-the-art natural language processing, 2020.