S25-022-D-vocalConcierge

Project Team

Mustafa Irfan 21I-0626
Farheen Malik 21I-0454
Hamza Mehmood 19I-0458

Session 2021-2025

Supervised by

Dr. Mehreen Alam

Department of Computer Science

National University of Computer and Emerging Sciences
Islamabad, Pakistan

December, 2025

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This document outlines the development of the AI-Powered Hotel Reception Voice Bot, VocalConsigere. The purpose of this system is to enhance guest interactions by automating customer service tasks in hotels, reducing staff workload while improving efficiency and customer satisfaction. The system utilizes advanced Natural Language Processing (NLP), Speech-to-Text (STT), and Text-to-Speech (TTS) technologies to facilitate seamless real-time communication between guests and hotel management. By integrating with existing hotel management systems, it streamlines reservation management, service requests, and general inquiries. In addition, the bot ensures multilingual support and provides analytics for guest interactions, contributing to better customer service strategies. [**?**].

1.1 Problem Statement

The traditional hotel reception network has many barriers, including communication difficulties and lack of operational efficiencies. Hotels suffering from these operational problems face lower guest satisfaction, increased staff workload, and decreased hotel income. The current hands-on reservation management query response slows operations down, while increasing the chance of mistakes.

1.2 Scope

This project aims to develop an AI-powered voice bot for hotel receptions, offering an advanced guest interaction system with:

1. **Hyper-Personalization:** AI adapts to guest preferences, offering personalized recommendations.

2. **Omnichannel Support:** Works across phone, website, and smart speakers.

3. **AI-Driven Intent Recognition:** Handles inquiries about room service, pricing, amenities, booking, check-in, checkout, and feedback.

4. **Automated Service Coordination:** Connects with room service, housekeeping, and concierge.

5. **Cloud and Edge Computing:** Ensures low-latency responses and offline capabilities.

6. **Voice Commands:** Supports contactless interactions via voice.

1.3 Modules

1.3.1 Module 1: Foundation Module

This module establishes the core technical infrastructure. It includes building the Speech-to-Text (STT) system for converting voice input into text, setting up automation pipelines (CI/CD), fine-tuning pre-trained models for domain-specific use, and creating an initial UI/UX prototype.

1. Voice-to-Text (STT)

2. CI/CD and Workflow Automation

3. Fine-Tuning Models

1.3.2 Module 2: Interactive Voice Module

This module focuses on enhancing user interaction capabilities. The Text-to-Speech (TTS) system enables natural-sounding voice responses, while UI/UX refinement improves the user interface for smoother interactions and better usability.

1. Text-to-Speech (TTS)

2. UI/UX Refinement

1.3.3 Module 3: Intelligent Understanding Module

This module adds intelligence to the bot by implementing AI-powered intent recognition and SMOL Agents to understand guest queries and extract meaningful entities such as dates, names and room preferences.

1. AI-Powered Intent Recognition
2. Entity Extraction

1.3.4 Module 4: Integration and Deployment Module

This module focuses on making the system fully functional in a real world environment. It involves deploying on cloud platforms for scalability, and connecting the bot to multiple devices for seamless user interaction.

1. Cloud Deployment
2. End-Devices Integration

1.4 User Classes and Characteristics

Table 1.1: User Classes and Description

User class	Description
Hotel Guests	Guests are the primary users who interact with the AI bot for various needs, including room bookings, check-in and check-out processes, service requests, and general inquiries about hotel amenities. The AI personalizes their experience by remembering preferences and offering tailored recommendations.
Receptionists	The AI bot assists receptionists by handling routine guest inquiries, such as room availability, pricing, and booking confirmations. This allows receptionists to focus on more complex guest requests and provide better in-person service.
Hotel Managers	Managers oversee hotel operations and use AI-driven analytics to monitor guest interactions, service efficiency, and customer satisfaction. They can adjust AI responses based on guest feedback and trends to improve overall service quality.
Housekeeping Staff	The AI bot streamlines service coordination by automatically forwarding guest requests for room cleaning, wake-up calls, transportation, and special assistance. This improves response time and operational efficiency.

Chapter 2

Project Requirements

This chapter describes the functional and non-functional requirements of the project.

2.1 Use-case

The following are the primary use cases for the **Hotel Voice Bot** system, categorized according to user interaction.

2.1.1 UC01: Booking and Reservation

- **Actor:** Guest
- **Type:** Primary
- **Main Flow:**
    1. The guest calls the hotel or talk to bot at reception to check for availability and pricing of the rooms. Then he/she requests a room according to their requirements.
    2. The voice bot captures and processes the request.
    3. The bot then confirms the reservation and add it to database.

2.1.2 UC02: Request Housekeeping and Room Service

- **Actor:** Guest
- **Type:** Primary

- **Main Flow:**

  1. Guest makes a voice request for housekeeping (e.g., "Please clean my room" or "I want to order a cheeseburger") or requests for room service (e.g., "I need extra towels").
  2. The voice bot processes the request and forwards it to the housekeeping or kitchen staff.
  3. The housekeeper schedules the service and notifies the guest.

2.1.3 UC03: Request Hotel Information

- **Actor:** Guest
- **Type:** Primary
- **Main Flow:**

  1. The guest asks the voice bot for hotel-related information (e.g., "What are the spa hours?").
  2. The voice bot retrieves relevant information from the hotel database.
  3. The bot provides a verbal response to the guest.

2.1.4 UC04: Feedback Collection

- **Actor:** Guest
- **Type:** Primary
- **Main Flow:**

  1. At the end of stay, during check-out process, the bot asks the guests to provide any feedback or suggestion.
  2. The guest replies accordingly and feedback is saved in database for future improvements.
  3. Another use of this is that if a guest wants to complain about something during the stay, the can call the bot and submit their issue or complaint.

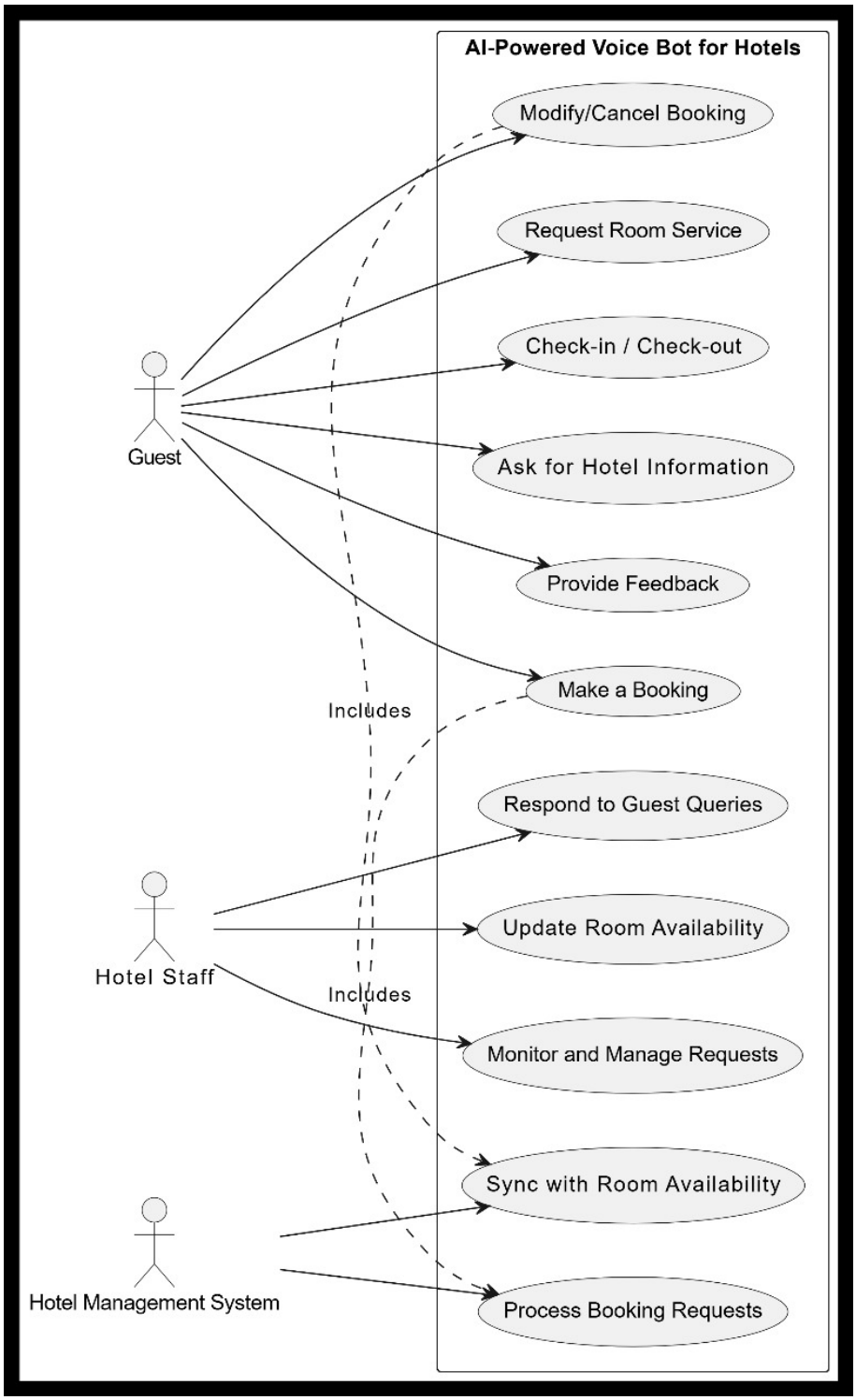

Figure 2.1: Use Case Diagram

2.2 Functional Requirements

This section describes the functional requirements of the Hotel Voice Bot, outlining the essential features and capabilities that the system must fulfill. The requirements are categorized into four core modules: Foundation Module, Interactive Voice Module, Intelligent Understanding Module, and Integration and Deployment Module. Each module plays a crucial role in ensuring the system's efficiency, accuracy, and user experience. The following sections detail the specific functional requirements associated with each module.

2.2.1 Module 1: Foundation Module

Following are the requirements for the Foundation Module:

1. FR1: The system will initialize and configure essential backend services.
2. FR3: The system will store and retrieve system configurations for optimal performance.

2.2.2 Module 2: Interactive Voice Module

Following are the requirements for Interactive Voice Module:

1. FR1: The system will capture and process guest voice commands in real time.
2. FR3: The system will apply noise reduction techniques to improve voice recognition accuracy.
3. FR4: The system will provide immediate voice feedback and acknowledgment of guest inputs.

2.2.3 Module 3: Intelligent Understanding Module

Following are the requirements for Intelligent Understanding Module:

1. FR1: The system will analyze guest voice inputs and classify them into predefined request types.
2. FR2: The system will be able to understand contextual meaning and intent.
3. FR3: The system will personalize responses based on guest history and preferences.
4. FR4: The system will handle ambiguous queries by asking clarifying questions.

2.2.4 Module 4: Integration and Deployment Module

Following are the requirements for Integration and Deployment Module:

1. FR2: The system will support cloud-based and on-premises deployment options.
2. FR3: The system will provide API endpoints for third-party service integrations.
3. FR5: The system will log interactions for analytics and continuous improvement.

2.3 Non-Functional Requirements

This section specifies nonfunctional requirements.

2.3.1 Reliability

1. The system will maintain an MTBF of at least 1000 hours, ensuring minimal failures and continuous operation.
2. The system will provide automatic failure detection and notify administrators within 10 seconds of an issue.
3. The system will support real-time monitoring of critical functions, including voice recognition accuracy and response latency.
4. The system will have automated backup mechanisms to prevent data loss and ensure service continuity.
5. The system will automatically attempt recovery within 60 seconds after detecting a failure.

2.3.2 Usability

1. The system will provide a natural and intuitive voice interface, allowing guests to make inquiries and requests without requiring prior training.
2. The system will respond to user queries within 3 seconds to ensure a smooth and efficient interaction.
3. The system will allow guests to correct misinterpreted voice inputs by rephrasing their request.

4. The system will provide clear audio feedback and confirmations after processing requests to minimize user errors.

2.3.3 Performance

1. The system will process and respond to voice queries within 3 seconds to ensure real-time interaction.
2. The system will handle at least 100 concurrent user interactions without performance degradation.
3. The system will achieve at least 95% accuracy in speech recognition under optimal conditions.
4. The system will process and analyze guest intents within 2 seconds to provide timely responses.
5. The system will efficiently manage background processes, such as logging and analytics, without impacting response time.

2.3.4 Security

1. The system will ensure that all voice and text interactions are encrypted to protect guest privacy.
2. The system will restrict unauthorized access to sensitive data, allowing only authenticated users to interact with administrative functions.
3. The system will log all interactions and access attempts to enable auditing and anomaly detection.
4. The system will sanitize all user inputs to prevent injection attacks and unauthorized system commands.
5. The system will ensure compliance with relevant data protection regulations by securely storing and processing guest information.

2.4 Domain Model

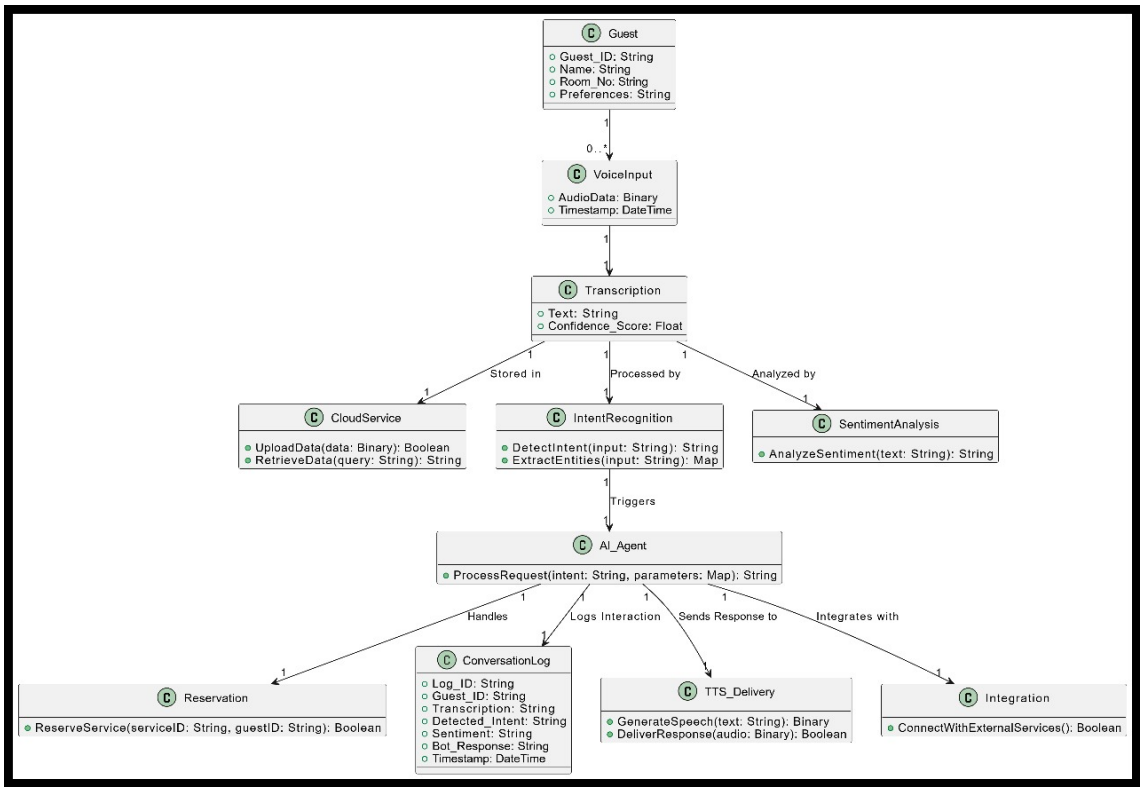

Figure 2.2: Domain Model

Chapter 3

System Overview

The AI-powered voice assistant for hotel guest services is designed to provide seamless and natural interactions, allowing guests to make inquiries, request services, manage bookings, and receive personalized assistance using voice commands. Leveraging advanced speech-to-text (STT) and text-to-speech (TTS) technologies, the system ensures accurate voice recognition and responsive audio feedback. It integrates with the hotel's management system to automate service requests such as housekeeping, dining, and concierge assistance, improving operational efficiency and reducing manual workload. AI-driven intent recognition enhances the system's ability to understand guest queries, while real-time data synchronization ensures prompt service fulfillment.

The system is built with a user-friendly interface that prioritizes ease of use and accessibility. With support for multiple accents and a fast response time, it ensures a smooth and efficient interaction experience for a diverse range of guests. Additionally, the system adapts to guest preferences over time, offering a more personalized service experience. Robust security measures are implemented to ensure data privacy and confidentiality, making it a reliable and secure solution for modern smart hotels.

3.1 Architectural Design

The AI-powered hotel voice assistant follows a modular, service-oriented architecture, designed for scalability, reliability, and efficiency. The system is structured into multiple layers, ensuring smooth guest interactions, accurate request processing, and seamless integration with hotel services.

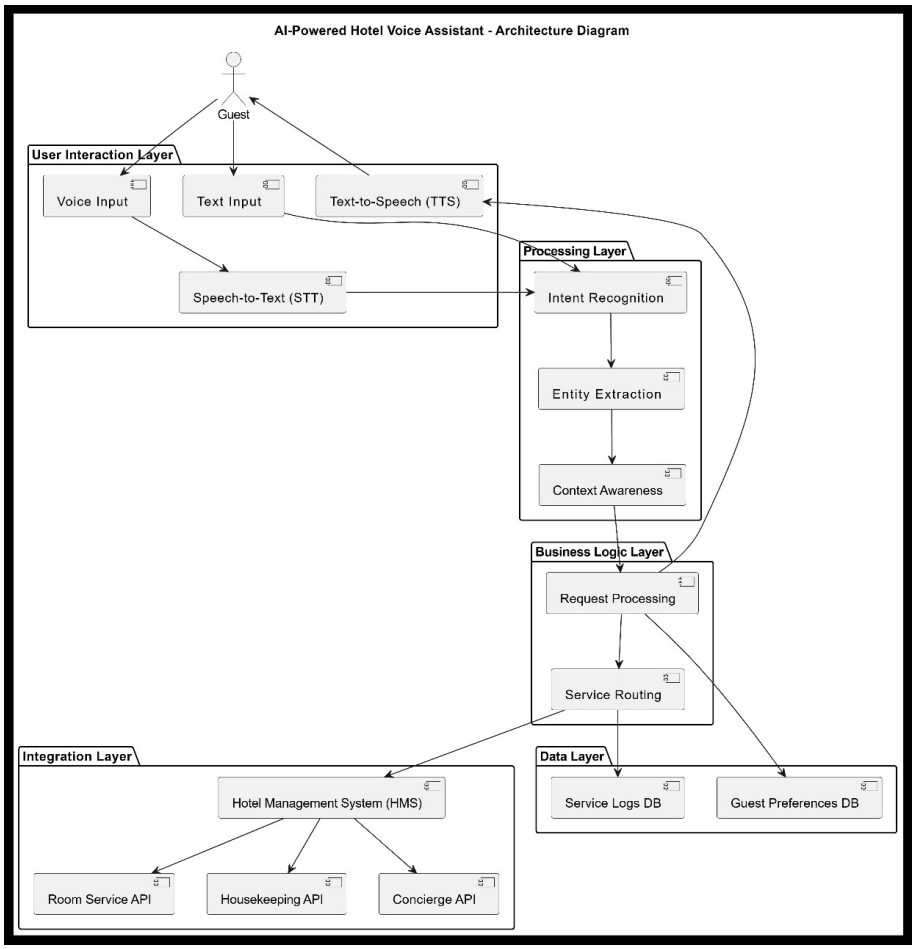

Figure 3.1: Architecture Diagram

3.1.1 System Architecture Overview

The system is divided into five main layers, each responsible for a specific aspect of functionality:

- **User Interaction Layer**
  - Handles guest interactions through voice and text-based inputs.
  - Provides feedback using Text-to-Speech (TTS) and UI components.
- **Processing Layer**
  - Converts voice commands to text using Automatic Speech Recognition (ASR).
  - Identifies user intent using Natural Language Processing (NLP) models.
- **Business Logic Layer**
  - Processes requests and determines the appropriate response.

  - Routes tasks to the appropriate hotel service (e.g., housekeeping, room service).

- **Integration Layer**
  - Connects with external hotel management systems (HMS) via APIs. (if needed)
  - Ensures smooth communication between different modules.
- **Data Layer**
  - Stores guest preferences, service history, and interaction logs.
  - Ensures secure data storage and retrieval.

3.1.2 System Modules and Their Roles

The system is composed of four key modules, each handling different aspects of the assistant's functionality.

Foundation Module (Core Infrastructure)

- **Speech-to-Text (STT) Engine**: Converts guest voice commands into text using ASR models.
- **Continuous Model Fine-Tuning**: Enhances accuracy for hotel-specific terminology and frequent requests.
- **Error Handling Mechanism**: Detects failures and ensures fallback responses to maintain reliability.
- **Role**: Processes raw voice inputs and converts them into structured text for further processing.

Interactive Voice Module (User Interaction)

- **Text-to-Speech (TTS) Engine**: Generates natural-sounding responses for guest queries.
- **User Interface (UI) Design**: Ensures intuitive interactions through both voice and text-based communication.
- **Role**: Manages user interactions and provides immediate feedback through voice or text responses.

Intelligent Understanding Module (AI Processing)

- **Intent Recognition**: Determines the meaning behind guest requests (e.g., ordering room service, asking for weather updates).
- **Entity Extraction**: Identifies relevant details such as dates, room numbers, and service types.
- **Context Awareness**: Understands follow-up questions to maintain a natural conversation flow.
- **Role**: Interprets guest requests and extracts relevant information before passing it to the business logic layer.

Integration and Deployment Module (Backend & Scalability)

- **Hotel Management System (HMS) Integration**: Connects with hotel services such as room service, housekeeping, and concierge.
- **Cloud-Based Deployment**: Ensures high availability, reliability, and performance across multiple hotel branches.
- **Performance Monitoring & Logging**: Tracks response times, error rates, and service uptime.
- **Role**: Manages backend operations, integrates with hotel services, and ensures system reliability.

3.2 Design Models

3.2.1 Activity Diagram

The activity diagram illustrates the workflow of VocalConcierge, an AI-powered hospitality assistant. It begins with audio input, which is processed and transcribed into text. The system then performs intent recognition and sentiment analysis to understand the user's request. Based on the analysis, an AI agent retrieves relevant information, processes reservations if needed, and logs interactions. The response is then generated using Text-to-Speech (TTS) and delivered to the user. If transcription fails, the system prompts the user to repeat their input. This structured flow ensures efficient and intelligent automation of hospitality services.

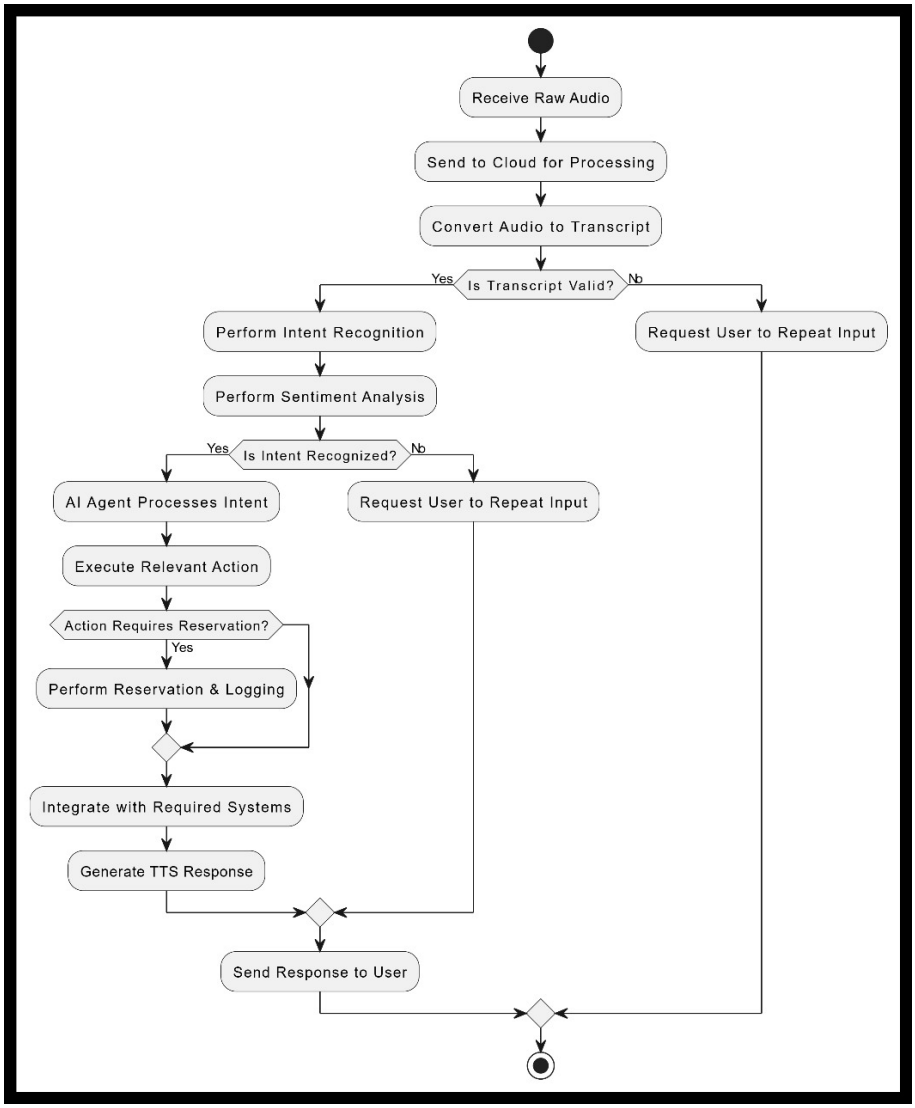

Figure 3.2: Activity Diagram

3.2.2 DataFlow Diagram

The Level 1 DFD (Context Diagram) provides a high-level view of the VocalConcierge System, illustrating how different actors interact with it. The primary users, including hotel guests and staff, communicate with the system via voice commands. The system processes these inputs and either responds directly to the users or forwards relevant information to the hotel staff for further action. The System Database plays a crucial role in storing and retrieving data such as room availability, booking details, and frequently asked questions. This diagram gives a broad understanding of the system's boundaries and external interactions.

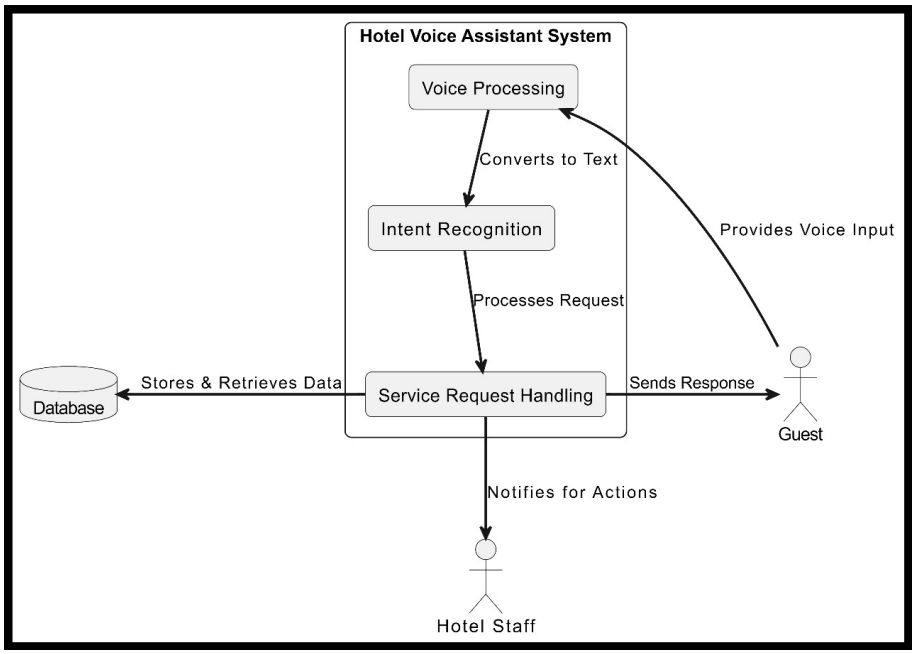

Figure 3.3: Level 1 DataFlow Diagram

The Level 2 DFD expands the system into its main functional components, offering a more detailed view of the data flow. When a user provides voice input, the Speech Processing Module converts it into text. This text is then analyzed by the Intent Recognition Module, which determines the user's request using Natural Language Processing (NLP) techniques. If needed, the Sentiment Analysis Module evaluates the user's tone and urgency to prioritize requests effectively. The AI Agent Processing Module then handles various tasks such as room booking, general inquiries, and service requests. Finally, the Response Generation Module either provides the appropriate reply to the user via text-to-speech conversion or forwards the request to hotel staff for manual processing.

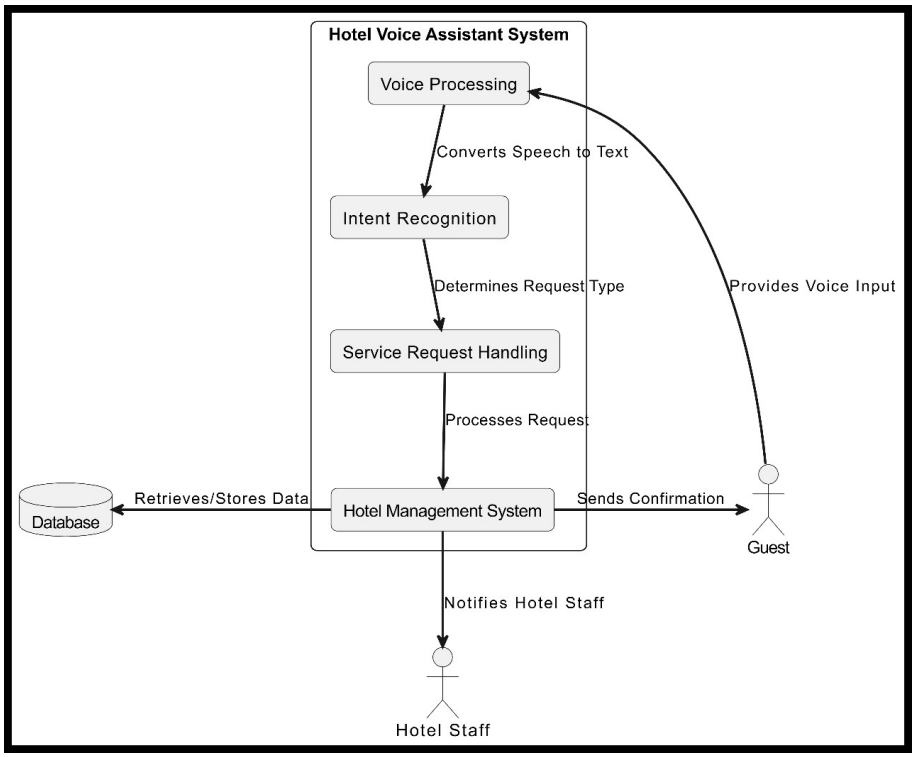

Figure 3.4: Level 2 DataFlow Diagram

The Level 3 DFD further decomposes each module into sub-processes, offering a deeper understanding of the system's internal workings. The Speech Processing Module consists of voice capture and speech-to-text conversion. The Intent Recognition Module breaks down into NLP processing and intent classification to determine the exact nature of the request. The Sentiment Analysis Module (if needed) is responsible for emotion detection and tone analysis, ensuring that urgent requests are prioritized. The AI Agent Processing Module includes steps for checking room availability, confirming bookings, and managing inquiries. Finally, the Response Generation Module ensures that responses are delivered accurately through text-to-speech conversion or staff notifications when necessary.

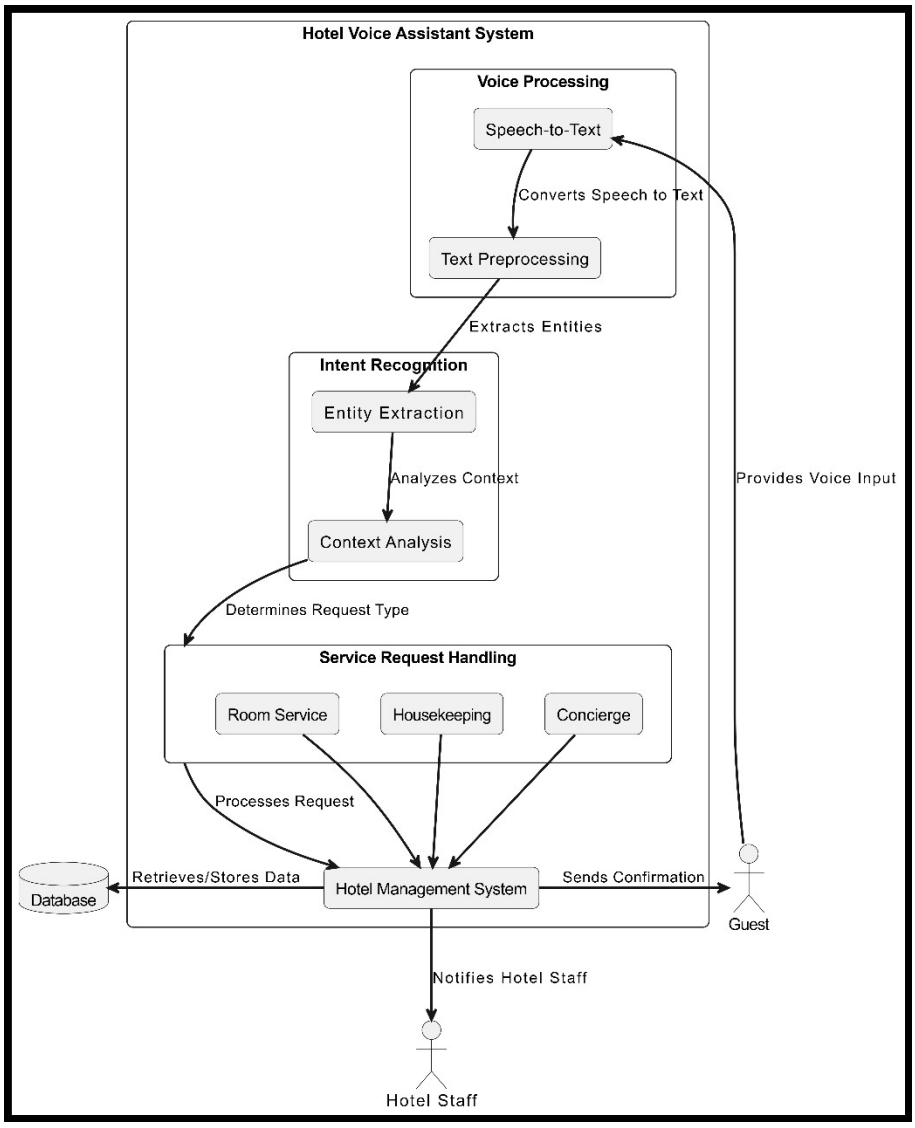

Figure 3.5: Level 3 DataFlow Diagram

3.2.3 System Level Sequence Diagram

The System-Level Sequence Diagram illustrates the interaction between external actors (guests, hotel staff) and internal system components. It depicts the step-by-step flow from user request to system response, involving speech recognition, NLP processing, business logic, and database operations.

The diagram shows how guest inputs are processed, intent is recognized, and appropriate hotel services (e.g., room service, housekeeping) are triggered. It also includes alternative flows for error handling and invalid requests.

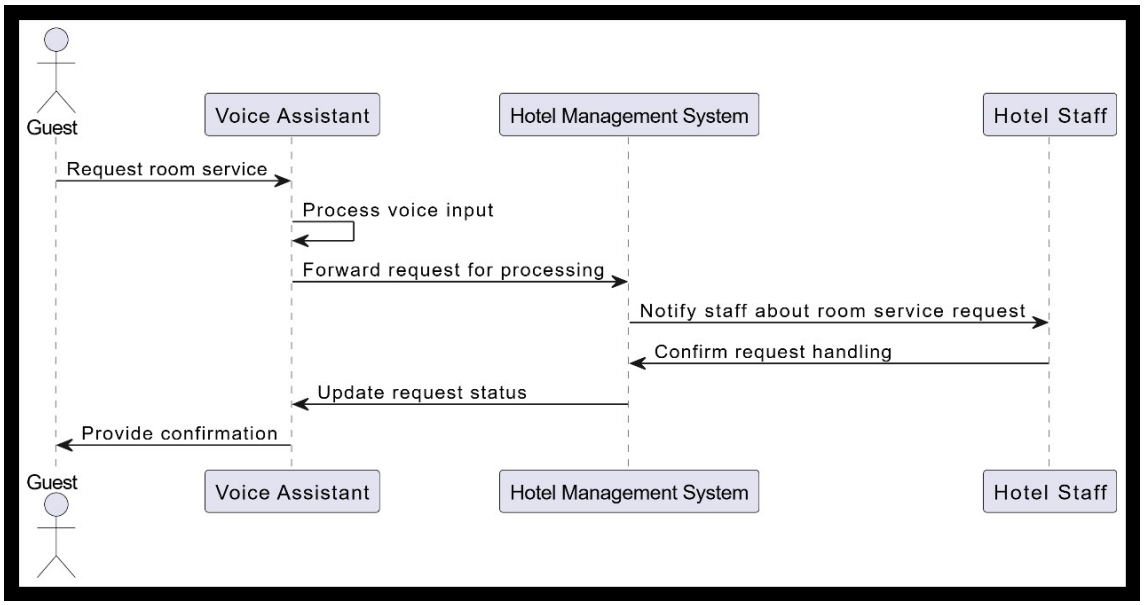

Figure 3.6: System Level Sequence Diagram

3.2.4 State Transition Diagram

The State Transition Diagram illustrates how the hotel voice assistant system progresses through different states in response to user interactions. The system begins in the Idle state, where it passively waits for user input. Once a user initiates interaction by speaking or sending a text command, the system transitions to the Listening state to capture the input.

After receiving the input, the system moves to the Processing state, where it applies speech recognition and natural language processing to interpret the user's intent. If the input is valid and recognized, the system transitions to the Action Execution state, where it performs the requested task, such as booking a room, requesting housekeeping, or retrieving hotel information.

Once the action is executed, the system enters the Response state, where it generates feedback and communicates the outcome to the user through text or voice output. If the system encounters an error, such as an unrecognized command or missing details, it moves to the Error Handling state, prompting the user for clarification or offering alternative responses.

Finally, after successfully delivering the response or resolving any errors, the system returns to the Idle state, ready for the next interaction. This structured transition ensures a seamless user experience while efficiently managing user queries and service request

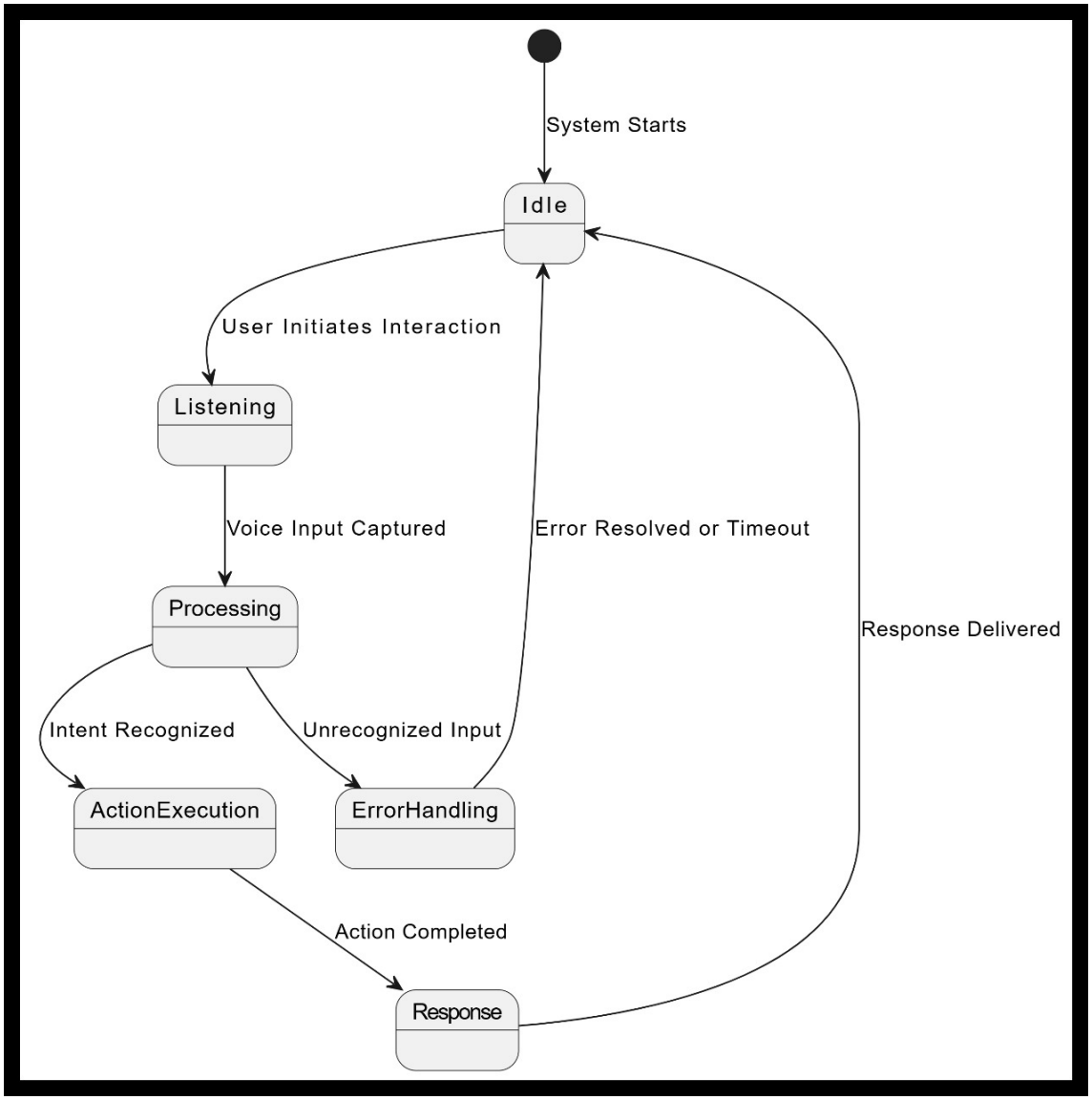

Figure 3.7: State Transition Diagram

3.3 Data Design

In the Hotel Voice Bot system, the information domain is transformed into structured data entities that facilitate voice-based interactions, service requests, and system automation. The data structures manage user requests, staff actions, service records, and integrations with hotel management systems. These entities are stored in a database and processed efficiently to ensure smooth operations.

3.3.1 Data Structures and Entities

3.3.1.1 Guest Requests

- **Data Structure:** Guest requests are stored as structured records containing request type (e.g., room service, housekeeping, reservations), timestamp, and status.
- **Storage and Processing:** Requests are stored in a relational database (e.g., MySQL or MongoDB). The voice bot processes the guest's voice input using NLP and maps it to predefined request categories.
- **Organization:** Each request is linked to a unique guest profile and room number for tracking and history.

3.3.1.2 Staff Task Assignments

- **Data Structure:** Tasks are assigned to hotel staff members based on request type, availability, and priority. Each task includes details such as request ID, assigned staff, and completion status.
- **Storage and Processing:** Task assignments are dynamically updated in the database, and staff receive notifications via the hotel's task management system.
- **Organization:** Tasks are categorized by department (housekeeping, room service, concierge) for better organization and tracking.

3.3.1.3 Voice Interaction Logs

- **Data Structure:** Voice interactions are logged as text transcriptions along with metadata such as time, guest ID, and confidence level of speech recognition.
- **Storage and Processing:** Logs are stored in a NoSQL database (e.g., MongoDB) for efficient retrieval and analysis.

- **Organization:** Interaction logs are used for improving NLP models, enhancing system accuracy, and generating service insights.

Bibliography

[1] Hadeel Sa'ad Al-Hyari, Hisham Mohammad Al-Smadi, and Sulaiman R Weshah. The impact of artificial intelligence (ai) on guest satisfaction in hotel management: An empirical study of luxury hotels. *Geo Journal of Tourism and Geosites*, 48:810–819, 2023.

[2] Sagina Athikkal and John Jenq. Voice chatbot for hospitality. *arXiv preprint arXiv:2208.10926*, 2022.

[3] Bitext. Bitext hospitality llm chatbot training dataset. Hugging Face, 2024.

[4] Dimitrios Buhalis and Iuliia Moldavska. Voice assistants in hospitality: using artificial intelligence for customer service. *Journal of Hospitality and Tourism Technology*, 13(3):386–403, 2022.

[5] Amma Liesvarastranta Haz, Evianita Dewi Fajrianti, Nobuo Funabiki, and Sritrusta Sukaridhoto. A study of audio-to-text conversion software using whispers model. In *2023 Sixth International Conference on Vocational Education and Electrical Engineering (ICVEE)*, pages 268–273. IEEE, 2023.

[4] [5] [1] [2] [3]