AI Voice Assistant for Financial Operations

1. Introduction

In the modern era of digital transformation, banks are continuously striving to make financial services more user-friendly, secure, and accessible. However, a significant segment of users—especially elderly citizens, rural customers, and people unfamiliar with technology—struggles to navigate complex banking applications or websites. To address this challenge, our team developed an **AI Voice Assistant for Financial Operations**, a smart, conversational, and secure banking interface that enables users to perform financial tasks through natural voice commands.

The system bridges the gap between human interaction and digital banking by combining **Artificial Intelligence (AI)**, **Natural Language Processing (NLP)**, **Speech Recognition**, and **Secure Authentication Mechanisms**. It allows users to check their account balance, make fund transfers, view transaction history, apply for loans, verify KYC, check credit scores, and perform other core banking operations — all through voice commands.

2. Objectives

The primary goal of this project is to create a **hands-free, intelligent banking experience** that enhances accessibility, security, and efficiency.

Key Objectives:

- Enable core banking operations through natural voice commands.
- Ensure secure authentication using **Voice Biometrics** and **OTP validation**.
- Provide intelligent, context-aware responses using **AI and NLP models**.
- Maintain a robust backend for transaction data and account management.
- Design a conversational user experience that supports diverse accents and languages.
- Implement a scalable and modular architecture suitable for future integration with real banking systems.
- Promote financial inclusion for all users, including those with disabilities or limited digital literacy.

3. Technology Stack

The project integrates multiple technologies across AI, web, and backend domains to deliver a secure and intelligent experience.

Frontend:

- **HTML5, CSS3, JavaScript:** For a responsive, user-friendly web interface.
- **Speech Synthesis & Recognition API:** To capture user voice commands and provide spoken feedback.
- **Bootstrap / Tailwind CSS:** For clean and accessible UI components.

Backend:

- **Python:** Core language for backend logic, speech processing, and data handling.
- **Pandas & NumPy:** For data manipulation and transaction record management.
- **Pathlib, Re, Time, Twilio, Pyttsx3:** Supporting libraries for file handling, regular expressions, OTP generation, and voice output.

AI & Voice Processing:

- **SpeechRecognition (Python Library):** Converts voice commands to text.
- **Voice Encoder:** Used for generating voice embeddings for biometric verification.
- **PyAudio / pyttsx3:** For audio input and voice feedback generation.
- **NLP Models:** For understanding user intent and providing meaningful responses.

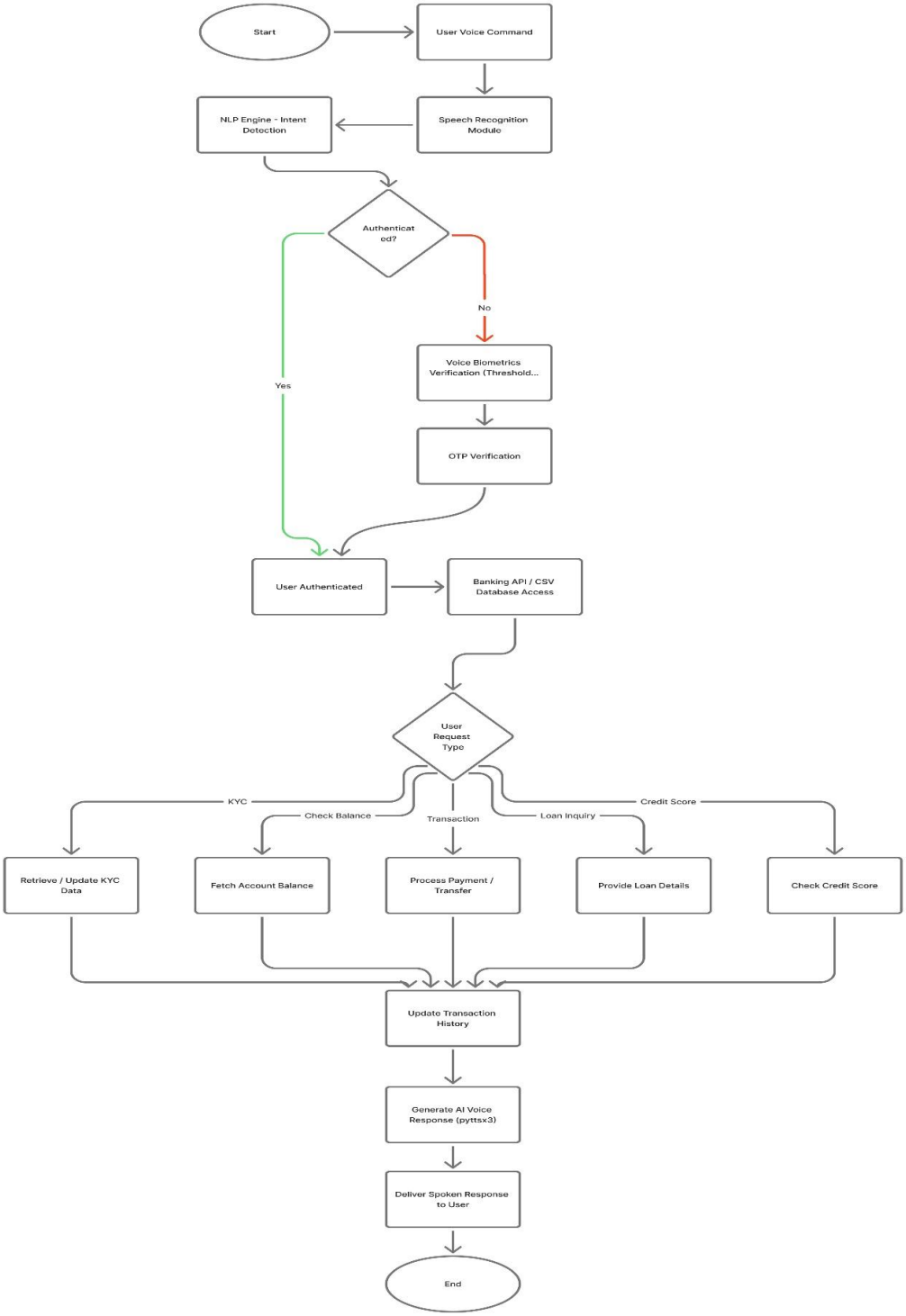
Database & Storage:

- **CSV-based Dummy Database:** Contains 100 records representing customer data (name, account number, balance, credit score, transaction logs, and KYC status).
- **Future Integration Option:** Can be migrated to SQL or Firebase for scalability.

APIs Used:

- **Twilio API:** For OTP generation and verification.
- **Mock Banking API (Custom):** For simulating balance updates, transfers, and transaction retrievals.

4. Detailed Flow Chart



5. System Architecture

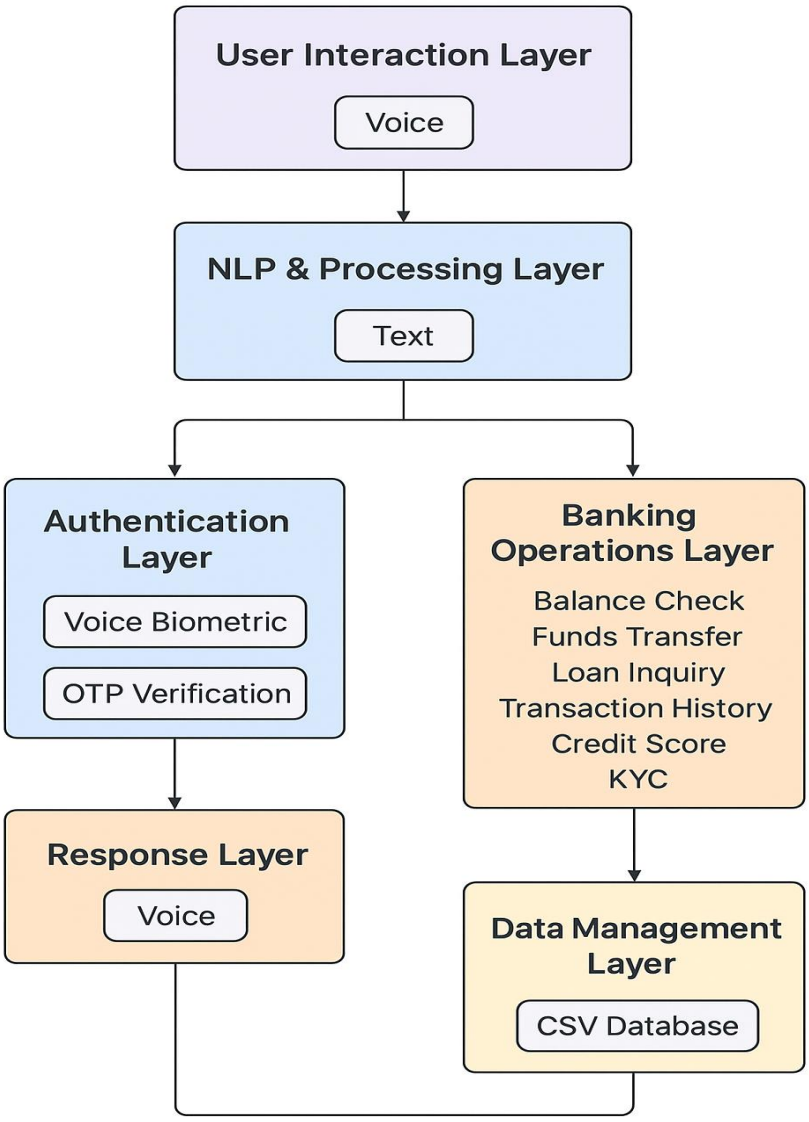
The **System Architecture** is designed for modularity, scalability, and data security. The flow of the architecture can be divided into the following layers:

a) User Interaction Layer

This is the **voice-based frontend interface** that allows users to speak directly to the assistant. The microphone captures voice commands and sends them to the backend via REST API.

Main Components:

- Voice Command Capture (SpeechRecognition API)
- Text-to-Speech Feedback (pyttsx3)
- Browser-based interface with microphone access



b) NLP & Processing Layer

This layer interprets the user's spoken query and determines the intent of the operation. For example, when the user says *"Transfer ₹500 to Rahul"*, the NLP engine identifies the action ("Transfer"), amount ("500"), and recipient ("Rahul").

Core Components:

- NLP model for intent detection
- Tokenization and keyword extraction
- Context maintenance for multi-turn conversations

c) Authentication Layer

This layer ensures that all transactions are secure and that the voice matches the registered customer.

Features:

- **Voice Biometric Authentication:**
  Compares the input voice embedding with stored voice data.
  If similarity ≥ 0.80 (on scale 0–1.00), access is granted.
- **OTP Verification:**
  Twilio-based OTP sent to registered mobile for secondary validation.

d) Banking Operations Layer

Handles all the logic for executing financial functions and updates the database.

Operations Supported:

- Account balance check
- Fund transfers and payment processing
- Loan inquiry and interest calculation
- Transaction history retrieval
- Credit score and limit checking
- KYC verification and update

e) Data Management Layer

A CSV-based database acts as the core storage system.

Database Schema:

Field	Description

Field	Description
Customer_ID	Unique identifier assigned to each customer.
Name	Full name of the customer.
Account_No	Unique bank account number linked to the customer.
Balance	Current available balance in the customer's account.
No_of_Transactions	Total number of transactions made by the customer.
Last_Payment	Date and amount of the most recent payment or transaction.
Loan_Status	Indicates whether the customer has an active or closed loan.
Interest_Rate(%)	Interest rate applicable on the customer's loan or savings.
Credit_Limit	Maximum credit amount assigned to the customer.
Reminder_Type	Type of reminders set (e.g., bill payment, EMI, due date).
Alerts_Enabled	Specifies whether alerts/notifications are activated.
Last_Transaction_Date	Date of the customer's most recent transaction.

f) Response Layer

After executing the requested operation, the assistant responds with a voice message confirming success or displaying relevant details on-screen.

5. Data Model & Storage

For this prototype, we utilized a **CSV-based data model** for simplicity and easy simulation of real banking data.

Key Features of the Data Model:

- 100 synthetic customer records for demonstration.
- Stores fields such as Name, Account ID, Balance, Credit Score, KYC, and Transaction History.
- Data is dynamically updated during fund transfers or payments.
- Each transaction is logged with timestamp, amount, and operation type.

Future Improvements:

- Migration to **SQL or Firebase** for secure and concurrent transactions.
- Addition of **NoSQL (MongoDB)** for handling voice logs and audit trails.
- Implementation of **encryption at rest** for sensitive financial data.

6. AI / ML / Automation Components

The intelligence of the system lies in its ability to listen, understand, and act.

a) Speech Recognition & Processing

- Converts user speech into text using the **SpeechRecognition** library.
- Integrates language models to handle diverse accents and natural variations in speech.

b) NLP (Natural Language Processing)

- Tokenizes voice input and extracts the **intent** (e.g., "check balance", "transfer money").
- Uses pre-trained language models for context-aware understanding.
- Ensures conversational flow continuity even across multiple turns.

c) Voice Biometrics

- Captures the unique **voice print** of each user.
- Compares it with stored embeddings to verify identity.
- Uses a **threshold accuracy (0.80)** to ensure security while minimizing false rejections.

d) Text-to-Speech (TTS) Engine

- The **pyttsx3** library is used to convert textual responses into natural voice output.
- Supports male and female voices, adjustable pitch, and multilingual tones.

e) Automation Features

- **Automated OTP verification** for transaction authentication.
- **Automatic balance updates** post successful transfers.
- **Real-time data synchronization** for transaction and KYC status.

7. Security & Compliance

Security is the most critical aspect of financial systems. The AI Voice Assistant follows multiple layers of protection to ensure privacy and compliance.

a) Authentication & Authorization

- **Two-Factor Authentication (2FA):** Voice + OTP verification.
- **Voice Biometrics:** Provides secure, password less authentication.
- **Session-based Tokens:** Prevent unauthorized reuse of access.

b) Data Privacy

- Customer data in CSV is anonymized and secured in restricted storage.
- All sensitive operations are logged for audit purposes.
- Data masking techniques used for partial display of account details.

c) Transaction Security

- OTP verification before executing fund transfers.
- Real-time validation against user account limits and status.
- Logs every transaction to ensure traceability.

d) Audit Logging and Governance: To enhance transparency, all interactions and transactions are recorded in immutable audit logs. These logs facilitate forensic analysis, fraud detection, and compliance reporting.

e) Regulatory Compliance: The AI Voice assistant aligns with Reserve Bank of India (RBI) digital banking directives, General Data Protection Regulation (GDPR) principles, and ISO/IEC 27001 standards. All operations are carried out with explicit user consent, following strict privacy-by-design principles to prevent misuse of sensitive information.

8. Scalability & Performance

To ensure the system can handle large volumes of users and queries, the architecture is designed to be modular and scalable.

a) Horizontal Scalability

- The backend modules (Speech Recognition, NLP, OTP services) can be hosted independently as microservices.
- Load balancers can distribute voice processing requests efficiently.

b) Caching and Performance

- Frequently accessed user data and configurations are cached.
- Response latency optimized by pre-loading models and limiting database I/O.

c) Modular Expansion

- Can integrate new banking services such as investment queries, bill payments, and insurance using APIs.

- Multilingual model integration for regional adaptation.

d) Deployment Flexibility

- Deployable on **cloud platforms (AWS / GCP / Azure)** or **on-premise servers**.
- Compatible with both **web and mobile** voice interfaces.

e) Future Enhancements

- **Edge AI Processing:** Running lightweight models on mobile for faster local responses.
- **AI Feedback Learning:** Continuously improves recognition accuracy from user interactions.
- **Integration with Core Banking APIs:** For real-time balance and transaction updates from actual systems.

Code Repository Link : https://github.com/prajesdas/FinSpeakAssistant.git

Demo Video Link: https://drive.google.com/file/d/1B7lMCwDlovtzTezydOuy-S7Y4bQhV0_J/view?usp=sharing