

PayDay – Multilingual Voice AI Architecture

1. Purpose

This document explains how the PayDay AI Assistant enables users to speak in their native Indian languages while the AI processes all logic in English for higher accuracy, and then responds back in the user's language with optional voice playback.

2. High-Level Idea

User speaks → System understands → AI thinks in English → User hears response in their language

3. Supported Capabilities

- Voice input in Indian languages
- Automatic language detection
- Translation between native language and English
- AI reasoning in English
- Text-to-Speech (TTS) voice responses
- Chat history with text and audio playback

4. Key Components

User Device: Records voice input

Speech-to-Text (STT): Converts voice to text and detects language

Translation Engine: Converts native language to English and back

AI Engine: Processes queries in English

Text-to-Speech (TTS): Converts responses into voice

Chat UI: Displays text and audio controls

5. End-to-End Workflow

1. User speaks in native language
2. Audio is recorded in browser
3. Speech converted to text with language detection
4. Text translated to English for AI
5. AI generates response in English
6. Response translated back to user language
7. Text converted to speech
8. Chat UI shows text, audio, and translation info

6. Visual Flow

User Voice → Audio Capture → STT → Translate to English → AI Processing → Translate to Native Language → TTS → Chat UI

7. Why This Architecture

For Users: Natural speech, no English dependency

For Business: Single AI language, scalable, accurate

For Engineering: Clean separation, maintainable, extensible

8. Data Stored

- Native transcript
- English translation
- AI response
- Language code
- Audio URL
- Timestamp and intent

9. Example Experience

User (Telugu): “■■■■■■■■■■ ■■■■ ■50 ■■■■■■ ■■■■■■?”

AI (Telugu): “■■ ■■■■■■■■■■ ■■■■■■■■■■■■ ■50 ■■■■■■ ■■■■■■ ■■■■■■...”

Audio playback and English info available

10. Summary

Users speak in their language, AI thinks in English, and responses come back in their language with voice support.