AI-Powered Multilingual Voice Bot – Summary

Objective

To build a real-time multilingual AI voice assistant that can understand and respond in both English and Hindi (extendable to other Indian languages), enabling natural customer interactions.

System Architecture

Pipeline:

Audio Input → STT (Deepgram) → LLM (Gemini) → TTS (Google/Gemini) → Audio Output

STT (Speech-to-Text) – Deepgram

- Converts live speech into text.
- Supports multilingual recognition with the nova-2 model.
- Provides high accuracy with real-time streaming.

LLM (Natural Language Understanding) – Google Gemini 2.0 Realtime

- Context-aware, intelligent response generation.
- Handles conversation flow, intent detection, and multilingual context.
- Configured with voice synthesis for English output.
- Instructions ensure: “Respond in the same language as the user.”

TTS (Text-to-Speech) – Google Gemini (English) + Google TTS API (Hindi)

- Gemini provides natural English voices (e.g., “Puck”).
- Combined approach ensures bilingual speech output.

LiveKit Agents – Real-time media engine

- Captures live audio from user via WebRTC.
- Streams bot responses back to the user with low latency.
- *Provides session management, room handling, and integration with plugins.*

Noise Cancellation – LiveKit BVC

- Filters background noise for cleaner input audio.
- Improves STT accuracy and user experience.

Key Features

Multilingual Support

- Understands and responds in Multilingual.
- Auto-detects spoken language and adapts response.

Real-time Interaction

- End-to-end latency target: ~2 seconds.
- Smooth conversation with natural turn-taking.

Demo Workflow

User speaks in English → “Hello, can you check my order status?”

- STT converts to text.
- Gemini generates a reply in English.
- Gemini voice speaks back naturally.

User speaks in Hindi → “नमस्ते, मेरा बिल स्टेटस बताइए।”

- STT converts to Hindi text.
- Gemini generates a reply in Hindi text.
- Google TTS speaks back in Hindi voice.