AI Voice Agent System - Summary Notes

1. Core Components

LiveKit: Real-time audio/video engine; used for rooms like Zoom. Works with SIP & WebRTC.

SIP (Session Initiation Protocol): Starts/ends real phone calls.

SIP Trunk = Virtual phone line via providers (Exotel, Twilio).

2. Language Support

LiveKit supports all audio. Add intelligence via:

- STT (Whisper, Gemini, Google)

- LLM (GPT-4o Mini, Gemini Pro)

- TTS (Silero, Cartesia, Google TTS)

3. Real-Time Conversation Flow

Caller -> SIP Trunk -> LiveKit Room

-> Audio -> Whisper (STT)

-> Text -> GPT-4o Mini / Gemini (AI Response)

-> Text -> TTS -> Back to Caller (via LiveKit)

Emergency? -> Call Doctor -> Resume Agent

4. SIP Provider Comparison

- Exotel: [Yes] Voice + [Yes] WhatsApp (Best for India)

- Telnyx: [Yes] Voice, [No] WhatsApp (Global)

- Plivo: [Yes] Voice, [No] WhatsApp

- Twilio: [Yes] Voice + [Yes] WhatsApp (Expensive)

5. LiveKit: Self-Hosted vs Cloud

AI Voice Agent System - Summary Notes

Self-Hosted: Full control, flat cost, own data

Cloud: Easy scaling, low latency via India Edge

6. Other Key Concepts

TURN/STUN: Ensures audio routing even in blocked networks.

Dispatch Rules: Control which region serves a room (e.g., Mumbai).

7. Tools Stack

- LiveKit: Real-time audio/video
- Whisper/Gemini STT: Transcribe
- GPT-4o/Gemini: Logic/Intent
- Silero/Cartesia TTS: Speak
- Exotel/Twilio: SIP & WhatsApp bridge