

Hassan Kazi

Voice AI Research Engineer · real-time voice agents · Indic speech

Bengaluru, India · open to remote (IST) · hassankazinasik@gmail.com · linkedin.com/in/erhassankazi · github.com/ANDROIDHASSAN · hassankazi.vercel.app

EXPERIENCE

- Product Developer — AI & Real-time Engineering · InCruiter, Bengaluru

Jul 2026 – Present

Lead the AI & real-time stack for InCBot (AI video interviewer) and IncScreen (AI phone & WhatsApp screening, which I built from scratch on Twilio and Exotel telephony with Vapi); scaled to 1,000+ concurrent interviews. Root-caused why multilingual interviews broke: English-only text guards failed on Devanagari (0.39 vs 0.5 match threshold) and Marathi STT hallucinated in 10 of 15 turns; drove script-aware, feature-flagged fixes.
- Software Engineer — AI & Automation · Anantkamal Software Labs, Nashik

Worked on WhatsApp automation and real-time voice agents. Full-stack MERN development (MongoDB, Express.js, React, Node.js) across multiple projects, including Raise Wademo, Hobbit and Ping Me: frontend implementation, backend API integration, database operations, debugging and feature development.
- Software Developer Intern · Fuzen.io, Nashik

Sep 2025 – Dec 2025

Shipped 5+ responsive React/JavaScript apps for 4+ clients, cutting page load time by 40%. Refactored 3 legacy codebases, raising SonarQube maintainability scores by 35%.
- AI Engineer · myRide, Boston, MA · remote

Mar 2025 – Aug 2025

Built a 5,000+ rule Boston parking-regulations dataset for autonomous-vehicle rule validation. Designed a real-time data pipeline and frontend used across 10+ AV models, with a US team.

PROJECTS

- Own Voice AI Models · In progress

PyTorch, ONNX, asyncio

My own Hindi/Marathi ASR and TTS models and a LoRA-tuned LLM, joined by a from-scratch streaming orchestrator (VAD, end-of-turn, barge-in); ONNX, INT8, speculative decoding.

Target: ~800 ms p50 voice-to-voice, ~5× lower cost/min than hosted APIs.
- InCBot — AI Video Interviewer · Shipped

Python, LiveKit Agents, Azure OpenAI GPT-4.1 / 4o

STT → LLM (tool calling) → TTS interviewer with feature-flagged guards; Docker CI/CD to LiveKit Cloud via GitHub Actions; runs in 42+ languages.
- IncScreen — AI Phone & WhatsApp Screening · Shipped

Twilio, Exotel, Vapi

Built from scratch: AI screening phone calls over Twilio and Exotel telephony, orchestrated with Vapi, plus WhatsApp interviews. Multilingual question sets, 7 Hindi voices, voicemail detection and fluency scoring. Hindi call-quality work: replies took ~5 s vs ~2 s in English, plus cut-off speech and interruptions.
- This site's AI voice clone · Shipped

LiveKit Agents, Sarvam, Deepgram

Talks in English, Hindi or Marathi and switches when you do. Three-deep fallback chains for STT, LLM and TTS; keyterm-prompted STT; per-reply voice routing (Hindi and Marathi share a script, so it reads the reply's words); false-interruption recovery; Krisp noise cancellation; hangs up on silent callers. Five Python files, deployed to LiveKit Cloud by GitHub Actions.
- Planned: IndicVoiceBench, Hinglish Turn Detector, indic-guard, Bhasha Bridge, Prompt-Injection Guard for Voice Agents, and 13 paper reimplementations in PyTorch. Targets, not results yet.

OPEN SOURCE · LIVEKIT/AGENTS

- #7740 feat(voice): add bit_rate and constant_bit_rate options to RecorderIO

Open · in review
- #7741 fix(voice): don't hang or buffer audio when the recorder encoder fails

Open · in review
- #7742 fix(tokenize): ignore stray closing tags in the XML depth check

Open · in review

SKILLS

LiveKit Agents · WebRTC · Vapi · Twilio · Exotel · Telephony (PSTN) · STT / TTS · Deepgram · ElevenLabs · Silero VAD · Sarvam AI · Azure OpenAI · Groq · gpt-oss · LLM tool calling · Prompt engineering · LLM guardrails · Model fallback routing · Python · TypeScript · Node.js · React · Docker · Kubernetes · PostgreSQL · GitHub Actions · PyTorch · learning

EDUCATION

- B.Tech, Computer Science & Engineering · Sandip University

2023 – Present
- Diploma, Computer Engineering · Guru Gobind Singh Polytechnic

2020 – 2023

ACHIEVEMENTS & CERTIFICATIONS

- 9× Hackathon Winner
- CodeChef 3-Star: Top 15% globally; 1st in October Long Challenge among 8,000+ coders
- 30+ merged open-source PRs: Across 15 repos, Hacktoberfest 2022–2024
- Technical Lead, Leads Community: 2024 – present · 5+ workshops for 200+ students; 4 hackathons with 1,500+ participants
- Certifications: Google Cloud — 5 Skill Badges, Deep Learning with PyTorch, Generative AI Concepts, AI Solutions in Business