

# RAHUL WALE

AI / ML Engineer | Full-Stack Developer

+91 93568 53041 • walerahul96@gmail.com • [github.com/Rahuwale123](https://github.com/Rahuwale123) • [linkedin.com/in/rahul-wale](https://linkedin.com/in/rahul-wale)

---

## SUMMARY

AI/ML Engineer with 2 years building production LLM and real-time voice systems. Shipped AI platforms serving 500+ concurrent users at 99% uptime, reducing operational workload by 60%. Strong end-to-end ownership across LLMs (Gemini, LangChain, RAG), real-time speech (Sarvam, Deepgram, Pipecat), FastAPI/Python backends, and Next.js/React/Flutter frontends. Currently leading an AI team and mentoring 50+ engineers.

---

## EXPERIENCE

### AI Engineer & Team Lead · The BAAP Company

Jul 2023 – Present

*Generative AI · Voice AI · Computer Vision*

- Architected and shipped 4+ production AI platforms (real-time voice agents, RAG chatbots, no-code AI tools, CCTV intelligence) serving rural India at 99% uptime under 500+ concurrent users.
  - Designed a real-time multilingual Voice-on-Call AI platform combining Gemini LLM, Sarvam STT, Deepgram TTS, and Pipecat over Twilio - cut manual call-centre workload by 60%.
  - Built a no-code AI chatbot platform on LangChain + Gemini with RAG over municipal knowledge bases, enabling local governments to self-serve 5,000+ daily citizen interactions with zero engineering support.
  - Engineered low-latency voice pipelines (sub-second response) with WebSocket audio, VAD, and barge-in - productionised STT, LLM reasoning, and TTS as a single conversational stack.
- 

## KEY PROJECTS

### 1) Voice-on-Call AI Platform Real-time multilingual voice agent for civic services

**Pipecat · Gemini · Sarvam AI · Deepgram · Twilio · FastAPI · React · WebSockets**

- Real-time AI voice assistant letting citizens lodge civic queries over a phone call in their native Indian language.
- Owned the full stack: ML pipeline, FastAPI backend, React dashboard, and Twilio voice integration; 99% uptime under 500+ concurrent calls.

### 2) SmartCCTV - AI Surveillance & Monitoring Platform Plug-and-play AI layer for existing IP cameras

**Python · YOLOv9 · ArcFace · OpenCV · FastAPI · React · PostgreSQL · Redis**

- Upgrades existing IP CCTV cameras into intelligent monitoring - no hardware replacement; connect a camera and the system becomes instantly intelligent.
  - Deployed across pilot sites with 20+ cameras processing 10,000+ daily face-recognition events; ArcFace attendance, unknown-person alerts, and live React monitoring dashboard.
- 

## TECHNICAL SKILLS :

**AI / ML:** LLMs, LangChain, RAG, Gemini, Hugging Face, Prompt Engineering, AI Agents, LLM Evaluation, Guardrails, Real-Time Voice Pipelines

**Voice & Speech:** Pipecat, Sarvam AI, Deepgram, Vosk, VITS, Twilio Voice, WebSocket audio streaming, VAD

**Computer Vision:** YOLOv9, ArcFace, OpenCV, real-time video pipelines, face recognition

**Backend:** FastAPI, Python, REST APIs, WebSockets, PostgreSQL, Redis

**Frontend & Mobile:** Next.js, React, Flutter (cross-platform iOS / Android)

**Infrastructure:** Docker, AWS (EC2, S3), Firebase, Git, CI/CD

---

## EDUCATION

**Bachelor of Computer Applications (BCA)**

2023 – 2026