Swapnil Chavan

Mumbai, Maharashtra, India +91 9175541890 portfolio swapnilchavan18901@gmail.com linkedin.com github

SUMMARY

Software Engineer with 2 years of experience building AI-powered backend systems across healthcare, B2B, and SaaS. Experienced in voice agents, RAG, and scalable microservices using Python, TypeScript, and AWS, with a 5x reduction in inference costs.

TECHNICAL SKILLS

Languages: Python, TypeScript, C++
Backend & Frontend: FastAPI, NestJS, Express.js, React.js, Next.js
AI & Voice: OpenAI, Gemini, Sarvam AI, Pipecat, RAG, LLM Tool Calling
Databases & ORMs: PostgreSQL, MongoDB, Redis, Qdrant, Pinecone, Prisma, Drizzle
Cloud & Infrastructure: AWS (EC2, S3, ECR), Docker, Kafka, RabbitMQ, Cloudflare Queues, Git
Integrations: Twillo, Razorpay, Google APIs, Socket.IO

EXPERIENCE

Onicares AI **May 2026 – Present**
Software Engineer *Remote — Full-Time*

- Built **DIDI**, a voice-to-voice maternal-health bot with decision-tree symptom assessment, complaint registration, emergency flagging, and intent-based mode switching at **$0.021/turn**.
- Built **Didi Plus**, a doctor-facing **STT + LLM** bot to manage diagnosis cards and select templates dynamically, with **1 sec response latency**.
- Built a **Sarvam AI + Gemini OCR pipeline** converting clinical reports into structured JSON at **$0.04/document** and **$0.01/additional page**.
- Built **Razorpay subscription autopay** with a provider-agnostic schema and webhook handling maintained platform cron jobs, monitoring, and bug fixes.

Klubit AI **Aug 2025 – May 2026**
Software Engineer *Remote — Full-Time*

- Built **9+ backend modules** and REST APIs using **NestJS, TypeScript, PostgreSQL, and Kafka** for posts, courses, events, and social features.
- Deployed **Docker** applications on **AWS (EC2, S3, ECR)** with secure S3 pre-signed URLs, health checks, and monitoring, achieving **99.9% uptime**.
- Built a responsive **TypeScript web app** integrating **Google APIs** and **Razorpay Autopay Marketplace** for authentication and payments.

Sharbo AI **Nov 2024 – Jul 2025**
Software Engineer *Remote | CONTRACT*

- Built a **B2B visitor intelligence platform** using **NestJS and TypeScript**, with real-time tracking, IP enrichment, and lead generation across **20+ microservice modules**.
- Designed an **LLM-powered company profiling pipeline** to enrich IP and firmographic data, automatically scoring and qualifying leads against **ICP criteria**.
- Integrated **LLM summarization and classification** to turn visitor and company data into sales-ready insights, reducing manual triage.

PROJECTS

Retrieva | *FastAPI, PostgreSQL, Prisma, RabbitMQ, AWS S3, QdrantDB, OpenAI, Next.js* **View**

- Enabled businesses to **automate document-based customer support** through embeddable **RAG chatbots**, making business knowledge accessible through natural-language queries.
- Optimized a **streaming RAG pipeline with Sarvam AI and Qdrant**, achieving **1.5 - 2.5 sec server time to first token** and **$0.00034 estimated LLM cost per query**.
- Engineered an **async FastAPI backend** with **AWS S3 and RabbitMQ** for PDF ingestion and **incremental indexing**, avoiding redundant embedding costs while supporting real-time chat through **Socket.IO**.

Audiva | *FastAPI, Pipecat AI, Pipecat Flows, Silero VAD, Sarvam AI, Attendee, Twillo, Alembic, Next.js* **View**

- Enabled clinics to **automate appointment booking, rescheduling, and cancellation** through a **real-time AI voice receptionist**, supporting natural-language queries for clinic information.
- Built a **streaming voice pipeline with Silero, Sarvam AI and Pipecat**, achieving **0.4 sec LLM time to first token** and **$0.00036 estimated LLM cost per request**.
- Engineered an **async FastAPI backend** with **WebRTC audio streaming**, speech recognition, speech synthesis, and **tool-based workflows** for appointment management.

EDUCATION

University of Mumbai **2025**
B.E. in Computer Engineering, Pillai HOC College of Engineering and Technology *GPA: 8/10*

MSBTE **2019 – 2022**
Diploma in Computer Technology, Government Polytechnic Pen *85.77%*