

NIHAL PARDESHI

Pune, Maharashtra | +91 7666815841 | nihalpardeshi12344@gmail.com

[LinkedIn](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

AI Solutions Engineer with hands-on experience building and deploying production-grade AI systems using LLMs, RAG, and cloud-native architectures. Strong in agentic workflows, voice AI, and end-to-end system deployment on AWS.

EXPERIENCE

AI Intern – ChatMaven.ai | Jun 2025 – Nov 2025

- Designed and deployed an AI-powered healthcare IVR platform using LangGraph and Retrieval-Augmented Generation on AWS Lambda.
- Built unified AI workflows integrating voice (Twilio), SMS (SNS), and email (SES), reducing dependence on manual telecalling.
- Owned architecture decisions across model orchestration, cloud deployment, and system reliability with minimal supervision.
- Translated operational requirements into scalable, production-ready AI services.

PROJECTS

RelayX – AI Voice Calling Platform

Python, FastAPI, Docker, React, TypeScript, Twilio, Groq LLM (LLaMA 3.1), Whisper, Piper TTS, PostgreSQL

- Built a full-stack AI voice agent platform handling outbound sales and appointment calls with real-time, natural conversations.
- Designed an asynchronous AI pipeline (Audio → STT → LLM → TTS → Call) with context tracking and fault tolerance.
- Reduced end-to-end response latency from 8.3s to 0.9s through streaming, prompt optimization, and token control.
- Implemented automated call analysis and lead qualification with outcome tagging and analytics dashboard.

RateMyProf India – Professor Review Platform

React, Next.js, FastAPI, PostgreSQL

- Developed a full-stack professor review platform serving 500+ professors with real-time ratings and voting.
- Built automated moderation and spam detection, reducing invalid content by 85%.
- Achieved 85+ PageSpeed score and top-6 SEO ranking through performance and structured

data optimization.

HouseGPT – Character-Driven Voice AI Chatbot

Python, Hugging Face, FAISS, XTTS

- Fine-tuned Flan-T5-Large using LoRA and prompt engineering for character-consistent, sarcastic responses.
- Integrated Retrieval-Augmented Generation to reduce hallucinations and improve response grounding.

SKILLS

- **Languages:** Python, C++, JavaScript, SQL
- **AI/ML:** PyTorch, Hugging Face, RAG (FAISS, Sentence Transformers), Reinforcement Learning
- **Systems & Cloud:** FastAPI, AWS Lambda, SES, SNS, API Gateway, Docker
- **Agentic & Tools:** LangGraph, MCP-style tool integration, Git/GitHub, Cursor, GitHub Copilot

EDUCATION

B.Tech in Computer Science (AI/ML Specialization) – Vishwakarma University, Pune

Aug 2023 – May 2027