

Aditya Pratap Singh

Portfolio | akkupratap323@gmail.com | +91-8005729753 | [linkedin.com/in/akkupratap323](https://www.linkedin.com/in/akkupratap323)

Education

Indian Institute of Information Technology, Nagpur
B.Tech in Electronics and Communication Engineering

May 2022 – May 2026

Experience

Software Developer Intern, Nester Labs (sunnyvale CA) Onsite Bangalore Nov 2025 – Present

- **NesterAI Conversational system:** Engineered an emotion-aware voice pipeline achieving **90% cost reduction** (\$0.33→\$0.01/min) by replacing commercial stacks with self-hosted **wav2vec2-MSP** (emotion) and **Deepgram Nova-3** and finetuned opensource models. [\[Engineering Case Study\]](#)
- Architected **GraphRAG** system using **Neo4j** on AWS EC2, enabling multi-hop retrieval ('Error'→'Cause'→'Solution') and reducing hallucinations by **25%** compared to vector-only baselines.
- Developed custom **Pipecat** processors ('EmotionForkProcessor') and a **Hysteresis Emotion Controller** to modulate **Resemble AI** TTS tone (CFG/Exaggeration) in real-time (< 200ms latency).
- Deployed scalable infrastructure on AWS Lightsail/ECS with Docker CI/CD, handling parallel audio processing (Librosa) and WebSocket streams for 20+ concurrent sessions.
- **Goonj AI Optimization:** Reduced retrieval latency by **82%** (1215ms→220ms) using profile-weighted retrieval and local embeddings (Sentence Transformers) with an 86.7% Redis cache hit rate.

AI & Full stack Intern, Responsenet

Sep 2025 – Nov 2025

Remote/US-based projects

- Integrated an AI assistant chatbot using RAG, capable of system platform-aware responses.
- Developed a real-time opportunity finder: shows nearby opportunities on a map with distance, detects current or user-set location worldwide, auto-generates search radius circles (5–100 km) for proximity analysis.

Full stack Summer Intern, NIAN Technologies

May 2024 – Aug 2024

- Built full-stack web solutions using **Next.js**, **React**, **Node.js**, **PostgreSQL**, **Prisma** in a distributed team.
- Implemented RESTful APIs and optimized queries, reducing response times by **35%**.

Achievements

- **Published article at Nester Labs:** Technical blog on voice AI architecture featured on the company's official website. [\[View Article\]](#)
- Achieved **8th** rank in Unstop coding Challenge. Certificate: [\[view\]](#)
- Global rank **1861** in LeetCode Biweekly Contest **153**. [\[View Profile\]](#)

Projects

AI Interview & Analytics Platform

[\[Code\]](#) [\[Live\]](#) [\[Demo\]](#)

- Built platform for automated interviews, analytics, and resume parsing.
- Voice-based interviews with Retell AI and OpenAI; dashboard with sentiment analysis.
- Tech Stack: Next.js 14, TypeScript, Tailwind, Supabase, OpenAI GPT-4, Retell AI, Clerk Auth, PostgreSQL

Ultron – Next-Gen Communication Platform

[\[Code\]](#) [\[Live\]](#)

- Scalable AI-powered social platform: real-time chat, video calling, and assistant personas.
- Integrated Stream Chat, ZEGOCLOUD, Lucia Auth, file uploads, infinite scroll, notifications.

Credit Approval System

[\[Code\]](#) [\[Video\]](#)

- REST API with Django: customer registration, loan eligibility, credit scoring.
- 5-factor scoring (history, volume, debt ratio), microservice architecture (Docker, Celery, Redis).

Technical Skills

AI/ML & Voice: OpenAI GPT-4, Gemini 2.5 Flash, GraphRAG (Neo4j), wav2vec2-MSP, Deepgram Nova-3, Resemble AI, Pipecat, Librosa, Sentence Transformers

Databases & Vector Stores: Neo4j, Pinecone, Redis, FAISS, PostgreSQL, Supabase

Backend: Python 3.13, FastAPI (WebSocket), Next.js API, Celery, Loguru

Cloud/DevOps: AWS (Lightsail, EC2, ECS), Docker, GitHub Actions CI/CD, Nginx

Frontend: React, Next.js 15, TypeScript, Tailwind, Vite