

Education

University of Petroleum and Energy Studies (UPES)
B.Tech in Computer Science Engineering (AI/ML)

August 2023 - Present
CGPA: 8.33/10

Technical Skills

LanguagesTypeScript, JavaScript, Python, C++, Rust

Frameworks/DBReact.js, Next.js, TailwindCSS, Express.js, MongoDB, PostgreSQL, PyTorch, Hugging Face

Cloud & ToolsAWS (EC2, S3, Lambda), GCP, Vercel, Docker, Git, WebSockets, Stripe, LoRA

Experience

Freelance Full-Stack DeveloperMay 2026 - July 2026
Upwork Client – Mom Gov Assistance PlatformRemote

- Spearheaded development of a new partner portal from inception, concurrently revamping the existing mother portal to serve U.S. mothers with 400+ government assistance programs
- Optimized AI-powered eligibility scans, reducing end-to-end processing time from over 30 seconds to under 2 seconds through state-based filtering, asynchronous AI processing, and database optimizations

Full-Stack Developer InternApril 2025 - October 2025
Classy EndeavorsRemote

- Implemented 20+ production features across 5+ client projects, spanning frontend and backend development
- Modernized existing React/Next.js codebases and backend services while implementing CI/CD pipelines across 3 client projects.
- Managed deployments and cloud infrastructure on AWS (EC2, Lambda, S3) and GCP, ensuring reliable releases and application availability

Projects

Sentinel - AI Observability & Incident Management Platform
Next.js, Express.js, PostgreSQL, Prisma ORM, Redis, Claude AI, Docker

- Built a self-hosted observability platform to monitor server telemetry, visualize metrics, explore logs, and manage incidents in real time
- Engineered an anomaly detection engine that automatically identifies CPU, memory, disk, error-rate, and agent failures to create evidence-backed incidents
- Integrated Claude AI to generate streaming root-cause analyses from logs, metrics, and historical incidents using Server-Sent Events (SSE)

SahayaAI - LLM Career Counseling Platformsahayaai.dev
Python, FastAPI, PEFT (LoRA), Transformers, Llama 3.1 8B

- Fine-tuned Llama 3.1 8B using PEFT (LoRA) on a 10K synthetic career-counseling dataset generated for domain-specific adaptation
- Engineered a FastAPI inference service to serve the fine-tuned model on Kaggle GPUs, with Gemini API fallback when local inference was unavailable
- Implemented rolling conversation memory by maintaining a running summary with recent chat history, enabling long conversations while reducing context length and inference latency

Achievements & Leadership

Technical HeadJune 2025 - Present
UPES Computer Society of India

- Leading technical initiatives, mentoring 10+ team members, and organizing college-wide technical events

Hackathon 4.0 FinalistFebruary 2025
UPES-CSA Hackathon - Top 5 out of 180+ Teams

- Developed a Smart Digital Coaching Platform with real-time chat and video calls