GAUTAM KUMAR

+919572898131 gautamraaz936@gmail.com LinkedIn Github Portfolio Leetcode GFG

Professional Summary

Full-Stack Developer with 2+ years of experience in Node.js, React.js, Langchain, and TypeScript, building scalable web and AI-powered applications. Experienced in backend APIs, RAG pipelines, AI agents, and cloud deployments, with a strong focus on **performance and reliability**.

Technical Skills

Programming Languages: JavaScript, TypeScript, Python
Frameworks & Libraries: Node.js, Express.js, React.js, Next.js, FastAPI, Redux, Zustand
AI & Retrieval: LangChain, RAG, AI Agents, Voice Agents, Vector Search, LiveKit, Hugging Face
Databases & Messaging: MongoDB, PostgreSQL, MySQL, Redis, RabbitMQ
Cloud & DevOps: AWS, GCP, Docker

Work Experience

Excellence Technologies **July 2024 – Present**
Full-Stack Developer *onsite*

- Delivered **15+** full-stack client projects across backend services, frontend systems, database design, AI integrations, and production rollout.
- Built RAG pipelines, voice-agent flows, and multi-agent systems with embeddings, vector databases, and prompt orchestration.
- Deployed open-source and Hugging Face models for inference and evaluation in production AI features.
- Owned delivery end to end, from requirements and architecture through implementation, deployment, and production support.
- Improved API latency, frontend responsiveness, and backend reliability through query tuning, caching, debugging, and service simplification.

Align InfoTech **Feb 2022 – March 2022**
Full-Stack Intern *Remote*

- Built web applications with JavaScript and React, and maintained server-side APIs using Node.js and Express.
- Worked with cross-functional teams to ship responsive, high-quality products with strong performance.

Projects

PatentWatch AI **Live Link**

- Built a patent infringement workflow that scans product data and generates claim charts for patent review.
- Implemented a RAG pipeline across **140M+ documents** using embeddings, vector search, and semantic retrieval to match products with relevant patent claims.
- Cut semantic search latency from **10s to 5s** by optimizing retrieval, adding Redis caching, and improving query handling.

GHSRK **Live Link**

- Designed a search and agent platform for personal and enterprise workflows, supporting **4+ input types**: voice, code, images, and documents.
- Connected Slack, Notion, and Google Drive APIs to enable cross-platform search and unified knowledge workflows.
- Delivered contextual memory, geo-aware responses, and white-label copilots through queue-based orchestration and real-time response handling.

QuickPostAI **Source Code**

- Developed a writing tool that turns long-form blog content into **280-character X posts** using LLM-based generation.
- Introduced tone controls and scheduling workflows to reduce manual edits and keep publishing consistent.
- Enabled contextual reply generation for social engagement with automated response workflows.

AceMock AI **Source Code**

- Developed a mock interview platform that generates role-specific questions from job details, tech stack, and experience level.
- Delivered real-time feedback, answer evaluation, and audio-based interview workflows for practice sessions.
- Implemented secure authentication, webcam interview flows, interview history tracking, and Stripe-based subscription management.

Education

Priyadarshini College of Engineering,Nagpur **2020 - 2024**
B.Tech in Computer Technology *8.45/10 CGPA*