

Udyan Upal

Senior Full-Stack Engineer — AI & API Integrations

udyan.upal@gmail.com | +880 1705 933999 | [linkedin.com/in/udyan-upal](https://www.linkedin.com/in/udyan-upal) | github.com/udyan-upal | Dhaka, Bangladesh

PROFILE

Senior Full-Stack Engineer with **4+ years** building and scaling SaaS platforms end-to-end — **Node.js/TypeScript** backend services, **REST APIs**, **PostgreSQL** data models, and **React/Next.js** frontends. Deep experience integrating **third-party APIs** and shipping production **LLM features** (OpenAI, Claude, LangChain) with **RAG and vector databases**, background workers, and queue-driven automation. Owns systems end-to-end: architecture, delivery, Dockerised deployment on **AWS/GCP**, and monitoring. Strong written English, high initiative, and a habit of thinking about the whole system rather than the ticket.

TECHNICAL SKILLS

AI Coding Agents	Claude Code, Codex, OpenCode, Pi Agent — agent-driven development, review and refactoring workflows
AI / LLM	OpenAI & Claude APIs, Function/Tool Calling, RAG & RAG Fusion, LangChain / LangGraph, LlamaIndex, CrewAI, AutoGen, prompt engineering, fine-tuning, embeddings
Languages	JavaScript (ES6+), TypeScript, Python, SQL
Backend	Node.js, Express.js, REST API design, GraphQL, FastAPI, Microservices, Background workers & job queues, WebSockets, Nginx
Frontend	React, Redux, Next.js, ReactFlow, HTML/CSS, Tailwind
Databases	PostgreSQL, MySQL, MongoDB, Redis (cache + queues), Qdrant, Weaviate
Integrations	Third-party REST/webhook APIs, OAuth 2.0, email & voice/telephony APIs (LiveKit, Google STT/TTS), payment & CRM-style data sync
Cloud / DevOps	AWS (EC2, S3, Route 53, CloudWatch), GCP, Docker, Kubernetes, Jenkins CI/CD, Git / GitHub, Grafana, ELK Stack

PROFESSIONAL EXPERIENCE

Senior Software Engineer — **Factiv.io** 04/2025 – Present | Toronto, Canada (Remote)

- Architected and shipped a multi-tenant **AI agent-builder SaaS platform**, owning the full stack from **Node.js/TypeScript** REST APIs and PostgreSQL schema design through to the React front end; currently serves **1,000+ API requests daily**.
- Built production **LLM features** on **OpenAI and Claude** with custom **RAG pipelines** over **Qdrant vector DB** (LangChain, LlamaIndex), including chunking, embedding and re-ranking strategies for multi-turn, context-aware responses.
- Cut hallucinations by **~90%** via multi-chain prompting, tool/function calling and optional live web-search retrieval; fine-tuned a GPT model for domain summarisation, improving coherence **15%**.
- Integrated multiple **third-party APIs** into the platform — LiveKit plus Google Cloud STT/TTS, VAD and turn detection — to deliver real-time bi-directional voice agents with sub-second latency.
- Built a **self-learning LLM feedback loop** that captures conversation outcomes and user signals from both the **chat and voice interfaces**, feeding them back into retrieval ranking, prompt templates and fine-tuning datasets so agent responses measurably improve with usage — lifting resolution quality and customer experience across both channels.
- Orchestrated **asynchronous background workers** for model-serving, ingestion and vector-store sync, keeping long-running AI jobs off the request path.
- Ran code reviews, set backend/AI architecture standards, and mentored junior engineers; continually evaluates emerging frameworks (LangGraph, RAG Fusion, agentic architectures) and brings the useful ones into production.

Backend / Full-Stack Engineer — **Luminous Labs** 11/2022 – 04/2025 | Dhaka, Bangladesh

- Designed and built scalable **Node.js / Express.js microservices** with **React + Redux** and Next.js clients for a customer-facing SaaS product; improved system scalability **40%** and reduced **REST API** response time **25%** through query tuning, caching and Dockerised deployments.
- Engineered AI/ML features into the platform: **RAG pipelines** with OpenAI, **Weaviate** and **Redis** that lifted retrieval accuracy **35%**, plus a visual **LLM orchestration tool** built with ReactFlow supporting fine-tuning workflows.
- Owned cloud infrastructure across **AWS and GCP** — EC2, Route 53 DNS, Cloud Storage and Nginx geo-routing — sustaining **99.9% uptime** and **50% faster** regional content delivery.
- Automated build and release with **Jenkins CI/CD** pipelines, cutting manual deployment effort **70%** and doubling release frequency while preserving database integrity and cache consistency.

EDUCATION

B.Sc. in Computer Science & Engineering — **North South University** 2018 – 2023 | Dhaka, Bangladesh