

Vivek Gubbala

Full Stack Software Engineer | GenAI & Cloud Platforms

Hyderabad, India | gvivek24111999@gmail.com | LinkedIn | GitHub | vivekgubbala.in

Summary

Full Stack Engineer with nearly **5 years** of experience building scalable web and AI platforms. Specialized in GenAI systems, RAG pipelines, and Kubernetes-based architectures. Proven track record of improving scalability **2–3x**, reducing infrastructure footprint, and delivering production-grade AI platforms end-to-end. Strong ownership mindset with a history of leading critical platform components in fast-paced environments.

Skills

Languages	JavaScript, TypeScript, Python
Frontend	React.js, Next.js
Backend	Node.js, FastAPI
Databases	PostgreSQL, Supabase
AI / ML	RAG, LLM Integration, Langflow, OpenAI APIs, Hybrid Search, BM25
DevOps	Docker, Kubernetes, Helm, Helmfile, GitHub Actions
Cloud & Infra	API Gateways, OAuth2 Proxy, Kubernetes Gateway API, Traefik
Observability	ClickHouse, HyperDX, ClickStack (logs, traces, metrics)
AI-Assisted Dev	Cursor, Claude Code, VS Code Cline

Experience

One Convergence

Software Engineer

Hyderabad, India

Feb 2023 – Present

Dkubex — Enterprise GenAI Platform

- Led core engineering of a GenAI platform enabling model deployment, RAG pipelines, and AI application hosting.
- Re-architected the RAG pipeline, reducing container size from ~10GB to ~2GB (**80% reduction**).
- Improved throughput from 30 to **50–80 concurrent requests per GPU** (>2x scalability).
- Enhanced response accuracy using hybrid search and BM25 indexing.
- Introduced multimodal ingestion, improving dataset quality and enabling complex data handling.
- Integrated Langflow for agentic workflows; built custom components for seamless platform integration.
- Unified authentication across services using OAuth2 Proxy.

Dkubex 2.0 — Platform Revamp & Next-Gen Architecture

- Led integration of multiple open-source AI tools into a lightweight unified platform.
- Built backend services with FastAPI and frontend with Next.js.
- Managed multi-service deployments using Helmfile and Kubernetes.
- Implemented observability stack with ClickStack, ClickHouse, and HyperDX.
- Designed scalable routing using the Kubernetes Gateway API and Traefik.
- Solved complex integration challenges across tools (routing, authentication, base paths) with minimal source modifications.
- Accelerated delivery using AI-assisted development tools (Cursor, Claude Code).

SecureLLM — LLM Gateway Platform

- Independently designed and built an **LLM proxy gateway** for secure API abstraction and routing.
- Developed backend with FastAPI and frontend with Next.js.
- Enabled dynamic routing across multiple LLM providers with API key abstraction.
- Used Supabase for authentication and database management.
- Containerized and deployed via Docker, Kubernetes, and Helm.

Infosys

Digital Specialist Engineer

Hyderabad, India

Aug 2021 – Feb 2023

- Worked as a Full Stack Developer on a client-facing web platform.
- Built responsive UI supporting both desktop and mobile environments.
- Developed backend APIs and designed database schemas using ORM tools.
- Wrote data migration scripts to support legacy system modernization.

Education

Vishnu Institute of Technology (Autonomous)

Bachelor of Technology, Mechanical Engineering

2017 – 2021