NIKUNJ KHITHA

njkhitha2003@gmail.com +91 9540234616 LinkedIn Github Portfolio

Profile

Software Engineer specialised in productionizing Generative AI within the enterprise ecosystem. Expert in architecting GraphRAG systems and multi-agent orchestrations for complex enterprise data.

Technical Skills

Languages: Python, Java, Go, TypeScript, JavaScript, SQL, Bash

Frameworks: FastAPI, Spring Boot, Gin, Express.js, Next.js, React

AI & LLM: RAG, Graph RAG, ETL Pipeline, LangGraph, Langchain, Vector Store, AWS Bedrock, Spring AI, MCP

Platforms: AWS, Microsoft Azure, Vercel, AWS Lambda, MongoDB Atlas, AWS S3, Auth0, Supabase, Neo4j, n8n

Tools: Git, GitHub, Jenkins, Docker, Kubernetes, Swagger, Grafana, Prometheus, Playwright, JUnit, PyTest, Nginx

Professional Experience

Armorcode **Jan 2025 – Present**

Associate Engineer (Platform & GenAI)

- Architected "Nexus", a hybrid GraphRAG engine using Python Fast API and Neo4j, integrating 5+ enterprise sources to improve retrieval precision by 40%.
- Core contributor to "Anya", ArmorCode's agentic AI security champion, delivering 80% MTTR reduction for Fortune 500 customers through intelligent vulnerability triage and remediation.
- Engineered a unified proxy gateway with load balancing for Gemini, Claude, and OpenAI, reducing token costs by 37% through prompt-template optimization, custom caching, and request pooling.
- Designed and implemented multiple MCP servers in n8n, Go and Python and integrations for internal platform workflows and customer-facing use cases, improving AI feature adoption and operational efficiency.
- Created and deployed various AI-based agents/automations in n8n and Python for SDLC, security orchestration, and documentation workfows, reducing manual efort and enabling scalable, repeatable processes.
- Received the first "AI Ninja Award" and recognized as the youngest recipient for pioneering AI automation, MCP-based tooling, and platform intelligence initiatives.

Xansr Media (Aiko) **Jun 2024 – Dec 2024**

SDE Intern (Backend/AI)

- Built scalable backend microservices for AIKO (voice-based sports media assistant) using FastAPI, Python, NestJS, and async task queues, achieving 96% accuracy in AI-generated commentary.
- Developed and maintained ETL pipelines and AI microservices integrating Azure AI Foundry, Azure AI Search (semantic RAG), Azure Speech SDK, and SQLAlchemy to deliver personalized sports highlight reels.
- Engineered Fantasy GPT, a multi-step reasoning sports agent using LangGraph, FastAPI, Express.js, Azure SQL (SQL RAG), and a fine-tuned LLM, achieving 98% resolution for complex sports queries.

Projects

CodeNex **Dec 2025**

- Architected microservices-based platform automating full-stack web application generation using LLMs, replicating functionality of industry leaders like Lovable, Emergent and v0.
- Developed custom Reverse Proxy (Node.js & Redis) enabling dynamic wildcard routing (e.g., *.app.domain.com), mapping 10,000+ ephemeral subdomains to internal cluster IP addresses with <10ms lookup overhead.
- Integrated tool calling, reducing context tokens by 50% and improving latency by 50%.

Link | *[Spring Boot, Spring AI, Open Router, Kubernetes, Redis, NodeJs, Kafka, Microservices]*

CodeNex AI API Gateway **Feb 2026**

- Architected a unified AI API gateway using Go/NodeJS and Gin/Express.js, enabling automatic protocol translation across OpenAI, Claude, and Gemini formats to standardize multi-agent workflows.
- Engineered resilient provider pooling system with real-time health tracking, LRU selection, and automatic failover.

Link | *[Redis, PostgreSQL, Gemini, OpenAI, Claude, NodeJS, Express.js, React, Go, Gin]*

Education

The NorthCap University **Aug 2021 – Jun 2025**

B.Tech CSE (CGPA: 8.16) *Gurugram, India*