

Sahil Singhai Jain

 Github  LinkedIn  Hackrank  +91-8839955166  Portfolio  Stack Overflow

 Sahil.singhai.jainn@gmail.com

Summary

GenAI Backend Engineer with 3+ years of experience building **scalable, high-performance systems** using **Python** and **Golang**. Experienced in **RAG, Agentic AI**, and **LLM integrations**, with a strong foundation in **microservices** and **distributed architectures**. Currently working on **IIoT cloud solutions**, designing **secure** and **scalable systems** for real-world industrial use cases. **A problem solver who makes complex systems work.**

Skills

Language — Python3, Golang1.24 | **Operating Systems** — Linux (Ubuntu), Windows | **Databases** — SQLite, MySQL, Redis, Postgres | **Technical Skills** — Git, Github/Gitlab, Docker (CLI), Kubernetes (CLI), Model Context Protocol (MCP), Prometheus, Grafana, Loki, Terraform (Basic), Prompt Engineering. | **Framework/Library** — Python: FastAPI, Flask, Django, Falcon, RAG, Gen AI • Go: Gin, Gorm | **Other** — RestAPI (Go & Python), GRPC (Go) | **Tools** — Micro-Service, GitHub, Gitlab, Postman, Dbeaver, VS Code, Amazon Web Services, Agentic AI, Neovim, Nvchad

Professional Experience

Associate Consultant, Capgemini Jan 2026 – Present | Noida, India

- Architected and developed **end-to-end IIoT data pipelines**, connecting **PLC → OPC-UA → Edge → UNS → AWS Cloud**, enabling **30%+ faster data processing** and scalable industrial data ingestion.
- Built a **dynamic orchestration platform** coordinating **10+ applications/services**, enabling user-defined workflows and real-time **OPC-UA data enrichment** with flexible data structures.
- Designed an **Edge Application Management platform** supporting **multi-instance deployments**, application configuration, and lifecycle management across **50+ edge instances**.
- Engineered **queue-based edge provisioning infrastructure**, enabling concurrent provisioning/unprovisioning and reducing manual edge setup effort by **40%+** through one-click **edge gateway cluster deployment**.
- Implemented centralized observability using **Prometheus, Grafana, and Loki**, providing monitoring and distributed log visibility across **100+ edge nodes/services**.
- Collaborated with **global engineering teams** to deliver **production-ready enterprise solutions**, improving system reliability and deployment efficiency by **30%+**.

Associate Software Engineer, Netlink Software Private Limited Apr 2023 – present | India

- Improved the **AskMe Engine** service performance by **50%** through backend optimizations, Gen AI-assisted tagging, and enhanced Highcharts visualizations using Python3.10.
- Leveraged **ChatGPT** and **prompt engineering** to enhance response quality, driving higher user engagement and more contextual answers.
- Implemented **Qdrant vector** search by chunking text, generating **embeddings**, and storing them for fast **semantic retrieval**.
- Developed a Django-based financial tool, increasing budget analysis efficiency by **15%**.
- Built and deployed **Lumenore-MCP** to extend **Lumenore's AI analytics** functionality to **LLMs and VLMs** ([GitHub](#)).

Projects

JsonYaml-Formatter, JSON & YAML Developer Tool, Currently Live Jul 2026 – Sep 2026

Built and deployed a browser-based JSON/YAML developer tool supporting **formatting, validation, minification, conversion, and comparison**, with a fully client-side architecture for fast and privacy-focused processing. • **JavaScript** • **JSON/YAML** • **Vite** • **Cloudflare** • **Monaco Editor** • **Web APIs**

Ask Me, Lumenore's Feature Apr 2023 – present

Enhanced **Lumenore's "Ask Me"** feature by integrating GenAI for intelligent tagging, data formatting, **RAG**, and dynamic visualizations. Built RESTful APIs powered by LLMs, optimized metadata services for contextual responses, and improved **auto-dashboards** using prompt engineering, caching, and AI-driven KPI generation. • **LangChain** • **OpenAI API** • **Prompt Engineering** • **Redis** • **Qdrant**, • **Vector Embeddings** • **RAG** • **Databases** • **Python3.10**.

Resume Parser, Generative and Agent Nov 2025 – Nov 2025

A resume parsing application built with **FastAPI**, **LangChain**, **LangGraph**, and **OpenAI GPT-4**. This application extracts structured information from PDF and DOCX resumes with high accuracy. • Utilized **LangGraph**, • **FastAPI**, • **Streamlit**, • **OpenAI API** • **Python3.10**.

Education

Bachelor of Technology, Oriental Institute of Science and Technology 2019 – May 2023

- CGPA - 7.88 / 78.8%
- Developed projects including YouTube Downloader and Bondly, demonstrating practical application of IT skills in web development and cloud technologies.