

Soham Ghosh

E-mail: ghoshsoham71@gmail.com | *Phone:* 9835029117 | [LinkedIn](#) | [Portfolio](#)

SUMMARY

Software Engineer specializing in Gen-AI systems, building LLM-powered workflows and scalable backend infrastructure. Experienced in Gen-AI frameworks and distributed systems, with a focus in performance optimization, automation and real-world AI Applications.

PROFESSIONAL EXPERIENCE

Software Engineer - Gen AI
Tata Consultancy Services

January 2025 - Present
Kolkata, India

- Developed a **Neo4j GraphRAG** pipeline to transform source code and technical documents into interconnected knowledge graphs using **AST**-semantic hybrid chunking, enabling dependency-aware retrieval and **LangChain**-based summarization.
- Developed a **rate limiter** that tracks and adapts to LLM max-token and cost limits and breaks down to compress large prompts into cached batches using **LLMLingua**, reducing latency and token usage by **20%**.
- Developed a hybrid **PII** detection and masking pipeline leveraging **GLiNER**-based NER, Microsoft **Presidio** analyzers, and custom regex validators to redact sensitive information from source code and enterprise documents.
- Developed **LangGraph**-based workflows for feedback generation and evaluation for **GraphRAG** systems, using **Ragas**, answer-faithfulness assessment, and custom graph-based metrics.

DevOps Engineer
Nokia Solutions and Networks

September 2023 - June 2024
Bengaluru, India

- Developed cron-scheduled **Jenkins** pipelines with parallel builds for artifact deployment, retrieval and migration achieving **55–60%** runtime optimization.
- Automated the report generation of **Nmap**-based TCP/UDP port scanning and OS security patching across **10+** isolated clusters using **Shell** scripts and **Groovy**.
- Developed **Cucumber** executable **Gherkin** scripts and leveraged UI Testing with **Selenium** Scripts for **20+** components using **Java**.

TECHNICAL SKILLS

- Programming Languages:** Java, Python, Shell Script
- Databases:** PostgreSQL, Neo4j, ChromaDB, Redis
- Web Development:** FastAPI, HTML, CSS, REST APIs, GraphQL, Celery
- AI:** LangChain, CrewAI, RAG, LangGraph, Transformers, PyTorch, FastMCP
- DevOps & Infrastructure:** Jenkins, Docker, Kubernetes, Ansible, Terraform, SonarQube, Git, AWS

PROJECTS

tokentaxi - 🌀 : Rate-limit-aware Python library published on PyPI for LLM routing

- Developed a BYOC (Bring Your Own Client) LLM router to solve vendor lock-in and prevent **429** rate-limit errors, providing a secure, transparent alternative to black-box proxy services.
- Architected a distributed state layer using **Python** and **Redis**, synchronizing provider health and usage metrics across server instances to coordinate reliable traffic at scale.
- Engineered adaptive **load balancing** and automated failover, utilizing real-time performance telemetry and predictive quota management to optimize inference across any fleet of AI providers.

Tech Stack- Python, Asyncio, Redis, Streamlit, Pydantic.

BurstDB: Synthetic Database Generator – 🌀 <https://burst-db.vercel.app/>

- Architected **Redis**-backed **Celery** orchestration to handle parallelized, high-throughput generative data synthesis pipelines for large-scale, heterogeneous enterprise synthetic databases.
- Developed a custom **ConstraintGraph** engine using Hierarchical Multi-table Analysis to preserve complex referential integrity and maintain multi-table statistical correlations.
- Implemented high-entropy generative models with differential privacy guarantees to synthesize data securely across diverse, heterogeneous database schemas.

Tech Stack- FastAPI, Next.js, Redis, Celery, Generative Models

EDUCATION

Bachelor of Technology - Kalinga Institute of Industrial Technology, Bhubaneswar [2020-24], **CGPA:** 9.19

AWARDS & ACHIEVEMENTS

TCS AI Hackathon - 2025: Achieved notable recognition for designing a code review assistant prototype that demonstrated measurable productivity gains and a strong practical impact.

TCS Young Entrepreneur Award - 2026: Recognized for innovative entrepreneurial thinking and outstanding business acumen among young leaders.