

Sheik Sadi Rohmotullah

Senior Software Engineer — AI Product

Dhaka, Bangladesh (open to Tokyo relocation) | +880-1973-908-532 | biis.saadi@gmail.com | github.com/SheikSadi | linkedin.com/in/sheik-sadi-rohmotullah

PROFILE

Backend-first software engineer with 4.5 years shipping production LLM and voice/NLP systems on cloud infrastructure, currently acting Head of Engineering at a GenAI SaaS company. Builds RAG pipelines, LLM-powered analytics, and quality-controlled agentic workflows — from architecture through evaluation (RAGAS, DeepEval, TruLens) to manage the inherent uncertainty of LLM output — while owning cost and latency trade-offs. Track record of tying AI systems to measurable business outcomes, including a recommendation-engine rebuild that lifted CTA/click-through by 50% via A/B testing. Prior contract experience building computer-vision and LLM-security systems in Tokyo; conversational Japanese speaker actively seeking to relocate to Japan.

PROFESSIONAL EXPERIENCE

Acting Head of Engineering — ELELEM AI (UK, Remote)

Jun 2026 – Present

- Assumed CTO-level responsibility for a GenAI SaaS company's product and engineering roadmap after the founding CTO's departure, reporting directly to the CEO.
- Own architecture, quality (evaluation/QA), and delivery for two live AI products used by enterprise clients including Sony/SIE.
- Lead a 3-engineer team while engaging directly with client stakeholders on technical strategy and delivery.

Lead Engineer — ELELEM AI

Nov 2025 – Jun 2026

- Built the Dhaka engineering team from scratch, hiring 2 of 3 current engineers; ran 2-week sprints end-to-end for 1.5 years.
- Owned the full GCP infrastructure stack (DevOps, networking, security) across two production products.

Software Engineer (Backend) — ELELEM AI

May 2024 – Nov 2025

- Built a RAG engine ("Snippets API") with hybrid search serving hundreds of live queries/day — real-time, cost/latency-constrained AI product work.
- Built LLM-based analytics: success/failure and CTA tracking, intent classification, and topic grouping.
- Scaled the platform to 8 active clients; built onboarding automation cutting manual overhead significantly.

Back-end Developer — ELELEM AI (formerly CONCURED)

Jul 2023 – Feb 2024

- Decomposed a monolithic GCP pipeline into two independently deployable microservices.
- Re-architected a recommendation engine using text embeddings; A/B testing showed a 50% lift in CTA/click-through.

Software Engineer (Backend, Contract) — EBLICT, Bangladesh Ministry of ICT

Jan 2024 – Jun 2024

- Led backend and ML delivery for a Virtual Private Assistant project, coordinating a 10-person cross-functional team.

AI Engineer (Contract) — Hiperdyne Corporation, Tokyo, Japan

Dec 2023 – Feb 2024

- Built a computer-vision face-privacy pipeline (YOLOv4/v8, YOLOX, DeepFace) for selective blurring/swapping.
- Researched LLM jailbreaking vulnerabilities to inform secure production LLM deployment.

Junior Back-end Developer → NLP Intern — CONCURED

Feb 2022 – May 2023

- Replaced a paid third-party keyword-extraction service with an in-house open-source solution, eliminating a recurring monthly cost.

TECHNICAL SKILLS

AI / GenAI	LLM production integration, RAG pipelines, LangChain, LangGraph, OpenAI SDK, Anthropic SDK, RAGAS, DeepEval, TruLens
Backend	Python (Expert, 4.5 yrs), FastAPI, Django, Flask, AsyncIO, microservice architecture
Cloud & Infra	GCP (Expert — Cloud Run, GKE, Pub/Sub, Cloud Build), AWS (personal projects), Docker, Kubernetes, GitHub Actions
Data	MongoDB, ElasticSearch, Redis, BigQuery, MySQL, PostgreSQL
ML / Data Science	TensorFlow, PyTorch, ONNX, Pandas, NumPy, Scikit-Learn, NLTK
Leadership	Sprint planning, roadmap ownership, budget management, stakeholder & client communication

EDUCATION

B.Sc. Electrical and Electronics Engineering
Bangladesh University of Engineering and Technology (BUET), 2021

LANGUAGES

Bengali — Native | English — Fluent | Japanese — Conversational