

Zahid Sarker

Senior Software Engineer — Full-Stack & AI/LLM (RAG · Agents · MCP)

Dhaka, Bangladesh (UTC+6) | zahid.sarker111@gmail.com

Portfolio: zahid-sarker-7.github.io | LinkedIn: linkedin.com/in/zahid-sarker7 | GitHub: github.com/Zahid-Sarker-7

SUMMARY

Senior software engineer (4+ years) and full-stack developer specializing in production Generative AI — LLM applications, RAG, multi-agent systems, MCP, and the REST APIs behind them. Led a remote MCP server to GA, re-architected product-wide documentation search on Google Vertex AI Search, and cut a production service's cost 98%. Top-5% performer.

EXPERIENCE

Senior Software Engineer (promoted Mar 2026), **Optimizely** — 1500+ person global software company May 2022 – Present

- Re-architected the RAG system behind Optimizely's entire public documentation onto **Google Vertex AI Search**, replacing an unreliable prior system — **significant accuracy & relevance gain**.
- Led a remote **MCP (Model Context Protocol)** server to general availability, enabling external AI agents to operate the product through its REST APIs — built **6 MCP tools** with audit logging, Datadog observability, and RFC 9728 OAuth discovery.
- Delivered the **full CRUD toolset** for the experimentation agent platform — covering **~90% of the experimentation API** across all entity types — and built **end-to-end testing** to prevent regressions.
- Built end-to-end MCP audit logging coordinated **across 3 teams** (API, backend, frontend) and user-level role authorization for the AI assistant **across 3 products**.
- Cut a production service's cost by **98%** by replacing brute-force polling with adaptive tiered workload reduction.
- Mentor junior engineers and conduct technical hiring interviews; **SPARK award** (top 5% of performers).

Teaching Assistant, BRAC University — Computer Science & Engineering May 2021 – Dec 2021

- Taught programming fundamentals and object-oriented programming to **hundreds of students** over the final two semesters — led lab classes, designed and graded assignments, and mentored students one-on-one.

OPEN-SOURCE PROJECTS

Askable — Production RAG engine (FastAPI · Next.js · Python) github.com/Zahid-Sarker-7/askable

- Chat-with-your-docs that answers **only** from source content with citations, and abstains when the answer isn't there. Hybrid retrieval (dense + BM25 + RRF) → cross-encoder reranking → compression → cited, streamed answers.
- Guardrails (faithfulness scoring), semantic cache, per-session isolation, pluggable backends. Gated by a **RAGAS** eval harness — **100% faithfulness + 100% context recall** on the eval set.

Converse — Multi-agent LLM engine (LangGraph · FastAPI · Next.js) github.com/Zahid-Sarker-7/converse

- LangGraph graph (director → role-player → **LLM-judge**) returning a structured rubric scorecard with verbatim evidence. Provider-agnostic port (OpenAI/Anthropic/Groq/Gemini swap by config); SSE streaming; automated test suite.

TECHNICAL SKILLS

AI / LLM: Generative AI · RAG (hybrid retrieval, reranking, RRF) · LLM agents & multi-agent (LangGraph, LangChain) · MCP · prompt engineering · evals (RAGAS) & guardrails · vector DBs (pgvector, ChromaDB, Elasticsearch, Upstash) · LLM APIs (OpenAI, Anthropic, Gemini, Groq) · function calling / tool use · SSE streaming

Languages & Frameworks: Python, JavaScript/TypeScript · FastAPI, Node.js · React, Next.js · REST API design · microservices

Cloud & Infra: Google Cloud Platform (Associate Cloud Engineer) · Docker · Terraform · CI/CD · PostgreSQL · Datadog/Grafana observability · TDD

AI Dev Tools: Claude Code · GitHub Copilot · Cursor

EDUCATION & CERTIFICATIONS

- **B.Sc. in Computer Science & Engineering**, BRAC University — CGPA 3.87/4.00; merit scholarship 8 consecutive semesters (2018 – 2021)
- **Google Cloud Platform — Associate Cloud Engineer** (certified 2024)