

Rushad Suresh Khandekar

+91 8485008201 | rushadkh2020@gmail.com | linkedin.com/in/rushad-khandekar | github.com/Rushadkh | Navi Mumbai, Maharashtra, India

Professional Summary

Software Engineer with experience building backend systems, AI-powered search pipelines, and multi-agent LLM workflows. I work with Python, FastAPI, LangGraph, OpenSearch, and RAG to ship systems that work in production. I enjoy solving real problems with AI and backend engineering.

Technical Skills

- Gen AI & Agentic AI: Generative AI, LLMs, LangGraph, LangChain, MCP Protocol, RAG Pipelines, Hybrid Search, Semantic Search, Embeddings, Multi-Agent Orchestration, Prompt Engineering
- Backend & APIs: Python (Expert), FastAPI, JSON, RESTful APIs, JWT Authentication, SSE, WebSockets, AsyncIO, Microservices
- Data & Search: OpenSearch, FAISS, Vector Databases, MySQL, MongoDB, SQLite, pandas, Flatbuffers, Event-Driven Architecture
- Cloud & DevOps: Google Cloud Platform (GCP), Cloud Run, Vertex AI, GitLab

Professional Experience

Software Engineer — Capgemini | Sep 2024 - Jul 2026

Project: iAssist (Testing Agent) | Nov 2025 - May 2026 | Technologies: Python, FastAPI, LangGraph, LangChain, OpenSearch, MySQL, RAG, Embeddings, GitLab

- Designed and implemented a hybrid search pipeline combining keyword and semantic search over OpenSearch — enabling support engineers to retrieve relevant logs, incidents, and SME knowledge using identifiers like E2E IDs, MDNs, and error messages.
- Built and tuned embedding-based semantic search that captures meaning rather than exact keywords — allowing the system to surface similar historical incidents even when described differently, significantly improving retrieval accuracy.
- Wired the retrieval pipeline into the LangGraph-orchestrated AI workflow, ensuring retrieved logs, incidents, and SME knowledge were correctly assembled as RAG context before being passed to the LLM for response generation.
- Developed FastAPI endpoints supporting the iAssist chatbot — handling input validation, OpenSearch queries, and coordinating the end-to-end AI response flow for production troubleshooting requests.
- Contributed to the knowledge synchronization pipeline — working with scheduled jobs that loaded resolved incidents and SME resolutions, regenerated embeddings, and refreshed OpenSearch indexes daily.

Project: AI-Powered Order Accuracy System (McDonald's Japan) | Jan 2025 - Nov 2025 | Technologies: Python, Flatbuffers, TCP/IP Sockets, HubHook POS, Tkinter

- Led backend delivery as primary engineer — designed and shipped a real-time comparison engine validating Edge AI camera detections against live POS transaction data.
- Achieved sub-millisecond serialization overhead by adopting Flatbuffers for binary telemetry streams from Sony IMX500 AI cameras, enabling zero-copy data access across the pipeline.
- Architected a TCP/IP socket listener for the HubHook POS, consuming event-driven streams (STORE, BUMP, SERVE) in real time with no buffering delay.
- Resolved complex async race conditions by designing state-synchronization logic that reconciles asynchronous POS events (voided/modified orders) with camera triggers, preventing false accuracy flags.
- Built a Tkinter crew-facing dashboard with instant Green/Red visual feedback per order line, reducing manual verification effort for kitchen staff.

Education

JSPM Rajarshi Shahu College of Engineering, Tathawade — Pune, Maharashtra, India

Bachelor of Technology in Electronics and Telecommunication Engineering | 2020 - 2024 | CGPA: 7.31/10.0

Certifications & Achievements

- Google Cloud Associate Cloud Engineer (GCP ACE) | Google Cloud | July 2025
- Internal Certifications: L2 & L1 Technology Certifications (Capgemini)