

RAHUL SOHANDANI

+1 201-423-0534 | rsohandani@gmail.com | [LinkedIn](#) | [Github](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Full-Stack Software Engineer with 3+ years owning backend systems and applied LLM features end-to-end — from prototyping and evaluation to production deployment. Experienced in **agent reliability, retrieval/RAG, evaluation pipelines, prompt & tool orchestration, containerization, data modelling, and distributed systems**, with a strong focus on **measured improvements in agent performance, cost, and correctness**.

SKILLS

- **Languages:** Python, TypeScript, Java, SQL, JavaScript
- **AI & LLM:** Prompt Engineering, RAG, Agent Orchestration, LangChain, OpenAI, PyTorch
- **Evaluation & Observability:** Prometheus, Grafana, OpenTelemetry, Model Evaluation (SFT, Offline/Online Evals)
- **Web & Backend:** FastAPI, Node.js, REST, PostgreSQL (Supabase), Redis, Docker, Kubernetes, Flask, Django, React
- **Big Data:** Apache Spark, Kafka, Hadoop, Databricks, Data Pipelines

WORK EXPERIENCE

Radius Technology Systems

Aug 2025 - Oct 2025

Fellow Software Developer

Cambridge, MA

- Architected production-grade MCP payment infrastructure with cryptographic request signing, enabling **autonomous agent tool-calling** and secure transaction execution.
- Built and shipped **agent orchestration tooling** (TypeScript SDK) enabling function/tool-calling workflows and faster agent iteration; reduced integration time by **60%**.
- Deployed **100+ production LLM workloads** across AWS and GCP with Docker, Kubernetes, and CI/CD, supporting **evaluation-driven iteration**, high availability, and rapid rollout.

LTIMindtree

Jul 2021 - Aug 2023

Senior Software Engineer

Bangalore, India

- Engineered real-time data pipelines using Spark Structured Streaming and Kafka, shrinking processing latency from **3 hours to near real-time** and restoring critical analytics SLAs.
- Led a cloud migration that increased throughput by 70% while optimizing **cost efficiency** and infrastructure performance.
- Implemented **observability and evaluation metrics** (latency, failure modes, throughput) using Prometheus and Grafana to debug and improve production ML systems.
- Introduced a **time-series forecasting pipeline** that improved capacity planning and delivered **94.3% prediction accuracy**.
- Scaled ML microservices via Kubernetes HPA, ensuring reliable **job orchestration** for high-volume serving.

Netwin Infosolutions

May 2019 - Jul 2020

Software Trainee

Nashik, India

- Automated ingestion and preprocessing of **10,000+ labeled samples** on AWS SageMaker to generate **high-quality training data** for AI models.
- Delivered end-to-end ML pipelines with 91% accuracy, managing the full **research-to-production** lifecycle on SageMaker.
- Eliminated manual setup by provisioning versioned, reproducible **infrastructure with Terraform** across 10+ data and AI services.

RELEVANT PROJECTS

Cerebro - Multimodal AI "Second Brain" | [GitHub Link](#) | [Video Demo](#)

- **Constructed** a multimodal intelligence system using **FastAPI, Next.js, and PostgreSQL (pgvector)** that enables conversational recall across audio transcripts, PDFs, and web content.
- **Implemented** a multi-stage hybrid information retrieval strategy by integrating semantic vector search with BM25 keyword precision and a **Cross-Encoder re-ranker** to ensure high-accuracy context grounding.
- **Streamlined** high-volume data ingestion by building an asynchronous pipeline with **Redis and Celery** to handle Whisper-based audio transcription and Vision-LLM captioning with sub-3s latency.

EDUCATION

Stevens Institute of Technology, Hoboken, NJ | MS - Computer Science | GPA 4.0/4.0

Aug 2023 - May 2025

Relevant Coursework: Big Data Technologies, Data Mining, Natural Language Processing, Data Structures, Deep Learning

University of Mumbai, Mumbai, India | BEng - Computer Engineering | GPA 3.58/4.0

Jul 2017 - May 2021

PUBLICATIONS AND ACHIEVEMENTS

- AWS Solution Architect Associate Certification | Amazon Web Services | [Certification Link](#)
- Identification of food oils and quality | IEMAICLOUD 2021 | SPRINGER | [Publication Link](#)