

RAHUL SAMANT

+91-9368884561 | rahulsamantcoc2@gmail.com | linkedin.com/in/ rahul-samant-kb37 | github.com/rahulsamant37 | Portfolio

OBJECTIVE

Aspiring Applied Scientist with expertise in machine learning, NLP, and deep learning. Proven experience developing innovative, data-driven solutions in AI, contributing to production-ready applications. Passionate about solving real-world problems at scale and advancing customer-facing technologies in complex environments.

TECHNICAL SKILLS

Programming & Frameworks: Python, FastAPI, LangChain, LangGraph
Data Science & ML: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Statistics, Probability
Deep Learning & NLP: PyTorch(Basic), Hugging Face, RAG, Natural Language Processing (NLP), BERT
Cloud & MLOps: AWS (SageMaker, EC2, Lambda, S3), Docker, Git, GitHub Actions
Core CS: Data Structures & Algorithms (DSA), System Design, Object-Oriented Programming (OOP)
Soft Skills: Communication and Interpersonal Skills, Team Leadership, Analytical Thinking

WORK EXPERIENCE

Generative AI Engineer Intern

Embedded Control Systems | Internship | Feb 2025 – June 2025 | Chandigarh, India (On-site)

- Architected intelligent meeting management system using LangGraph multi-agent workflow that automates scheduling, conducts real-time transcription, and generates actionable summaries, reducing meeting overhead by 85% for 5-member team.
- Built production-grade infrastructure on AWS EC2 with MongoDB and FastAPI, serving 3 revenue-generating MVPs with ₹1k MRR and 99.99% uptime.
- Engineered automated Google Maps scraper using LangGraph agents, extracting 50k+ business listings with 95% accuracy and reducing manual research time from 8 hours to 15 minutes.
- Built AI tool recommendation system that suggests productivity tools to non-technical team members based on their workflow patterns, reducing manual task completion time by 45% and improving cross-departmental efficiency.

AI Engineer Intern

Jan 2025 – Feb 2025 | Xzayogn (EdTech) | Remote

- Developed end-to-end job application automation reducing user effort from 6 hours/week to 8 minutes while improving interview rates from 3% to 18%.
- Implemented memory-augmented search with 90%+ relevance filtering across 8 major job boards, eliminating 97% of irrelevant listings.
- Built ATS optimization engine achieving 95%+ ResumeWorded scores through intelligent keyword integration and quantified achievement matching.
- Deployed scalable microservices on AWS Lambda/BeanStalk with TypeScript frontend, serving 50+ beta users.

Python Developer Intern

Infosys Springboard | Sept 2024 - Jan 2025 | Remote

- Engineered Scrapion multi-agent system monitoring 15 global e-commerce sites with 50k+ weekly SKU extraction and RAG-powered enrichment.
- Achieved 5x throughput improvement through async Playwright implementation with hourly data freshness for Fortune-500 retail clients.
- Built natural language query interface using Streamlit and LangChain, delivering sub-second retrieval for business analysts.

PROJECTS

LLM-from-Scratch - End-to-End Large Language Model | Project Developer | [GitHub Link](#)

- Built a GPT-style Large Language Model from the ground up, implementing core Deep Learning and NLP principles for architecture design, dataset curation, pretraining, and fine-tuning.
- Improved learning and generalization by carefully adjusting model settings during practice runs, systematically comparing word encoding approaches, and adding intuitive checkpoints to measure progress and guide troubleshooting.
- Impact: Enabled end-to-end experimentation in LLMs, improved benchmarking efficiency for model iterations by 70%, and showcased reproducibility for foundational LLM research.

AI-Powered E-commerce Assistant | Lead AI Engineer | [GitHub Link](#)

- Developed a containerized, AI-driven product assistant using an agentic RAG workflow, achieving 92% accuracy in contextual recommendations and reducing LLM hallucinations by 80%.
- Technologies: Python, LangChain, LangGraph, FastAPI, AstraDB, Docker, Kubernetes, AWS, GitHub Actions, Selenium, Streamlit.
- Established a robust MLOps evaluation framework using Ragas, improving the RAG pipeline's Context Precision score from 0.78 to 0.95 through iterative retriever optimization.

EXTRA CURRICULAR ACTIVITIES & ACHIEVEMENTS

- LeetCode:** Solved 300+ problems with a 90% acceptance rate; current rating: 1976
- Competitive Programming:** Achieved a Codeforces rating of **2005(HR)** and a CodeChef rating of **1761(HR)**.
- Top Hackathon:** Top-10 Finalist in the AMD × IIT Bombay Hackathon 2025 for an AI scheduling agent.
- Recognitions:** Shortlisted for Amazon ML Summer School 2025 and Infosys SpringBoard 2025.
- Open Source:** Contributed bug-fixes and small features to LangChain, LangGraph, Gradio, Agno and Google ADK enhancing their functionalities.