

NILOY SAHA

New York | sahaniloy389@gmail.com | (347)-609-8284 | Portfolio: <https://niloy-saha84-github-io.vercel.app>

LinkedIn: <http://www.linkedin.com/in/niloysaha35> | GitHub: <https://github.com/NiloySaha84>

EDUCATION

Hofstra University

B.S. in Computer Science

New York, USA

Expected Graduation, Dec 2027

- o **GPA:** 3.38/4.00, *Dean's List*
- o **Related Coursework:** Data Structures & Algorithms, Computer Architecture & Programming, Software Engineering, Object-Oriented Programming, Applied NLP and Text Mining, Operating Systems, Building RAG Agents with LLM (Course By **NVIDIA**), Intro to Artificial Intelligence

EXPERIENCE

Optimum Communications INC.

New York, USA

Software Developer Intern (Infrastructure Observability Team)

Jan 2026 - July 2026

- **Automated observability** across **5,000+** servers, measured by **10+ real-time system** and **container metrics per host**, by configuring **Zabbix agents** and **SNMP polling**.
- **Reduced manual monitoring configuration** by **70%**, measured across **5,000+ servers**, by engineering dynamic **Low-Level Discovery pipelines**.
- **Reduced alert detection time** by over **50%** during internal testing by designing **trigger-based alerting pipelines**.

ACHIEVEMENTS & ADDITIONAL EXPERIENCE

- **3rd Place, Google Cloud Rapid Agent Hackathon (GitLab Track)** | [Demo & Submission](#) — Built **ContribPilot**, an AI-powered open-source contribution coach leveraging **Gemini**, **Google Cloud Agent Builder**, and **GitLab MCP**.
- **IT Technician, Maurice A. Deane School of Law at Hofstra University** — Provided technical support for faculty, staff, and students, troubleshooting hardware, software, and classroom technology issues.
- Solved **300+ LeetCode problems** across **DSA**, strengthening **problem-solving skills** in **Arrays, Graphs**, and **DP**.

PROJECTS

MarketMind (Distributed Market Research & Competitor Analysis Platform)

GitHub: <https://github.com/NiloySaha84/MarketMind> | <https://marktmind.duckdns.org>

- Built a distributed backend for asynchronous market research workflows, using **PostgreSQL**, **Redis**, **BullMQ**, transactional outbox, dead-letter queues, **Docker Swarm**, and **Google Cloud**, achieving support for **1,400+** virtual-user sessions with sub-200 ms latency and **99% API request success rate** during load test..
- **Reduced average user-facing research latency** to **~35 ms**, measured during **load testing**, by shifting work to **asynchronous background processing** and **cached services**.
- **Secured access across 7 PostgreSQL tables**, measured by **eliminating application-layer filtering**, by enforcing **Row-Level Security** and **role-based database permissions**.

My Jarvis (Autonomous Local AI Agent)

GitHub: <https://github.com/NiloySaha84/myJarvis> | [Live Demo](#)

- Built a local AI desktop assistant using **LangGraph**, **FastAPI**, **Electron**, and **Deepgram**, engineering **real-time streaming voice interactions** and instrumenting **40+ operational metrics** across **agent**, **API**, **retrieval**, and **voice subsystems**.
- Engineered a **low-latency streaming voice pipeline** by implementing real-time **STT/TTS**, **transcript filtering**, **microphone muting**, and **concurrency-safe turn synchronization**.
- **Reduced retrieval latency** from **~2.8s** to **<1s per query**, measured in **workflow-aware RAG retrieval**, by combining **ChromaDB**, **BM25**, and **FlashRank** reranking.

SKILLS

Languages: Python, Java, TypeScript, JavaScript, SQL, C++

Frameworks/Libraries: FastAPI, Next.js, Node.js, LangChain, LangGraph, Electron

Databases: PostgreSQL, MongoDB, Redis, ChromaDB, FAISS

Cloud/DevOps: Docker, Oracle Cloud, Google Cloud, Vercel, CI/CD, Git/GitHub

Core: Distributed Systems, Microservices, REST APIs, RAG, GenAI, OOP, DSA, Observability, Debugging, SDLC