SOHAM VISHNU SONAR

United States ◇ +1 (312) 975-7439 ◇ ssonar2@hawk.illinoistech.edu ◇ LinkedIn ◇ Github ◇ Portfolio

EDUCATION

Master of Computer Science August 2023 - May 2025
Illinois Institute of Technology, Chicago, IL
Relevant Coursework: Machine Learning, Natural language processing, Advanced Operating Systems, Cloud Computing

Bachelor of Computer Engineering (Honors in Data Science) August 2018 - July 2022
Savitribai Phule Pune University
Relevant Coursework: Data structures and algorithm, Object oriented programming, Advanced database organization, Big data.

EXPERIENCE

Research Assistant February 2025 - Present
Gnosis Research Center - Illinois Institute of Technology *remote*

- Developed **agentic AI platform** leveraging **multi agent orchestration**, **LLM fine tuning**, and conversational AI to automate end-to-end workflows across 40+ node clusters, enabling **autonomous task execution** and **intelligent workflow coordination**.
- Enhanced the performance of **open source projects (IOWarp, Chronolog)**, by integrating an intuitive **natural language assistant** for **data analytics** and **AI driven operations**, reducing average data retrieval latency by 60%.
- Conducted research on testing LLM based applications development (cursor, claude), **building risk assessment frameworks** and evaluating best practices to ensure robustness, reliability, and compliance in **enterprise scale AI systems**.

Machine Learning Intern January 2025 - April 2025
Vosyn Inc. *Chicago, IL*

- Designed and optimized machine learning models using **Vertex AI** to improve real-time multilingual voice synthesis accuracy by 35%, ensuring seamless contextual translation across global markets.
- Integrated **10+ AI voice features** into customer-facing applications through automated **CI/CD pipelines** for continuous model deployment and **A/B testing**, enabling real time support and improving usability for non-technical users.
- Deployed ML models for real-time voice localization using **Kubernetes & Cloud Run**, optimizing inference via **CPU/GPU benchmarking** and reduced latency by 20% and enabled scalable cross-platform integration.

Executive March 2023 - June 2023
Hexaware Technologies *Mumbai, India*

- Achieved a 60% improvement in platform performance for a healthcare software solution by debugging and optimizing using **Python, SQL**, and database tuning techniques.
- **Automated data entry workflows**, reducing manual workload and improving data processing efficiency by 30% through Python scripting and **workflow automation**.
- Collaborated with **cross functional** teams in an **Agile Scrum** environment, and led backlog grooming and sprint planning across software engineering and QA teams, reducing post-deployment defects by 30%.

SKILLS

Programming Languages	Python, Java, C++, SQL, C#, Shell Scripting.
AI Development	LLMs, Google GenAI sdk, RAG Models, Langchain, Langgraph, MCP Server.
Machine Learning	Tensorflow, Pytorch, Scikit-learn, Supervised & Unsupervised Learning, Model Distillation.
Software Dev. & Database	Linux, Git, Agile, Scrum, Jupyter, React, MySQL, MongoDB, PostgreSQL.
Cloud & Big Data	AWS (EC2, S3), GCP (Vertex AI, GKE, GCE), Docker, Kubernetes, Hadoop, Spark.

PROJECTS

Enterprise IO Automation Framework [Link]

- Led the development of the **Model Context Protocol (MCP) server framework**, including Pandas, Parquet, Plot and HDF5 MCP servers, to **automate I/O** and **filesystem workflows** for local and cloud environments.
- Designed a **custom LLM client** using **Google Gen AI sdk** to coordinate **120+ simulation pipelines,** processing **multi-terabyte datasets** and significantly **reducing data access latency** across distributed systems.

Intelligent Security Operations Center (SOC) [Link]

- Built a **hybrid log classification system** and transformed it into enterprise grade SOC platform using ensemble **ML (BERT + Groq/Llama 3.1)** with real-time threat detection, event correlation for attack pattern identification.
- Implemented MCP based **Agentic AI framework orchestration** with **Slack** (instant threat alerts), **JIRA** (automated incident tickets), and **Grafana** (real-time security dashboards) while incorporating analyst feedback loops for continuous model improvement, reducing mean time to detect by 70% and false positives by 60%.