

VINAYAK SAHU

vinayak1672006@gmail.com • [GitHub](#) • [LinkedIn](#) • [PORTFOLIO](#) • Contact: +918858838007

Technical Skills

- **Languages & Frameworks:** Python, SQL, Shell Scripting, FastAPI, Git/GitHub, Gradio, Docker, React
- **Machine Learning / AI Libraries & Concepts:** PyTorch, TensorFlow (Model Training & Evaluation), Scikit-learn (Feature Engineering, Statistics), Transformers, SentenceTransformers, CLIP, FAISS (Vector Search), YOLOv8 (Object Detection), LangChain, LangGraph, LlamaIndex (RAG Pipelines), n8n (Workflow Automation), CI/CD, Docker
- **Big Data & Query Engines:** Hadoop, HiveQL, HDFS, Cloudera VM
- **Data Analytics & Visualization:** Pandas, NumPy, Matplotlib, Tableau, Power BI, Advanced Excel

Experience

AI Research Associate — Vidur Research (Subsidiary of Dreamskrin)

Oct 2025 – Present / Remote

- Architected Subjective-RAG system for financial analysis achieving **90%+ claim precision** and **40% hallucination reduction** through **multi-agent verification**, processing 5-10 whitelisted sources per query with **92% pipeline success rate**
- Engineered production scraping infrastructure with **99% uptime** across Cloudflare-protected domains (Yahoo Finance, Moneycontrol, Economic Times), **implementing intelligent caching** that **reduced fetch latency by 25%**
- Optimized worker-checker validation loop **reducing verification attempts from 3→1** and **end-to-end latency by 30%** through advanced chunk ranking, deduplication (**60% noise reduction**), and strict claim-matching rules
- Deployed **scalable RAG framework** with PyTorch, LangChain, and FastAPI via **CI/CD pipelines**, implementing comprehensive structured logging **enabling 100% query traceability** and real-time performance monitoring

Projects

PrepGraph - RAG-Based Course Chatbot | Personal Project [\[GitHub\]](#) [\[LINK\]](#)

- Engineered a **Retrieval-Augmented Generation (RAG)** system combining BM25 and FAISS to enhance query **accuracy by 30%** and reduce response **latency by 25%**.
- Developed a full-stack chatbot with FastAPI backend and React frontend, **integrated via CI/CD pipelines** for automated builds and scalable deployment.
- Optimized context retrieval and **chunking pipeline (400+80 overlap)** to **minimize LLM token usage by 40%** while maintaining high response relevance.

Command Copilot - Offline AI Assistant for Shell Automation (Under Active Development) [\[GitHub\]](#)

- Automated system administration tasks using an **AI-powered terminal assistant** that converts natural language queries into **Linux shell commands, eliminating 75% of manual lookups**.
- **Fine-tuned Phi-2 with Low-Rank Adaptation (LoRA)** on a 150-sample custom dataset, **improving command accuracy by 35%** in benchmark evaluations.
- Deployed within a **VirtualBox Linux VM** for fully offline execution with **no cloud dependency**.

SecuFlow – AI-Powered Security Automation Workflow [\[GitHub\]](#)

- Built a real-time intrusion detection workflow using **n8n** automation and Groq LLM, implementing **MLOps-style** alerting and risk classification with **90% less manual review**.
- Flagged IPs with **>3 failed login attempts** in **<2 min**; triggered alerts and automated UFW firewall blocking, cutting breach exposure **by >80%**.
- Developed Telegram bot for **human-in-the-loop** approval; enabled one-click firewall actions and **reduced response time** from minutes to seconds.

Achievements

- **First Prize – FED Hackathon:** Built and integrated an intelligent chatbot with **MoneyMate**, a **finance tracker**, enhancing user interaction and financial management.

Education

Kalinga Institute of Industrial Technology (KIIT), BTech in Computer Science | 2023 – 2027