VINAYAK SAHU

Diving Into AI/ML
vinayak1672006@gmail.com • github.com/07Codex07 • linkedin.com/in/vinayak-sahu-8999a9259
Portfolio: portfolio-delta-two-15.vercel.app/

Technical Skills

- Languages & Frameworks: Python, SQL, Shell Scripting, FastAPI, Git/GitHub
- ML & AI Libraries: PyTorch, TensorFlow, Scikit-learn, Hugging Face Transformers, SentenceTransformers, CLIP, FAISS, YOLOv8, LangChain, LangGraph, LlamaIndex
- Big Data & Query Engines: Hadoop, HiveQL, HDFS, Cloudera VM
- Deployment & Interfaces: Gradio, Streamlit, FastAPI, Jupyter Notebook
- Data Analytics & Visualization: Pandas, NumPy, Matplotlib, Tableau, Power BI, Advanced Excel

Projects

- **PrepGraph - RAG-Based Course Chatbot | Personal Project [GitHub]**

- Developed a **Retrieval-Augmented Generation (RAG) chatbot** for exam prep, combining **6 subjects' syllabus + 200+ PYQs** into one knowledge base.
- Implemented **LangGraph + Groq API** with FAISS + SentenceTransformer embeddings for semantic retrieval and short-context answers.
- Used by **10+ students** during semester exams, answering **100+ domain-specific queries** with **>85% accuracy**.
- Deployed via **Gradio interface**, making it accessible as a 24/7 study assistant for revision.
- Tech Stack: Python, LangGraph, LangChain, Groq API (LLaMA3-8B), FAISS, SentenceTransformer, Gradio

- **Linux Command Copilot - Offline AI Assistant for Shell Automation** (in progress) **[GitHub]**

-Built a **terminal-native AI assistant** that converts natural language into Linux shell commands for system automation.
- Fine-tuned **Phi-2 with LoRA** on a custom dataset of size 150, improving command accuracy by **35% on benchmark evaluations**.
- Deployed inside a VirtualBox Linux VM for **fully offline execution with no cloud dependency**.
- Automated operations like user management, file handling, and networking using **text-to-bash pipeline**.
- Stack: Hugging Face Transformers, Python, Shell Scripting, Low-resource LLM Inference, Prompt Engineering.

- **Big Data Analytics on MovieLens – *Hive + Hadoop Query Engine Project [GitHub]***

- Designed and implemented a **scalable ETL pipeline** for large-scale movie datasets using Hadoop & Hive.
- Created external Hive tables and optimized queries with **partitioning & bucketing**, reducing query execution time by **~40%**.
- Analyzed viewing trends, top-rated movies, and user behavior across **multi-million record datasets**.
- Managed distributed querying using **Cloudera QuickStart VM**, enabling reproducible big data analysis.
- Stack: Hadoop, Hive, HiveQL, HDFS, Cloudera VM, Linux CLI, Big Data Analytics

- **Reel2Retail – AI Fashion Video-to-Product Matching [GitHub]**

- Built an AI system to **detect fashion items in social media reels** and match them with products in a catalog.
- Applied **YOLOv8** for frame-wise detection, **CLIP + FAISS** for similarity search, and **NLP** for vibe classification.
- Achieved **~85% retrieval accuracy** on 100+ frames from the reel samples, while reducing latency via async caching and frame differencing.
- Delivered a **real-time GPU-based pipeline**, bridging fashion video content with retail product discovery.
- Tech Stack: Python, YOLOv8, OpenAI CLIP, FAISS, NumPy, OpenCV, Scikit-learn, Pandas, NLP, Asynchronous Programming, GPU Inference

Education

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