

Subandhu Agravanshi

+91-8171687666 · subandwho@gmail.com

LinkedIn: Subandhu Agravanshi · GitHub: subandwho · LeetCode: subandwho

Work Experience

Stealth Startup (Fintech), ML and Gen AI Engineer

Aug 2024 - Present

- **Engineered** agentic workflows, integrating a hybrid **RAG (Graph+LanceDB)** to process **1M** financial prospectuses and launched a financial analysis agent using **AWS Bedrock**, **Groq**, and **LangGraph**.
- **Developed 5** unique POCs for **RAG** and conversational AI algorithms, including **HYDE RAG**, **CRAG**, **ReRankers** and **Reflexion Agents**.
- Introduced an **Equity Research Agent** with **7+** functionalities for insights, **web search**, and visualization, integrating **REPL API**, **OpenFIGI**, and **QnA** chains.

Arcesium (DE Shaw Group), Machine Learning Engineer, SDE

Mar 2023 - Aug 2024

- **Code Translator with LLMs:** Automated migration of **350** Perl functions to **Python** using **Vicuna13b** with **75%** accuracy; created a **UI** for real-time translation and benchmarked **15+** LLMs, including **LLaMa** and **GALACTICA**.
- **Tax Character Prediction:** Implemented an image classifier achieving **98%** accuracy on **250+** documents; analyzed **1M+** rows to identify tax trends. Trained **XGBoost** Classifier at **0.9 ROC-AUC** score.
- **Synthetic Data Generation:** Boosted test coverage from **32%** to **80%** with automated framework generating **24,000** test cases using **LLama2**, **Mistral7b**, and **CodeLLama**.
- **Text2Code and Text2SQL:** Collaborated on the company's flagship **Gen AI** product in the **VP's AI squad** to build pipelines across **150+** database tables for **SparkSQL** text-to-query generator with an **auto-correct API**.
- **LLM Operations Setup:** Configured solutions for **data retrieval**, **vectorization**, **prompts** and **AWS KnowledgeBase** sync; established LLM-Ops observability for **10,000+** **LangChain** flows via **Langfuse**.
- **Guardrails Integration:** Deployed guardrails to reduce token usage (saving **\$3** per **10** calls) and secure data via anonymization across **50+** **LangChain** flows.

KPMG Global Services, Analyst, Technical Consulting

Jan 2022 – June 2022

- Implemented Microsoft Power Apps for analytics and client research. Oversaw API integration and developed POCs.

Education

- **UC Berkeley - Plaksha University**, Post Graduation (Applied AI & Product), CGPA: 8.5 2022-23
- **Kalinga Institute of Industrial Technology**, B.Tech Computer Science and Engineering, CGPA: 8.01 2018-2022
- **St. Paul's Church College**, 12th — ISC Board, 10th — ICSE board, Scores: 95%, 96%

Projects

- **Video Game Recommendation System:** Processed **30,000+** text game reviews to generate cosine similarity-based vector matching for title recommendations, implemented **Word2Vec** with feature engineering.
- **Phrase Matching with BERT:** Leveraged **Auto Tokenizer** and **DeBERTa Transformer** models for sequence classification, aligning US Patent data with **12** target categories.
- **Intelligent Question Recommendation Engine:** Built an **Ollama-driven LLMRecSys** to cluster and recommend **600+** JEE practice questions across **5** parameters.
- **LLM-driven Data Extractor:** Created skive.ai, a **multi-agent AI** workflow employing **Claude**, to extract and structure data from **7,000 unstructured KYC documents**, reducing verification time by **10 hours**.

Skills

Data Science: Machine Learning (Predictive Analytics, Classification), Statistical Analysis, Deep Learning (ANN, CNN, RNN), Transfer Learning, Data Visualization, Web Scraping, NLP (LSTM, Transformers), Model Evaluation, LLMs, LangChain, OpenAPI, RAG, CREW.AI, LangGraph, Langfuse, Agents, MLOps

Engineering and Tools: Python (Pandas, NumPy, Sklearn, TensorFlow, NLTK, HuggingFace), Linux, Git, AWS Bedrock, Flask, Figma, Streamlit, Gradio, Microsoft GraphDB, MLFlow, ChatBots

Publications

- “An Analytical Perspective of Machine Learning applied to Cybersecurity”
- “Understanding CNN-based models applied to interpreting ECG data”