VIKRANTH BANDARU

bandaruvikranth@gmail.com — +1(469)-439-0911 — Buffalo, NY, USA — LinkedIn — GitHub — Portfolio

Professional Summary

AI/ML professional with 2+ years of experience building end-to-end ML pipelines, RAG systems, and full-stack apps with Python, PySpark, Azure/GCP, LangChain, and React; proven gains in LLM fine-tuning, automation, and engagement.

Professional Experience

AI Product Development Intern, Vaspian - Python, LangChain, MCPs, RAG, LLMs *Sep 2025 – Present*

- Developed an AI-powered **FAQ automation RAG system** using LLMs (**Llama 3.3 70B, Llama 4 Scout**) and **Qdrant** to convert **12000+** raw call transcript JSON files into structured problem–solution knowledge, enabling Vaspian's chatbot to deliver instant, accurate answers and reduce customer problem-resolution time by **50% (from 6 minutes to 3 minutes).**
- Integrated MCP servers and Dockerized services with Git-based version control, improving deployment reliability and iteration speed.
- Built an ElevenLabs conversational AI agent linked to Airtable via webhooks, automating webinar invites and customer response tracking.

Data Scientist, Genpact — Python, Linear Mixed Models, GCP, Spark *Jul 2023 – Aug 2024*

- Assisted in engineering an ML pipeline for "Price-Pack Architecture" using **PySpark**, **LLMs**, and **elasticity modeling** to simulate over **1,000** pricing scenarios across **25+** markets and **6,000+** SKUs for **Nestlé Purina**, achieving **86%** forecast accuracy.
- Developed **Data Structurizer** for invoice query processing: automated **OCR** extraction, performed **ETL**, stored in **vector DB**, and utilized **GPT-4o API** with **LangChain** for efficient invoice search, achieving **80%** query-response accuracy.
- For Genworth and Unilever, built a self-serve **invoice QA chatbot**: extracted and structured invoice data using **GCP Document AI**, enabling rapid, accurate answers from invoice content and reducing manual lookup effort.

Software Engineering Intern, Genpact — Python, MongoDB, React, Firebase *Mar 2023 – Jul 2023*

- Designed and deployed a web-based mentor-mentee platform for **Hopeworks NGO** (with Genpact) using **React**, **SQL**, and **Python**, increasing **matching engagement by 30%**.
- Led development of a full-stack legal services platform using **React**, **Node.js**, **MongoDB**, and **Firebase**, integrating versioned docs, authentication, chat, and analytics.

Product and Engineering Intern, HighRadius — Python, Flask, CI/CD, MongoDB *Jan 2022 – Apr 2022*

- Built and deployed an AI-enabled fintech cloud application with **React**, **MongoDB**, and Python-based ML models (**XGBoost**, **Random Forest**), improving **efficiency by 30%**.
- Transformed raw, **unstructured invoices** into structured data, engineered features, and implemented ML models to identify the true invoice date among multiple fields, enabling seamless integration into the application.

Skills

- **Languages & Databases:** Python, R, JavaScript, C/C++, MongoDB, SQL (PostgreSQL, MySQL), Firebase
- **Artificial Intelligence and ML Libraries:** Machine Learning, Deep Learning, Reinforcement Learning, NLP(SpaCy, NLTK, HuggingFace Transformers), LLMs, CV, CUDA, Pandas, NumPy, PyTorch (CNNs, RNNs, LSTMs), Keras, Scikit-learn, TensorFlow
- **Tools & Technologies:** VS Code, Power BI, GCP tools (vertex AI, Document AI), Anaconda, Microsoft Office, Excel, Agile,
- **Certifications:** Microsoft Azure Fundamentals; Machine Learning; Data Science and Deep Learning with Python; NLP with Python; Elements of AI (Google); Building RAG Agents with LLMs (NVIDIA)

Education

University at Buffalo — SUNY, *Master's, Artificial Intelligence* *Aug 2024 – Dec 2025*
SRM Institute of Science and Technology, *B.Tech, Computer Science (AI/ML)* *Jun 2019 – May 2023*

Projects

- **LLMs in Social Sciences** — Pandas, Sentiment Analysis, Text Scraping, Selenium *Mar 2025 – May 2025*
  - Developed a multilingual policy analysis system using **Python**, **Pandas**, **NLTK**, **DistilBERT**, **RoBERTa**, and **GPT-O2** to process English/Hindi sources; sentiment and stance detection with **87%** accuracy for USA–India–Canada policy trends.
- **Data-Driven Wikipedia Q&A Chat-bot** — AI, GCP, RAG, Streamlit, Data Indexing *Nov 2024 – Dec 2024*
  - Developed and deployed this chatbot using **Python, Streamlit, RAG with OpenAI GPT, and Apache Solr** to index **50K+** documents, achieving **92% response** accuracy with a chat interface and analytics dashboard.

Publications

"Person Re-identification from Video using Hybrid Approach," *ICAECT 2023, IEEE Xplore.* *May 2023*

- The paper uses GANs and computer vision for person re-identification in video, improving matching accuracy in noisy settings.