

Vikranth Bandaru

📞 +1 (469) 439-0911 📩 bandaruvikranth@gmail.com 🌐 vikranthbandaru.github.io 💬 vikranthbandaru 🎙️ vikranthbandaru

Professional Summary

AI/ML professional with 2+ years of experience building end-to-end **ML pipelines**, **RAG systems**, and full-stack applications using **Python**, **PySpark**, **Azure/GCP**, **LangChain**, and **React**; demonstrated impact in **LLM fine-tuning**, automation, and customer engagement.

Professional Experience

AI Product Development Intern, Vaspian — Python, LangChain, MCPs, RAG, LLMs Sep 2025 – Dec 2025

- Developed an AI-powered FAQ automation **RAG system** using **LLMs** (Llama 3.3 70B, Llama 4 Scout) and Qdrant to convert 12,000+ raw call transcript JSON files into structured problem–solution knowledge, reducing 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** connected to Airtable via webhooks, automating webinar invites and customer response tracking.

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

- Engineered an **ML pipeline** for *Price-Pack Architecture* using **PySpark**, **LLMs**, and elasticity modeling to simulate 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**, **ETL**, vector database storage, and **GPT-4o-powered search** using **LangChain**, achieving **80% query-response accuracy**.
- Built self-serve invoice QA chatbots for Genworth and Unilever using **GCP Document AI**, significantly reducing manual invoice 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 using **React**, **SQL**, and **Python**, increasing matching engagement by **30%**.
- Led development of a full-stack legal services platform with document versioning, authentication, chat, and analytics features.

ML Engineer Intern, HighRadius — Python, Flask, CI/CD, MongoDB, ML Models Jan 2022 – Apr 2022

- Built and deployed an AI-enabled fintech cloud application using **XGBoost** and **Random Forest models**, improving operational efficiency by **30%**.
- Transformed raw invoice data into structured formats and engineered features to identify true invoice dates among multiple fields.

Skills

Programming Languages: Python, R, JavaScript, C/C++, SQL, TypeScript

Databases: MongoDB, PostgreSQL, MySQL, Firebase, Vector Databases (Qdrant, Pinecone)

Machine Learning & AI: Machine Learning, Deep Learning, Reinforcement Learning, LLMs (GPT-4o, Llama), RAG, Agentic AI

NLP & Text Processing: Natural Language Processing, SpaCy, NLTK, HuggingFace Transformers, LangChain, sentiment analysis

ML Frameworks: PyTorch, TensorFlow, Keras, Scikit-learn, XGBoost, MLFlow

Data Processing: Pandas, NumPy, PySpark, ETL pipelines, OCR

Cloud & DevOps: GCP (ADK, Vertex AI, Document AI), Azure, AWS, Docker, CI/CD, Git

Web Development: React, Flask, FastAPI, Streamlit, RESTful APIs

Computer Vision: CUDA, OpenCV, GANs, image processing

Tools & Methodologies: VS Code, Power BI, Anaconda, Excel, Agile, MCP servers, n8n

Education

University at Buffalo — SUNY Aug 2024 – Dec 2025

Master's in Artificial Intelligence

SRM Institute of Science and Technology Jun 2019 – May 2023

B.Tech in Computer Science (AI/ML)

Projects

Multilingual Policy Analysis using LLMs

- Developed a multilingual policy analysis system using **DistilBERT**, **RoBERTa**, & **GPT-based models** to process English & Hindi policy documents, achieving **87% sentiment & stance detection accuracy** for USA–India–Canada policy trends.

Data-Driven Wikipedia Question Answering System

- Built and deployed a **RAG-based chatbot** using **Python**, Streamlit, **OpenAI GPT**, and Apache Solr to index 50K+ Wikipedia documents, achieving **92% response accuracy** with an interactive chat interface and analytics dashboard.

Publications

Person Re-identification from Video using Hybrid Approach

ICAECT 2023, IEEE Xplore May 2023

- Proposed a hybrid **GAN-based computer vision** approach for person re-identification in video streams, improving matching accuracy in noisy and occluded environments.

Certifications

Microsoft Azure Fundamentals — Machine Learning, Data Science and Generative AI with Python (Udemy) — Text Analysis and Natural Language Processing With Python (Udemy) — Building RAG Agents with LLMs (NVIDIA) — Elements of AI (Google)