SATYA SAI PRAKASH KANTAMANI

Data Scientist & Gen AI Engineer

571-752-2051 | prakashkantamani90@gmail.com | Richardson, TX | linkedin.com/in/prakash-kantamani | github.com/kantamaniprakash | kantamaniprakash.github.io

PROFESSIONAL SUMMARY

Results-driven Data Scientist & Gen AI Engineer with 4+ years of experience designing end-to-end RAG pipelines, autonomous LLM agents, and production-grade ML systems. Deep expertise in LangChain, OpenAI GPT-4o, ChromaDB, FAISS, and transformer fine-tuning (BERT, T5), combined with robust cloud data engineering on Snowflake and AWS. Proven track record of delivering measurable business impact — including a 48× ML runtime reduction and 35% retrieval improvement — by engineering scalable, intelligent systems that translate complex data into actionable decisions.

EXPERIENCE

AAA National · **Data Scientist, Business Solutions Group** — Aug 2024 – Present | Richardson, TX

- Reduced ML model runtime from 24 hrs to 30 min (**48× speedup**) by migrating analytical workflows to R + Posit Workbench on Snowflake, directly accelerating underwriting decision cycles and business reporting throughput.
- Engineered enterprise-scale data pipelines processing **240M+ rows across 312 columns**, optimizing feature engineering and model training workflows to support production ML deployments.
- Automated SQL Server → Snowflake migration using BCP + AWS S3 with modulo-based partitioning, eliminating manual bottlenecks and ensuring zero data loss across a large-scale cloud migration initiative.
- Built end-to-end ETL workflows via Automate BPA and authored comprehensive data documentation in Confluence, improving cross-team onboarding speed and operational transparency across business solutions groups.
- Managed multi-track analytical project deliverables via Jira, ensuring on-time delivery, stakeholder alignment, and consistent sprint velocity across concurrent data science initiatives.

Imperium Data · **Data Analyst Intern** — Aug 2023 – Dec 2023 | FL

- Prototyped a production-ready **RAG pipeline** (LangChain + FAISS + OpenAI embeddings) for semantic enterprise document search, reducing information retrieval time by **35%** and improving answer precision and relevance.
- Engineered **LLM-augmented NLP pipelines** with zero-shot classification and Sentence Transformers, improving unstructured text categorization accuracy by **30%** across large enterprise document corpora.
- Developed K-Means + PCA customer segmentation models on Azure Databricks, driving a **25% lift** in targeted marketing effectiveness and cutting anomaly detection false positives by **40%**.
- Optimized SQL performance via partitioning and materialized views (**35% faster query execution**); built PyTorch deep learning classification models achieving **85% validation accuracy**.

Skilalogy · **Data Analyst** — Jan 2020 – Dec 2021 | India

- Fine-tuned **T5 transformer** for abstractive text summarization on equipment fault documentation, reducing document review time by **40%** and accelerating maintenance response and reporting workflows.
- Applied **BERT transfer learning** with custom NER on fault logs, improving anomaly classification precision by **28%** and enabling early identification of critical equipment failure patterns.
- Developed a computer vision defect-detection pipeline achieving **95% accuracy**, reducing manufacturing false positives by 20% and boosting overall production line efficiency by 15%.
- Automated ML data preparation via enhanced SQL pipelines and applied GridSearchCV + K-Fold CV hyperparameter tuning, cutting data prep time by 20% and improving model generalization across production datasets.

PROJECTS

Financial RAG Chatbot | *LangChain · OpenAI GPT-4o · ChromaDB · FAISS · Streamlit* | github.com/kantamaniprakash/financial-rag-chatbot

- Architected an end-to-end conversational RAG system over financial PDFs with MMR retrieval, persistent vector store, multi-document ingestion, and citation-grounded answers — enabling accurate, auditable financial Q&A at scale.

Data Analysis AI Agent | *LangChain ReAct · OpenAI GPT-4o · pandas · Streamlit* | github.com/kantamaniprakash/data-analysis-agent

- Built an autonomous ReAct agent with 4 custom tools that interprets natural language queries, executes sandboxed Python analysis, and delivers dynamic visualizations and actionable business insights from uploaded datasets.

SKILLS & CERTIFICATIONS

Gen AI & LLMs: LangChain, OpenAI GPT-4o, RAG Pipelines, ReAct Agents, Prompt Engineering, HuggingFace Transformers, ChromaDB, FAISS, Pinecone, Sentence Transformers, LlamaIndex, Vector Databases

ML / Deep Learning: PyTorch, TensorFlow, Keras, Scikit-learn, XGBoost, CNN, RNN, LSTM, BERT, T5, NLP, Random Forests, SVM, K-Means, PCA, Gradient Boosting, Ensemble Methods

Languages: Python, R, SQL **Cloud:** AWS (S3, SageMaker), Azure Databricks, Snowflake, Posit Workbench **DevTools:** Git, Docker, Jira, Confluence

Databases: MySQL, PostgreSQL, MongoDB, SQL Server **Visualization:** Tableau, Power BI, Looker, Matplotlib, Seaborn **Packages:** NumPy, pandas, NLTK, spaCy

Certifications: Google Data Analytics Professional Certificate · Applied Machine Learning — University of Texas at Dallas