Chaitanya Dasari

chaitanyadasari09@outlook.com | LinkedIn | GitHub

Professional Summary

Full Stack Data Scientist and AI Engineer with 5+ years of experience architecting scalable Generative AI and ML systems for Microsoft and Fortune 500 clients. Expert in building end-to-end Agentic RAG pipelines and autonomous systems using GPT-4, LangGraph, and Vector Databases handling 10TB+ of enterprise data. Proven track record of bridging the gap between research and production—deploying Deep Learning models via Docker/Kubernetes and robust CI/CD pipelines to serve 100,000+ daily users. Passionate about optimizing inference latency (sub-200ms) and driving AI safety in high-stakes environments.

Skills

Programming Languages
- Python; SQL; R

Machine Learning & AI
- Supervised/Unsupervised Learning, NLP, Recommendation Systems, Information Retrieval, Reinforcement Learning, LLMs(GPT-4, Claude, LLaMA)

Frameworks
- PyTorch, TensorFlow, Keras, XGBoost, Scikit-learn, Spark MLlib, LangChain, LangGraph, Hugging Face, Jupyter notebooks

Data Tools & Technologies
- Pandas; NumPy; REST APIs; ETL Pipelines; Automation; Feature Engineering; Dimensionality Reduction (PCA); SQL (MySQL, BigQuery, Snowflake, Redshift)

Statistical Methods
- Regression (Linear, Logistic, Ridge), Hypothesis Testing, ANOVA, Bayesian Inference, Time Series (ARIMA, Prophet), Clustering (K-means, DBSCAN, Hierarchical)

DevOps & MLOps
- CI/CD Pipelines, Git, Unit Testing, Automation, Model Deployment & Monitoring.

Cloud & Big Data
- AWS (SageMaker, Redshift), GCP (Vertex AI, BigQuery), Azure, Snowflake, Databricks

Data Visualization & BI
- Tableau, Power BI, AWS QuickSight (familiar), Matplotlib, Seaborn, Interactive Dashboards

Professional Strengths
- Cross-functional Collaboration, Leadership, Technical Communication, Innovation, Agile/Scrum, Mentorship, Data Storytelling

Professional Experience

AI/ML Engineer,
Microsoft - Redmond, WA | July 2025 – Present

- **Architected and deployed** Agentic RAG pipelines for Microsoft 365 Business Copilots using **Azure OpenAI (GPT-4o)**, enabling reasoning over **10TB+** of enterprise data while maintaining **95% retrieval accuracy.**
- **Engineered autonomous AI agents** for API orchestration using **Semantic Kernel** and Azure Functions, automating workflows to reduce manual user steps by **40%**.
- **Optimized production inference** by implementing high-performance vector indexing on **Azure AI Search** and Pinecone for **50M+ documents**, achieving **sub-200ms query latency** at global scale.
- **Reduced hallucination rates by 30%** by developing hybrid search strategies combining dense embeddings (Ada-002) and sparse keyword retrieval.
- **Established CI/CD testing pipelines** for Generative AI applications using **Prompt Flow**, increasing test coverage to **90%** and identifying critical edge cases weeks prior to deployment.
- **Implemented AI Safety guardrails** using Azure AI Content Safety, successfully blocking **99%** of adversarial prompt attacks in alignment with Responsible AI principles.

Data Scientist,
Piper Sandler, USA | Sep 2024 – July 2025

- **Developed an Agentic NLP workflow** using **LangGraph** on Azure ML Studio to summarize unstructured financial reports, reducing analyst time-to-insight by **50%**.
- **Modeled complex entity relationships** using **Graph Theory (Neo4j)** and **GraphRAG**, statistically reducing hallucination rates in automated summaries compared to baseline vector-only approaches.
- **Designed Quantitative Trading Strategies** using **Reinforcement Learning** agents, conducting rigorous backtesting on Azure Compute to demonstrate a **15% increase** in risk-adjusted returns (Sharpe Ratio).
- **Engineered dynamic context retrieval** strategies from Azure SQL Database to feed LLM contexts, ensuring high data fidelity for downstream investment reporting.

Graduate Research Assistant,
Concordia University St. Paul | Mar 2024 – Aug 2024

- **Built a Student Retention System** from scratch: engineered SQL pipelines in **Amazon RDS**, trained an **XGBoost** classifier (**88% accuracy**), and served predictions via a **FastAPI** endpoint on **AWS EC2**.
- **Deployed a scalable NLP pipeline** using **Docker** containers on **AWS Lambda**, applying Hugging Face Transformers for real-time sentiment analysis of social media data.
- **Co-authored a $50,000 AI Ethics grant**, utilizing **Bayesian Inference** and R to simulate algorithmic bias, defining the roadmap for fair healthcare AI deployment.

Data Scientist & ML Engineer
Infosys, India | Jun 2022 – Aug 2023

- Served as an ML Engineer and Data Scientist on engagements with primary healthcare and financial services clients, delivering advanced AI/ML solutions.

Client Project: CVS Aetna (Health Care)

- **Architected Big Data pipelines** using **Snowflake** and **PySpark** on **Azure Databricks**, processing TB-scale claims data for **5M+ patients**.
- **Containerized** a BERT + XGBoost ensemble model using **Docker** and deployed on **Azure Kubernetes Service (AKS)**, reducing deployment cycles from weeks to hours via CI/CD.
- **Ensured HIPAA compliance** by engineering PII masking and network isolation within Azure Virtual Networks (VNet).

Client Project: Charles Schwab (Financial Services)

- Designed a centralized Feature Store on Databricks (Delta Lake) to ensure consistency between training and production data for financial indicators.
- **Optimized inference latency to <100ms** per transaction using **ONNX runtime** and efficient vectorization, ensuring zero lag for real-time payment fraud detection.
- **Developed a Hybrid Fraud System** combining **PyTorch (LSTM)** for temporal patterns and Isolation Forest, reducing false positives by **20%**.

Data Analyst/Data Scientist
Twilight Software Solutions, Hyderabad, India | April 2020 – May 2022

- Developed a Customer Churn Model using AWS SageMaker (XGBoost) that predicted attrition with 85% accuracy, saving $1.5M in annual revenue.
- **Transformed operational reporting** by applying **Time-Series Forecasting (ARIMA/Prophet)** to predict network congestion, enabling proactive bandwidth allocation 2 hours in advance.
- **Conducted A/B Testing** and ANOVA analysis to validate network optimization strategies, presenting results to leadership via **Tableau**.
- **Automated ETL processes** using Python and AWS S3, saving 15 hours of manual work per week.

Education

- Master's in Data Analytics | Concordia University, St. Paul, MN | **GPA: 3.91/4.00**
- Bachelor's in Science | Nizam College, Hyderabad, India **| GPA: 8.22/10.00**

Certifications

- Machine Learning (Stanford)
- Data Science & Cloud Computing (Infosys)
- Data Analytics (Cisco)