

Ramasaireddy Chintham

Ashburn, Virginia | 813-825-7918 | cramasaireddy@gmail.com

SUMMARY

Software Engineer with 5+ years of experience building production-grade machine learning and Generative AI systems using Python. Specialized in designing LLM-driven applications, Retrieval-Augmented Generation (RAG) pipelines, and agentic workflows using LangChain and LangGraph. Experienced in developing scalable AI systems with vector databases, embeddings, and MLOps pipelines on AWS and Azure. Strong focus on explainable AI, hybrid ML + rule-based systems, and real-time AI deployment.

SKILLS

Programming: Python, SQL, PySpark, JavaScript, Scala

Generative AI: LLMs, RAG Pipelines, Prompt Engineering, Hugging Face, LangChain, GitHub Copilot, Cursor, LangGraph

Agentic AI: Multi-Agent Systems, Tool Calling, Workflow Orchestration

Vector Databases: FAISS, Chroma, Embeddings (Sentence Transformers, OpenAI)

Machine Learning: Scikit-learn, XGBoost, OpenCV, Transformers (BERT), AutoML, K-Means, Random Forest, Anomaly Detection

Deep Learning & NLP: PyTorch, TensorFlow, NLTK, CNNs, RNNs, Sentiment Analysis, Text Classification.

MLOps & CI/CD: MLOps, CI/CD pipelines, Azure DevOps, Jenkins, Git, Docker, Kubeflow, ECS, Airflow, Model Monitoring

Cloud Platforms: Azure (ADF, Key Vault, Synapse, Monitor), AWS (Functions, Lambda, S3, EC2, EMR, SageMaker, Bedrock), GCP

Certifications: AWS Certified Machine Learning – Specialty, IBM Generative AI Engineering with LLMs

PROFESSIONAL EXPERIENCE

Argano, Tampa, FL | Software Developer - AI

July 2024 – Present

- Designed and implemented **LLM-driven applications using Retrieval-Augmented Generation (RAG)** with LangChain, enabling intelligent querying over enterprise knowledge bases with high contextual accuracy.
- Built **agentic AI workflows using LangGraph**, orchestrating multi-step reasoning pipelines for query routing, context enrichment, and response generation and integrated **TensorFlow**-based classification models
- Developed **embedding pipelines and vector search systems (FAISS/Chroma)** to support semantic retrieval and context grounding for LLM responses.
- Deployed **Generative AI pipelines using AWS Bedrock and SageMaker**, combining foundation models, embedding-based retrieval, and scalable model hosting for real-time inference and enterprise applications.
- Implemented OCR pipelines using **OpenCV** and **PyTorch** via **AWS Lambda**, enabling real-time document classification and reducing human verification costs by 60% across operational workflows

American Express, Bangalore, India | Software Engineer

January 2019 - August 2022

- Built a **fraud risk scoring system** using **machine learning**, improving detection accuracy by 27% and enabling real-time decisioning via Flask APIs and AWS Lambda.
- Created scalable feature engineering pipelines with **AWS Glue** and **PySpark** to process 5TB daily, transforming transactional data into training-ready formats and supporting downstream machine learning pipelines with clean, structured outputs.
- Applied machine learning techniques including **classification models (Logistic Regression, Random Forest)**, and **anomaly detection** to build a fraud risk scoring system, improving detection accuracy and enabling real-time decisioning.
- Built scalable **ML pipelines on Azure ML + Databricks**, processing 3TB+ data and enabling automated retraining using MLflow.
- Established an **MLOps** lifecycle within **Azure ML**, automating the retraining and versioning of the Python model to ensure its predictive accuracy and relevance.

EDUCATION

Master of Science in Artificial Intelligence and Business Analytics

December 2023

University of South Florida, Tampa, FL

Bachelor of Technology in Computer Science

May 2019

Sree Vidyanikethan Engineering College, India

PROJECTS

AI Clinical Note Validator

- Built an end-to-end **RAG** pipeline using sentence-transformer embeddings, NLP Parsing and vector search (Chroma) to retrieve domain-specific clinical guidelines and integrated **LLM (FLAN-T5)** for contextual reasoning
- Implemented a custom **NLP** section parser to extract structured components such as diagnosis, medications, and follow-ups from unstructured notes and **rule-based validation + LLM reasoning** for explainable output generation

Agentic Transaction Risk Investigator

- Built an AI-powered transaction risk investigation platform to detect anomalous financial transactions, apply weighted risk rules, generate grounded case summaries, and support analyst review workflows with dashboard analytics.
- Designed agentic workflows using **LangGraph**, modeling investigation steps such as context loading, evidence evaluation, severity assignment, and report generation and implemented **hybrid risk scoring (ML + rules + LLM)**