

# Adithya Karumuri

Fort Collins, CO | +1 (970) 689 0873 | [adithyakarumuri@gmail.com](mailto:adithyakarumuri@gmail.com) |

## SUMMARY

Results driven AI/ML Engineer with 5 years of experience architecting and deploying production-scale machine learning and generative AI solutions across finance, healthcare, and enterprise domains. Proven success in building LLM-powered applications, RAG pipelines, fraud detection, and personalization systems that delivered double-digit efficiency gains, multi-million-dollar cost savings, and improved customer experiences. Skilled in Python, PySpark, SQL, and cloud platforms (AWS, Azure, GCP) with deep expertise in MLOps, CI/CD, Docker/Kubernetes, and model observability with MLflow. Recognized for driving responsible AI adoption with a strong foundation in NLP, GenAI (GPT, LLaMA, LangChain), A/B testing, and time-series forecasting, consistently delivering compliant, scalable, and business-aligned AI solutions.

## SKILLS

**Programming & Tools:** Python, R, SQL, MATLAB, Shell Scripting, Pandas, NumPy, Scikit-Learn, TensorFlow, PyTorch, Keras, Flask, REST APIs, Postman, JSON, Parquet

**Machine Learning & AI:** Supervised/Unsupervised Learning, Regression & Classification, Clustering, Ensemble Models, XGBoost, Transformers, CNN, RNN, LSTM, Transfer Learning, Time-Series Forecasting, A/B Testing

**Generative AI & LLMs:** GPT, LLaMA, Claude, Hugging Face, LangChain, Bedrock, RAG, Vector DBs (Pinecone, FAISS, Weaviate, ChromaDB), Prompt Engineering, Tokenization (BPE, WordPiece), RLHF

**Agentic AI:** CrewAI, LangGraph, AutoGen, multi-agent workflows, memory management, tool orchestration, task delegation

**MLOps & Cloud:** AWS (SageMaker, Glue, Bedrock, Feature Store, Redshift), Azure ML, GCP AI, MLflow, Weights & Biases (W&B), CI/CD (Jenkins, CodePipeline, GitHub Actions), Docker, Kubernetes, Cloud Architecture, Model Monitoring, Responsible AI (Bias/Fairness, Explainability, Governance)

**Big Data & ETL:** PySpark, Apache Spark, Hadoop, Hive, Kafka, HBase, AWS Glue, Spark Streaming, Data Lakes & Warehousing

**Databases:** MySQL, PostgreSQL, SQL Server, MongoDB, Cassandra, Teradata, Amazon RDS, Redshift

**Visualization & Analytics:** Tableau, Power BI, AWS QuickSight, Plotly, Seaborn, Matplotlib, Excel (Pivot, VLOOKUP)

## EXPERIENCE

**Morgan Stanley, USA | AI/ML Engineer**

**Nov 2024 - Present**

- Architected LLM-powered agentic systems with LangChain, LangGraph, and CrewAI, enabling inter-agent memory and communication, and improving workflow task completion 30%
- Fine-tuned LLMs (Hugging Face, LangChain) for clinical/legal summarization, improving accuracy 27% and reducing analyst review time by 15 hrs/week.
- Optimized AWS compute spend by analyzing usage patterns and implementing SageMaker instance lifecycle tuning and right-sizing strategies, reducing costs by 18% (~\$200K annually) while maintaining model performance.
- Deployed GenAI solutions on AWS Bedrock using Claude and GPT-4 for document summarization, client communication analysis, and regulatory filings, cutting processing time by 45%.
- Built RAG pipelines with Pinecone & FAISS, boosting retrieval accuracy 35% and saving ~\$400K/year in manual lookup costs.

**Intex Technologies, India | AI/ML Engineer**

**Dec 2020 - Dec 2023**

- Conducted A/B testing of LLM prompts using LangChain and evaluation metrics (BLEU, ROUGE), optimizing chatbot responses and improving resolution accuracy by 18%
- Developed XGBoost-based failure prediction models, reducing system downtime 15% and preventing ~\$1.2M in annual lost revenue.
- Designed deep learning models (CNN, RNN, LSTM, Transformers) for anomaly detection, achieving greater than 90% accuracy and reducing MTTR by 25%.
- Developed scalable ML pipelines with PySpark, AWS Glue, and Redshift for real-time analytics, boosting business insight delivery by 60%.
- Mentored engineers on LoRA fine-tuning, PEFT, and Responsible AI practices, ensuring GDPR/HIPAA-compliant deployments.

**Groovy Web, India | ML Engineer**

**Jan 2019 - Nov 2020**

- Automated ETL pipelines using Python, SQL, and SAS to streamline feature engineering and model ingestion, boosting training throughput by 60% and cutting data preparation time from days to hours, enabling faster model iteration and deployment.
- Developed time-series forecasting models (DeepAR, XGBoost) to predict financial KPIs and transaction volumes, improving forecast accuracy 22%.
- Built and deployed classification models (Random Forest, SVM, XGBoost) with F1-scores >90%, driving fraud detection improvements.
- Established MLOps pipelines with Docker, Kubernetes, and MLflow for real-time deployment and monitoring of financial models, ensuring 99.9% uptime and reducing model release cycles by 35%.
- Built interactive dashboards and advanced visualizations (Matplotlib, Seaborn, Plotly, Power BI) to perform EDA on financial and customer data, uncovering risk patterns and behavioral trends that informed executive decision-making and improved reporting efficiency by 35%.

## PROJECTS

**Enterprise Multi-Agent RAG Copilot for Financial Workflows**

- Designed an agentic AI system with CrewAI, LangGraph, LangChain, and RAG pipelines (Pinecone, GPT-4) to coordinate LLM-powered agents for finance and IT workflows. Enabled context-aware decision-making via RAG from regulatory docs, compliance databases, and IT knowledge bases. Automated regulatory summarization, KYC/AML checks, and IT ticket resolution, boosting workflow completion by 30%, cutting manual review by 40%, and saving ~\$1.5M annually while maintaining SOX/AML compliance.

**Real-Time Fraud Detection & Risk Scoring Pipeline**

- Designed and deployed a real-time fraud detection platform using XGBoost, PySpark, and AWS SageMaker, capable of delivering sub-second predictions on millions of daily transactions. The solution improved detection accuracy by 42%, reduced false positives by 30%, and helped prevent multi-million-dollar annual losses, while ensuring compliance with SOC2 and HIPAA standards.

## EDUCATION

**Master's in Computer Information Systems** | Colorado State University, Fort Collins, CO

**May 2025**

**Bachelor of Technology in Mechanical Engineering** | M.V.G.R College of Engineering, Vizianagaram, India

**Jun 2020**

## CERTIFICATIONS

**Mastering Generative AI with OpenAI, LangChain, and LlamaIndex V2 (Ineuron.ai)**

**Microsoft Azure Machine Learning (Microsoft)**

**AWS Certified Solution Architect (AWS)**