

Amitha Akepati

(469) 318-8841 | akepati.amitha@gmail.com | [GitHub](#) | [Portfolio](#)

SUMMARY

Forward-Deployed AI Engineer / Data Scientist with 4+ years of experience building production-grade AI/ML pipelines, ETL workflows, and data analytics solutions. Experienced in LLMs, Generative AI, Agentic AI, and multi-agent systems. Skilled at translating business requirements into actionable insights, automating workflows, and delivering scalable AI solutions that reduce manual effort and improve operational efficiency.

EDUCATION

UNIVERSITY OF TEXAS AT ARLINGTON

Arlington, TX

MASTER OF SCIENCE IN DATA SCIENCE

Jan 2023 - May 2024

Concepts: Machine Learning, Neural Networks, Deep Learning, Probability, Statistics, Data Analytics, Visualization, Artificial Intelligence

ANNAMACHARYA UNIVERSITY

Andhra Pradesh, IN

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING

Jun 2017 - Jul 2021

Concepts: Python, Database Management, Big Data, Cloud Computing, Machine Learning, Web development, Data Science, Data Structures and Algorithms, Thesis: Object Detection with Deep Learning, Computer Vision and dual stage Convolutional Neural Network

WORK EXPERIENCE

Data Scientist, SMD Technosol - Thermo Fisher Scientific, Dallas, TX

Dec 2024 - Present

- Built and maintained production data pipelines integrating multiple structured and semi-structured data sources, preparing curated datasets for analytics and ML workflows.
- Orchestrated end-to-end pipelines using Apache Airflow with monitoring, alerting, and dependency management to ensure reliable scheduling and data availability across teams.
- Developed and deployed production ML forecasting models (time-series and regression) for capacity planning, integrating model outputs into dashboards for data-driven operational decisions.
- Designed AI-driven LLM workflows for automated report generation, summarization, and dashboard, reducing manual reporting time by 40% and saving an estimated \$50K+ annually.
- Implemented LLM-powered assistants and autonomous agents using LangChain and Azure OpenAI for operational workflows, improving accessibility to analytics insights.
- Collaborated with product, data, and business teams to implement anomaly detection and synthetic data pipelines, improving data quality, and reducing operational errors by 20%.

Data Analyst, Bright Mind Enrichment & Schooling, Irving, TX

Sep 2024 - Dec 2024

- Analyzed and cleaned 10,000+ sponsorship and beneficiary records supporting education, food programs, and services for underprivileged and homeless communities using Python, SQL to improve reporting accuracy.
- Designed and maintained Power BI dashboards to track funding sources, allocation patterns, and program reach, while standardizing data definitions and KPIs across initiatives to ensure accuracy and consistency in reporting.

Research Assistant, University of Texas at Arlington, Arlington, TX

Oct 2023 - May 2024

Advisor: Dr. Aera Kim LeBoulluec

- Developed and optimized CNN, LSTM, and GAN models for crop yield prediction, improving model accuracy and supporting data-driven agricultural insights.
- Built multimodal ML pipelines combining sensor, soil, and weather data for robust crop performance modeling.
- Studied transformer and LLM architectures, fundamentals, gaining expertise in attention mechanisms, tokenization, and embeddings.
- Performed feature engineering, statistical analysis, and model optimization to improve model stability and generalization.

Data Research Intern, Tomorrow's Leaders Today, Frisco, TX

Feb 2024 - Apr 2024

- Prepared and analyzed 10,000+ grant records, using machine learning for prioritization and targeting improvements.
- Built NLP and LLM-based summarization pipelines, reducing manual review effort by 45% and increasing throughput of reports.
- Developed Power BI dashboards visualizing trends, insights, and recommendations for stakeholder teams.
- Implemented data validation and quality checks across analytics to ensure consistency, reliability across workflows.

Software Engineer, DXC Technology, Bangalore, IN

Jun 2021 - Aug 2022

- Worked on HP Service Bus (HPSB) to manage and monitor secure file transfers across multiple systems, reducing operational incidents and improving delivery reliability.
- Migrated legacy, file-based workflows into structured ETL pipelines using Python and SQL, improving processing efficiency by 25%.
- Automated end-to-end ETL workflows using Apache Airflow, implementing alerting to detect and resolve pipeline issues, reducing manual intervention by 30 to 40%.
- Built Power BI dashboards for operational and business teams to track data availability, pipeline status, SLA compliance and key metrics.

KEY SKILLS

Programming: Python, SQL, PySpark

Machine Learning & AI: Supervised/Unsupervised Learning, Deep Learning (CNN, RNN, LSTM, Transformers), NLP, Computer Vision, Generative AI, Agentic AI, Multi-Agent Systems, RAG, Prompt Engineering, LLM Fine-Tuning (LoRA, QLoRA, PEFT)

Frameworks & Libraries: PyTorch, TensorFlow, Keras, Scikit-Learn, XGBoost, LangChain, LangGraph, LlamaIndex, Pandas, NumPy, Matplotlib, Seaborn, OpenCV, FastAPI

Data Engineering, & Cloud: Data Modeling, ETL/ELT Pipelines, Databricks, Apache Airflow, Docker, Kubernetes, CI/CD, Spark, AWS, Azure, GCP, Vector Databases (FAISS, Chroma, Pinecone), Delta Lake

Visualization & Tools: Power BI, Tableau, Excel, Jupyter, Git/GitHub, VS Code, Jira, Agile/Scrum

APPLIED AI & ML SYSTEMS

AI Log & Error Explainer Agent

Streamlit application using LLMs to analyze system logs, identify root causes, and provide corrective guidance.

Patient Readmission Predictor with AI Appeal Letter Generator

End-to-end ML pipeline predicting patient readmission risk; integrates LLMs to auto-generate insurance appeal letters in PDF format.

Intelligent Business Analytics Agent (Multi-Agent System)

Built multi-agent analytics system automating data ingestion, statistical analysis, insights generation, and report creation.

CERTIFICATION

- Microsoft Azure Fundamentals - AZ-900
- Complete Guide to Databricks for Data Engineering
- PowerBI by PwC
- Python for Data Science by IBM
- Hackerrank Problem Solving Certificate
- Hackerrank Python Certificate
- Sales Data Exploration Certificate - HiCounselor