

Durga Vikas Thota

AI/ML Engineer

vikasthota09@gmail.com | [Linkedin](#) | (773) 848 3509 | Detroit, MI

AI/ML Engineer with 4+ years of experience in designing, developing, and deploying scalable machine learning and NLP solutions in finance and enterprise environments. Proven expertise in transformer-based models (BERT, T5, GPT), MLOps, and real-time AI systems, with a strong focus on optimizing performance, reducing latency, and improving accuracy. Adept at end-to-end ML pipeline development, from data pre-processing and model training to deployment and monitoring using Docker, Airflow, CI/CD, and cloud platforms.

PROFESSIONAL EXPERIENCE

AI-ML Engineer | Citigroup

Jan 2024 – Present

- Architected and deployed large-scale NLP systems for document intelligence, processing over 2 million financial documents monthly, yielding \$1.2M annual operational cost reduction through 45% workload automation.
- Led implementation of deep learning-powered, transformer-based (BERT, T5) conversational AI solutions for customer service automation, resulting in 27% faster resolution and 18% higher CSAT scores.
- Pioneered and deployed a real-time fraud detection system leveraging Graph Neural Networks (GNNs) and XGBoost, preventing \$4.8M in fraudulent transactions annually.
- Developed a model performance monitoring dashboard with Prometheus and Grafana, adopted by 12 engineering teams, reducing incident response time by 65%.
- Collaborated on model explainability (XAI) requirements, integrating interpretable AI techniques (e.g., SHAP, LIME) into production models for enhanced transparency and regulatory compliance.
- Optimized GPU utilization and distributed training strategies for deep learning models, reducing training time by 35% and improving resource efficiency.
- Facilitated to a secure data governance framework for financial data, ensuring compliance with industry regulations (e.g., GDPR, CCPA) within ML pipelines.

AI-ML Engineer | Sage Softtech

Apr 2021 – Jul 2022

- Spearheaded and refined a computer vision solution for autonomous vehicle diagnostics, reducing client false positives in defect detection by 38%.
- Deployed a federated learning framework for manufacturing clients, improving model accuracy by 22% while ensuring data privacy across distributed operations.
- Optimized Natural Language Processing (NLP) models for in-vehicle voice assistant platforms, decreasing query latency from 2.1s to 1.4s, enhancing user experience.
- Authored and enacted company-wide MLOps standards and best practices, reducing model drift incidents by 45% for internal projects and client engagements.
- Directed the Proof-of-Concept (POC) for a predictive maintenance system, demonstrating a 27% reduction in unplanned downtime for industrial clients.
- Implemented model versioning and data lineage tracking for all solutions, enhancing reproducibility and auditability of delivered ML assets for client transparency.

AI-ML Engineer | Aspire Technolab

Aug 2018 – Feb 2021

- Developed and deployed a multilingual conversational AI chatbot framework supporting 8 languages, adopted by 14 clients and handling over 50,000 monthly queries.
- Innovated and applied data augmentation techniques that improved small dataset model performance by 20-25%.
- Containerized Machine Learning (ML) workflows using Docker, reducing development environment setup time from 8 hours to 35 minutes.
- Engineered an intent classification system with 92% accuracy across 142 business domains.

EDUCATION

MS in **Computer Science**, Chicago State University

BE in **Electrical and Electronics Engineering**, Andhra University

Aug 2022 – Dec 2023

May 2014 – May 2018

TECHNICAL SKILLS

- **Programming & Scripting:** Python, SQL, UNIX Shell Scripting, Bash, PySpark, Spark SQL, Hive
- **Data Engineering:** ETL Pipelines, Data Lake Architecture, Data Quality Automation, Apache Spark, Apache Kafka, Apache Hadoop, Apache Airflow, AWS Glue, AWS DMS, AWS Kinesis
- **Cloud Platforms & Data Warehousing:** AWS (S3, Lambda, Redshift, Glue, Lake Formation, CodePipeline), IBM Cloud, Snowflake
- **Databases:** MySQL, PostgreSQL, Oracle SQL, SQL Server, Amazon Redshift
- **Visualization & BI Tools:** Tableau, Power BI, Amazon QuickSight, Looker
- **Workflow Orchestration & CI/CD:** Apache Airflow, AWS Step Functions, AWS CodePipeline, AWS CodeDeploy, Jenkins
- **Data Governance & Metadata Management:** AWS Glue Data Catalog, Apache Atlas, Data Lineage Management
- **Real-Time Data Streaming:** Apache Kafka, AWS Kinesis Data Streams, Spark Structured Streaming
- **Version Control & Collaboration:** Git, GitHub, Bitbucket, JIRA, Confluence
- **DevOps & Automation Tools:** Docker, Jenkins, AWS Systems Manager