Vikas Reddy

Data scientist and NLP engineer with 2+ years building LLM-powered solutions—experienced in prompt engineering, model evaluation, and end-to-end MLOps.

(240)-326-3889 | vvikasreddy675@gmail.com | linkedin.com/in/vvikasreddy | github.com/vvikasreddy | Portfolio

Education

University of Maryland Baltimore County (UMBC) — Baltimore, MD
Master's in Data Science; GPA: 4.0 — *Aug 2023 – May 2025*

Institute of Aeronautical Engineering (IARE) — Hyderabad, India
Bachelor's of Technology in Computer Science and Engineering; GPA: 3.41 — *May 2019 – May 2023*

Research Publications

Deep fusion of neurophysiological and facial features for enhanced emotion detection. (n.d.). IEEE Journals & Magazine — IEEE Access. **https://ieeexplore.ieee.org/abstract/document/10945364**

NLP Projects

Terminology-Enforced Machine Translation | *Python, PyTorch, MarianMT* — GitHub

- Developed a lexically-constrained machine translation model using beam search with bigram constraints; validated model outputs using BLEU and ROUGE metrics on the WMT Turkish-English corpus—demonstrated capability in constraint-aware sequence modeling.
- Applied efficient decoding and optimization techniques to control output variance, simulating use-cases in document standardization and regulatory compliance.

LLM Fine-Tuning for Legal Text Summarization | *Python, PyTorch, Hugging Face* — GitHub

- Fine-tuned FLAN-T5-small on legal contracts and regulatory documents; implemented full ML pipeline (tokenization, validation, model checkpointing) using HuggingFace Transformers and AdamW optimization.
- Achieved high summary fidelity using ROUGE-1/2/L and BERTScore; project demonstrates scalable fine-tuning workflows and domain adaptation applicable to financial document summarization.

Experience

Brain-Machine Interface Lab at UMBC — Baltimore, USA
ML Researcher & AI Solutions Engineer — *Jan. 2024 – Present*

- Developed a production-grade PyTorch + Spark pipeline for multi-modal EEG + facial data, achieving 97% classification accuracy and managing full model lifecycle in Conda.
- Implemented AWS data ingestion using Kafka and Spark Streaming for real-time inference, sustaining 80%+ live accuracy under load.
- Prototyped LLM-assisted annotation scripts to automate EEG event labeling, reducing manual effort by 20%.
- Built Python evaluation frameworks to track model performance and usage metrics, informing iterative prompt and dataset optimization.
- Utilized Named Entity Recognition (NER) and cosine similarity to automatically cluster and group experiment descriptions, streamlining the discovery of related protocols and accelerating analytical workflows.

Epam Systems — Hyderabad, India
Data Science & Software Engineer Intern — *Jan. 2023 – Jun. 2023*

- Engineered Python /Java test automation frameworks with SOLID principles, cutting deployment failures by 30%
- Built Spark + Kubernetes pipelines to process 50 GB+ of logs, delivering near real-time dashboards for operational monitoring.
- Integrated AWS (Lambda, S3, Athena) into Jenkins CI/CD workflows to automate test artifact deployment and reporting.

IndicWiki Project — Hyderabad, India
Data Analyst Intern — *Mar. 2022 – Jun. 2022*

- Designed Spark ETL pipelines to curate 10K+ multilingual Wikipedia articles into PostgreSQL, improving query latency by 30%.
- Optimized tokenization and context-window strategies to support downstream NER and summarization tasks.
- Developed Python scripts to evaluate and visualize summarization quality, boosting relevance metrics by 20%.

Technical Skills

Languages: Python, SQL (PostgreSQL), R
Frameworks & Libraries: Conda, PySpark, H2O.ai, scikit-learn, TensorFlow, PyTorch, Hugging Face Transformers
Cloud & Infrastructure: AWS (S3, EC2, Lambda, Athena), Kafka, Docker, Kubernetes, CI/CD Pipelines
Data Engineering: Spark, ETL, Data Warehousing, Stream Processing (Kafka + Spark Streaming)
Modeling Techniques: Classification, Clustering, Time Series, Deep Learning, Statistical Modeling, Model Evaluation (ROC, Confusion Matrix)
Dev Tools & Workflow: Git, Jenkins, Agile (Scrum/Kanban), Reproducible Environments (Conda)
Agentic tools : LangChain, LangGraph, LangSmith, Vector Databases, RAG, Knowledge Graphs