SATHVIK ADDICHARLA

Denton, TX 76207 | sathvikaddicharla@my.unt.edu | +1 (940) 597-5745 | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Experienced ML Engineer with 3+ years of hands-on experience in building, fine-tuning, and benchmarking LLMs and deep learning systems. Proven track record in implementing reproducible ML pipelines, optimizing model performance, and delivering production-ready solutions. Strong expertise in Python engineering, Docker-based development, and detailed technical planning for complex ML tasks. Published IEEE researcher with demonstrated ability to translate natural language requirements into executable code implementations.

TECHNICAL SKILLS

- **Machine Learning & Benchmarking:** Model evaluation, reproducibility, Kaggle-style competition workflows
- **LLMs & NLP:** Transformers, HuggingFace, LangChain, GPT-style models, RAG pipelines
- **Model Optimization:** Quantization, pruning, ONNX, TorchScript
- **Data Engineering:** Large-scale preprocessing (500GB+), FAISS vector search, feature extraction
- **Infrastructure & Tools:** Docker, AWS, GitHub, Flask/FastAPI
- **Frameworks & Libraries:** PyTorch, TensorFlow, NumPy, SciPy, Scikit-learn
- **Other:** Detail-oriented technical documentation, compliance-driven development

EDUCATION

Master's in Data Science | University of North Texas | 04/2024 – 08/2026 | GPA: 3.83/4.0
Bachelor's in Computer Science – Data Science | VNR Vignana Jyothi Institute of Technology, India | GPA: 8.60/10

EXPERIENCE

AI/ML Engineer Intern | Fintech Solutions LLC | Irving, Texas | 03/2025 – 07/2025

- Designed and executed end-to-end ML workflows, drafting executable plans and implementing them in Python for customer prediction, fraud detection, and RAG-powered search.
- Benchmarked 3 PyTorch models, improving prediction accuracy by **85%** and cutting inference time by **40%**.
- Built scalable preprocessing pipelines for **500GB+** data, reducing processing time by **60%** through automation.
- Integrated models into **AWS + Docker** environments for reproducible deployments.
- Developed a RAG system with LangChain, boosting query accuracy by **75%** and halving response time.
- Created real-time fraud detection system with **95%** detection accuracy within 100ms, preventing ~$500K/month in losses.

PROJECTS

Retrieval-Augmented Generation (RAG) Based Large Language Model Chatbot

- Developed an **AI-powered chatbot** leveraging **LLMs and RAG** for improved diagnosis of **physical and mental health** conditions.
- Implemented **NLP-based question-answering models** with real-time **retrieval and summarization** using Langchain.
- Built LLM-powered applications using LangChain and LangGraph to enhance retrieval-based NLP models. Published in **IEEE (09/2024)**. This also includes Model Training & Evaluation, Statistical Analysis.

AI YouTube Summarizer (NLP, LLM, FAISS Vector Search)

- Developed an AI-powered YouTube summarizer using OpenAI's **Whisper** model to transcribe audio into text with high accuracy.
- Generated vector embeddings using a lightweight **transformer model** from Hugging Face for semantic understanding.
- Integrated a **FAISS vector database** to enable efficient query-based retrieval of relevant content, achieving ~92% relevance accuracy.
- Reduced manual summarization time by over **75%**, enhancing content accessibility and user experience.
- Tools & Tech: **Python, Whisper, Transformers, FAISS, NLP, Vector Search**

Speech Emotion Detection (LSTM)

Developed an LSTM-based deep learning model to detect emotions in speech. Preprocessed audio using MFCCs and noise reduction. Fine-tuned model to handle varied speech patterns and accents. Experimented with attention mechanisms to improve interpretability.

CERTIFICATIONS & TRAINING

- Deep Learning – NPTEL (IIT Ropar)
- Python for Data Science and Machine Learning Bootcamp – Udemy
- Training LLMs with Own Data – DeepLearning.ai
- Google Data Analytics Certification- Coursera

PUBLICATIONS

Retrieval-Augmented Generation Based Large Language Model Chatbot for Improving Diagnosis for Physical and Mental Health

• 09/2024 IEEE

Transfer Learning based Autism Detection in Children

•08/2024 IEEE