Ramachandra Rahul Taduri

Secaucus, NJ | 6148434715 | rahul.taduri@gmail.com

Summary

Machine Learning Engineer with a decade-long engineering background and hands-on experience in LLMs, model deployment, MLOps, and deep learning projects. Skilled in building and deploying ML solutions (BERT, GPT-2, RAG, Agentic AI) using Python, PyTorch, Hugging Face, FastAPI, and Docker. Passionate about applying AI to real-world problems in healthcare, finance, and automation.

Skills

ML/DL Frameworks & Libraries: TensorFlow, PyTorch, Keras, Hugging Face, Scikit-learn, OpenCV

ML & AI Techniques: Deep Learning, Computer Vision (Object Detection, Segmentation, OCR), NLP, Reinforcement Learning, LLM Fine-tuning, Prompt Engineering, RAG, Agentic AI (LangChain)

Model Architectures: FCN, SegNet, YOLOv5/v8, Residual U-Net, Faster-RCNN, Vision Transformers (ViT), GPT-4, LLaMA, DeepSeek (MoE, MLA, GRPO)

MLOps & Deployment: DVC, MLFlow, Git, CI/CD, Docker, Kubernetes, Airflow, HPO (Ray Tune),Evidently

Cloud & Big Data: AWS, Microsoft Fabric, GCP, Azure

Projects

Droplet Detection System – Embedded CV Application | OpenCV, Yolov5 · LiDAR + IR Sensors · TensorRT| Developed and deployed a real-time object detection and segmentation pipeline for windshield droplet classification in autonomous vehicles using YOLOv5 and OpenCV. Achieved 95% detection accuracy and reduced inference latency by 40% on Jetson Nano edge devices. Integrated REST API for remote inference improving ADAS reliability by 25%.

Custom Decoder-only LLM – LLM from scratch with instruction Fine-tuning | Transformers · GPT-2 | Constructed a decoder-only LLM with grouped-query attention and classification head. Integrated pre-trained weights (TensorFlow with Keras) GPT-2 with custom BPE tokenizer; fine-tuned on 500K+ instruction-style records. Self-hostel model on AWS with RestAPI.

Patient Readmission Prediction – ML Pipeline Deployment | Classical ML · MLflow · FastAPI · CI/CD | Engineered, trained, and deployed models to predict diabetic patient readmissions. Achieved 88% recall score by combining tree-based and neural models with hyperparameter optimization via Ray Tune. Built CI/CD pipeline with MLflow model registry and Evidently AI for drift detection.

Resume Coach – AI-Powered Coaching Assistant | LLaMA-3 (Self-hosted on AWS SageMaker) · LangChain · Streamlit · RAG · Prompt Engineering | Developed a resume optimization tool using self-hosted LLaMA-3 on AWS SageMaker, integrated with LangChain and RAG architecture for context-aware retrieval. Delivered tailored resume feedback aligned to job descriptions, improving alignment scores from 5/10 to 9/10. Enabled a chatbot interface via Streamlit and RESTful APIs using prompt engineering for personalized coaching.

Predictive Analytics for Project Funding | Classical ML & NLP

Tackled the challenge of highly imbalanced textual data to automate project funding predictions, integrating PCA for dimensionality reduction and leveraging BERT models for classification. Achieved a strong F1 score of 0.82, demonstrating effective model performance for critical decision support.

Publications

He, Z., Mongeau, L., Taduri, R., and Menicovich, D., "Feedforward Harmonic Suppression for Noise Control of Piezoelectrically Driven Synthetic Jet Actuators," SAE Int. J. Adv. & Curr. Prac. in Mobility 6(2):945-952, 2024, https://doi.org/10.4271/2023-01-1042.

Professional Experience

ML Engineer - Acoustics & Signal Processing | Actasys Inc, Brooklyn **Oct 2021 - Present**

- Led design and deployment of deep learning-based novel active noise cancellation technology, reducing cabin noise by 50%+.
- Developed real-time computer vision pipelines (object detection, segmentation) for autonomous vehicle sensor cleaning, achieving 95% detection accuracy.
- Built and served REST APIs for ML inference using FastAPI and Docker on AWS.
- Automated data pipelines and model monitoring with MLflow, DVC, and Evidently.

Engineer III - Product Development & Data Analysis| ZF USA, Livonia, MI **July 2015 - Oct 2021**

- Automated data analysis pipelines for acoustics and vibration datasets using Python and MATLAB.
- Designed experiments and validated simulation models to optimize damping system behavior..
- Applied DOE and statistical modeling to reduce part failure rates by 30%, saving $300K annually.

Assistant Manager | NTPC Ltd., India **July 2010 - July 2013**

- At India's largest power utility, improved processes increasing annual revenue by $0.7 million.

Education

Interview Kickstart March 2024 - Present

Machine Learning program covering mathematical foundations and practical applications in classical ML, deep learning, NLP, computer vision, LLMs/ Generative AI (fine-tuning, RLHF, RAG, Agentic AI)

McGill University Jan 2023 - June 2023, Visiting Scholar, Montreal, QC, Canada

Developed and tested 3D-printed acoustic enclosure prototypes with novel materials, and programmed code to collect data from various sensors for prototype evaluations.

The Ohio State University Aug. 2013 - Aug. 2015, Master of Science, Columbus, OH USA

Developed mathematical models to optimize vehicle exhaust system design. (Link)

National Institute of Technology Karnataka Aug. 2006 - May 2010, B. Tech, Surathkal, India

Designed and tested vertical axis wind turbines for power generation in areas with low wind speeds.

Certifications

- Erdös Institute : Data Science Bootcamp, Deep Learning Bootcamp
- Microsoft Fabric : Data Analyst (DP600) Certified
- IBM Data Warehouse Engineer (Coursera)