

MUPPIDI RAHUL REDDY

940-745-9717 ♦ Sunnyvale, CA, 94085

rahulreddy3620@gmail.com ♦ [linkedin.com/muppidi-rahul-reddy](https://www.linkedin.com/in/muppidi-rahul-reddy)

OBJECTIVE

AI Engineer with experience in developing and deploying machine learning and generative AI solutions, including LLMs, using Python, PyTorch/TensorFlow, and cloud platforms like AWS. Proven ability to improve model performance in production and integrate AI into workflows. Seeking to leverage expertise in a fast-paced insurtech startup environment.

EDUCATION

Master of Science in Computer Science, University of Texas at Dallas Aug '2022 – Dec '2023 Relevant Coursework: GPA 3.61. Courses: Statistical Methods for Data Science, Machine Learning, Computer Vision, Big Data Analytics, Design and Analysis of Algorithms, Web Programming Languages, Database Design.

Bachelor of Technology in Information Technology, Indian Institute of Information Technology Allahabad Jul '2018 - May '2022 Relevant Coursework: Courses: Data Structures and Algorithms, Operating Systems, Automata Theory, Database Management, Machine Learning, Computer Networks, Data Mining, Graph Theory, Object Oriented Programming, Wireless Network Security

SKILLS

Technical Skills Python, PyTorch, TensorFlow, AWS (S3, EC2, RDS, Lambda, Textract, OpenSearch Service, SageMaker, Bedrock)

EXPERIENCE

Software Development Engineer – RAG, NLP, LangChain Jan '2024 – Present
Applab Systems *Dallas, TX*

- Improved information retrieval accuracy by 70\
- Built a highly accurate (95\
- Optimized retrieval precision by 40\
- Integrated GPT-4 with Vision via API calls from AWS compute services to analyze and interpret engineering drawings, enhancing multimodal RAG capabilities and boosting usability by 60\
- Leveraged the GPT-4 API from AWS services for generating accurate, context-aware answers based on technical document chunks retrieved from Amazon OpenSearch Service.

Machine Learning Engineer – Pytorch, YOLOv5, Python Dec '2021 – Jul '2022
Coral Innovations Pvt. Ltd *Hyderabad, India*

- Engineered a Meter Board Reader application using YOLOv5, automating meter reading with 78\
- Improved dataset preprocessing, enhancing pipeline efficiency by 30\
- Implemented data augmentation techniques, increasing model robustness and generalization.
- Fine-tuned object detection models, achieving higher accuracy with optimized hyperparameters.

PROJECTS

Mini Unix Kernel Simulation – C, Inodes, File-handling, Page tables Built a simplified Unix-like kernel supporting process creation, scheduling, and system calls (e.g., fork(), exec(), wait()). Implemented a basic file system layer with support for block-level access, inodes, and file metadata management. Designed an interrupt handling mechanism and incorporated basic virtual memory management with page tables and swapping.

Scalable Search Engine Using Elasticsearch – Elasticsearch, Python, Kibana Indexed large datasets (e.g., Wikipedia dump) and exposed a REST API for fast, ranked queries. Tuned shard allocation, replication, and indexing strategies to improve query throughput and fault tolerance. Integrated Kibana dashboards to visualize usage stats and search heatmaps.

Predicting Safety Probabilities by Location and Day in Dallas – Pandas, Streamlit, sklearn, Llama-2 Created a Streamlit dashboard to analyze crime severity using Random Forest classification. Integrated Llama-2 NLP model for text-based severity analysis.

EXTRA-CURRICULAR ACTIVITIES

- Demonstrated problem-solving and technical skills through active participation in hackathons and data science competitions.
- Managed social media engagement and event coordination for a university dance crew, increasing audience interaction.