

Praneeth Posina

716-468-9996 | praneethposina@gmail.com | linkedin.com/in/praneethposina
github.com/praneethposina | praneethposina.github.io

EDUCATION

University at Buffalo, Buffalo, NY

Masters in Artificial Intelligence

Aug 2023 - Dec 2024

KL University, Hyderabad, India

Bachelor of Technology, Electronics and Communication Engineering

Jun 2019 - May 2023

Relevant Courses: Artificial Intelligence, Machine Learning, Math for Computing, Algorithms Analysis and Design, Deep Learning, Information Retrieval, Data Science & Big data, Data structures, Object Oriented Programming.

Certifications: Azure AI fundamentals, Google Cloud Skill Badges, AWS Machine Learning Foundations, Oracle Architect Associate.

SKILLS

Languages: Python, C, C++, Java, SQL.

Frameworks & Libraries: TensorFlow, Keras, Scikit-Learn, NLTK, OpenCV, PyTorch, Hadoop, MapReduce, Tableau, Power BI.

Other Skills: Deep Learning, MLOps, MLFlow, CUDA, CI/CD, Git, Version Control, CLI, Computer Vision, Large Language Models, Natural Language Processing, Computer Science, Data Analysis, Cloud Computing, Docker, Containerization, Kubernetes.

EXPERIENCE

ML Research Assistant

KL University

Dec 2021 - Jan 2023

- Led research on mediastinal lymph node malignancy detection, developing models that increased diagnostic accuracy to 98.2% using reinforcement learning techniques.
- Authored two IEEE-published papers [1] [2], introducing innovative methods that outperformed existing models by 3% and 4% respectively in accuracy.
- Engineered and validated custom deep learning algorithms, including a DQN policy and an ensemble learning strategy, achieving a 98.56% accuracy rate.
- Analyzed extensive medical image datasets, ensuring robust model performance and contributing to significant advancements in medical diagnostics.

PROJECTS

Customer Support Chatbot

LLaMA 3.1 8B, Docker, Flask API, AWS, Python

- Engineered and deployed an end-to-end customer support chatbot using the LLaMA 3.1 8B model with LoRA fine-tuning and 4bit, 8bit, and 16bit quantization for optimized inference via Ollama and Flask API.
- Leveraged Docker for containerization and AWS ECS with Fargate to deliver a scalable, highly available solution. Automated CI/CD pipelines with AWS CodePipeline for seamless deployment, while implementing comprehensive monitoring and logging through AWS CloudWatch to ensure performance and reliability.

LLM Powered Mobile Assistant

LLaMA 3.1 8B, Appium, Groq, Python

- Developed an LLM-powered mobile assistant using a fine-tuned LLaMA 3.1 8B model to achieve complex task execution with over 90% accuracy. Leveraged Appium for real-time app UI analysis and action automation, enabling dynamic interaction and error recovery, allowing the assistant to adapt and navigate unfamiliar apps seamlessly.

RAG Chatbot

OpenAI API, TF-IDF, Semantic Search, NLTK, Python

- Built a Retrieval-Augmented Generation (RAG) chatbot integrating web scraping, indexing, and query handling, leveraging Sentence Transformers for embedding 60,000+ Wikipedia documents, TF-IDF and Cosine Similarity for retrieval and re-ranking, and OpenAI GPT API for precise, context-aware responses.

Text Generative AI

Gen AI, LLMs, Python

- Developed a text-generative AI system using transformers to generate fiction stories, Trained over the vocabulary size of 42610 with 43 million parameters. Optimized the model architecture and hyperparameters to achieve a 4.02% testing loss in generating coherent and diverse storylines.

Multi-Agent Reinforcement Learning System

RL, PyTorch, Python

- Architected and implemented a complex multi-agent game environment, leveraging DQN and A2C algorithms to train competitive AI agents. Achieved 99% target rate and 80% improvement in agent efficiency over 2000 training episodes, significantly outperforming random baseline models.

Deepfake Detection System

OpenCV, Deep Learning, PyTorch, Python

- Engineered a state-of-the-art deepfake detection model using ResNeXt101 architecture and Bidirectional LSTM, achieving 91.30% validation accuracy and 86.96% test accuracy on a 477GB dataset. Outperformed baseline CNN by 17.95% and matched leading models like EfficientNet in accuracy while providing superior interpretability through heatmap visualizations.

ACHIEVEMENTS

- Secured 13th Rank in India's nationwide Machine Learning hackathon challenge 2021.
- Awarded the Prime Minister's scholarship for exceptional academic performance throughout undergrad studies in India.