GAURAV TADKAPALLY

Los Angeles, CA | (213) 913-7899 | gaurav.tadkapally@usc.edu | linkedin.com/in/gauravreddy08 | gauravreddy08.github.io/portfolio

EDUCATION

University of Southern California California, United States
Master of Science in Computer Science: 3.7/4.0 June 2023 - December 2024

- **Coursework**: Analysis of Algorithms, Applied Natural Language Processing, Machine Learning
- Served as a Teaching Assistant (TA) for the graduate course Applied Machine Learning for Natural Language Processing (ITP 459)

Vellore Institute of Technology Andhra Pradesh, India
Bachelor of Technology in Computer Science and Engineering: 8.94/10 May 2019 - May 2023

EXPERIENCE

Pitney Bowes Connecticut, United States
Data Science Intern June 2024-August 2024

- Designed agentic software testing with fine-tuned CodeLlama and GPT4o, leveraging JaCoCo code coverage tool to **agentically** generate test suites and increase code coverage by 15% **(Demo)**
- Optimized LLM inference by exclusive decoding strategies (**Speculative Decoding**), accelerating speed by 3x, and **Abstract Syntax Tree based agentic retrieval** for relevant codebase context and indexing
- Implemented **Direct Preference Optimization (DPO)** and **4-bit QLoRA** quantization, improving model's generative performance

MUKHAM Andhra Pradesh, India
Machine Learning Engineer Intern October 2022-May 2023

- Optimized facial recognition model for edge deployment (mobile application), leveraging **knowledge distillation**, **Post-training Quantization (8-bit quantization) and Automatic Mixed Precision**, decreasing model size by 75%
- Designed a Presentation Attack Detection system (facial spoof detection) utilizing the Lucas Kanade algorithm for motion analysis, achieving a 80% success rate in identifying spoofed faces

MUKHAM Pvt Ltd Andhra Pradesh, India
Research Assistant October 2022 - May 2023

- Developed a UAV-based wildfire detection algorithm utilizing the EfficientNetB0 architecture, incorporating **Neural Architecture Search (NAS)** for model optimization, resulting in a 98% precision rate
- Engineered smart glasses with a Continual Object Detection model (Incremental Learning) for visually impaired, leading 78% navigational accuracy

SKILLS AND CERTIFICATIONS

Languages: Python, TypeScript, JavaScript
ML Stack: PyTorch, Tensorflow, HuggingFace, LangChain, Keras, OpenCV, Scikit-learn, Pandas, NumPy, Matplotlib
Tools & Technologies: AWS (Cloud Practitioner), Azure (AI Fundamentals), MySQL, MongoDB, Selenium

ACADEMIC PROJECTS

AK15: Agentic Kubernetes Middleware (Github)

- Devised an LLM-based middleware that automates Kubernetes cluster read queries, achieving a 93% reduction in contextual token usage through agentic function calling and context retrieval
- Implemented 15 specialized API functions enabling the LLM to perform human-like, context-aware interactions with Kubernetes, optimizing and reducing API costs by leveraging targeted data retrieval strategies

GlancyAI: Consumer Product Research Assistant (Github)

- Developed an AI agent using GPT-4 and Agentic Retrieval Augmented Generation (RAG), with a vector database for optimized query retrieval, automating the extraction of data from YouTube
- Integrated summarization module condenses extensive online information into concise insights, streamlining the product recommendation process and significantly reducing user research time

Original Vision Transformer Implementation from Scratch (Github)

- Implemented ViT components including MultiheadAttention, Image Patch Embedding, and MLP layers, achieving a one-to-one parameter match (86 million) with the original proposed model

PUBLICATIONS

- Sethuraman, S. C., Reddy Tadkapally, G. et al. **Simplymime: A dynamic gesture recognition and authentication system for smart remote control**. IEEE Sensors Journal (2024). https://doi.org/10.1109/JSEN.2024.3487070
- Sethuraman, Sibi C., Gaurav Reddy Tadkapally, et al. ***iDrone: IoT-Enabled Unmanned Aerial Vehicles for Detecting Wildfires Using Convolutional Neural Networks***. Springer Nature Computer Science (2022). https://doi.org/10.1007/s42979-022-01160-7