Ruben Varkey Alias

ruben.alias715@gmail.com | Open to Relocate | Linkedin | Website

SUMMARY

AI Engineer who has delivered AI/ML solutions at T-Mobile, Delta, and the New York Yankees. Specialized in RAG, LLMs, and computer vision with impact on latency, cost, and accuracy.

EXPERIENCE

Associate Software Engineer **May 2025 – Present**
T-Mobile

- Deployed an **AI Agent Evaluation platform** running on an **NVIDIA DGX BasePod**, using LangChain and MLFlow to evaluate off-the-shelf and **fine-tuned Agent** performance on use-case specific benchmarks
- Engineering **ML Pipelines** for alarm correlation, processing fault data using **PyTorch and scikit-learn** to reduce MTTR for incidents across **60,000 devices**; early results show effective and explainable correlations being made
- Prototyped internal **RAG** service for live email search, improving **precision@k by 45%** vs baseline semantic search on live NOC email activity data, while reducing email processing **token cost by ~75%** via efficient preprocessing
- Mitigated a **critical production outage** while covering for the Lead Architect, diagnosing that our Azure PTU deployment of **GPT-4o** faced overload from a batch workflow test, halting the test and collaborating on fixes
- Deployed a scheduled **ansible automation** to manage storage utilization across 30+ Linux machines, reducing alarms generated from excessive storage utilization due to dangling docker images to zero

Agentic AI Intern **Jan 2025 – May 2025**
Delta Air Lines

- Prototyped an **agentic app** leveraging **Llama 3.2**, **LangChain**, and **OpenAI-Whisper** for **SQL-RAG** based on natural-language voice commands, featuring OpenAI-style live status for **improved UX**
- Utilized **Chain-of-thought** SQL querying, where the model iteratively fetches, analyzes, and decides if the results match the user's requested query while applying personalization of responses based on their user profile
- Implemented **server-side media caching** to reduce API latency, **reducing average response time by 95%**

Machine Learning Associate **May 2024 – August 2024**
New York Yankees

- Delivered a **CV** model achieving **93% accuracy** in classifying what camera position game footage was taken from, **accelerating timelines** in training a **YOLOv8** Batting Keypoint Detection Model achieving **96% precision**
- Designed a swing extractor which utilized key-points gathered from batting footage to identify with **98% accuracy** the swing window and bat trajectory at both bat ends for use in automated swing analysis from video footage

Application Developer **Oct. 2022 – May 2025**
Rutgers OIT

- Launched and supported myRutgers, a cross-platform mobile application serving **70,000+ users**

PROJECTS

FrameX Image Super Resolution

- Increased SSIM by **30%** and PSNR by **90%** using an EDSR model on Minecraft frames

Multimodal Action Classifier

- Trained and evaluated multiple **multimodal sequence-to-classification models** to classify physical action state (walking, running, jumping) based on sensor data collected from their phone, achieving **95% classification accuracy**

RASCAL - Self-Driving Robot Car

- Trained a CNN model deployed on a NVIDIA Jetson Nano to drive a robot car in a small-scale city environment
- Designed a **modular software architecture in ROS2**, enabling rapid hot-swapping of control from human to AI

CUDA Kernel - Mandelbrot Set Generator

- Achieved **11x throughput improvement** (17s to 1.6s) by writing **CUDA kernels** for rendering the Mandelbrot set

LoreAdapter

- Build an end to end **LLM fine-tuning pipeline** for style and persona enforcement, built to run locally on an RTX 4060, leveraging LoRA for execution on constrained hardware and minimizing model drift
- Synthetically generates dialogue examples based on character and environment information provided to the system

TECHNOLOGY/SKILLS

Python, Prefect, SQL, LangChain, Docker, PyTorch, Linux, CUDA, C++, LLMs, RAG, Azure, MLFlow, Flutter, React, FastAPI

EDUCATION

Rutgers University - School of Engineering **May 2025**
B.S. Electrical and Computer Engineering *Highest Honors*