Atharv Ramesh Nair

@ atharv.ramesh2003@gmail.com | @ a3nair@ucsd.edu | linkedin.com/in/Atharv-Ramesh/
| atharvrn.github.io/AtharvRamesh.github.io | github.com/AtharvRN | +1 858 241 4500

RESEARCH INTERESTS : Medical Imaging, Reinforcement Learning, AI Safety, Mechanistic interpretability.

Education

University of California San Diego — La Jolla, USA
M.S. in Machine Learning and Data Science (ECE) ***GPA: 4.0/4.0*** — *Sep 2025 - Jun 2027*
- **Relevant coursework:** Reinforcement Learning, Deep Generative Models, Efficient AI, Statistical Learning, Linear Algebra

Indian Institute of Technology Hyderabad (IITH) — Hyderabad, India
B.Tech. in Electrical Engineering ***GPA: 3.91/4*** — *Nov 2020 - May 2024*
- **Relevant coursework:** ML, DL, NLP, CV, Matrix Theory, Probability, Information Theory, Algorithms, Convex Opt.

Research Experience

Trustworthy ML Lab (Prof. Lily Weng's Group) — Concept Bottleneck Models UCSD, Nov 2025 - Present
- Training **CheXpert** abnormality classifiers and making them explainable with Concept Bottleneck Models by auto-generating concept annotations and localizations using **CheX** (text-conditioned detector) and VLMs (**CheXAgent/MedGemma/RadVLM**), with **CheX-DETR** concept detectors and CBM layers (BiomedCLIP for LF-CBM, CheX for VLG-CBM), plus BiomedCLIP similarity scoring, VLM/LLM concept checks, and Grad-CAM saliency.

Deep Learning for OCT Imaging (Dr. Kiran Vupparaboina) University of Pittsburgh, Jan 2024 - Jun 2024
- Co-authored a **Bioengineering 2024** paper on RETFound-based retinal OCT feature detection. We fine-tuned a foundation model pretrained on **1.6M** OCTs using **1,770** labeled B-scans (SRF/IRF/drusen/PED) and benchmarked single-task, multi-task, and ResNet-50 baselines, reaching **0.75-0.80 AUC-ROC** and **0.75-0.77** accuracy.
- Explored artificial OCT scan generation with Pix2PixGAN using a RETFound decoder and built a MONAI Generative AI pipeline to run synthesis experiments.

Publications

Du, K.; **Nair, A.R.**; et al. *Detection of Disease Features on Retinal OCT Scans Using RETFound. Bioengineering*, 2024, 11, 1186. https://doi.org/10.3390/bioengineering11121186

Select Projects

Prompt-Length Optimization via Reinforcement Learning
- We studied AdvBench adversarial suffix optimization where the goal is a target completion with minimal suffix length.
- We trained a lightweight RL agent to add, delete, or keep suffix tokens and passed the updated suffix to the Greedy Coordinate Gradient optimizer.
- We trained the agent with GRPO, PPO, and REINFORCE and kept GRPO because it was more stable in training.
- We achieved 43.8 percent compression from 16 to 9 tokens while improving per-token log likelihood from -1.64 to -1.12.

StealthRL: RL paraphrasing for detector robustness
- We trained a paraphrasing LLM (**Qwen3-4B-Instruct**) with **GRPO + LoRA** via the Tinker API and preserved semantics at **0.896** E5 similarity. Robustness evaluation showed **0% TPR @ 1% FPR** on RoBERTa, Fast-DetectGPT, and Binoculars.
- These results show that current detectors are brittle under adaptive paraphrasing and motivate more robust detection methods.

LLM Test-Time Scaling using Process Reward Models
- Adapted DreamPRM to Lean4 and trained initial PRMs to test PRM-style scoring for formal math.
- Built the evaluation and training stack with vLLM, FlashAttention, and PEFT fine-tuning.
- Ran A100 Kubernetes experiments to assess feasibility and identify scaling bottlenecks.

Industry Experience

Netradyne – *Software Engineer — Machine Learning (Edge / Perception)* — Bengaluru, India - *Jun 2024 – Aug 2025*
- **ML perception systems:** Designed multi-camera, multi-model perception for driver/road monitoring; trained, evaluated, and deployed on resource-constrained edge devices.
- **Pipeline re-architecture (-17% latency):** Decoupled ingestion/preprocess/inference/publish with scheduled execution for temporal video models, reducing contention and enabling new features on a limited-memory Qualcomm SKU.
- **Reliability & debugging:** Owned a lightweight Qualcomm SKU; profiled memory/concurrency issues, built diagnostics, and monitored production devices for stable real-world performance.

Silicon Labs – *Software Engineer Intern* — Hyderabad, India - *May - Jul 2023*
- Built adaptive rate control for embedded Wi-Fi; tuned throughput-range from telemetry; shipped firmware.

Alog Tech – *Robotics Software Developer* — Hyderabad, India - *May - Aug 2022*
- Built an autonomous mobile robot with YOLO perception and ROS navigation (global/local) for real-world deployment.

Awards

1st Place — IEEE Signal Processing Cup 2024 (ICASSP, Seoul, South Korea) — Far-Field Speaker Verification.

2nd Runner-Up — IEEE Video and Image Processing Cup 2023 (ICIP, Kuala Lampur, Malaysia) — OCT Biomarker Detection.

Skills

ML	PyTorch, TensorFlow, Transformers, vLLM, LoRA, ONNX, TensorRT, Quantization
Programming	Python, C/C++, MATLAB, SQL, Bash, HTML/CSS/JS
MLOps / Infra	Linux, Docker, Kubernetes, Git, CI/CD, AWS (EC2, S3), LangChain, LlamaIndex
Data	NumPy, Pandas, SciPy, Matplotlib, PostgreSQL, MongoDB, Streamlit