

# 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 :** Mechanistic interpretability, Medical Imaging, Reinforcement Learning, AI Safety.

## 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

- Exploring chest X-ray interpretability with **CheXAgent/MedGemma** VLMs, **CheX-DETR** concept detectors, and CBM layers (BiomedCLIP for LF-CBM, CheX for VLG-CBM); comparing concept presence/accuracy via VLM/LLM queries and visualizing activations (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; fine-tuned a foundation model pretrained on **1.6M OCTs** using **1,770** labeled B-scans (SRF/IRF/drusen/PED) and benchmarked single- vs multi-task vs ResNet-50, reaching **0.75-0.80 AUC-ROC**.
- Explored OCT scan synthesis with GAN/Pix2Pix and MONAI latent diffusion models for data augmentation (exploratory).

## 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

- Formulated discrete prompt optimization as RL (AdvBench, Pythia-70M); GRPO-style updates shrink suffixes while preserving log-likelihood—e.g., 16→9 tokens (**43.8%** compression) with per-token log-likelihood improved from **-1.64** to **-1.12**; ablated entropy/credit assignment and REINFORCE/PPO baselines.

### LLM Test-Time Scaling using Process Reward Models

- Built an **LLM test-time reasoning prototype** with CoT + PRM-style scoring for Lean4 tasks; implemented **vLLM/FlashAttention** inference, KV-cache batching, and evaluation harnesses (exploratory).
- Launched **Kubernetes** jobs on shared A100-80GB clusters; fp16 inference, memory-capped dataloaders; PVC-mounted datasets; containerized **PyTorch/PEFT (LoRA)**; automated checkpoints and W&B logging.

### StealthRL (Exploratory)

- Prototyped **StealthRL**: GRPO + LoRA paraphrasing policy tuned against detector feedback with semantic/fluency and **ESL** fairness penalties; reduced Fast-DetectGPT scores (0.587 → 0.458) while holding similarity **0.944** at best checkpoint.

### IEEE SP Cup 2024 (ICASSP) — Far-Field Speaker Verification

- **1st place globally**: adapted **ERes2Net** with targeted augments (RIR, MUSAN, speed) and robot-ready scoring (cosine + adaptive s-norm); final leaderboard **minDCF 0.67, EER 8.93**.

## 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 Lumpur, Malaysia) — OCT Biomarker Detection.

## SKILLS

**ML Programming** PyTorch, TensorFlow, Transformers, vLLM, ONNX, TensorRT, Quantization

**MLOps / Infra** Python, C/C++, MATLAB, SQL, Bash, HTML/CSS/JS

**Data** Linux, Docker, Kubernetes, Git, CI/CD, AWS (EC2, S3), LangChain, LlamaIndex

NumPy, Pandas, SciPy, Matplotlib, PostgreSQL, MongoDB, Streamlit