VARAD VIJAY SURYAVANSHI

MSCS @ NYU | ML Infrastructure • Foundation Models & World Models • Data Pipelines
(551) 229 6385 | vs3273@nyu.edu | LinkedIn | Github | Portfolio

EDUCATION

New York University | CGPA: 3.76/4.0 — New York, NY, US
Master of Science, Computer Science — *September 2024 - May 2026*

University of Mumbai | CGPA: 3.71/4.0 — Mumbai, India
Bachelors in Engineering, Computer Science — *September 2020 - June 2024*

SKILLS

Languages: Python, C++, SQL, Bash, JavaScript

ML & Foundation Models: Generative Modeling, Diffusion-Based World Models, Reinforcement Learning, Reward Modeling, Learning from Demonstration, Model Distillation, Self-Supervised Learning, Robust Evaluation, PyTorch, JAX

ML Infrastructure & Data: Large-Scale Data Pipelines, ETL, Apache Spark, Apache Kafka, Distributed GPU Training (A100s), FSDP, Mixed Precision, Data Curation, Multimodal Datasets, CI/CD for ML

Systems & Practices: Distributed Systems, REST APIs, FastAPI, Docker, AWS, Linux, Git, Evaluation Framework Design, Ablation Studies, Technical Documentation

EXPERIENCE

AI4CE Lab, NYU (Prof. Chen Feng) — NYU, USA
Research Assistant — World Models — *June 2026 - Present*

- Curating data and building **data pipelines** to improve dataset quality for the **Wanderland world-model project**, supporting **data-centric improvement** of models for autonomous navigation research.
- Developing processes for **multimodal data collection and curation** to enable more reliable world-model training on real-world scenes.

CILVR Labs NYU *(ICLR 2026)* — NYU, USA
Research Assistant under Prof. Sherry Yang — *May 2025 - October 2025*

- Built **large-scale data pipelines and distributed training infrastructure** for a **diffusion-based video world model**, processing **5TB+ of multimodal data** across **100K+ rollouts** on **2–4 A100 GPUs with FSDP and mixed precision**. *(ICLR 2026)*
- Shepherded the model **from research prototype to a robust evaluated component**, deploying **GPT-4o** as an automated judge across **100K+ runs** for **robust evaluation**, achieving $\mathbf{r \approx 0.78}$ with real-world outcomes.
- Developed a **multimodal perception pipeline** integrating **LLaVA-v1.5 (7B)** and **DeepSeek-R1**, with rigorous testing and code review across a shared research codebase.

NTT DATA BUSINESS SOLUTIONS — Hyderabad, India
Software Engineering Intern — *August 2023 - September 2023*

- Built **production ML pipelines** using **LangChain** and **GPT-3.5** with embedding-based retrieval and **REST APIs** on **AWS**, improving accuracy by **40%** and optimizing inference latency to ∼**1.8s**.

PROJECTS

Reinforcement Learning Research (DreamerV3): (Website)

- Developed **model-based RL** with **reward modeling** and exploration strategies across **32 distributed environments**, using iterative plan–retrain loops for scalable policy improvement.
- Validated an **architectural fix via systematic ablation** under constrained compute, demonstrating rigorous experimental methodology.

Representation Learning & Scaling Research (DINOv2): (GitHub)

- Studied **scaling and generalization** via **DINOv2** distillation with **VICReg regularization**, analyzing how **data diversity and compute** affect representation quality, achieving **+5.1%** and **+4.8%** successive gains.
- Benchmarked **SimCLR, BYOL, and SwAV** through controlled experiments and ablation.

Foundation Model Post-Training (CogniLLM): (GitHub)

- Fine-tuned **DeepSeek-R1 (14B)** using **QLoRA** and **synthetic data generation**, building **post-training pipelines** with **multi-solver evaluation**, improving accuracy **8%** → **29%**.
- Designed **evaluation harnesses** studying how data composition and training techniques affect generalization on novel distributions.