HARSH SUTARIA

+1 (201) 522 2179 ◇ New York, USA
h.sutaria@nyu.edu ◇ linkedin.com/in/harsh-sutaria ◇ github.com/harsh-sutariya ◇ leetcode.com/theharshsutariya

EDUCATION

Masters of Science - Computer Science Sep 2024 - May 2026
New York University - Courant Institute of Mathematical Sciences

B.Tech. Computer Science Engineering Sep 2020 - May 2024
Vellore Institute of Technology

PROFESSIONAL EXPERIENCE

Graduate Researcher Jan 2025 - Present
Planning with Latent Dyanmics Models (Advised by Prof. Yann LeCun)

- Extending PLDM to solve complex tasks by training a **hierarchical JEPA world model**.
- Implemented **Transformer with Markov assumption** as predictor. Trained using teacher forcing with **VICReg** regularization to **prevent representation collapse** and reduce embedding redundancy.
- **Optimal control (MPC)** planning on the latent dynamics model shows best **generalization to new tasks**.

Student Leader Sep 2024 - Present
NYU Self Drive | AI4CE Lab

- **Led the team** to develop **autonomous navigation** with **world models as visual inertial state estimators** for **localization and controls**. Achieved comparable results against Kalman Filter based VIOs.
- Built **visual place recognition** using YOLO, DINOv2 and VLAD-BuFF improved image matching by 30%.
- Participated at the joint competition by NASA & Johns Hopkins University's **Lunar Autonomy Challenge**.

Deep Learning Engineer Apr 2023 - Jul 2024
Jaipur Robotics Sagl (Formerly, WiSort)

- Performed image classification and segmentation by **fine-tuning YOLO and CLIP for downstream tasks**.
- Used **Segment Anything** and **LabelStudio** to create datasets resulting in reduction of human work by 60%.
- Created resilient backend architectures by containerizing applications with Docker and Google Cloud Platform.

PROJECTS

Attention-Aware DPO for Reducing Hallucinations in Multi-Image QA) Nov 2024

- Engineered DPO loss incorporating **cross attention penalties** to reduce hallucinations in multi image QA for Large Vision Language Models (LVLMs), **improving target image focus by 33.93%** (vs. 29.43% baseline).
- Performed inference time optimization by **confidence based attention scaling**, boosting accuracy by 10%.
- Trained on **LLaVA665k augmented datasets** using **LoRA fine-tuning**, achieving 3.20 average accuracy (vs. 2.95 baseline) across **PixMo benchmarks** with 0.393 attention ratio.

Generative Fashion Design Agent Jan 2024

- Finetuned **SDXL** diffusion model using **LoRA** and **VAE** to generate textile patterns inspired from base image.
- Leveraged Meta's **SAM model for better image layer extraction**, enabling diverse textile patch creation.
- **20+ Daily active users** (sketch artists) boost their productivity by over 75% at the industry level.

SKILLS

- **Framework and Libraries** : PyTorch, TensorFlow, ROS, Flask, Streamlit, Scikit-learn, OpenCV, Pillow.
- **Tools** : AWS, Google Cloud, Git, Docker, Firebase, ComfyUI, PowerBI, Tableau, QuickSight, RapidMiner.
- **Programming Languages** : Python, C++, C, Go, Java, JavaScript, SQL, Racket, PLAI-Typed.

ACHIEVEMENTS

- Ranked **3rd in Deep Learning Challenge** hosted by Prof. Yann LeCun at NYU in PhD level course.
- Scored A+ Grade in On-Campus Deep Learning Program at **National University of Singapore** (NUS).
- Organized Riviera'23, South India's largest fest with budget of **$1 million** and hosting **30,000+ participants**.