

Yash Thube

Pune, India | thubeyash09@gmail.com | github.io

Research Interests and Technologies

- **Multimodal Learning** (VLMs, MLLMs, representation learning)
- **Computer Vision** (Deep learning for vision, 3D scene understanding, Few/zero shot learning, video understanding & long-horizon prediction)
- **Reinforcement and Open-ended Learning** (World models, VLAMs, Embodied AI)

Tools - PyTorch, OpenCV, Huggingface (Transformers, TRL, Diffusers, PEFT), NumPy, TorchVision, TransformerLens, Pillow, Scikit-learn, Matplotlib, MLOps (Docker, Kubernetes, AWS).

Experience

- **Jr. Machine Learning Engineer**, Hudl India – Pune, MH 05/2024 – 11/2024
Enhanced sports video performance tracking accuracy by 35% through deep learning model development. Automated video classification using computer vision (SVM, CNN), reducing manual review time by 60% and streamlining workflows.
- **Machine Learning Collaborator**, Omdena – Remote, Open Source 12/2023 – 03/2024
Analyzed social media's mental health impact and subsequently designed/implemented solutions for healthier online interactions using Language models, RAG, prompt engineering, NLP, audio processing, and deployment.
- **Technical Writer**, InPlainEnglish | Towards AI – Remote (Freelance) 04/2023 – Present
Distilled complex technologies into clear and accessible content with primary focus on AWS, ML and Serverless technologies.
- **Member (Communities)**
Cohere Labs | Eleuther AI | Hugging Face | ML Collective (MLC) | MLOps Community | AI Accelerator Institute

Projects

- **MATS (arXiv preprint)** – A behavioral audit toolkit to detect pathological truth bias in Vision-Language Models (VLMs), experiments include activation patching to causally localize failures in cross-attention layers and pooled representations across LLaVA, CLIP, and Qwen-VL architectures.
- **Multimodal/VLMs Research Hub** – A technical resource for researchers exploring Vision-Language Models (VLMs) and Multimodal Learning, featuring seminal papers/models, datasets, benchmarks, ethical challenges, and research directions.
- **Task-aware SAM LoRA** – PyTorch pipeline that uses a hypernetwork to generate task-specific LoRA adapters for Meta's Segment Anything Model from natural language prompts, targeted segmentation on COCO instances and benchmarked mIoU via pycocotools.

Github

Education

Savitribai Phule Pune University (SPPU)
B.E. Computer Science | Pune, India

2021 – 2025