

Shrikar Madhu

New York, NY | sm12761@nyu.edu | 347-866-1492
github.com/Shrikar17 | linkedin.com/in/shrikar-madhu

Education

New York University - Courant Institute of Mathematical Sciences, New York, NY *Aug 2025 – May 2027*
M.S. in Computer Science (Concentration: Artificial Intelligence) GPA: 3.7/4.0
PES University, Bangalore, India *Aug 2019 – May 2023*
B.Tech. in Computer Science and Engineering (Data Science Specialization) GPA: 3.8/4.0
Relevant Coursework: Algorithms, Machine Learning, Computer Vision, Reinforcement Learning, Deep Learning, Statistics

Technical Skills

Programming Languages: Python, C/C++, Java, JavaScript, SQL, Shell Scripting
AI & ML: LLMs, Agentic AI, Reinforcement learning, deep learning frameworks (PyTorch, TensorFlow, JAX, Keras)
AI Systems & Tools: LangChain, LlamaIndex, RAG, multimodal AI, model fine-tuning, prompt engineering, foundation models, agent orchestration
Development & Systems: Distributed training, performance optimization, benchmarking, compiler optimization, CUDA
Cloud Platforms: AWS, GCP, CI/CD pipelines, Git, Docker, Kubernetes

Research & Professional Experience

New York University - Computer Vision Lab, Graduate Researcher with Prof. Saining Xie *Sep 2025 – Present*

- Architected long-context video understanding system using distributed PyTorch training across 8 A100 GPUs with custom CUDA kernels, achieving 7% mAP model performance improvement and 40% memory reduction over baseline.
- Designed spatial reasoning architectures by implementing sparse transformers, conducting 50+ ablation experiments across 3 benchmark datasets, improving inference speed by 2.1x through architecture optimization.
- Built Pico-LLM using Gated Linear Attention to achieve $O(N)$ complexity, outperforming Transformers by 8–16x on 5k-token sequences. Retained 94% reasoning accuracy via GRPO based RL post-training, achieving 0.98 reward stabilization.
- Optimized training pipelines for 7B+ parameter models using gradient checkpointing and mixed-precision training, reducing training time by 35% on multi-node clusters.

Kotak IISc AI-ML Research Center, Pre-Doctoral Scholar *Jun 2024 – Jul 2025*

- Developed VISTA-CLIP continual learning framework using visual prompt tuning with <1M parameters, achieving SOTA 62.3 mIoU on Cityscapes panoptic segmentation and preventing 53% catastrophic forgetting (**CVPR 2025**).
- Engineered edge deployment pipeline for autonomous systems using selective parameter adaptation and knowledge distillation, reducing inference latency by 3.2x while maintaining 98% accuracy on embedded devices.
- Built multi-sensor EO-IR dataset with 50K+ aligned pairs; trained Pix2Pix GANs achieving 0.85 SSIM for cross-modal synthesis, improving all-weather perception by 28% in autonomous navigation tasks.

Physi.Fit, Inc., Co-Founder & AI Research Engineer *May 2022 – Jul 2025*

- Deployed CV platform across 17 healthcare facilities serving 100+ patients using attention-based GNN achieving 92% motion assessment accuracy and 15% improvement in patient recovery outcomes.
- Built production ML pipeline processing 30 FPS real-time pose estimation with Mobile MediaPipe and custom CNNs, Video-LLaVA, reducing therapy session time by 25% and enabling automated progress tracking for 10K+ sessions.
- Led end-to-end ML systems development from data collection to Docker/Kubernetes deployment of a multimodal Health Chart-VLM achieving 99.7% uptime; **inventor on granted patent India, US PTO.**

Indian Institute of Science (IISc) - AI & Robotics Lab, Research Assistant *Jul 2022 – Sep 2023*

- Designed spiking neural network using temporal coding and STDP learning, outperforming CNN baseline by 4.35% on CIFAR-10 while reducing energy consumption by 60% for neuromorphic edge AI applications.

Selected Publications & Patents

- D. Manjunath, **S. Madhu**, et al. *VISTA-CLIP: Visual Incremental Self-Tuned Adaptation for Efficient Continual Panoptic Segmentation*, **CVPR 2025**.
 - D. Manjunath, **S. Madhu**, et al. *IndraEye: Infrared Electro-Optical UAV-based Perception Dataset*, Submitted to **ICRA 2025**.
 - **S. Madhu**, Y. Mahamuni, K. Suresh. *Method and system for conducting interactive rehabilitation sessions with continuous monitoring*, **U.S. Patent Application 18/352,300**.
 - **S. Madhu**, et al. *Detection of Conversational Health in a Multimodal Conversation Graph*, **ACM ICMVA 2023** (Oral).
-

Honors & Leadership

- Qualcomm AI Hackathon 2025 (Finalist – built RAG system with 85% retrieval accuracy), VMware Hackathon 2023 (National Finalist)
- IIT Bombay Robotics Cup (Runner-up), PESU Venture Labs × ACM-W Hackathon 2021 (Winner)
- Times Invest BLR (Top 10 AI Startups), First-author CVPR 2025 paper and 4 peer-reviewed publications with citations