Sidhartha Reddy Potu

+1 (917) 914-1526 | sp7835@nyu.edu | GitHub | LinkedIn | Portfolio

Education

New York University, Tandon School of Engineering — New York, NY
Master of Science in Computer Science; GPA: 3.83/4 — *Sep 2023 – May 2025*

Indian Institute of Technology (IIT), Indore — Indore, IN
Bachelor of Technology in Electrical Engineering, Minor in Astronomy; GPA: 8.49/10 — *Jul 2019 – May 2023*

Experience

SWE Intern — Mountain View, CA
Granica AI — *Sep 2025 – Present*

- Implemented a dual-lane FIFO router for memory-bound workers, doubling throughput (**10** to **20**+ req/worker).
- Prevented runaway Kubernetes API calls under **20k** concurrency by caching scaler state in Go.
- Fixed lock contention and races in load balancer/consumer goroutines, improving throughput (**28** to **31–32** cores).
- Built scale-validation tests for multi-consumer routing and integration checks for compression pipelines.

MTS Intern — Mountain View, CA
Neopolis AI — *Aug 2025 – Sep 2025*

- Refined LiveKit-based LLM voice AI Agent prompts to improve instruction-following and downstream tool inputs.
- Used Langfuse observability to inspect agent outputs and guide prompt iteration.
- Added session inactivity timeouts to prevent idle conversations and improve reliability.

Graduate Student Researcher, CILVR, NYU — New York, NY
3D Computer Vision under Prof. David Fouhey — *Jan 2024 – May 2025*

- Integrated COLMAP SfM with HaMeR to recover global 3D hand trajectories on Epic-Kitchens and EgoExo4D.
- Optimized hand-motion trajectories using temporal smoothing with velocity and jerk regularization.
- Recovered full-perspective 3D from weak-perspective predictions using SfM intrinsics and frame transforms.
- Developed visualization workflows with Rerun Viewer and scaled experiments on NYU HPC using SLURM.

Course Assistant, NYU — New York, NY
Computer Vision, Deep Learning, LLVM courses — *Sep 2024 – May 2025*

- Led office hours, grading, and project support across CV/ML courses; supported **450**+ students.
- Course staff for DL (Prof. Chinmay Hegde) and LLVM (Prof. Saining Xie); Advanced CV (Prof. David Fouhey).

B.Tech Thesis, PRIA Lab, IIT Indore — Indore, IN
Multimodal Crowd Counting using Vision Transformers [Report] — *Jun 2022 – Dec 2022*

- Trained a ViT-based model in a weakly supervised setting, achieving MAE 15.78 and MSE 30.70.
- Designed a Pyramid Vision Transformer (PVT v2) model, achieving MAE 19.218 and MSE 32.940.
- Ranked **top 10 of 250+** B.Tech theses; nominated for Best Thesis award at IIT Indore.

Projects

Fine-Tuning Medical QA Models — GitHub — Oct 2024 – Dec 2024

- Optimized Meta LLaMA 3.2 3B for medical QA using LoRA, QLoRA, and GaLore.
- Achieved best USMLE accuracies: Step 1 - **42**% (GaLore), Step 2 - **38**% (QLoRA), Step 3 - **45**% (GaLore).
- Improved MedQuAD F1 scores to **70**% (QLoRA) and 69% (GaLore) with efficient fine-tuning.

VioletPass: Ticket Booking Platform — GitHub — Oct 2024 – Dec 2024

- Designed a robust, scalable ticket booking platform on AWS to prevent double booking.
- Implemented Redis distributed locks with TTL for seat reservations and PostgreSQL for finalization.
- Integrated QR code-based authentication for secure and efficient event check-ins.

AnimeVerse: Recommendation System — GitHub — Apr 2024 – May 2024

- Built a personalized anime recommendation system; analyzed broadcast types, genres, and scores.
- Analyzed **300k** user-anime interactions with PySpark and trained ALS, TF-IDF, and ChromaDB models.

Technical Skills

Languages: Python, Go, C/C++, SQL
Systems: Linux, Git, Docker, Kubernetes, AWS, GCP, CI/CD, Weights & Biases (W&B)
ML/Data: PyTorch, Transformers (Hugging Face), TensorFlow, NumPy, Pandas, Scikit-learn, PySpark, OpenCV