

# Shitij Govil

U.S. Citizen | [sgovil9@gatech.edu](mailto:sgovil9@gatech.edu) | [linkedin.com/in/shitij-govil](https://www.linkedin.com/in/shitij-govil) | [github.com/sgovil5](https://github.com/sgovil5) | 980-269-9422

## EDUCATION

---

### Georgia Institute of Technology

Bachelor of Science in Computer Science (GPA: 3.95)

Atlanta, GA

May, 2027

- Accepted paper at ICML 2025 workshop on [finetuning Chain of Thought in LLMs](#)

## EXPERIENCE

---

### Machine Learning Research Intern

PAIR Lab - Georgia Institute of Technology

Aug 2025 – Present

Atlanta, GA

- Developed a robotics system for learning motion representations from 10K+ hours of unlabeled videos; targeting ICML
- Engineered a multi-GPU (DDP) preprocessing and dataloading pipeline for multi-terabyte video datasets, supporting many datasets, camera pose estimation and depth normalization, and 3D point-trackers.
- Built a modular research platform integrating 3+ point-tracking models and multiple VLM backbones, enabling rapid ablations and cross-domain transfer with minimal task-specific supervision

### Machine Learning Research Intern

GCOM Lab - Georgia Institute of Technology

May 2024 – Present

Atlanta, GA

- Scaled particle tracking to 250k+ points and achieved 35x faster inference by first-authoring a hashing-based point transformer (PyTorch) and leading a 10-person research team; accepted in [NeurIPS](#) workshop. Targeting Nature Physics.
- Delivered a 16% accuracy improvement over SOTA by engineering an end-to-end particle tracking pipeline with a MaskFormer-style segmentation backbone, FlashAttention, and multi-GPU support.
- Reduced GPU memory footprint by 30%+ by pioneering point cloud serialization techniques that transform irregular geometric data into hardware-friendly sequences, enabling transformer deployment at scale.

### Machine Learning Engineer Intern

Kitware

May 2025 – Aug 2025

Minneapolis, MN

- Proposed and led implementation a diffusion-based augmentation module, architecting its core API, plugin system, and data-flow pipeline for long-term maintainability in an open-source library
- Built a scalable image-generation pipeline capable of producing 10K+ balanced samples per run, using modular components, batched GPU scheduling, containerized environments, and automated validation.
- Reduced per-sample latency by 20% via noise scheduling and increased reliability with failure-resilient batch processing.

### Software Engineering Intern

Albertsons Companies

June 2023 – August 2023

Charlotte, NC

- Streamlined codebase management across 300+ repositories by developing a React-based dashboard and integrating CI/CD pipelines to enforce SonarQube Exceptions with automated deployment.
- Built RESTful APIs to fetch SonarQube Exceptions via OpenAPI specifications, improving maintainability.
- Improved API data retrieval speed by 60% through in-memory caching with node-cache in Node.js, reducing backend latency

## PROJECTS

---

### Winner-Takes-All Sparse Autoencoders for Interpretability - WTA SAE | *Python, Pytorch*

2025

- Designed and implemented a modular PyTorch autoencoder with a custom sparsity mechanism enforcing per-feature activation budgets, improving feature stability by 35% and eliminating dead-unit collapse
- Built a reproducible training pipeline with automated evaluation scrips, interpretability analysis tests, and CI workflows

### CUDA Genetic Image Recreation - PolyGenetic | *CUDA, C, C++, AWS, Docker*

2024

- Achieved 125x faster runtime by developing a high-speed genetic algorithm in CUDA to replicate images with randomly generated polygons, outperforming the original C++ baseline.
- Optimized GPU utilization with parallel programming techniques like thread pooling, privatization, and memory alignment.
- Deployed by containerizing applications with Docker and managing execution on AWS EC2 ad ECR

### Autonomous Rover Navigation - GT RoboJackets | *C++, Python, ROS, Nav2, CMake*

2024

- Reduced obstacle detection latency by 60% by architecting a real-time C++ perception pipeline that processes raw LiDAR at 500k points/sec with a novel voxel filtering method.
- Cut traversal times by 15% and achieved 95% navigation success in dynamic, unstructured environments by engineering a hybrid A\*/DWA planner for robust path planning.

## TECHNICAL SKILLS

---

**Languages:** Python, C/C++, CUDA, Java, Javascript, Bash, SQL

**Frameworks:** Pytorch, ROS, React, Next.js, Flutter, Tailwind

**Developer Tools:** Git, AWS, Docker, Jupyter Notebook, HPC (PACE, NERSC), Linux

**Libraries:** Tensorflow, Numpy, Pandas, Scikit-Learn, vLLM, Matplotlib