Shivam Singh

shivamsingh3458@gmail.com | LinkedIn | GitHub | shivhere.github.io | Google Scholar | Phone:+91 8130870896

AI Researcher with 3+ years of experience in Vision-Language Models, multimodal reasoning, video understanding, and embodied AI.

EDUCATION

M.S. -- Electronics and Communication Engineering | IIIT Hyderabad (2023–2026) CGPA: 8.33/10
B.Tech -- Electronics and Communication Engineering | RKGIT, Ghaziabad (2017–2021) CGPA: 8.07/10

RESEARCH & PROFESSIONAL EXPERIENCE

PIXELSTWO | Robotics Deployment Lead Remote | Sep 2026 - Present
Focus: Video Data Quality & Real-to-Sim / Sim-to-Real Systems

- Performing **quality control of video data**, ensuring recordings are reliable and consistent for downstream training and evaluation.
- Working on **real-to-sim and sim-to-real systems**.

INTUITION.DEV | Collaboration Remote | Aug 2026 - Sep 2026
Focus: Real-World Video Pipelines

- Label-free Segmentation and Grounded Description of Egocentric Work Videos: *2026*
  - Built a **geometry-first, label-free pipeline** that converts raw body-camera recordings of manual work into **clean, standalone sub-task clips**: removed inactive/noisy footage with classical detectors and detected **task-transition points** from metric camera trajectories using purely classical image-processing heuristics, without relying on large models for segmentation and prior annotations.
  - Structured raw videos for downstream **VLM-based annotation** by isolating relevant task segments before captioning, **reducing VLM inference cost** and enabling more consistent, task-level annotations under a fixed JSON schema with grounded actions. [Technical Report ↗]

AMD INDIA PRIVATE LIMITED | Research Fellow AI Models & Agents Team | Jun 2025 - July 2026
Focus: VLMs, Video Reasoning & Spatial Understanding

Supervisors: Saptarshi Majumder, Pratik Prabhanjan Brahma

- Spatial Video Understanding from Raw RGB Videos: *2026*
  - Proposed *SlotMem*, a **memory-augmented** transformer framework for object tracking, spatial understanding, and long-horizon video reasoning from raw monocular RGB videos.
  - Developed a **scalable pipeline** combining Depth-Anything-3, YOLO-World, and Vision-Language Models to reconstruct 3D scenes, estimate camera trajectories, and generate supervision for spatial, temporal, and goal-oriented reasoning.
  - Trained SlotMem using our automatically generated supervision together with diverse videos from VLM-3R and VICA-322K, achieving **state-of-the-art** performance on OST-Bench (**44.4%**) and outperforming all prior baselines.
- Pause and Think: Video-Grounded Assistive Action Suggestion: *2025*
  - Proposed a reasoning-centric framework for video-grounded assistive action suggestion by introducing *Pause-and-Think-T* (∼**10k** structured reasoning QA pairs) with explicit reasoning supervision and *Pause-and-Think-B*, a **300-sample** benchmark for grounded contextual understanding and goal-oriented planning across Epic-Kitchens, Assembly101, and Ego4D.
  - Built a scalable reasoning-data pipeline and fine-tuned a compact **4B** VLM on 8×AMD Instinct MI325 GPUs, achieving **58.0%** accuracy—matching Qwen3-VL-235B (**58.9%**, **59× larger**) and GPT-5.2 on scene understanding, outperforming GPT-4o, with strong zero-shot generalization on EgoThink and TempCompass.
  - Deployed the optimized model on an **AMD RyzenTM AI** platform for real-time on-device inference; as **First Author**, led the research, algorithm design, dataset creation, model training, evaluation, deployment, and manuscript preparation. [Code ↗ | Paper ↗]

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY - HYDERABAD | M.S. Researcher
Robotics Research Centre, Hyderabad | Jan 2023 - May 2025
Focus: Embodied AI & Robot Planning

Direct Supervisor: Prof. Madhava Krishna
Collaborators: Dr. Mohan Sridharan, Dr. Krishna Murthy Jatavallabhula, Dr. Brojeshwar Bhowmick

- **Hybrid Planning with LLMs & Probabilistic Models** (*ECMR 2025*) *2025*
  Mentor: Guided framework design and ablation experiments for task planning under uncertainty with probabilistic effects in multi-agent settings. [Paper ↗]
- **LLM & Knowledge Graphs for Task Decomposition** (*IEEE ICRA 2025*) *2025*
  Project Lead: Designed a framework that integrates Large Language Models (LLMs), Knowledge Graphs (KGs), and Human-in-the-Loop (HITL) feedback to enable embodied agents to perform long-horizon planning for unseen tasks and **progressively adapt** to new environments. [Website ↗ | Paper ↗]
- **Task Anticipation for Multi-Agent Systems** (*ICRA & RSS 2024 Workshops*) *2024*
  Project Lead: Extended LLM-based task anticipation to long-horizon planning in **multi-agent settings**, enabling robotic agents to anticipate domain-specific tasks, adapt to unexpected changes in human behavior, and handle environmental uncertainties for effective human–robot collaboration. [Website ↗ | Paper ↗]
- **Task Anticipation using LLMs** (*IEEE ICRA 2024*) *2023*
  Co-author & Project Lead: Developed an intelligent household agent that performs **long-horizon planning** by anticipating user actions through learning task execution patterns from user behavior and preferences, leveraging Large Language Models (LLMs) and Planning Domain Definition Language (PDDL)-based task and action representations.[Website ↗ | Paper ↗]

ALSOENERGY – INDIA | Technical Support Engineer April 2021 | Dec 2021

- PowerTrack platform support for commercial solar clients: troubleshooting, data visualization, hardware integration.

PUBLICATIONS & WORKSHOPS

- **Pause and Think: A Dataset and Benchmark for Video-Grounded Assistive Action Suggestion:** ↗ **Shivam Singh**, S Majumder, P Prabhanjan, Z Liu, E Barsoum.
  Accepted at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2026 ↗ ***(A*)***. *Work done at AMD (Advanced Micro Devices, Inc.).*
- **Anticipate, Adapt, Act: A Hybrid Framework for Task Planning:** ↗ N Dash, A Kaura, **Shivam Singh**, R Singh, S Banerjee, M Sridharan, M Krishna.
  Accepted at IEEE European Conference on Mobile Robots (ECMR) 2025.
- **AdaptBot: Combining LLM with Knowledge Graphs and Human Input for Generic-to-Specific Task Decomposition and Knowledge Refinement:** ↗ **Shivam Singh**, K Swaminathan, N Dash, R Singh, S Banerjee, M Sridharan, M Krishna.
  Accepted at IEEE International Conference on Robotics and Automation (ICRA) 2025 ↗ ***(A*)***, *Atlanta, US.*
- **Anticipate & Collab: Data-driven Task Anticipation and Knowledge-driven Planning for Human-robot Collaboration** ↗: **Shivam Singh***, K Swaminathan*, R Arora*, D Das, S Banerjee, M Sridharan, M Krishna. *Accepted in RSS2024 Workshop on Task Specification for General-Purpose Intelligent Robots* ↗, *Delft, Netherlands*
- **Anticipate & Act: Integrating LLMs and Classical Planning for Efficient Task Execution in Household Environments:** ↗ R Arora*, **Shivam Singh***, K Swaminathan, A Datta, S Banerjee, B Bhowmick, KM Jatavallabhula, M Sridharan, M Krishna.
  Accepted at IEEE International Conference on Robotics and Automation (ICRA) 2024 ↗ ***(A*)***, *Yokohama, Japan.*
- **DaTAPlan: Data-driven Task Anticipation and Knowledge-driven Planning for Human-robot Collaboration:** ↗ **Shivam Singh***, K Swaminathan*, R Arora*, R Singh, A Datta, D Das, S Banerjee, M Sridharan, and M Krishna.
  Accepted in ICRA 2024 Workshop on Cooking Robotics: Perception and motion planning ↗, *Yokohama, Japan.*

PATENTS

- **Method for Automatic Knowledge Refinement and Task Execution** ↗: **Shivam Singh**, S Banerjee, K Swaminathan, N Dash, R Singh, M Sridharan, and M Krishna. *US Patent App. 19420895*, published July 30, 2026.
- **Method and System for Task Anticipation by Integrating Large Language Models and Classical Planning** ↗: **Shivam Singh**, S Banerjee, B Bhowmick, R Arora, A Datta, K Swaminathan, M Krishna, M Sridharan, and KM Jatavallabhula. *US Patent App. 19089689*, published November 6, 2025.

TECHNICAL SKILLS

Research Areas: LLMs, VLMs, VLAs, Reasoning Models, Computer Vision, Multimodal Learning, Agentic AI, Embodied AI
Training & Inference: PyTorch, Hugging Face, LLaMA-Factory, SFT, RLHF, LoRA, DeepSpeed, vLLM, ROCm
Data & Infrastructure: Multimodal Data Curation, Synthetic Data, Benchmarking, Docker, Slurm, OpenCV, Open3D
Programming & Robotics: Python, C++, C, PDDL, RDDL, ROS, Gazebo, AI2-THOR