

SIDDHARTH PATKI

About

I'm a researcher on the GenAI team at Meta where I work on improving capabilities of Llama. Overall, my research deals with building multi-modal foundation models for agentic and robotic systems focusing on language grounding and perception.

Education

Ph.D in Electrical Engineering, University of Rochester, USA [2017 - 2023]

Dean's Fellow, Hajim School of Engineering | Advisor: Prof. Thomas Howard

M.S. in Electrical Engineering, University of Rochester, USA [2015 - 2017]

B.Tech. in Electronics Engineering, University of Pune, India [2009 - 2013]

Work Experience

Research Scientist, Meta, CA [Oct 2024 - Now]

- Post-training Meta's next generation multi-modal foundation models, Llama.
- Leading efforts on enabling Multilingual and Agentic capabilities in Llama. This allows Llama to natively use tools for effectively helping users with their queries.
- Expertise in curating data-mixes and training recipes for improving performance of multi-modal LLMs for various capabilities.

Postdoctoral Researcher, Meta, CA [April 2023 - Oct 2024]

- Developing a large-scale simulation benchmark for human-robot collaboration, consisting of 45,000 everyday tasks specified via natural language, paired with performance evaluating functions across 60 indoor environments from HSSD.
- Delivered a ReAct based multi-agent task planning baseline for coordinated execution of long-horizon everyday tasks in the aforementioned benchmark.
- Fine-tuning Llama3-8B model on collaboration data collected using a Human-In-The-Loop tool and with oracle based planners for learning efficient task decomposition, agent coordination, and theory of mind.

Applied Scientist, Amazon, CA [Oct 2022 - April 2023]

- Developed vision based damage detection model for a robotic arm deployed to filter damaged products passing on the packaging conveyor belts in AFCs.
- Designed annotation noise robust loss functions which boosted the model performance by ~5% helping the team meet their EOY goals. This enabled data collection at a larger scale by relaxing the dependency on expert annotators.

Graduate Research Assistant, University of Rochester [2017-2023]

- Conducted an arc of research on developing joint models of language and perception for efficient robot instruction following in cluttered dynamic spaces.
- Published 6+ peer-reviewed papers in top-tier robotics conferences and journals including ICRA, CoRL, IROS, RoMAN, SigDial and Field Robotics.
- 5+ years of hands on experience in developing 3D perception stacks for real world robot autonomy problems such as 3D object detection, 6DoF pose estimation, multi-object tracking, scene recognition, semantic segmentation, SLAM, sensor selection and calibration, custom dataset curation etc.

Contact

San Jose, USA
+1 585-434-6584
patkisiddharth@gmail.com

Links

[Personal Webpage](#)
[Google Scholar](#)
[Github](#)
[PARTNR](#)

Programming Skills

C++, Python, Matlab
PyTorch, OpenCV, PCL
ROS, CMake, Catkin

Robots and Sensors

Boston Dynamics Spot
Rethink Robotics Baxter
Universal Robotics UR5 arm
Clearpath Robotics Husky UGV
Intel Realsense 415, 435
Velodyne Lidar

Professional Service

Workshops Organized:
Causal and Object Centric
Representations for Robotics,
CVPR 2024

Program Committee:
CoRL 2020

Reviewer:
RSS 2022, 2020, 2017
IROS 2019, 2023
ICRA 2019
ACL 2020
EACL 2020

Public Outreach:
Robotics demonstrator at the
Rochester Museum and Science
Center 2016, 2019, 2022

Teaching Assistant:
Introduction to C++
Digital Image Processing

Select Publications

- **Siddharth Patki** et al. "PARTNR: A Benchmark for Planning and Reasoning in Embodied Multi-agent Tasks." *arXiv preprint arXiv:2411.00081* (2024).
- **Siddharth Patki** et.al. "Language Guided Temporally Adaptive Perception for Efficient Natural Language Grounding in Cluttered Dynamic Worlds" 2023 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*
- **Siddharth Patki** et.al. "Inferring Compact Representations for Efficient Natural Language Understanding of Robot Instructions" In *IEEE ICRA 2019*
- **Siddharth Patki**, E. Fahnstock, T.M. Howard, and M. Walter, "Language-guided Semantic Mapping and Mobile Manipulation in Partially Observable Environments," In *Conference on Robot Learning*. PMLR, Oct. 2019, vol. 100, pp. 1201-1210
- **Siddharth Patki** and T.M. Howard, "Language-guided Adaptive Perception for Efficient Grounded Communication with Robotic Manipulators in Cluttered Environments," In *19th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Jul. 2018
- Arkin, Jacob, **Siddharth Patki**, Joshua D. Rosser, and Thomas M. Howard. "An Efficient Algorithm for Visualization and Interpretation of Grounded Language Models." In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 266-273. IEEE, 2022.
- Walter, Matthew R., **Siddharth Patki**, Andrea F. Daniele, Ethan Fahnstock, Felix Duvallet, Sachithra Hemachandra, Jean Oh, Anthony Stentz, Nicholas Roy, and Thomas M. Howard. "Language Understanding for Field and Service Robots in a Priori Unknown Environments." In *Field Robotics 2021*
- E. Fahnstock, **S. Patki**, and T.M. Howard, "Language-guided Adaptive Perception with Hierarchical Symbolic Representations for Mobile Manipulators," In *6th AAAI Fall Symposium Series on Artificial Intelligence for Human-Robot Interaction*. Nov. 2019
- A. Boteanu, J. Arkin, **S. Patki**, T.M. Howard, and H. Kress-Gazit, "Robot-Initiated Specification Repair through Grounded Language Interaction," In *AAAI Fall Symposium on Natural Communication for Human-Robot Collaboration*. Nov. 2017

Mentored Students

Ethan Fahnstock, Currently a PhD student at MIT

AUGUST 2018 – DECEMBER 2019

Nikola Raicevic, Currently a PhD student at UCSD

AUGUST 2021 – AUGUST 2022