Thomas F. Kollar

CONTACT INFORMATION

Address:
2001 St. Emilion Way, Mountain View, CA 94043

Voice: (865) 250-2317
E-mail: tkollar@pm.me

EDUCATION

MASSACHUSETTS INSTITUTE OF TECHNOLOGY (MIT), Cambridge, MA USA

Ph.D., Electrical Engineering and Computer Science **June 2011**

- Title: Learning to Understand Spatial Language for Robotic Navigation and Mobile Manipulation
- Area: Robotics, Grounded Language Acquisition, Human-Robot Interaction
- Committee: Nicholas Roy (Adviser), Leslie Kaelbling, Antonio Torralba, Dieter Fox

M.S., Electrical Engineering and Computer Science **May 2007**

- Title: Optimizing Robot Trajectories using Reinforcement Learning
- Area: Probabilistic Robotics
- Adviser: Nicholas Roy

UNIVERSITY OF ROCHESTER, Rochester, NY USA

B.S., Computer Science **June 2004**
B.A., Mathematics **June 2004**

- Minor: Psychology as a Social Science
- *Cum Laude* with highest honors
- Research Advisers: Professor Chris Brown and Professor Steven Gonek

EXPERIENCE

TOYOTA RESEARCH INSTITUTE, Los Altos, CA

Research Manager **September 2022 to present**

Senior Research Scientist **July 2018 to September 2022**

- **Overview**: I led a 13-15 person team with three primary sub-thrusts: to develop capable large language models, create vision-language models (with an eye toward embodied settings), and lay the groundwork for embodied foundation models (or robot foundation models).
- **Large Language Models**: Training was scaled to 7B, 34B and 70B parameter language models on a local DGX cluster and SageMaker, a 7B parameter Mamba large language model was released and a competitor to Llama-3 at 7B parameter scale was developed, which has yet to be released. In addition, active research topics included linear attention, scaling laws and RLAIF.
- **Vision-Language Models**: Performant models and a hackable codebase was released for building vision-language models, called Prismatic.
- **Embodied Foundation Models**: I have co-led and released the large scale DROID dataset and corresponding Vision-Language-Action models. I was involved in Open X-Embodiment through DROID's inclusion; the former won Best Paper at ICRA this year.
- **3D**: I have worked on 3D detection and reconstruction (including articulated objects) with SimNet, SHAPO and CARTO. Much of this was applied to home robotics.
- **Robotics**: I worked on the mobile manipulation team, including langauge-centric teaching and interactions, and on articulated and deformable object manipulation.

AMAZON INC., Sunnyvale, CA

Senior Research Scientist **April 2017 to April 2018**

Research Scientist **July 2015 to April 2017**

- Key areas of innovation include knowledge graphs, semantic parsing, neural network models for natural language understanding and hierarchical classification. I developed some of the initial Alexa domains, performed research in semantic parsing (via encoder-decoder style models), and was involved at the inception of the Astro home robot.

APPLE INC., Cupertino, CA

Research Scientist **July 2013 to July 2015**

- Developed machine learning algorithms for natural language understanding to improve the accuracy and functionality of Siri, including on-device language understanding. Key areas of innovation included structured prediction, natural language understanding, semantic parsing, domain classification, weakly supervised learning and data analytics.

CARNEGIE MELLON UNIVERSITY, Pittsburgh, PA

Postdoctoral Fellow **September 2011 to July 2013**

- Conducted research in grounded language acquisition, knowledge representation, multi-modal human-robot interaction and human-robot dialog. Designed and implemented machine learning algorithms and a human-robot dialog system to enable a service robot to understand speech commands. The approach, called Logical Semantics with Perception (LSP), used weakly-supervised training to learn to identify objects and relations in images from highly variable natural language phrases.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY (MIT), Cambridge, MA

Research Assistant **June 2004 to September 2011**

- Designed and implemented algorithms to enable a micro-air vehicle, a 6,000 lb. robotic forklift and a robotic wheelchair to understand and execute natural language commands. The software infrastructure included machine learning, natural language processing, computer vision, data visualization and feature libraries as well as a dataset that has been used by other research groups.

Teaching Assistant **Fall 2007**

- Instructor: Professor Patrick Winston
- Course: Artificial Intelligence (6.034)
- Instructed five weekly tutorials, graded lab reports, and wrote/graded exams.

Teaching Assistant **Spring 2007**

- Instructor: Professor Nicholas Roy
- Course: Real-Time Systems and Software (16.35)
- Held office hours, reviewed assignments, and graded bi-weekly assignments.

UNIVERSITY OF ROCHESTER, Rochester, NY

Undergraduate Researcher and Teaching Assistant **September 2002 to May 2004**

- Led a Undergraduate Robot Research team, which created a service robot that delivered hors-d'oeuvres to conference participants at AAAI.
- Courses: Created and lectured in a robotics recitation for an Artificial Intelligence course.

SERVICE WORK

HRI PIONEERS WORKSHOP AT THE 2011 CONFERENCE ON HRI, Lausanne, Switzerland

General Chair **Fall 2010-Spring 2011**

- Won an NSF grant (1115939) for student funding to bring together 30+ participants from 10+ countries around the topic of Human-Robot Interaction.

WORKSHOP ON GROUNDING HUMAN-ROBOT DIALOG FOR SPATIAL TASKS, Los Angeles, CA

Organizer **July 2011**

- Organized workshop that brought together researchers in perception, natural language understanding and dialog with the goal of building robust dialog systems for robots.

MIDDLE EAST EDUCATION THROUGH TECHNOLOGY (MEET), Jerusalem, Israel

Year 3 Program Director and Instructor **Summer 2007-Summer 2011**

- Directed the year 3 instructor team in the preparation of the summer curriculum and projects.
- Instructor for the third-year students in a program that brings Israeli and Palestinian youth together around a curriculum of Computer Science and business.
- Developed a web application to organize procedures for Nesher, an Israeli concrete company.

Patents or Patent Applications

- **Thomas Kollar**, "Motion guidance and natural language commands based robotic systems," No. 17/840,070, 2023.
- **Thomas Kollar**, Kevin Stone, Michael Laskey, Mark Edward Tjersland, "System and method for unknown object manipulation from pure synthetic stereo data," No. 17/839,193, 2023.
- **Thomas Kollar**, Kevin Stone, Michael Laskey, Mark Edward Tjersland, "System and method for 3d object perception trained from pure synthetic stereo data," No. 17/839,201, 2023.
- Muhammad Zubair Irshad, **Thomas Kollar**, Michael Laskey, Kevin Stone, "Systems and methods for single-shot multi-object 3d shape reconstruction and categorical 6d pose and size estimation," No. 17/895,224, 2023.
- Cengiz Erbas, **Thomas Kollar**, Avnish Sikka, Spyridon Matsoukas, Simon Peter Reavely, "Processing complex utterances for natural language understanding," Patent 11,410,646, 2022.
- Lambert Mathias, **Thomas Kollar**, Arindam Mandal, Angeliki Metallinou, "Fine-grained natural language understanding," Patent 10,304,444, 2019.

Journal and Magazine Publications

- Joseph A Vincent, Haruki Nishimura, Masha Itkina, Paarth Shah, Mac Schwager, **Thomas Kollar**, "How Generalizable Is My Behavior Cloning Policy? A Statistical Approach to Trustworthy Performance Evaluation." Submitted, 2024.
- Vittorio Perera, Robin Soetens, **Thomas Kollar**, Mehdi Samadi, Yichao Sun, Daniele Nardi, Rene van de Molengraft and Manuela Veloso, "Learning Task Knowledge from Dialog and Web Access," Robotics, 2015.
- Krishnamurthy, J. and T. Kollar, "Jointly Learning to Parse and Perceive: Weakly-Supervised Grounded Language Acquisition," The Transactions of the ACL, 2013.
- Tellex, S., **Kollar, T.**, Dickerson, S., Walter, M., Banerjee, A., Teller, S. and N. Roy, "Approaching the Symbol Grounding Problem with Probabilistic Graphical Models." AI Magazine, 2011.
- **Kollar, T.** and N. Roy, "Trajectory Optimization using Reinforcement Learning for Map Exploration." The International Journal of Robotics Research 27 (2), 175-196, February 2008.

Conference Publications

- Ludwig Schmidt, Vaishaal Shankar, Achal Dave, Jeffrey Li, Alex Fang, Georgios Smyrnis, Maor Ivgi, ... Jean Mercat, Igor Vasiljevic, Sedrick Keh, Kushal Arora, **Thomas Kollar** ..., "DataComp-LM: In search of the next generation of language model training sets," submitted to NeurIPS Datasets and Benchmarks Track, 2024.
- Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Fosterm Pannag Sanketi, Quan Vuong, Siyuan Feng, **Thomas Kollar**, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, Chelsea Finn, "OpenVLA: An Open-Source Vision-Language-Action Model," submitted to CORL 2024.
- Jean Mercat, Igor Vasiljevic, Sedrick Keh, Kushal Arora, Achal Dave, Adrien Gaidon, **Thomas Kollar**, "Linearizing Large Language Models," submitted to CoLM 2024.
- Archit Sharma, Sedrick Keh, Eric Mitchell, Chelsea Finn, Kushal Arora, **Thomas Kollar**, "A Critical Evaluation of AI Feedback for Aligning Large Language Models," submitted to NeurIPS, 2024.
- Alexander Khazatsky, Karl Pertsch, Suraj Nair, Ashwin Balakrishna, Sudeep Dasari, Siddharth Karamcheti, Soroush Nasiriany, Mohan Kumar Srirama, ... **Thomas Kollar**, Sergey Levine, Chelsea Finn, "Droid: A large-scale in-the-wild robot manipulation dataset," Robotics: Science and Systems, 2024.
- Siddharth Karamcheti, Suraj Nair, Ashwin Balakrishna, Percy Liang, Dorsa Sadigh, **Thomas Kollar**, "Prismatic vlms: Investigating the design space of visually-conditioned language models," ICML 2024.
- Samir Yitzhak Gadre, Georgios Smyrnis, Vaishaal Shankar, Suchin Gururangan, Mitchell Wortsman, Rulin Shao, Jean Mercat, Alex Fang, Jeffrey Li, Sedrick Keh, Rui Xin, Marianna Nezhurina, Igor Vasiljevic, Jenia Jitsev, Alexandros G Dimakis, Gabriel Ilharco, Shuran Song, **Thomas Kollar**, Yair Carmon, Achal Dave, Reinhard Heckel, Niklas Muennighoff, Ludwig Schmidt, "Language models scale reliably with over-training and on downstream tasks," submitted to NeurIPS, 2024.
- Open X Collaboration, **Thomas Kollar**, "Open X-Embodiment: Robotic learning datasets and RT-X Models," ICRA 2024. **Best Paper.**
- Aviv Adler, Ayah Ahmad, Shengyin Wang, Wisdom C Agboh, Edith Llontop, Tianshuang Qiu, Jeffrey Ichnowski, Mehmet Dogar, **Thomas Kollar**, Richard Cheng, Ken Goldberg, "The Teenager's Problem: Efficient Garment Decluttering With Grasp Optimization," on Arxiv.

- Vainavi Viswanath, Kaushik Shivakumar, Mallika Parulekar, Jainil Ajmera, Justin Kerr, Jeffrey Ichnowski, Richard Cheng, **Thomas Kollar**, Ken Goldberg, "HANDLOOM: Learned Tracing of One-Dimensional Objects for Inspection and Manipulation," CORL 2023.
- Lawrence Yunliang Chen, Baiyu Shi, Roy Lin, Daniel Seita, Ayah Ahmad, Richard Cheng, **Thomas Kollar**, David Held, Ken Goldberg, "Bagging by Learning to Singulate Layers Using Interactive Perception," IROS 2023. **Best Industrial Robotics Research for Applications Paper Nominee**.
- Muhammad Zubair Irshad, Sergey Zakharov, Katherine Liu, Vitor Guizilini, **Thomas Kollar**, Adrien Gaidon, Zsolt Kira, Rares Ambrus, "NeO 360: Neural Fields for Sparse View Synthesis of Outdoor Scenes," ICCV 2023.
- Nick Heppert, Muhammad Zubair Irshad, Sergey Zakharov, Katherine Liu, Rares Andrei Ambrus, Jeannette Bohg, Abhinav Valada, **Thomas Kollar**, "Carto: Category and joint agnostic reconstruction of articulated objects," ICCV 2023.
- Lawrence Yunliang Chen, Baiyu Shi, Daniel Seita, Richard Cheng, **Thomas Kollar**, David Held, Ken Goldberg, "Autobag: Learning to open plastic bags and insert objects," ICRA 2023.
- Kaushik Shivakumar, Vainavi Viswanath, Anrui Gu, Yahav Avigal, Justin Kerr, Jeffrey Ichnowski, Richard Cheng, **Thomas Kollar**, Ken Goldberg, "Sgtm 2.0: Autonomously untangling long cables using interactive perception," ICRA 2023.
- Siddharth Karamcheti, Suraj Nair, Annie S Chen, **Thomas Kollar**, Chelsea Finn, Dorsa Sadigh, Percy Liang, "Language-driven representation learning for robotics," Robotics: Science and Systems, 2023. **Best Paper Nominee**.
- Muhammad Zubair Irshad, Sergey Zakharov, Rares Ambrus, **Thomas Kollar**, Zsolt Kira, Adrien Gaidon, "Shapo: Implicit representations for multi-object shape, appearance, and pose optimization," ECCV 2022.
- Lawrence Yunliang Chen, Huang Huang, Ellen Novoseller, Daniel Seita, Jeffrey Ichnowski, Michael Laskey, Richard Cheng, **Thomas Kollar**, Ken Goldberg, "Efficiently learning single-arm fling motions to smooth garments," ISRR 2022.
- Muhammad Zubair Irshad, **Thomas Kollar**, Michael Laskey, Kevin Stone, Zsolt Kira, "Centersnap: Single-shot multi-object 3d shape reconstruction and categorical 6d pose and size estimation," ICRA 2022.
- **Thomas Kollar**, Michael Laskey, Kevin Stone, Brijen Thananjeyan, Mark Tjersland, "Simnet: Enabling robust unknown object manipulation from pure synthetic data via stereo," CORL 2022.
- Max Bajracharya, James Borders, Dan Helmick, **Thomas Kollar**, Michael Laskey, John Leichty, Jeremy Ma, Umashankar Nagarajan, Akiyoshi Ochiai, Josh Petersen, Krishna Shankar, Kevin Stone, Yutaka Takaoka, "A mobile manipulation system for one-shot teaching of complex tasks in homes," ICRA 2020.
- **Thomas Kollar**, Danielle Berry, Lauren Stuart, Karolina Owczarzak, Tagyoung Chung, Lambert Mathias, Michael Kayser, Bradford Snow, Spyros Matsoukas, "The Alexa meaning representation language," NAACL 2018.
- Vittorio Perera, Tagyoung Chung, **Thomas Kollar**, Emma Strubell, "Multi-task learning for parsing the alexa meaning representation language," AAAI 2018.
- Vittorio Perera, Robin Soetens, **Thomas Kollar**, Mehdi Samadi, Yichao Sun, Daniele Nardi, RenÃ© Van de Molengraft, Manuela Veloso, "Learning task knowledge from dialog and web access," Robotics, 2015.
- **Kollar, T.**, Krishnamurthy, J. and G. Strimel. "Toward Interactive Grounded Language Acquisition." Proceedings of Robotics: Science and Systems (RSS), 2013.
- Duvallet, F., **Kollar, T.** and T. Stentz, "Imitation Learning for Natural Language Direction Following Through Unknown Environments," Proceedings of the International Conference on Robotics and Automation (ICRA), 2013 **Best Cognitive Robotics Paper Nominee**.
- **Kollar, T.**, Perera, V., Nardi, D. and M. Veloso, "Learning Environmental Knowledge From Task-Based Human-Robot Dialog," Proceedings of the International Conference on Robotics and Automation (ICRA), 2013.
- **Kollar, T.**, Vedantham, A., Sobel, C., Chang, C., Perera, V. and M. Veloso, "A Multimodal Approach for Natural Human-Robot Interaction," Proceedings of the International Conference on Social Robots (ICSR), 2012.
- Samadi, M., Kollar T. and M. Veloso, "Using the Web to Interactively Learn to Find Objects," Proceedings of the 26th Conference on Artificial Intelligence (AAAI), 2012.
- Tellex, S., Thaker, P., Deits, R., Simeonov, D., **Kollar, T.** and N. Roy, "Toward Information Theoretic

Human-Robot Dialog," Proceedings of Robotics: Science and Systems (RSS), 2012.

- **Kollar, T.**, Tellex, S., Dickerson, S., Walter, M., Banerjee, A., Teller, S. and N. Roy, "Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation," Proceedings of the 25th Conference on Artificial Intelligence (AAAI), 2011.
- Hemachandra, S., **Kollar, T.**, Roy, N. and S. Teller, "Following and Interpreting Narrated Guided Tours," Proceedings of the International Conference on Robotics and Automation (ICRA), 2011.
- **Kollar, T.**, Tellex, S. and N. Roy, "Grounding Verbs of Motion in Natural Language Commands to Robots." Proceedings of the 12th International Symposium on Experimental Robotics (ISER), 2010.
- Tellex, S., **Kollar, T.**, Shaw, G., Roy, N. and D. Roy, "Grounding Spatial Language for Video Search." Proceedings of the Eleventh International Conference on Multimodal Interfaces (ICMI), 2010 **Best Student Paper**.
- Huang, A., Tellex, S., Bachrach, A., **Kollar, T.**, Roy, D. and N. Roy, "Natural Language Command of an Autonomous Micro-Air Vehicle," Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2010.
- **Kollar, T.**, Tellex, S., Roy, D. and N. Roy, "Toward Understanding Natural Language Directions," Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2010.
- Espinace, P., **Kollar, T.**, Soto, A. and N. Roy. "Indoor Scene Recognition Through Object Detection," Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2010.
- **Kollar, T.** and N. Roy. "Utilizing object-object and object-scene context when planning to find things," Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2009.
- Wei, Y., Brunskill, E., **Kollar, T.** and N. Roy, "Where to Go: Interpreting Natural Directions Using Global Inference". Proceedings of the International Conference on Robotics and Automation (ICRA), 2009.
- **Kollar, T.** and N. Roy, "Efficient optimization of information-theoretic exploration in SLAM." Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI), Physically Grounded AI track. pp. 1369-1375, 2008.
- Brunskill, E., **Kollar, T.** and N. Roy, "Topological Mapping Using Spectral Clustering and Classification." Proceedings of the International Conference on Intelligent Robots and Systems (IROS), pp. 3491-3496, 2007.
- Doshi, F., Brunskill, E., Shkolnik, A., **Kollar, T.**, Rohanimanesh, K., Tedrake, R. and N. Roy, "Collision Detection in Legged Locomotion using Supervised Learning." Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). San Diego, 2007.
- **Kollar, T.** and N. Roy, "Using Reinforcement Learning to Improve Exploration Trajectories for Error Minimization." Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 3338-3343, 2006.

WORKSHOP PUBLICATIONS

- **Kollar, T.**, Samadi, M. and M. Veloso, "Enabling Robots to Find and Fetch Objects by Querying the Web," the 11th International Conference on Autonomous Agents and Multiagent Systems, 2012 (Extended Abstract).
- **Kollar, T.**, Tellex, S. and N. Roy, "A Discriminative Model for Understanding Natural Language Route Directions", AAAI Fall Symposium Series, 2010.
- **Kollar, T.** et. al., "Mabel: Extending Human Interaction and Robot Rescue Designs", Journal of Undergraduate Research, v. 2, no. 2, pp. 9-13, 2004.
- **Kollar, T.** et. al., "Mabel: Extending Human Interaction and Robot Rescue Designs", AAAI Mobile Robot Competition Workshop, Acapulco, Mexico, TR WS-03-01, pp. 20-29, 2003.
- Schmid, J., **Kollar, T.** et. al, "Mabel: Building a Robot Designed for Human Interaction." AAAI Mobile Robot Competition Workshop, Edmonton, Alberta, TR WS-02-18, p.24-32, 2002.