

San Jose, CA

Voice: (865) 250-2317

E-mail: tkollar@pm.me

PROFILE

Research leader who builds embodied foundation models and the teams and data engines behind them. Head of AI (Foundation Models) at **Wayve**, leading the **Wayve foundation model** project. Previously built and led a 20-person organization at **Toyota Research Institute** that produced OpenVLA, DROID, Prismatic VLMs and Datacomp-lm - open releases that became default baselines across robot learning and multimodal pretraining. Earlier shipped learned language understanding into **Siri** and **Alexa**. MIT Ph.D.; 13,900+ citations; Best Paper, ICRA 2024; AAAI Classic Paper Award, 2026.

IMPACT

- **Organization building** - grew two research organizations **from 4–5 people to 20 (TRI) and 25 (Wayve) in roughly a year each**, both high-performing and both delivering the flagship results below.
- **Models in production** - my group's work is **Wayve's baseline driving model**, adopted as the company landed **Nissan, Stellantis, and a third major manufacturer** as customers. Earlier, learned models shipped in Siri (Apple) and Alexa (Amazon), reaching hundreds of millions of users.
- **OpenVLA** - co-led the first widely adopted open-source vision-language-action model. **3,600+ citations** in two years and the de facto baseline for the field; nominated for Best Paper, CORL 2024.
- **Open X-Embodiment** - **Best Paper, ICRA 2024**; 2,000+ citations. The cross-institution dataset standard for robot learning, seeded in part by DROID.
- **DROID** - co-led the largest in-the-wild robot manipulation dataset at release; 1,000+ citations and now a standard training corpus.
- **Prismatic VLMs / Datacomp-lm** - open, reproducible recipes for vision-language and language-model pretraining; 400+ and 500+ citations.
- **Portfolio** - 45+ peer-reviewed publications, **8 granted U.S. patents and 5 pending applications**, h-index 42, i10-index 64, the **2026 AAAI Classic Paper Award**, and six best-paper awards or nominations across ICRA, RSS, CORL, ICMI and IROS.

EXPERIENCE

WAYVE AI, Sunnyvale, CA

Head of AI (Foundation Models)

September 2024–present

- Lead the **Wayve foundation model** project, **grown from 5 to ~25 people in about a year** across five workstreams, one led through a senior manager. Lead all pretraining; contribute to post-training and RL.
- **Modeling** (KV caching, flow matching, long context), **data** (data curation at scale), **evaluation** (foundational pretraining evaluations), **language** (captioning, VLMs), and **spatial intelligence** (learned visual odometry, Wayve's best). Public team output includes **Rig3R (NeurIPS 2025)** and **LA-Pose (CVPR 2026)**.
- The group's data curation, the evaluations built around it, the language work, and the resulting model became Wayve's baseline driving model; the same models power search at Wayve and are the company's de facto embeddings.

TOYOTA RESEARCH INSTITUTE, Los Altos, CA

Research Manager

September 2022 to July 2024

Senior Research Scientist

July 2018 to September 2022

- Built the team **from 4–5 people to ~20 in about a year**, across three thrusts: large language models, vision-language models, and embodied foundation models.
- **Owned the compute strategy**: managed the training compute directly and made the on-prem versus AWS cloud tradeoff.
- **Embodied foundation models**: co-led DROID and the corresponding vision-language-action models, and drove TRI's contribution to Open X-Embodiment (Best Paper, ICRA 2024). OpenVLA released open source.
- **Language models**: released state-of-the-art language models, scaling training up to 70B parameters. Work in linear attention, scaling laws and RLAIIF.
- **Vision-language models**: released Prismatic, a performant and hackable codebase for VLMs, now a common starting point for VLA work.

- **3D and robotics:** SimNet, ShAPO and CARTO for 3D detection and reconstruction, applied to home robotics and mobile manipulation.

AMAZON INC., Sunnyvale, CA

Senior Research Scientist, Research Scientist

July 2015 to April 2018

- Developed some of the initial Alexa domains and authored the Alexa Meaning Representation Language, the semantic layer behind Alexa’s natural-language understanding. Research in semantic parsing, knowledge graphs and hierarchical classification; involved at the inception of the Astro home robot.

APPLE INC., Cupertino, CA

Research Scientist

July 2013 to July 2015

- Built machine learning for natural language understanding in Siri, including on-device language understanding, structured prediction, domain classification and weakly supervised learning.

CARNEGIE MELLON UNIVERSITY, Pittsburgh, PA

Postdoctoral Fellow

September 2011 to July 2013

- Grounded language acquisition and human-robot dialog. Built Logical Semantics with Perception (LSP), a weakly-supervised approach that let a service robot ground natural language in perception.

EDUCATION

- **Massachusetts Institute of Technology** - Ph.D. (2011) and M.S. (2007), Electrical Engineering and Computer Science. Thesis: *Learning to Understand Spatial Language for Robotic Navigation and Mobile Manipulation*. Adviser: Nicholas Roy.
- **University of Rochester** - B.S. Computer Science and B.A. Mathematics, 2004. *Cum Laude* with highest honors.

SELECTED PUBLICATIONS

45+ publications in total; the full list is in the complete CV.

- Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, . . . **Thomas Kollar**, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, Chelsea Finn, “OpenVLA: An Open-Source Vision-Language-Action Model,” CORL 2024. **Nominated for Best Paper**. 3,600+ citations.
- Open X-Embodiment Collaboration, **Thomas Kollar**, “Open X-Embodiment: Robotic learning datasets and RT-X models,” ICRA 2024. **Best Paper**. 2,000+ citations.
- Alexander Khazatsky, Karl Pertsch, . . . **Thomas Kollar**, Sergey Levine, Chelsea Finn, “DROID: A large-scale in-the-wild robot manipulation dataset,” RSS 2024. 1,000+ citations.
- 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. 400+ citations.
- Ludwig Schmidt, Vaishal Shankar, Achal Dave, . . . **Thomas Kollar** . . . , “Datacomp-lm: In search of the next generation of language model training sets,” NeurIPS Datasets and Benchmarks Track, 2024. 500+ citations.
- Tellex, S., **Kollar, T.**, et al., “Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation,” AAAI 2011. **AAAI Classic Paper Award, 2026**. 1,075 citations.

PATENTS

Eight granted U.S. patents and five pending applications: robot manipulation and 3D perception from synthetic data, natural-language robot control, video-language understanding, efficient attention architectures, and natural language understanding at Amazon. Full list in the complete CV.

SELECTED SERVICE AND AWARDS

- **AAAI Classic Paper Award, 2026** - honoring the most influential papers from AAAI 2011. **Best Paper**, ICRA 2024 (Open X-Embodiment). Best Paper nominee, CORL 2024 (OpenVLA) and RSS 2023 (Voltron); Best Cognitive Robotics Paper nominee, ICRA 2013; Best Student Paper, ICMI 2010.
- **General Chair**, HRI Pioneers Workshop, 2011 (won NSF grant 1115939). **Program Director and Instructor**, Middle East Education through Technology (MEET), Jerusalem, 2007–2011 - a program bringing Israeli and Palestinian youth together around a CS and business curriculum.