From Tools to Partners:
In-the-Wild Learning and Deployment of Human-Centered Robotics

For robots to perform meaningful tasks that improve people's lives, they must move beyond tool-like roles and instead collaborate with us as partners. However, the widespread deployment of robot partners comes with three key challenges: (1) Humans communicate their diverse intentions in various ways; (2) Interactions among multiple humans and their surrounding environment are complex and highly dynamic; (3) Human environments continually evolve beyond the training domains of robots. When collaborating with each other, humans can handle such variabilities with ease. **For robots to achieve human-like partnership, my research treats humans, the end-users and beneficiaries of robots, as essential stakeholders in robot learning.** In particular, robots need strong capabilities to interpret human behavior and adjust their actions to collaborate, while remaining adaptive to changes after deployment. In pursuit of this vision, my research is centered around **learning and deployment of robots that infer and support human needs, operate alongside human crowds, and adapt through interactions with humans.** My research is organized into the following three thrusts:

1. *Inferring human behavior from diverse observations:* To assist humans, robots must infer human needs by grounding diverse human expressions, such as speech and motion, in scene observations and robot capabilities. My work harnesses vision–language foundation models to infer a wide range of user intended tasks from diverse input modalities, leading to trustworthy assistive navigation and mobile manipulation systems that serve users with disabilities [1, 2, 3, 4, 5, 6, 7].
2. *Robust planning in crowded and cluttered environments:* To work alongside humans, robots must operate robustly among human crowds and cluttered obstacles. My work introduces a spatio-temporal graph for robots to reason about heterogeneous interactions among agents and obstacles. Based on the graph, my work proposes a principled approach to design structured policy networks, which demonstrate safe and proactive robot behavior in unstructured spaces [8, 9, 10, 11, 12, 13, 14].
3. *Data-efficient adaptation from in-situ user teaching:* Domain shifts in the open world require robots to adapt to new tasks and environments. Leveraging data which is natural and intuitive for non-experts to provide, such as speech instructions and human demonstrations, my work develops data-efficient adaptation algorithms that refine robot policies in-situ after deployment [15, 16, 17].

Impact: My work lays the foundation for robot partners that address real human needs and operate robustly in human spaces. My work was nominated as a **Best Paper Award Finalist** at the Conference on Robot Learning (CoRL) 2023. I was selected into the **RSS Pioneer** cohort, which features 33 of the world's top early-career researchers in robotics. I was selected as a **Rising Star in EECS**, which features 70 of the world's top early-career researchers in EECS.

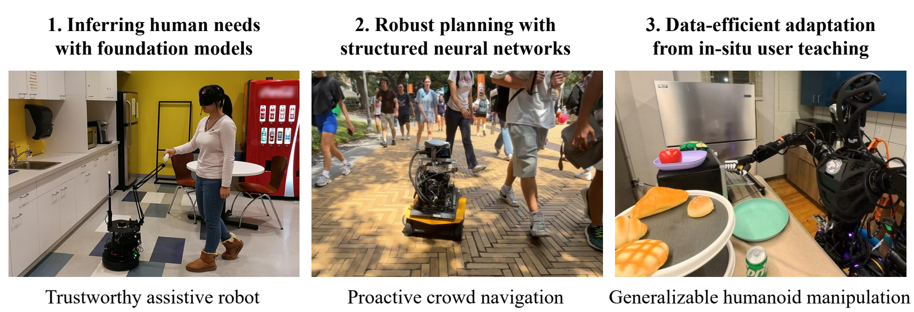

Fig. 1. Research overview: In-the-wild learning and deployment of robot partners.

Agenda: I aim to lead an agenda at the intersection of robotics and artificial intelligence, with a focus on developing algorithms and systems for capable and socially aware autonomy. Future research directions include expanding mobile manipulation capabilities for human-centered tasks, developing a data flywheel for lifelong learning of human values, and improving the explainability of robot behaviors.

Past and Ongoing Work

Thrust 1: Inferring Human Behavior from Diverse Observations.

When robots serve humans, they must cooperate with diverse human behaviors, which are subtle, difficult to categorize, and only partially observable. Rising to this challenge, my work builds effective and trustworthy human behavior models from diverse observations [1, 2, 3, 4, 5, 6, 7].

Prior work typically assumes a limited set of possible tasks that humans need robots to perform. Such assumptions restrict generalization when the desired tasks involve novel objects unseen during training or when humans use free-form expressions to communicate their needs. To address this limitation, for the first time, my work *harnesses the open-vocabulary perception, vision–language grounding, and commonsense reasoning capabilities of vision-language models (VLMs) to infer the needs of people with disabilities* who rely on specific modalities for communication (e.g. It is easier for blind users to say their needs rather than click on an unfamiliar graphic interface). Building on my algorithm for audio-commanded robots [4, 5], DRAGON uses a VLM to infer where the user wants to go by associating their language commands to semantic landmarks on a map [1]. My work integrates the VLM task inference module into a robotic sighted guide that leads blind people from place to place. Another work, CASPER, infers an intended task that a user with motor disabilities attempts to do from their teleoperation trajectories [2]. Casper is integrated into an assistive mobile manipulator that can transport objects across different rooms. *User studies show that our assistive robots are easy to use, reduce user mental and physical burdens, and preserve their decision-making authority.* As a result, my work not only unlocks new robotic applications for historically marginalized user groups, but also improves the trustworthiness of robots.

Another line of my work uncovers nuanced behavior patterns, such as aggressive or cooperative driving styles, from driver trajectories [3, 6]. For example, my work *learns latent representations of human traits using variational autoencoders without ground-truth trait labels* [3]. Compared with supervised learning, my work reduces the cost of labeling data and avoids oversimplification of nuanced human traits. The trait representation enables autonomous vehicles to navigate adaptively in T-intersections alongside different types of human drivers, even in the absence of traffic signals or stop signs. My work transforms abundant but unlabeled human data into knowledge essential for safe and efficient robot decision-making.

Thrust 2: Robust Planning in Crowded and Cluttered Environments.

Real-world environments contain abundant spatial and temporal interactions among agents and entities. These interactions are heterogeneous, dynamic, and difficult to reason about, making robot operation in such environments especially difficult. To address this problem, my work develops structured neural networks to learn robust robot policies in dynamic and human environments [8, 9, 10, 11, 12, 13, 14].

Prior work on robot crowd navigation often models these subtle interactions with simple heuristics or black-box neural networks. Although these approaches succeed in structured or low-variability settings, the robot exhibits unsafe or overly conservative behaviors when crowd densities increase and when layouts or pedestrian behaviors change. To address this problem, my work pioneers *a spatio-temporal graph as a formal representation of interactions among crowded agents and cluttered static obstacles* [8, 9, 10]. The spatio-temporal graph differentiates various types of interactions (robot–human, human–human, and obstacle–agent) that have different effects on robot navigation. From the graph, my work introduces *a principled approach to derive a modular yet end-to-end trainable robot policy network*: Attention networks assign more weights to spatial interactions that the robot needs to pay more attention to (e.g. a person approaching the robot), and a recurrent neural network reasons about historical interactions to encourage long-sighted decisions. Crucially, each type of interaction shares the same weights in the policy network, and thus the total number of parameters remains fixed regardless of the number of humans or obstacles.

This design makes the *structured robot policy robust to variations in crowd density and complexity of human behavior without sacrificing scalability.* Our method demonstrates strong zero-shot generalization in simulation when crowd densities and layouts shift, and has been *successfully deployed in both indoor and outdoor environments with real pedestrians.* This result demonstrates the power of injecting structures into neural networks. By doing so, complex multi-agent problems are decomposed into smaller components, making them easier to solve. Our work has spurred a new line of research that expands the use of graph attention networks in interactive navigation tasks.

Besides crowd navigation, I have also investigated modeling physical interactions between objects and robots in challenging manipulation tasks such as stowing and bi-manual coordination [13, 14]. Together, these contributions illustrate the effectiveness of reasoning about interactions among agents and entities for robust robot decision-making in complex real-world environments.

Thrust 3: Data-Efficient Adaptation from In-Situ User Teaching.

Robots deployed in the real world inevitably face domain shifts: Layouts are rearranged, objects are replaced, and users introduce new tasks. However, most adaptation algorithms are not human-centered: they rely on data intentionally collected by hired workers and often overlook what is natural and easy for non-experts to provide. In contrast, everyday interactions between robots and their users during deployment generate abundant data that contain rich cues about human values. *My work exploits in-situ teaching of non-expert users as a scalable source of training data, enabling the robot to generalize its skills under distribution shifts in the real world [15, 16, 17].*

My research explores two complementary approaches that enable data-efficient post-deployment generalization. First, to enable instant adaptation without retraining, MimicDroid allows a humanoid to acquire new manipulation skills from just a few demonstrations using in-context learning (ICL) [16]. For the first time, MimicDroid leverages *human play videos as a scalable and diverse training data source* that can be easily collected from non-experts and naturally covers a wide range of tasks, making it ideal for data-hungry ICL policies. At test time, the robot instantly learns to manipulate unseen objects in novel environments with only one to three human demonstrations. Second, to customize robot behaviors for a specific target domain, Diff-VAR develops a data-efficient finetuning method for instruction following robots [15]. When the domain shift happens, users can use their smartphones to provide photos and audio to finetune a visual-audio representation. Since the finetuned representation generates reinforcement learning (RL) reward for robot skills, the robot can then *use RL to self-improve its policy in the real world without any demonstration data*. After approximately an hour of RL finetuning, the robot is able to pick up correct objects among many based on user instructions in a novel real-world environment.

In summary, these contributions estabilish a new robot learning paradigm from data that can be easily obtained from everyone without expertise or specialized equipment. My work democratizes robot learning by reducing data collection costs and improving scalability, paving the way for long-term autonomy in human-centered environments.

Future Directions

Looking ahead, I envision **a human-centered robot learning paradigm that empowers capable and socially aware robot partners to serve people in their daily lives.** Advancing human-centered robot learning requires expanding robot capabilities in human-centric tasks, training robots with data that reflect human values, and fostering trustworthy human–robot communication. Thus, my future research will invest in the following three directions.

Whole-Body Control from Human Priors: Mobile manipulators (MoMA) can operate across extended workspaces and are well-suited for human-centered tasks. However, existing systems often decouple navigation and manipulation, leading to slow and awkward motions that hinder robot collaboration with people. To enable smoother coordination, it is crucial to learn whole-body skills that control the arm and base simultaneously. To this end, I plan to leverage human prior knowledge, such as spatio-temporal graphs, to decompose whole-body control into local controllers of the arms, the base, and the head. Building on my work on structured policy network [10], we can train whole-body skills to solve tasks such as opening a door by reasoning about spatial and temporal coordination of all local-controllers effectively.

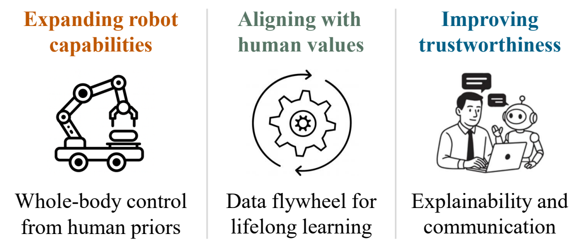

Fig. 2. Future research: Human-centered robot learning for capable and socially aware robot partners.

Data Flywheel for Human-Robot Alignment: Many misalignments between robot behaviors and human expectations only become apparent after real-world deployment, making lifelong learning important for long-term autonomy. Most robot learning methods use curated expert data, which fails to capture the diverse and constantly evolving human values. In contrast, data from people's natural interactions with robots, such as "I prefer the robot to grasp the toy from its body, not from its head," encodes individualized preferences that can only be obtained from end-users in real life. Building on my previous work on in-situ user teaching [15, 16], I will transform these daily robot-human interactions into a self-sustaining data flywheel for lifelong learning. Using algorithms such as reinforcement learning from human feedback and in-context learning, we can continuously refine robot skills and simulators for skill learning in a lifelong manner. As an initial step, *I am co-leading an NSF proposal for a research center on human–robot co-adaptation,* which will further support my work in this direction.

Explainable Robot Behavior through Communication: To collaborate with humans in an effective and trustworthy manner, robots must communicate their confidence, limitations, and intentions in ways that humans can intuitively understand. Building on my previous work on two-way language communication protocols in assistive navigation [1] and interpretable attention graph networks in crowd navigation [10], I will explore the following two directions: (1) To improve the transparency of robot behaviors and facilitate failure detection, I will equip robots with chain-of-thought reasoning capability, so that robots can express their reasoning process to human users. This can be done by methods such as probing the hidden layers of robot policy networks. (2) I will explore uncertainty quantification methods such as conformal prediction and anomaly detection to reveal robot confidence during task execution, so that users can understand the limitations of robots and take early actions to avoid severe failures.

Funding: There are numerous potential funding sources that align with my research. Within NSF, the Foundational Research in Robotics (FRR) program supports research on robotic systems with high computational and physical complexity, while the Cyber-Physical Systems (CPS) and Robust Intelligence (RI) programs provide opportunities for developing intelligent and interactive robot systems. In addition, programs in Army Research Laboratory (ARL), such as Artificial Intelligence for Maneuver and Mobility (AIMM), support adaptive autonomy in dynamic environments. Besides government funds, I will seek industry collaborations and funding from organizations such as Amazon, Google, and Samsung.

Summary Statement: I am a researcher who bridges robot learning and human-centered robotics by enabling robots to understand, adapt, and collaborate with people in the real world. I aim to lead research toward a future where trustworthy robot partners improve our productivity and enrich our lives.

References

[1] **S. Liu**, A. Hasan, K. Hong, R. Wang, P. Chang, Z. Mizrachi, J. Lin, D. L. McPherson, W. A. Rogers, and K. Driggs-Campbell, "Dragon: A dialogue-based robot for assistive navigation with visual language grounding," *IEEE Robotics and Automation Letters*, 2024.

[2] H. Liu, R. Shah, **S. Liu**, J. Pittenger, M. Seo, Y. Cui, Y. Bisk, R. Martín-Martín, and Y. Zhu, "Casper: Inferring diverse intents for assistive teleoperation with vision language models," in *Conference on Robot Learning (CoRL)*, 2025.

[3] **S. Liu**, P. Chang, H. Chen, N. Chakraborty, and K. Driggs-Campbell, "Learning to navigate intersections with unsupervised driver trait inference," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

[4] P. Chang, **S. Liu**, H. Chen, and K. Driggs-Campbell, "Robot sound interpretation: Combining sight and sound in learning-based control," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[5] P. Chang, **S. Liu**, and K. Driggs-Campbell, "Learning visual-audio representations for voice-controlled robots," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[6] H. Chen, T. Ji, **S. Liu**, and K. Driggs-Campbell, "Combining model-based controllers and generative adversarial imitation learning for traffic simulation," in *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2022.

[7] M. J. Munje, C. Tang, **S. Liu**, Z. Hu, Y. Zhu, J. Cui, G. Warnell, J. Biswas, and P. Stone, "Socialnav-sub: Benchmarking vlms for scene understanding in social robot navigation," in *Conference on Robot Learning (CoRL)*, 2025.

[8] **S. Liu**, P. Chang, W. Liang, N. Chakraborty, and K. Driggs-Campbell, "Decentralized structural-rnn for robot crowd navigation with deep reinforcement learning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[9] **S. Liu**, P. Chang, Z. Huang, N. Chakraborty, K. Hong, W. Liang, D. Livingston McPherson, J. Geng, and K. Driggs-Campbell, "Intention aware robot crowd navigation with attention-based interaction graph," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[10] **S. Liu**, H. Xia, F. C. Pouria, K. Hong, N. Chakraborty, Z. Hu, J. Biswas, and K. Driggs-Campbell, "Height: Heterogeneous interaction graph transformer for robot navigation in crowded and constrained environments," *arXiv preprint arXiv:2411.12150*, 2024.

[11] Y.-J. Mun, M. Itkina, **S. Liu**, and K. Driggs-Campbell, "Occlusion-aware crowd navigation using people as sensors," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[12] Z. Hu, C. Tang, M. J. Munje, Y. Zhu, A. Liu, **S. Liu**, G. Warnell, P. Stone, and J. Biswas, "Composablenav: Instruction-following navigation in dynamic environments via composable diffusion," in *Conference on Robot Learning (CoRL)*, 2025.

[13] H. Chen, Y. Niu, K. Hong, **S. Liu**, Y. Wang, Y. Li, and K. R. Driggs-Campbell, "Predicting object interactions with behavior primitives: An application in stowing tasks," in *Conference on Robot Learning (CoRL)*, 2023.

[14] H. Chen, J. Xu, L. Sheng, T. Ji, **S. Liu**, Y. Li, and K. Driggs-Campbell, "Learning coordinated bimanual manipulation policies using state diffusion and inverse dynamics models," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2025.

[15] P. Chang, **S. Liu**, T. Ji, N. Chakraborty, K. Hong, and K. R. Driggs-Campbell, "A data-efficient visual-audio representation with intuitive fine-tuning for voice-controlled robots," in *Conference on Robot Learning (CoRL)*, 2023.

[16] R. Shah, **S. Liu**, Q. Wang, Z. Jiang, S. Kumar, M. Seo, R. Martín-Martín, and Y. Zhu, "Mimicdroid: In-context learning for humanoid manipulation from human play videos," *arXiv preprint arXiv:2509.09769*, 2025.

[17] H. Chen, C. Zhu, **S. Liu**, Y. Li, and K. R. Driggs-Campbell, "Tool-as-interface: Learning robot policies from observing human tool use," in *Conference on Robot Learning (CoRL)*, 2025.