My research develops lifelong machine learning (ML) models and algorithms that enable robots to accumulate knowledge over long-term deployments by leveraging modularity and compositionality. Home assistant robots that handle a breadth of household chores must have a broad range of capabilities, such as fetching cutlery, moving furniture, navigating a teenager's cluttered bedroom, and cleaning countertops. These skills should also be specialized for individual users' varied needs, preferences, and environments, which will moreover likely *change over time*. Consequently, robots will need the ability to *adapt over time by learning from data*. But most ML is anything but cumulative: typical supervised methods train on a fixed dataset and remain frozen in time, and most reinforcement learning (RL) methods focus on one fixed reward function.

My work has developed the key insight that the forms of modularity that power robotics solutions simplify the problem of lifelong (or continual) learning. The predominant model for lifelong learning is monolithic: one neural net that contains all knowledge [16]. While successful in non-lifelong *stationary* settings, where the system makes predictions on the same distribution that generated the training dataset, monolithic approaches struggle to identify which weights remain relevant when the distribution shifts. This leads to *catastrophic forgetting*: new gradient updates overwrite previous knowledge [20]. While existing works have sought to address this shortcoming, a general solution is not within sight. Even worse, avoiding forgetting would not (necessarily) permit leveraging the accumulated knowledge to accelerate the learning of new tasks.

Modularity offers robots in such *nonstationary* settings three key benefits: 1) each part of the model encodes a self-contained unit of knowledge, which the agent can use and update individually; 2) the agent can introduce new modules to handle broader sets of problems; and 3) the agent can combine modules in novel ways to solve new problems compositionally.

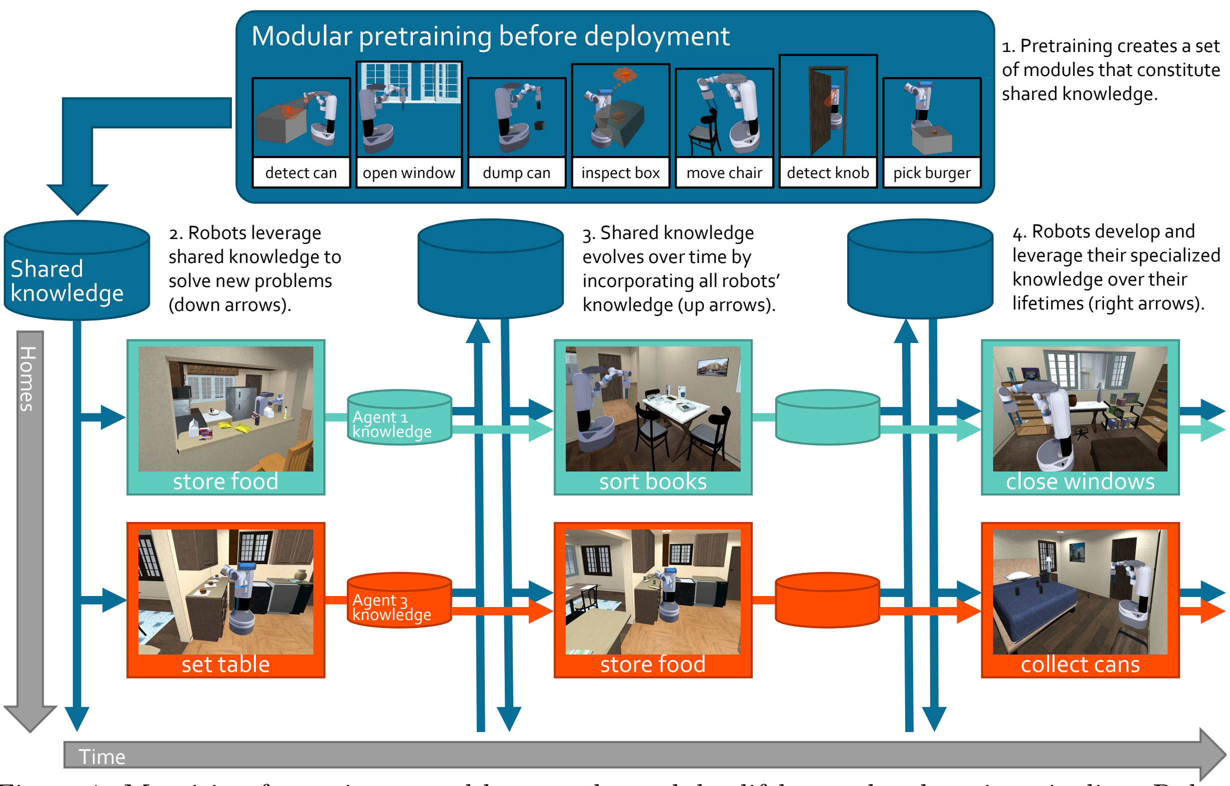

Figure 1: My vision for an integrated large-scale modular lifelong robot learning pipeline. Robots will pretrain a collection of perceptual, reasoning, and motor modules on massive simulated and real data. After deployment, the robots will continue refining and expanding their knowledge over time with their own data and data from other robots, while preserving user specialization and privacy.

My research has leveraged this insight to develop lifelong learning methods for supervised learning, RL, and a novel learning setting for task and motion planning. I have contributed more broadly to areas including deep learning, multitask learning, vision, and language. My work has been the first to permit robots to improve their capabilities as they face dozens of diverse tasks in sequence, including in the challenging Meta-World [23] and BEHAVIOR [22] benchmarks. **My future work will develop lifelong learning algorithms that bring us closer to deploying a robot in every household, starting from a competent and safe system and progressively expanding each robot's capabilities by leveraging both local and global data (Figure 1).** This will require fundamental advances, including pretraining methods that facilitate future lifelong updates and distributed learning techniques that maintain both user specialization and privacy.

1 PRIOR WORK

My work has made contributions toward lifelong robot learners in these three major directions.

1.1 Deep learning architectures and algorithms for modular lifelong learning

To build mechanisms that learn continually from varied data, we will need a fundamental understanding of the ways ML agents can decompose knowledge into pieces that combine in multiple ways to solve many tasks [6]. My research developed a framework—agnostic to the representational choice for each module, the mechanism to compose them, and the method to retain knowledge over time—that discovers compositional structures in a lifelong supervised setting [5]. The agent faces tasks in sequence (e.g., classify digits, then clothes, then bird species), and incrementally determines the number of modules, the modules themselves, and how to combine them to solve all tasks.

The key idea is to split the training process for each new task into 1) an assimilation stage that determines how to best compose existing modules, and 2) an accommodation stage that improves the chosen modules given the current composition of modules and adds any new modules needed to solve the task. An evaluation of 14 algorithms, 3 architectures, and 9 datasets demonstrated a **relative improvement in accuracy of up to 82.5% over non-modular lifelong approaches**. This work was the **first to study the intersection of lifelong and compositional learning**.

1.2 Compositional problem graph for lifelong learning in robots

To transfer these benefits to robotics, my work has more explicitly investigated what forms of modularity arise in robotics tasks. One key result is the *compositional problem graph*, which enables engineers to specify the relationships among robot problems [10]. We can view the behavior of a robot as a sequence of functional transformations from observation to action (e.g., detect object poses $\rightarrow$ determine next gripper pose to achieve goal $\rightarrow$ actuate motors to reach target gripper pose). We can then construct modules for pose detection of various objects, trajectory planning for various goals, and motor control for various robots, and combine them in multiple ways to solve many tasks. The nodes in the graph are the modules, while the paths are solutions to the tasks.

Following this graph perspective, my work developed an RL evaluation benchmark of 256 diverse robotics tasks (e.g., throw a box in the trash or place a plate on the shelf) [12, 4]. This benchmark enables the community to evaluate RL algorithms for *compositional generalization*, which will become increasingly important as we employ RL to solve a wide range of problems. My research developed an approach that uses small neural nets to represent each node in the graph and trains the overall graph end-to-end, **achieving up to 240% higher success rate than baselines and solving 80% of the *unseen* tasks** by constructing new paths in the graph [12].

My research also created a lifelong RL method within the framework of §1.1. The agent faces tasks sequentially, training one path in the graph at a time and collecting knowledge in the nodes [10]. On difficult 2D benchmarks, this method outperforms per-task training with **6.67× less data.**

1.3 Lifelong learning of long-horizon robot behaviors in task & motion planning

One downside of the functional composition approach in §1.2 is that current RL methods focus primarily on short-horizon problems. To handle longer-horizon tasks that require many steps (such as household chores), we need to complement the *functional* modularity discussed in §1.2 with *temporal* modularity. Task and motion planning (TAMP) [19] possesses both properties. It uses symbolic programs (i.e., software) as modules executed sequentially in *function space* (e.g., perceive objects, plan a collision-free path, control the motors to follow that path). TAMP further exploits temporal modularity to build long-horizon plans composed of high-level actions executed sequentially *in time* (e.g., go to an object, pick it up, go to a target location, place the object).

One key challenge in deploying TAMP-based robots is the difficulty of engineering such systems. To reduce this burden and enable robots to improve over time, my work developed the **first lifelong learning method for TAMP**. The approach learns a *continuous parameter generator* to interface between high-level actions (e.g., grasp spoon) and low-level motor execution (e.g., move gripper to this pose). The idea is to train a generative model to produce candidate parameters, and then use the TAMP system to check the feasibility of executing the action with the chosen parameters. Concretely, the learner uses *diffusion* models for diverse types of inputs (e.g., grasp large boxes vs. small spoons) by creating a mixture distribution that leverages both large-scale data from diverse types and small-scale data from individual types. This method solved a sequence of **10 families of tasks from the yet-unsolved BEHAVIOR benchmark** [22], tackling new tasks progressively more efficiently. While this work applied to TAMP specifically, mixtures of specialized and generic distributions would be useful for settings where some process (automated or manual) will *verify* the generated outputs (e.g., topology optimization for various heat transfer devices), which requires both broad coverage (because we have little data from each individual type or device) and specialization.

1.4 Additional lines of work

I have also worked on lifelong policy optimization [11 – best paper at Lifelong ML @ ICML-20, 1], inverse RL [9], theory of transfer and multitask learning [13, 14], open-world learning [3, 2], and RL for dialog [7]. I will continue exploring broad topics and their use for long-term robot deployments.

2 ONGOING AND FUTURE WORK

This section details three lines of research that I plan to pursue along with my future students, which I envision will transform our capabilities in lifelong robot learning.

2.1 Expanding single-robot lifelong learning

There are numerous opportunities to improve the ability of robots to accumulate knowledge. One major line of work is to more broadly explore the interplay between TAMP and ML. Developing ever-improving TAMP systems would result in robots that can robustly and safely handle a range of complex user requests in unstructured environments, without the need for vast amounts of data but leveraging large data when available. Learning can enhance TAMP systems in two key ways: improving their efficiency and broadening the scope of problems that they can solve. Directions to increase planning efficiency include further improving parameter generators, learning domain-specific heuristics, and learning *bypass policies* via RL for recurring chains of skills (e.g., move → pick → move → place). Solving a broader set of problems requires *discovering new skills*, because it is infeasible to design a set of skills that covers the spectrum of problems a robot will ever face.

My *current* work is taking steps in both directions. Master's students I supervise are extending generative models to yield parameterized actions that not only succeed at the current step (e.g., go to the fridge) but also set up the robot for future actions (e.g., open the fridge vs. push the fridge to inspect the wiring). I am also exploring the learning of a new skill from one single demonstration.

The main difficulties are learning in what situations the robot can execute the skill (the *preconditions*) and how to execute the skill (the *controller*). My research is developing the first approach that learns *preconditions* of a new TAMP skill from one demonstration in complex environments with many objects. The challenge is that the robot only observes the state in which the demonstrator executed the new skill, but not which aspects were necessary (e.g., when cleaning a plate, the demonstrator was holding a sponge and a spoon was in the sink). My work leverages inductive biases (e.g., the spoon is unlikely to be significant for cleaning the plate) to explore possible precondition sets when facing new problems *without demonstrations*. Preliminary results demonstrate that the approach finds increasingly accurate precondition sets, leading to shorter plans that avoid unnecessary setup (e.g., placing a spoon in the sink). Future work to learn the *controller* could use meta-learning to learn from small data, or alternatively leverage the fact that robots require a relatively small set of motion primitives (navigation, grasping, placing, contact motion . . .) and reduce the learning problem to identification, parameterization, sequencing, and RL tuning of known primitives.

2.2 Large-scale modular pretraining for future lifelong adaptability

Pretraining on large datasets has revolutionized generalization in language [15] and vision [21]. However, current pretraining approaches are not compatible with the continual adaptation paradigm of a lifelong learning robot. While recent works have explored how to update pretrained models, they typically rely either on finetuning (which is expensive and prone to forgetting) or prompting (which is not cumulative). I believe that the key to leveraging large-scale data is to design pretraining methods that themselves prepare the model for future continual updates. One promising avenue is to pretrain large *modular* models, such that the agent can add new modules later and improve existing modules via targeted updates. My prior work has demonstrated that (functional) modules generalize better as they are trained on more combinations with other modules [10, 12], which makes large-scale modular pretraining a promising avenue. We could pretrain modules (e.g., parameter generators, high-level actions, or perception models), compose these modules in a TAMP system (which is compositional by design), and improve the system over time via §2.1 during deployment.

2.3 Sharing modular knowledge across robots

Another means to scale up a robot's capabilities is to accumulate knowledge across multiple robots in multiple homes. This would enable one robot to leverage novel capabilities learned by another robot. Two central challenges to overcome are 1) maintaining specialization to each user's needs, preferences, environment, and robot hardware, and 2) preserving the privacy of each user's data. Regarding specialization, in future work I plan to explore extensions to the compositional problem graph of §1.2 to handle the variations across users (e.g., by incorporating user specialization nodes). In terms of data privacy, one major challenge is that differential privacy (a strong guarantee that bad actors cannot reverse-engineer a model to discover *any* information about an individual user) degrades as actors make more queries to the algorithm that generates the model [18]. The never-ending nature of the lifelong setting implies that there is no upper bound on the number of queries. We need new theory to understand the implications of lifelong learning on privacy and the resulting performance trade-offs. Modularity could plausibly enhance privacy guarantees without catastrophic performance loss, by decomposing models into privacy-sensitive and privacy-insensitive modules.

Leveraging modularity to develop ML methods that accumulate knowledge over time will enable continually deployed robots to become increasingly versatile. The algorithms I have developed have allowed robots to learn for long periods, becoming increasingly proficient at achieving complex goals. If successful, my future lines of work will result in a large-scale, integrated system that leverages engineering, pretraining, and lifelong multiagent training to progressively improve the capabilities of a collection of robots. See `bit.ly/LifelongRobotsBlog` for a deeper dive into these points.

References to my work

[1] M. M. Baker, A. New, M. Aguilar-Simon, Z. Al-Halah, S. M. R. Arnold, E. Ben-Iwhiwhu, A. P. Brna, E. Brooks, R. C. Brown, Z. Daniels, A. Daram, F. Delattre, R. Dellana, E. Eaton, H. Fu, K. Grauman, J. Hostetler, S. Iqbal, C. Kent, N. Ketz, S. Kolouri, G. Konidaris, D. Kudithipudi, E. Learned-Miller, S. Lee, M. L. Littman, S. Madireddy, **JMM**, E. Q. Nguyen, C. D. Piatko, P. K. Pilly, A. Raghavan, A. Rahman, S. K. Ramakrishnan, N. Ratzlaff, A. Soltoggio, P. Stone, I. Sur, Z. Tang, S. Tiwari, K. Vedder, F. Wang, Z. Xu, A. Yanguas-Gil, H. Yedidsion, S. Yu, and G. K. Vallabha. A domain-agnostic approach for characterization of lifelong learning systems. *Neural Networks*, 160, 2023.

[2] A. Ejilemele and **JMM**. Continual improvement of threshold-based novelty detection. *arXiv:2309.02551*, 2023.

[3] M. Gummadi, C. Kent, **JMM**, and E. Eaton. SHELS: Exclusive feature sets for novelty detection and continual learning without class boundaries. In *CoLLAs*, 2022.

[4] M. Hussing*, **JMM***, A. Singrodia, C. Kent, and E. Eaton. Robotic manipulation datasets for offline compositional reinforcement learning. *arXiv:2307.07091*, 2023.

[5] **JMM** and E. Eaton. Lifelong learning of compositional structures. In *ICLR*, 2021.

[6] **JMM** and E. Eaton. How to reuse and compose knowledge for a lifetime of tasks: A survey on continual learning and functional composition. *TMLR*, 2023.

[7] **JMM**, A. Geramifard, M. Ghavamzadeh, and B. Liu. Reinforcement learning of multi-domain dialog policies via action embeddings. In *3rd Conversational AI Workshop at NeurIPS*, 2019.

[8] **JMM**, L. P. Kaelbling, and T. Lozano-Pérez. Embodied lifelong learning for task and motion planning. In *CoRL-23*, 2023.

[9] **JMM**, S. Shivkumar, and E. Eaton. Lifelong inverse reinforcement learning. In *NeurIPS*, 2018.

[10] **JMM**, H. van Seijen, and E. Eaton. Modular lifelong reinforcement learning via neural composition. In *ICLR*, 2022.

[11] **JMM**, B. Wang, and E. Eaton. Lifelong policy gradient learning of factored policies for faster training without forgetting. In *NeurIPS*, 2020.

[12] **JMM***, M. Hussing*, M. Gummadi, and E. Eaton. CompoSuite: A compositional reinforcement learning benchmark. In *CoLLAs*, 2022.

[13] B. Wang, **JMM**, M. Cai, and E. Eaton. Transfer learning via minimizing the performance gap between domains. In *NeurIPS*, 2019.

[14] B. Wang, **JMM**, C. Shui, F. Zhou, D. Wu, G. Xu, C. Gagné, and E. Eaton. Gap minimization for knowledge sharing and transfer. *JMLR*, 24(33), 2023.

References to the work of others

[15] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language models are few-shot learners. In *NeurIPS*, 2020.

[16] R. Caruana. Multitask learning. *Machine Learning*, 28, 1997.

[17] O. X.-E. Collaboration et al. Open X-Embodiment: Robotic learning datasets and RT-X models, 2023.

[18] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4), 2014.

[19] C. R. Garrett, R. Chitnis, R. Holladay, B. Kim, T. Silver, L. P. Kaelbling, and T. Lozano-Pérez. Integrated task and motion planning. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1), 2021.

[20] M. McCloskey and N. J. Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. volume 24 of *Psychology of Learning and Motivation*. 1989.

[21] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning transferable visual models from natural language supervision. In *ICML*, 2021.

[22] S. Srivastava, C. Li, M. Lingelbach, R. Martín-Martín, F. Xia, K. E. Vainio, Z. Lian, C. Gokmen, S. Buch, K. Liu, S. Savarese, H. Gweon, J. Wu, and L. Fei-Fei. BEHAVIOR: Benchmark for everyday household activities in virtual, interactive, and ecological environments. In *CoRL*, 2021.

[23] T. Yu, D. Quillen, Z. He, R. Julian, K. Hausman, C. Finn, and S. Levine. Meta-World: A benchmark and evaluation for multi-task and meta reinforcement learning. In *CoRL*, 2019.