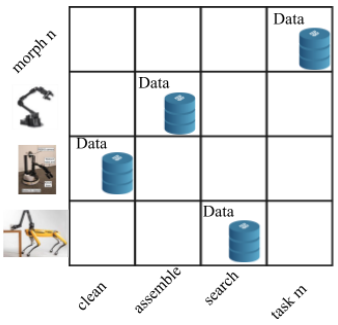
Dr. Glen Berseth

Sample-Efficient and Robust, Multi-Task Learning in the Real World

The grand challenge in creating useful real-world agents (robots) is to construct them such that they can solve a breadth of tasks when provided with unstructured commands. This challenge hinges on the (1) generalizability of the trained agent, (2) the agent's ability to search and find more optimal solutions and (3) the capability to interact and learn complex skills in the real world.

Aim 1: Generalization in Decision Making:

Training large models over increasing amounts of data is proving to be a recipe for success in creating better generalizable systems and is recently being adapted to reinforcement learning tasks [1–5] for increasing skills via multi-task training. However, these methods still struggle to exploit the collected experience properly, showing large gaps between the model's performance and the best data in the dataset [6,7]. **How can we ensure our learning systems are extracting the best knowledge from the available data?** Developing methods that can generalize across tasks, goals, data modes, and morphology to enable generalist agents that work in novel environments/tasks can be accomplished by learning to be invariant to task-unrelated changes and by training a model to combine portions of information in the data. This ability to combine experience is a type of combinatorial generalization that is achieved by piecing together experience without needing to see the desired exact complete trajectory.



Creating this multi-task model may sound onerous, but this model creation challenge is a blessing in disguise, as it allows us to **create a model that can use data from any robot on any task** to increase the dataset size by an order of magnitude. **This goal is akin to building a single main policy that benefits from data from many sources to create a better model, similar to the highly successful progression of training language translation models across more languages to increase performance.** Other potential benefits are better use of existing robot data and scaling training, which is currently a bottleneck for robot learning compared with LLMs.

Aim 2: Discovering New Knowledge:

Artificial intelligence is in the process of accelerating science, but struggles with the vast combinatorial options to evaluate. A promising solution in this area is to start from good pretrained models that encapsulate the knowledge of many experts and use them with advanced exploration algorithms to make discoveries to add to our collective knowledge. A key tool in this space is to expand on the capabilities of reinforcement learning methods that have shown they can discover masterful policies on complex planning problems in Go [8], and with the help of LLMs, can learn mathematics [9]. However, there are two hurdles to success in this space: (1) exploration is complex, and (2) even if the agent explores well, deep learning models struggle to learn under non-stationary distributions [6]. To overcome these issues, my research aims to scale deep reinforcement learning algorithms to larger models while coping with distribution change [10–12] and to increase exploration methods for RL agents [13,14], including methods for molecule discovery [15].

Aim 3: Learning Skilled Behaviours Autonomously

What objective function is necessary to incentivize an agent to become a multi-skilled generalist? While recent work has been able to extract reward functions from LLMs [16,17], where do LLMs, or the people who typed up all the data that LLMs train on, get their reward functions? Current algorithms still do not yet result in agents that learn diverse skills due to poor world models and limitations on learning optimal policies. My research improves diverse skill learning by building on surprise minimization methods and connects these objectives to physical and information-theoretic measures to outline the principled behaviour the agent should learn [18–20]. These information-theoretic connections of these methods allow us to understand the expected optimal behaviour, which is not typically well understood for the average extrinsic reward function. In addition, to create agents that will learn in the real world, it is best to find objectives that can be computed locally on embodied hardware.

References:

1. Reed, S. *et al.* A Generalist Agent. *TMLR* (2022).
2. Jang, E. *et al.* BC-Z: Zero-shot task generalization with robotic imitation learning. *CoRL* **abs/2202.02005**, (2022).
3. Brohan, A. *et al.* RT-1: Robotics Transformer for real-world control at scale. *arXiv [cs.RO]* (2022).
4. RT-2: New model translates vision and language into action. *Google DeepMind* https://www.deepmind.com/blog/rt-2-new-model-translates-vision-and-language-into-action?utm_source=twitter&utm_medium=social&utm_campaign=rt2.
5. Bousmalis, K. *et al.* RoboCat: A self-improving generalist agent for robotic manipulation. *arXiv [cs.RO]* (2023).
6. Nikishin, E., Schwarzer, M., D'Oro, P., Bacon, P.-L. & Courville, A. The Primacy Bias in Deep Reinforcement Learning. in *Proceedings of the 39th International Conference on Machine Learning* (eds. Chaudhuri, K. et al.) vol. 162 16828–16847 (PMLR, 17--23 Jul 2022).
7. Seohong, P., Kevin, F., Benjamin, E. & Sergey, L. OGBench: Benchmarking Offline Goal-Conditioned RL. *arXiv [cs.LG]* (2024).
8. Silver, D. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
9. AI achieves silver-medal standard solving International Mathematical Olympiad problems. *Google DeepMind* https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/.
10. Berseth, G., Xie, C., Cernek, P. & Van de Panne, M. Progressive reinforcement learning with distillation for multi-skilled motion control. *arXiv [cs.LG]* (2018).
11. Berseth, G., Zhang, Z., Zhang, G. & Finn, C. CoMPS: Continual Meta Policy Search. *Conference on Learning ...* (2021).
12. Tang, H. & Berseth, G. Improving deep Reinforcement Learning by reducing the chain effect of value and policy churn. *arXiv [cs.LG]* (2024).
13. Castanyer, R. C., Romoff, J. & Berseth, G. Improving Intrinsic Exploration by Creating Stationary Objectives. arXiv: 2310.18144 (2024).
14. Jain, A., Lehnert, L., Rish, I. & Berseth, G. Maximum state entropy exploration using predecessor and successor representations. *Neural Inf Process Syst* **abs/2306.14808**, (2023).
15. Ghugare, R., Miret, S., Hugessen, A., Phielipp, M. & Berseth, G. Searching for high-value molecules using reinforcement learning and transformers. *ICLR 2024* **abs/2310.02902**, (2023).
16. Ma, Y. J. *et al.* Eureka: Human-Level Reward Design via Coding Large Language Models. in *The Twelfth International Conference on Learning Representations* (2023).
17. Klissarov, M. *et al.* Motif: Intrinsic Motivation from Artificial Intelligence Feedback. *Models for Decision ...* (2023).
18. Berseth, G. *et al.* SMiRL: Surprise Minimizing Reinforcement Learning in Unstable Environments. *arXiv [cs.LG]* (2019).
19. Rhinehart, N. *et al.* Information is power: Intrinsic control via information capture. *Adv. Neural Inf. Process. Syst.* **34**, 10745–10758 (2021).
20. Hugessen, A., Castanyer, R. C., Mohamed, F. & Berseth, G. Surprise-adaptive intrinsic motivation for unsupervised reinforcement learning. *arXiv [cs.LG]* (2024).