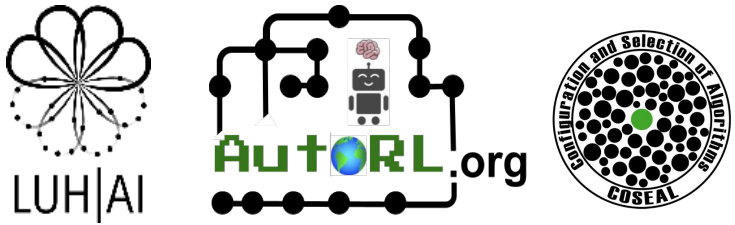
RL Algorithms Are Misbehaved Black Boxes

Theresa Eimer



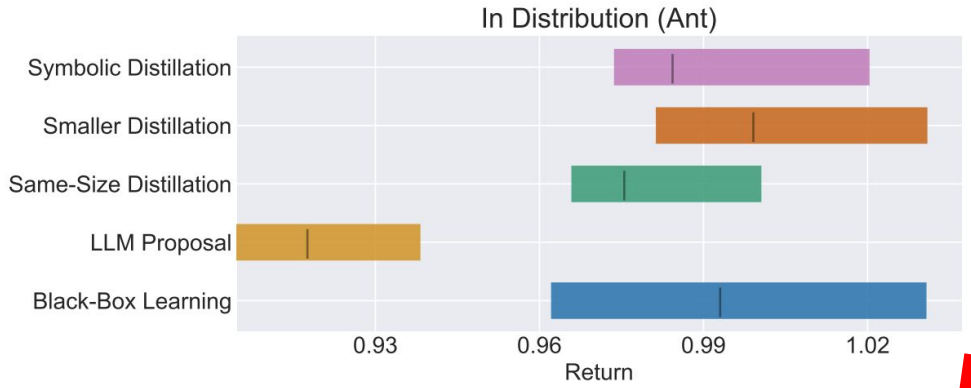
What is a Black Box and Why Would I Care?

What is a Black Box and Why Would I Care?

- Environment
- Model
- Hyperparameters
- ...

RL Algorithms As Black Boxes

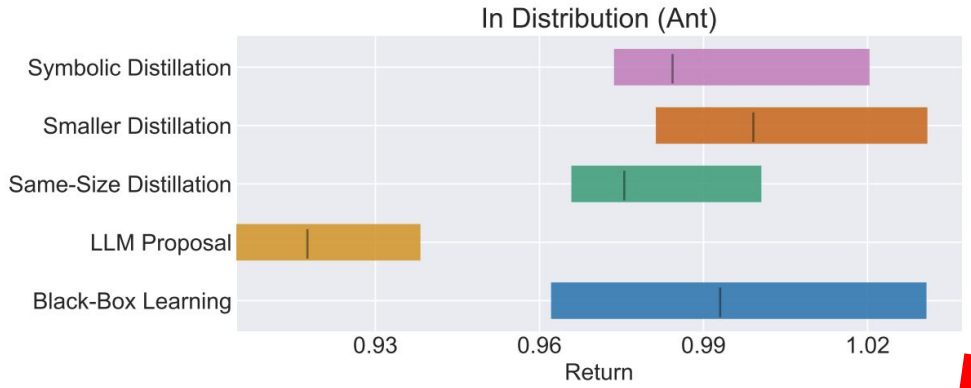
Meta-Learning Algorithms [Goldie et al. 2025]



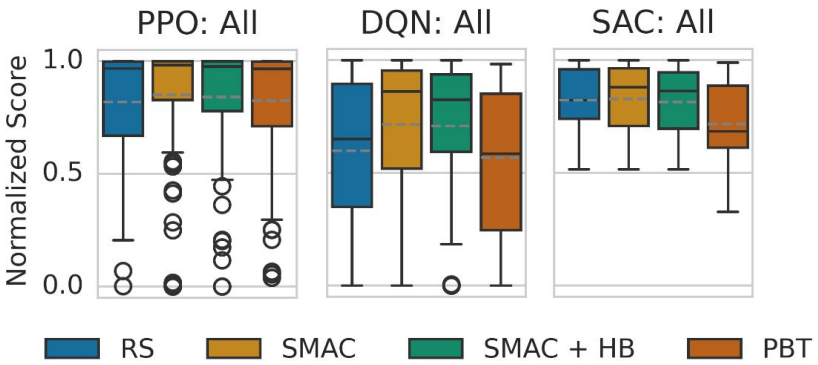
Outperformed by Distillation

RL Algorithms As Black Boxes

Meta-Learning Algorithms [Goldie et al. 2025]



HPO [Becktepe & Dierkes et al. 2024]

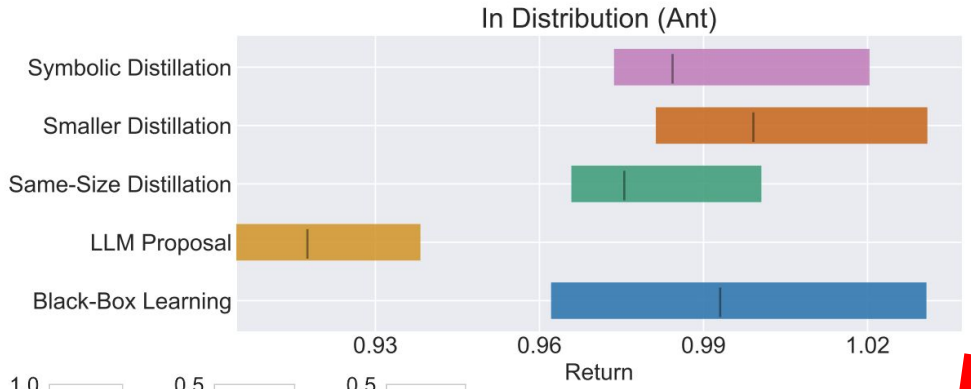


Outperformed by Distillation

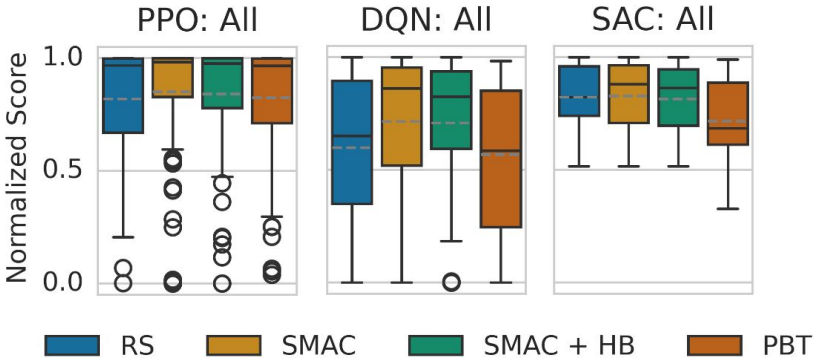
Too close to Random Search

RL Algorithms As Black Boxes

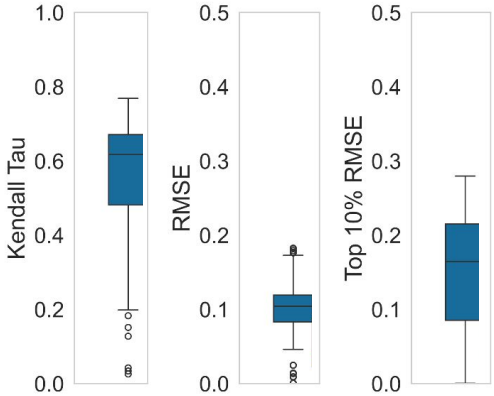
Meta-Learning Algorithms [Goldie et al. 2025]



HPO [Becktepe & Dierkes et al. 2024]



Performance Prediction [Dierkes et al. 2025]



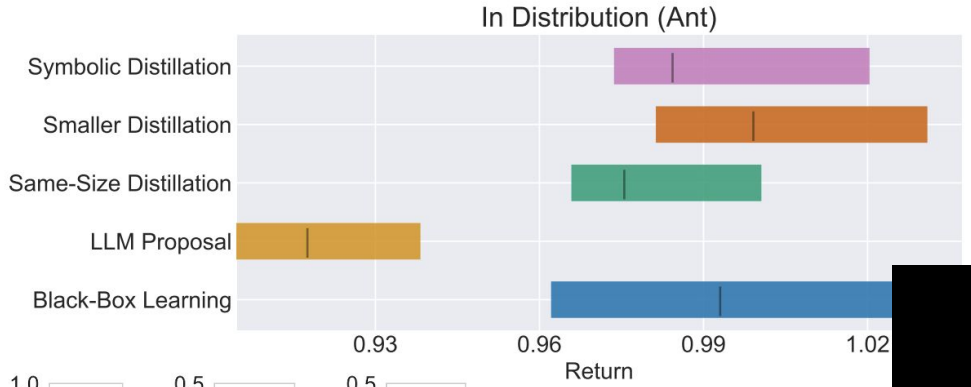
Outperformed by Distillation

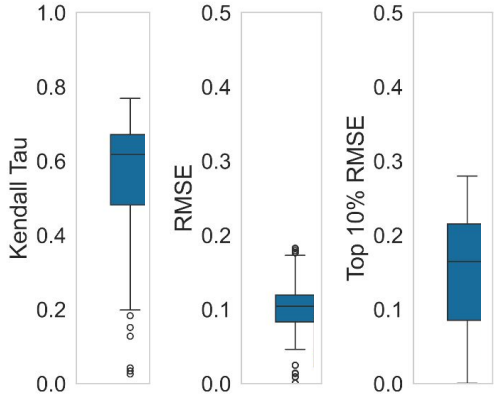
Too close to Random Search

Terrible Scores

RL Algorithms As Black Boxes

Meta-Learning Algorithms [Goldie et al. 2025]

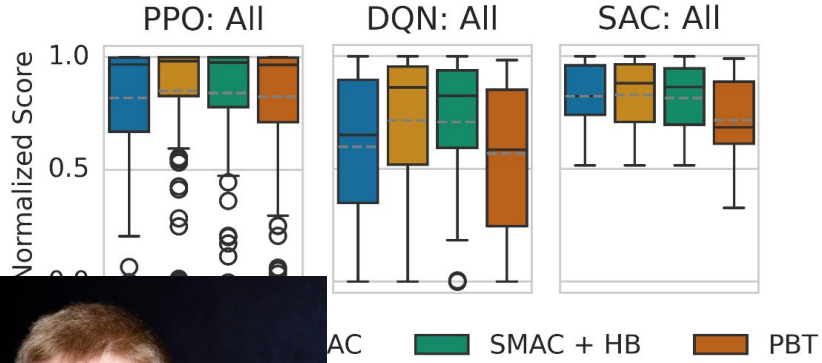




Performance Prediction [Dierkes et al. 2025]

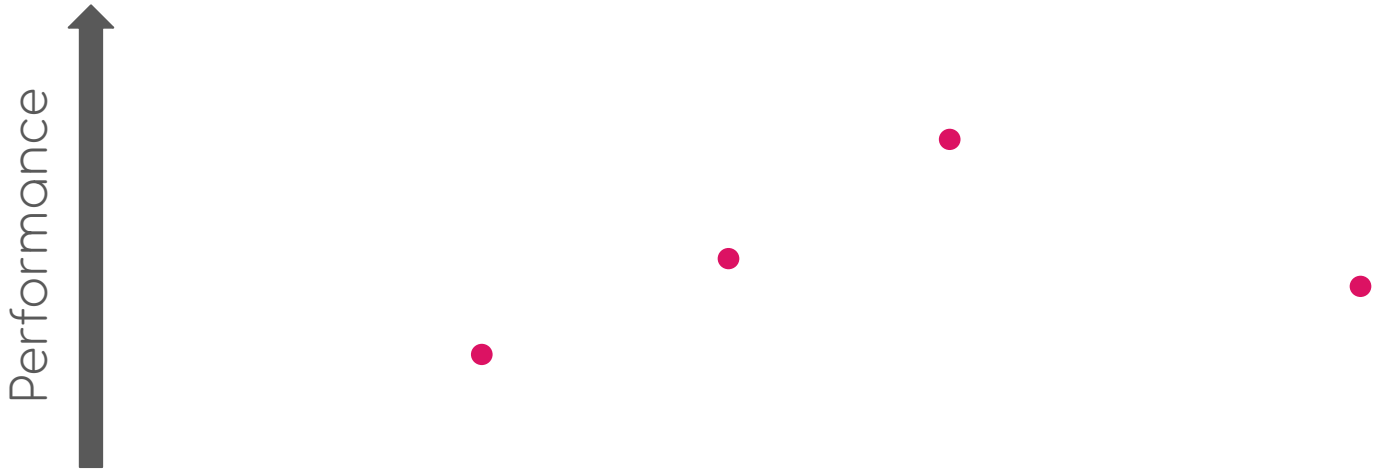
Terrible S

HPO [Becktepe & Dierkes et al. 2024]

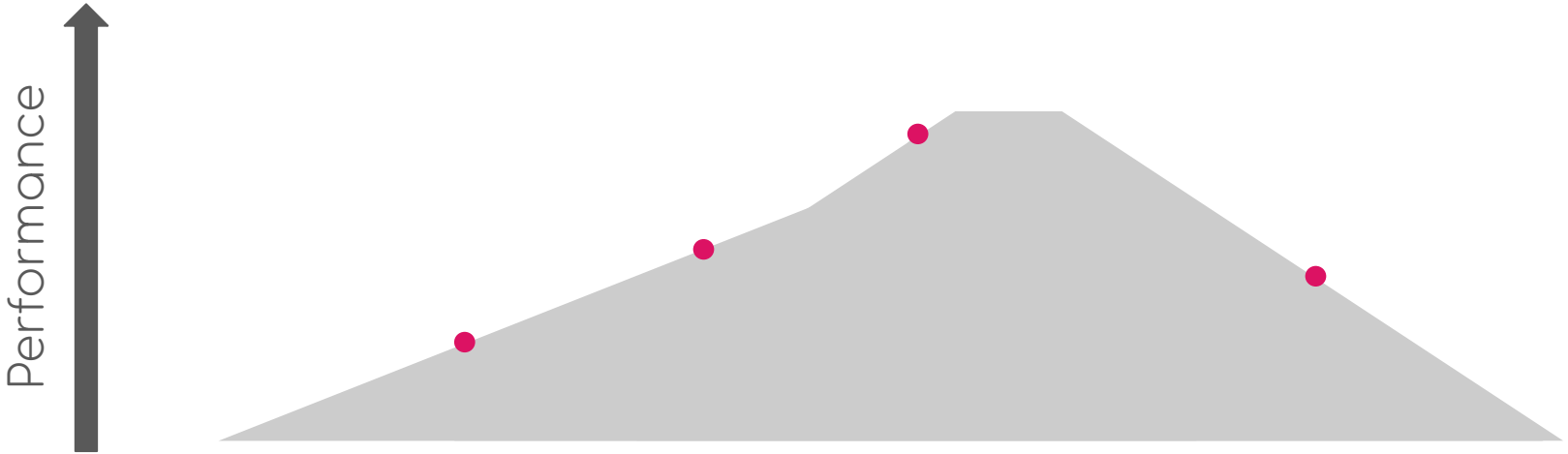


Too close to Random Search

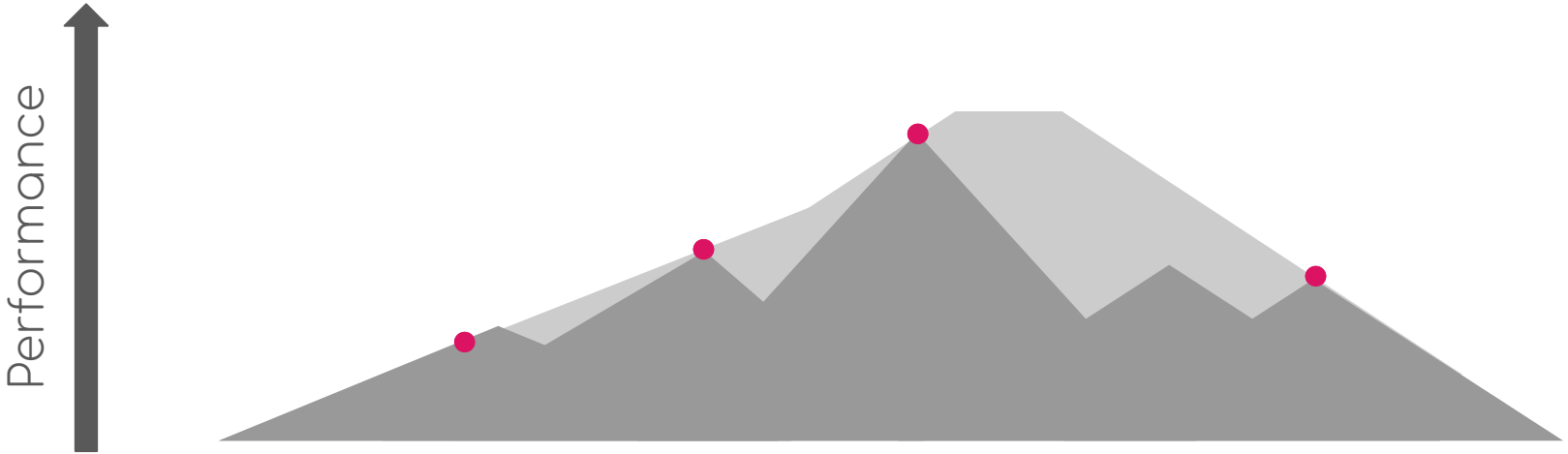
Explanation Attempt 1: Rugged Landscapes?



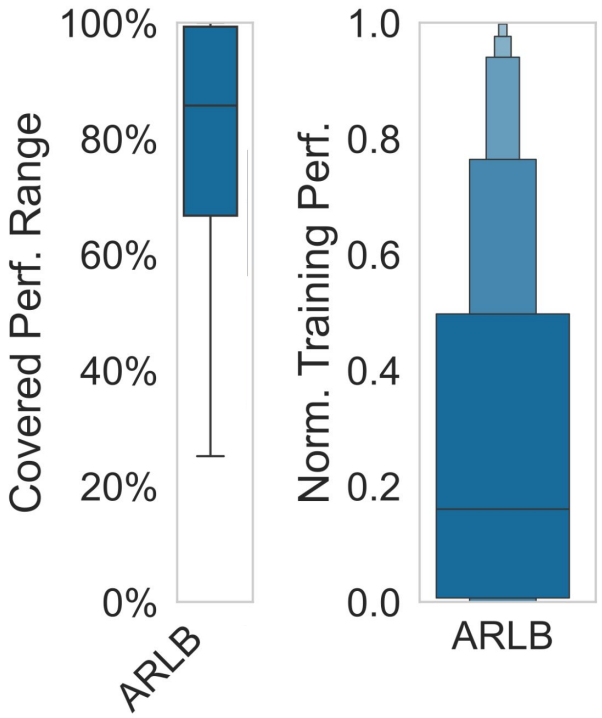
Explanation Attempt 1: Rugged Landscapes?



Explanation Attempt 1: Rugged Landscapes?

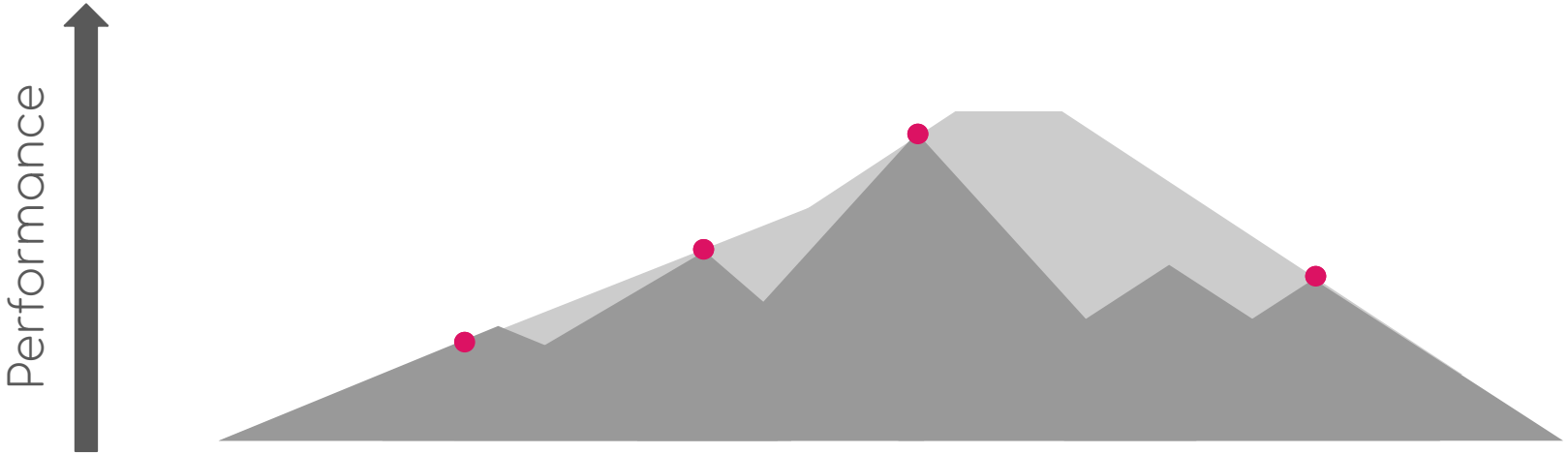


Performance Across Hyperparameters

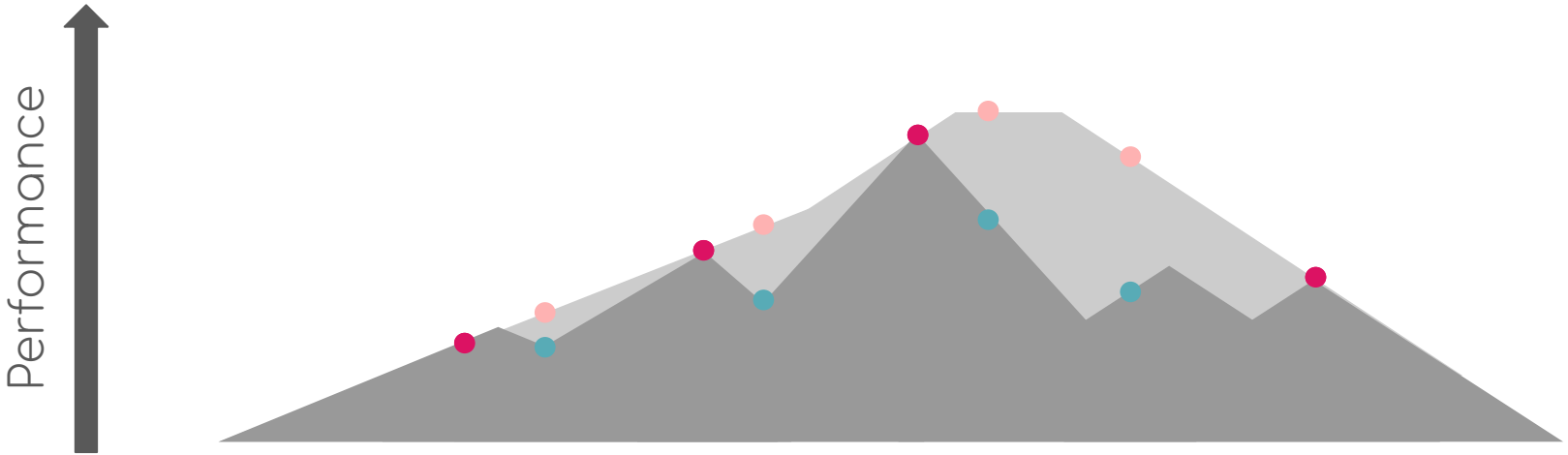


- ARLBench [Becktepe & Dierkes et al. 2024] Dataset
- DQN, PPO and SAC
- 21 Environments in total
- 256 randomly sampled hyperparameter configurations
- 10 seeds

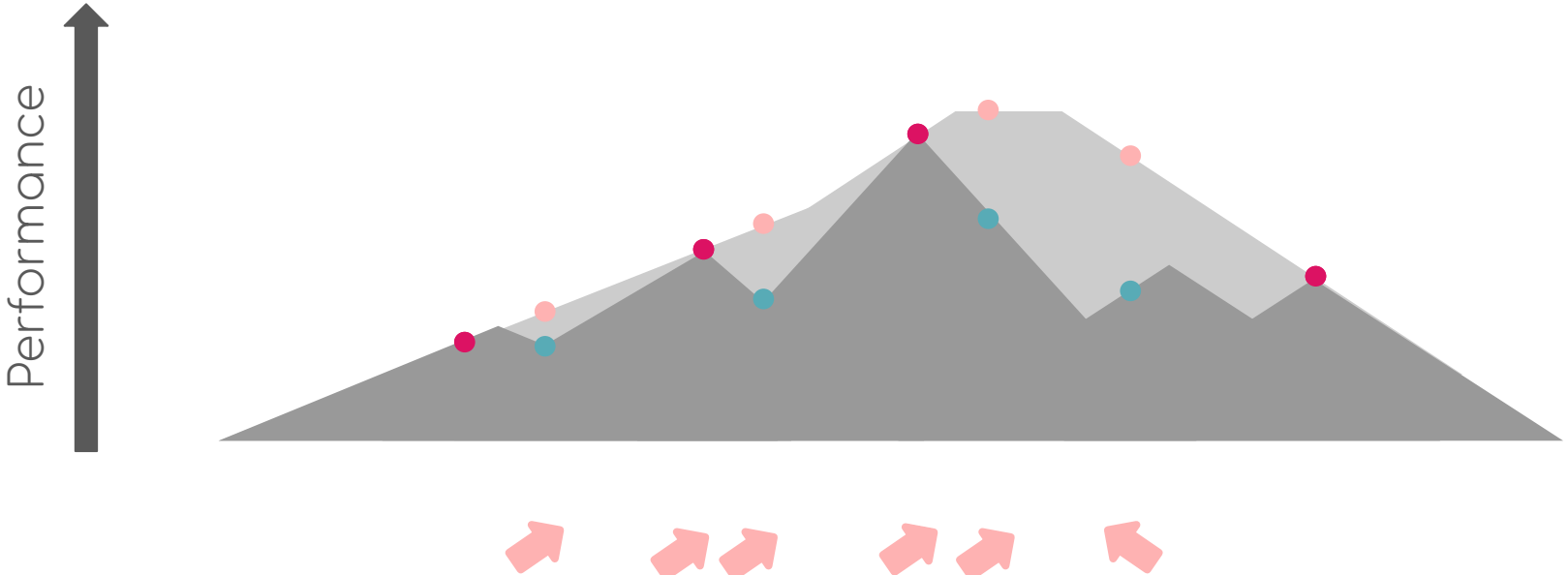
Unimodality



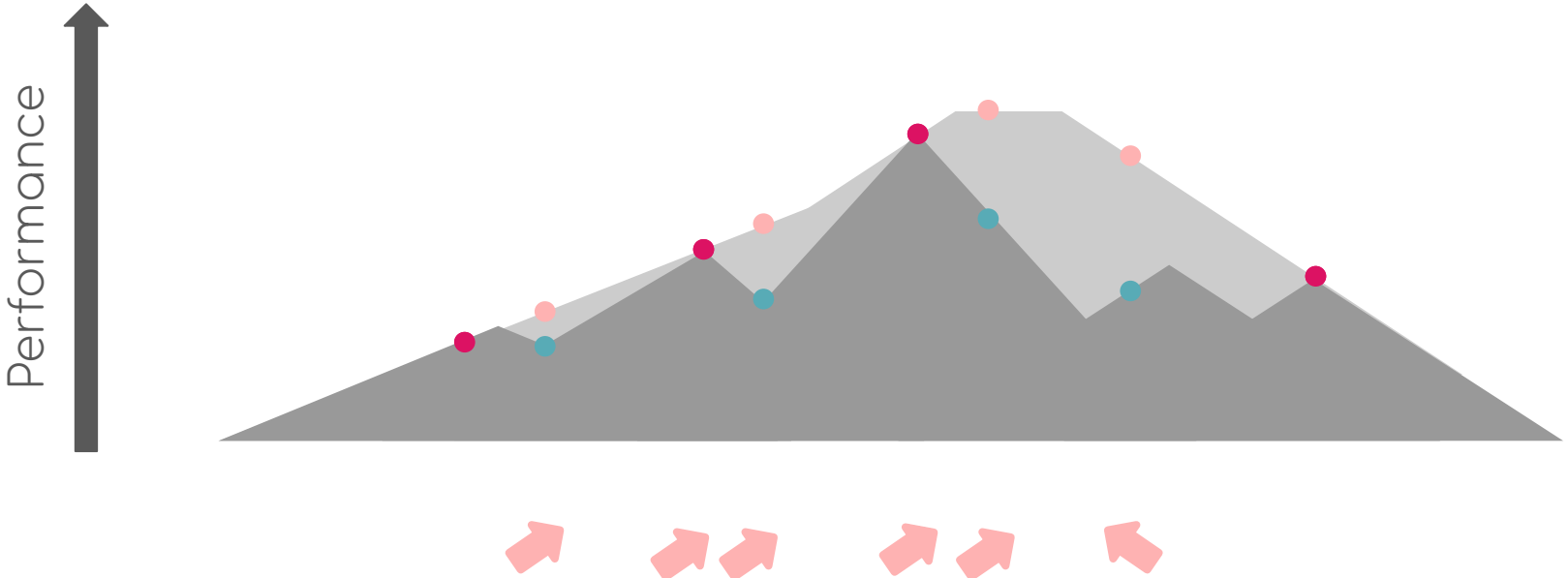
Unimodality



Unimodality

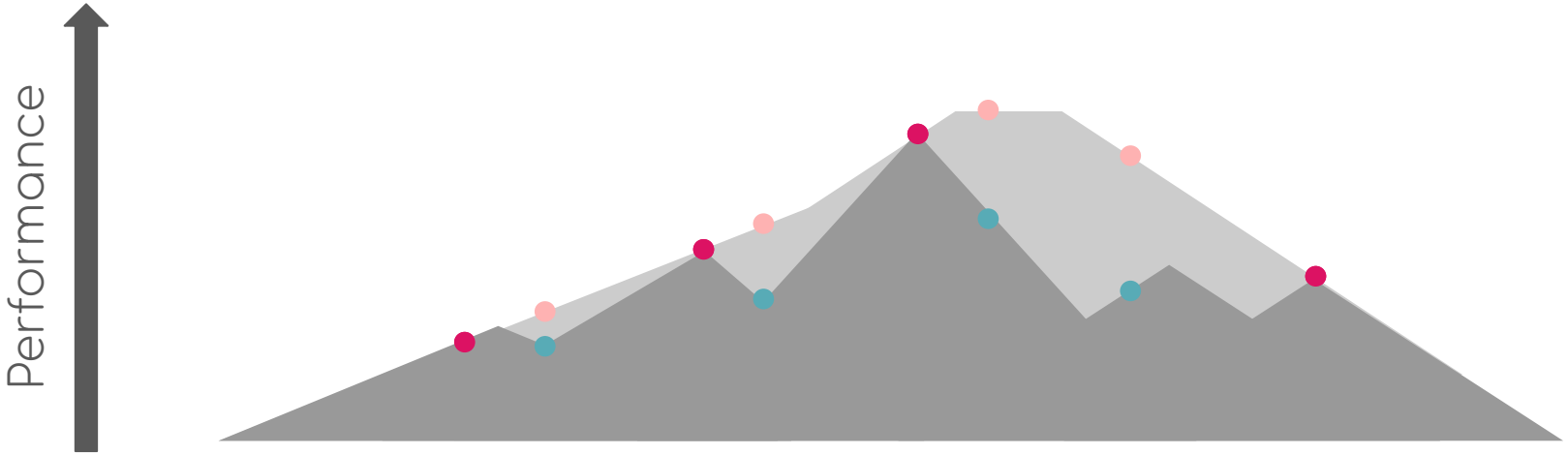


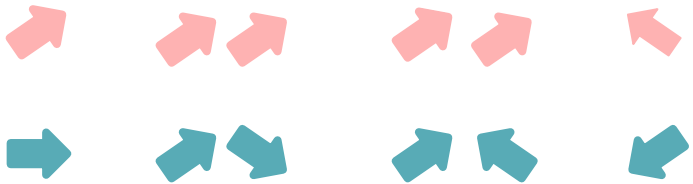
Unimodality



We can reach the optimum from every point with an increasing path!

Unimodality



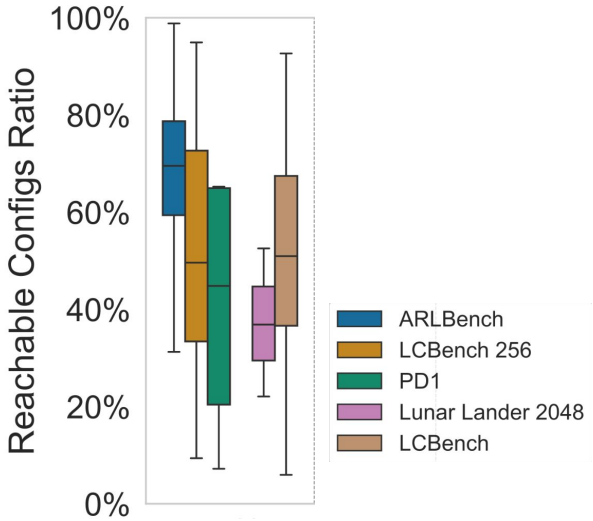


100% reachability ratio

29% reachability ratio

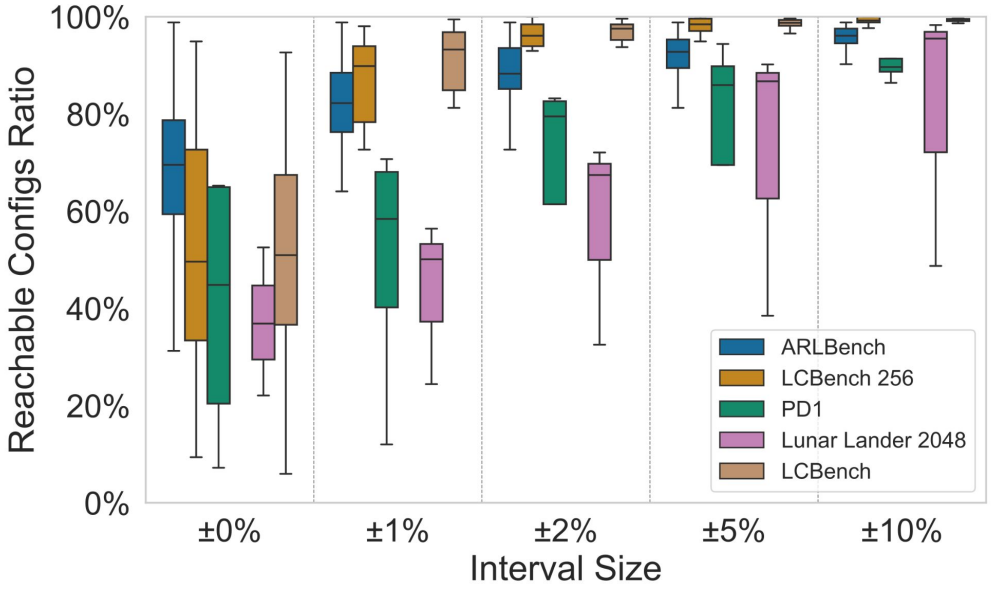
Unimodality

- **ARLBench**: RL Data
- **LCBench** [Zimmer et al. 2021]: supervised learning for tabular data
- **PD1** [Wang et al. 2024]: supervised learning for computer vision
- **LunarLander 2048**: larger PPO on LunarLander dataset



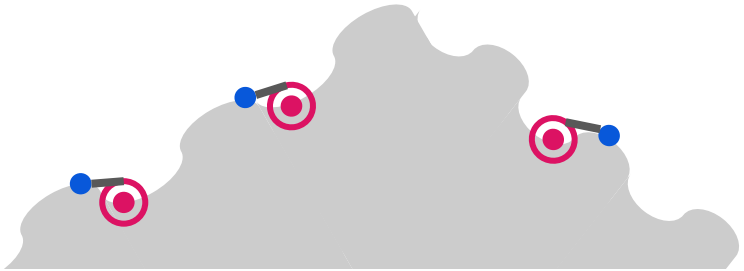
Unimodality

- ➔ **ARLBench**: RL Data
- ➔ **LCBench** [Zimmer et al. 2021]: supervised learning for tabular data
- ➔ **PD1** [Wang et al. 2024]: supervised learning for computer vision
- ➔ **LunarLander 2048**: larger PPO on LunarLander dataset



LCBench

LCBench



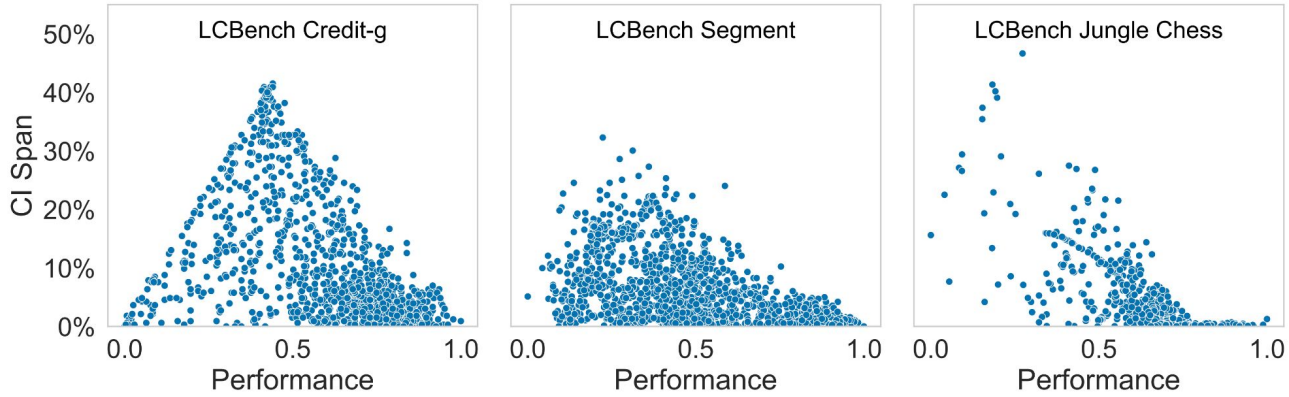
LCBench

ARLBench

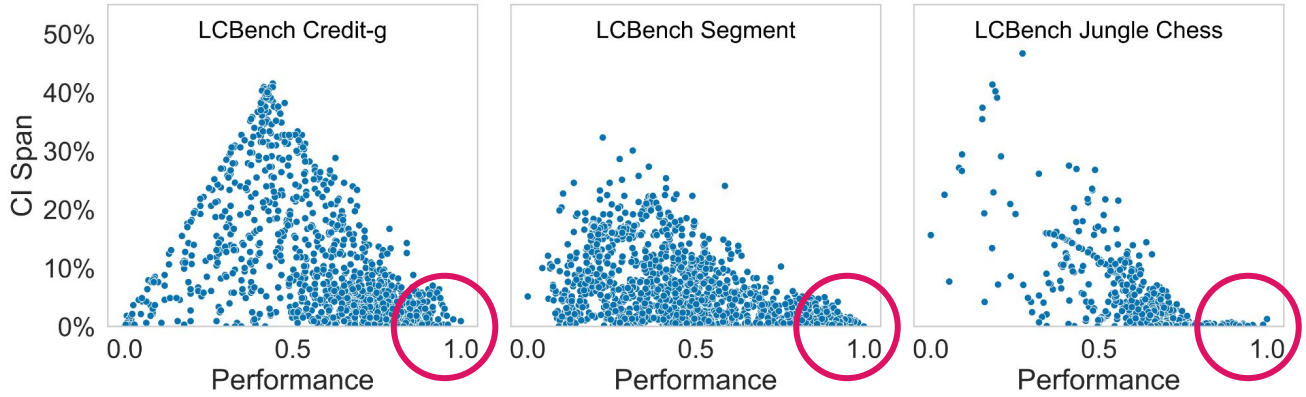
LCBench

ARLBench

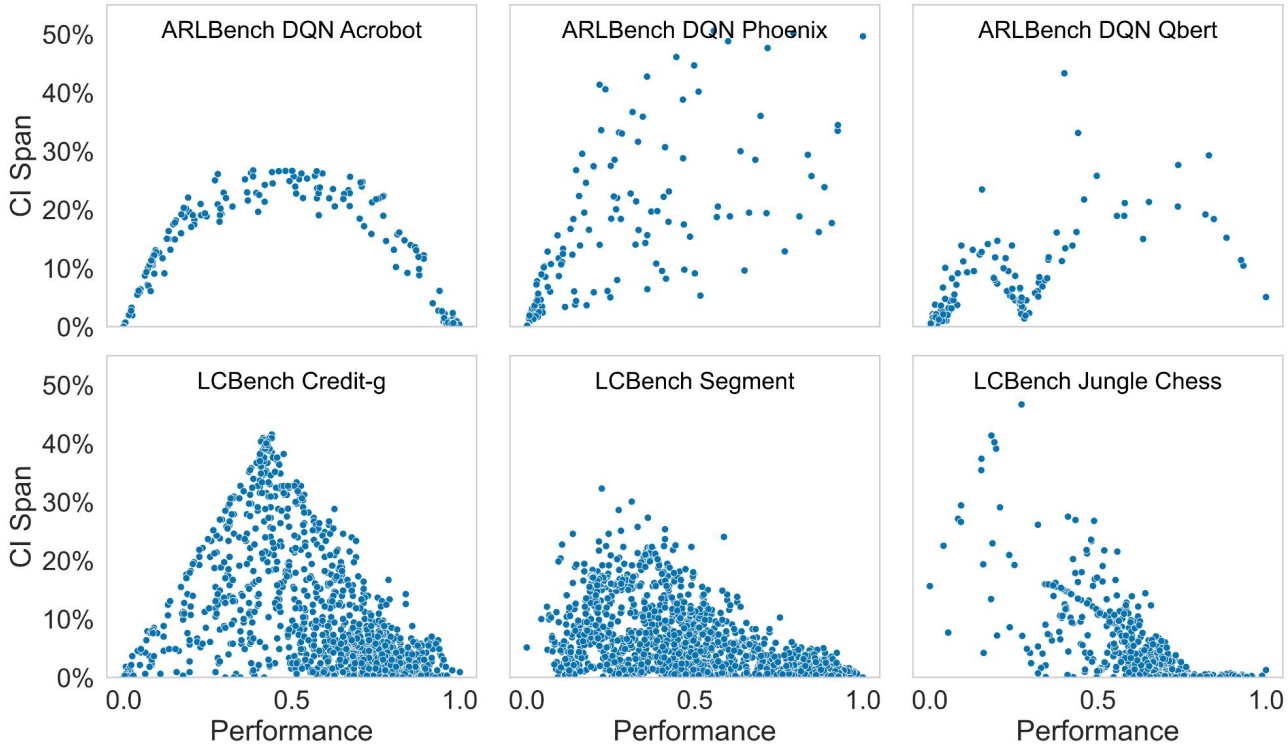
Explanation Attempt 2: Differences Between Runs?



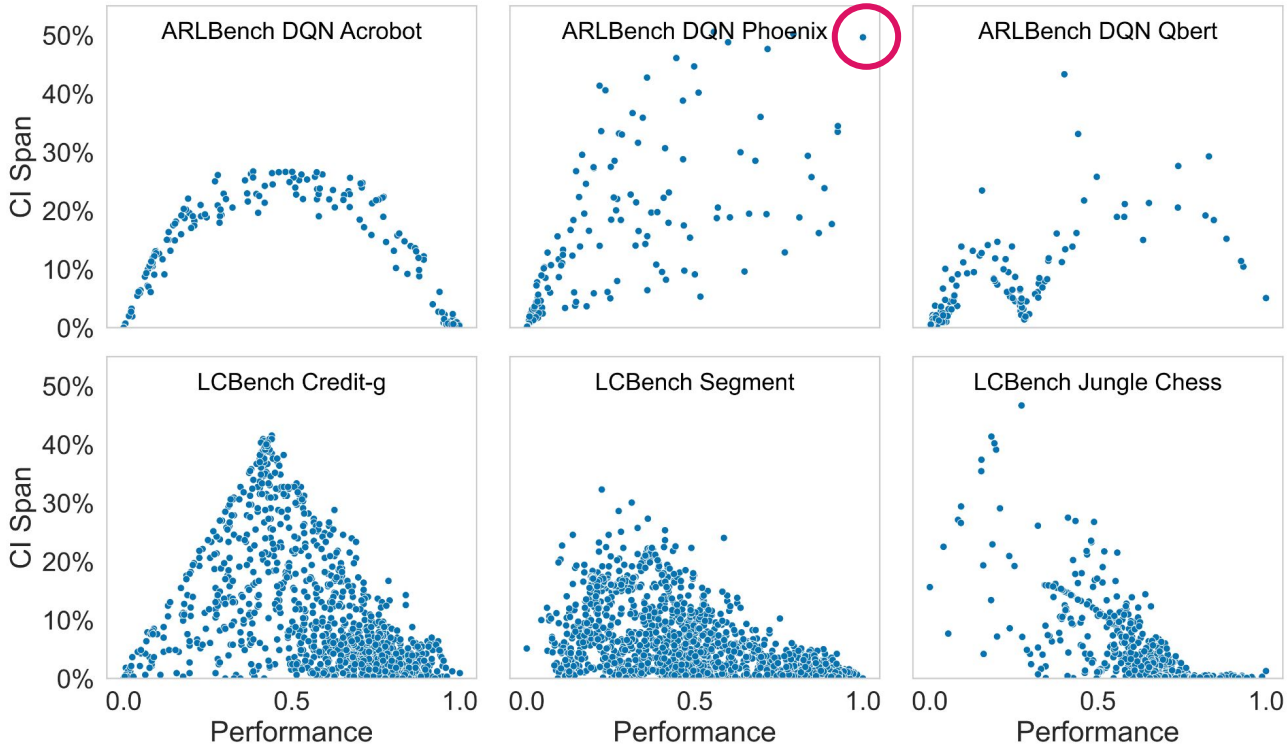
Explanation Attempt 2: Differences Between Runs?



Explanation Attempt 2: Differences Between Runs?

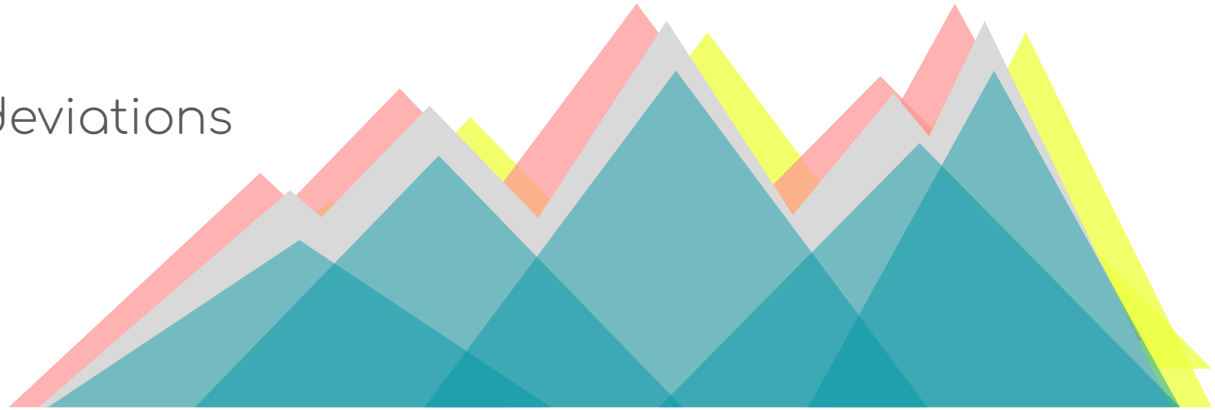


Explanation Attempt 2: Differences Between Runs?

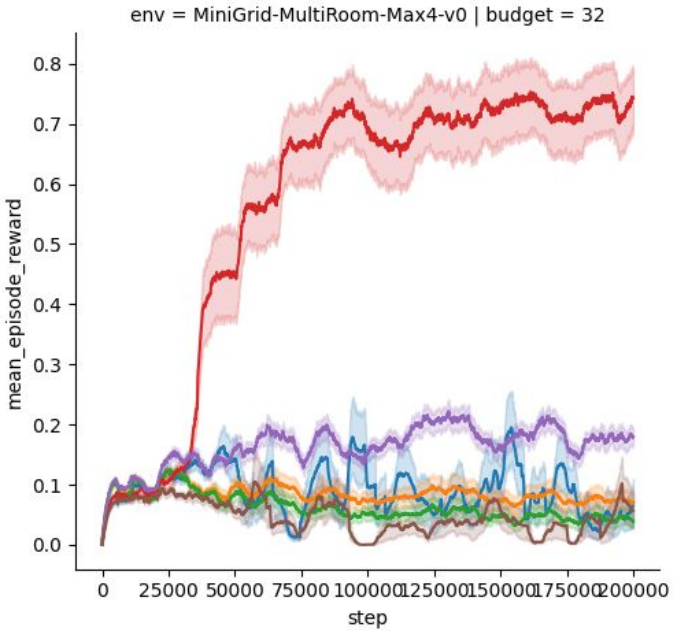


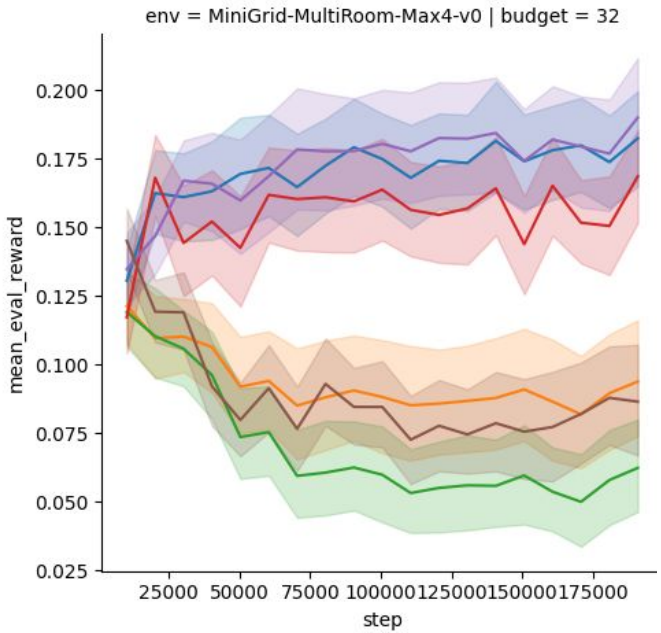
The RL Performance Landscape

- ➔ Bottom-heavy performance distribution
- ➔ Tendency towards fewer, but large landscape features
- ➔ Inconsistent, hard-to-model deviations between runs

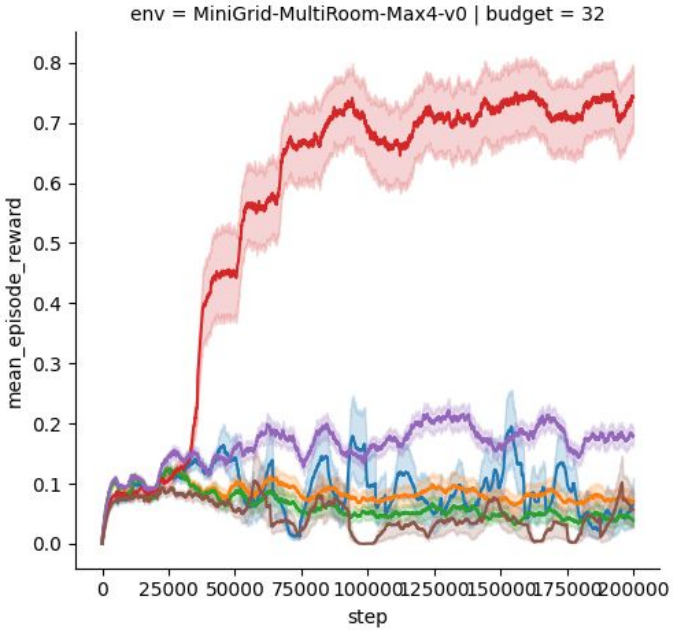


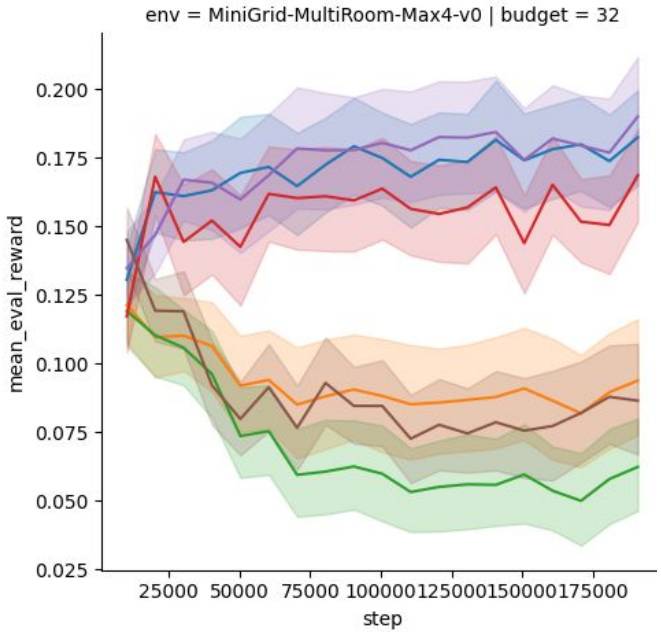
The Good News: It's Free!





The Good News: It's Free!





1. HPO
2. Training Across Seeds
3. Evaluation

The Good News: It's Free!