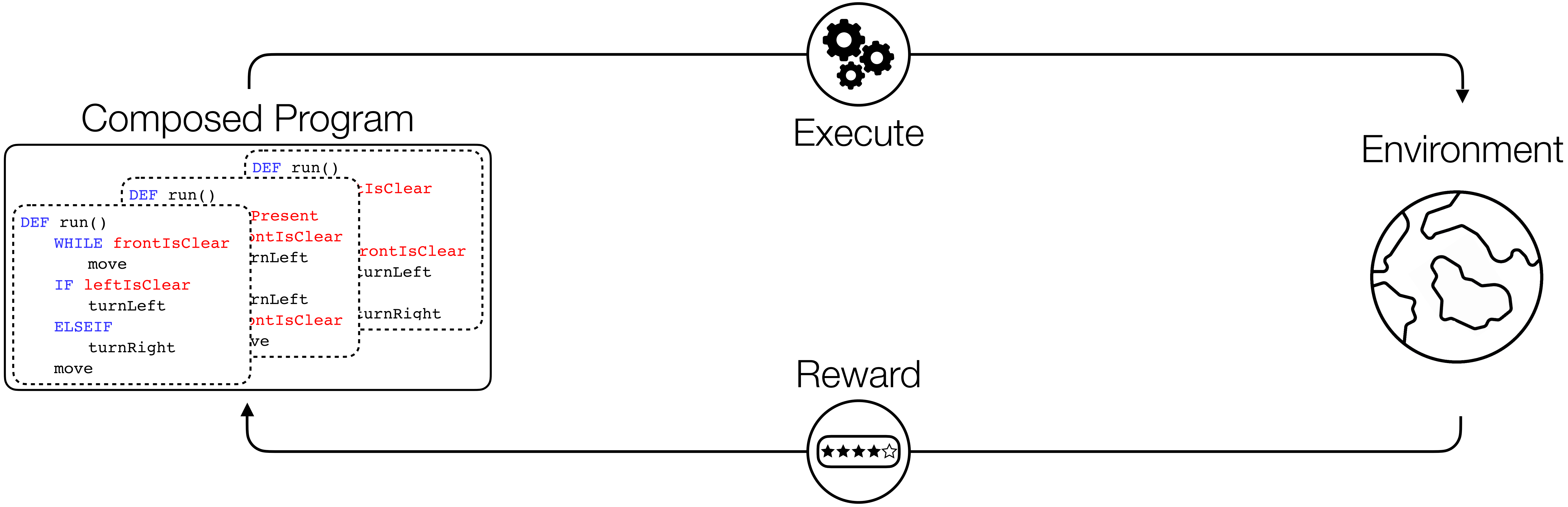
Hierarchical Programmatic Reinforcement Learning via Learning to Compose Programs

ICML 2023




Guan-Ting Liu*

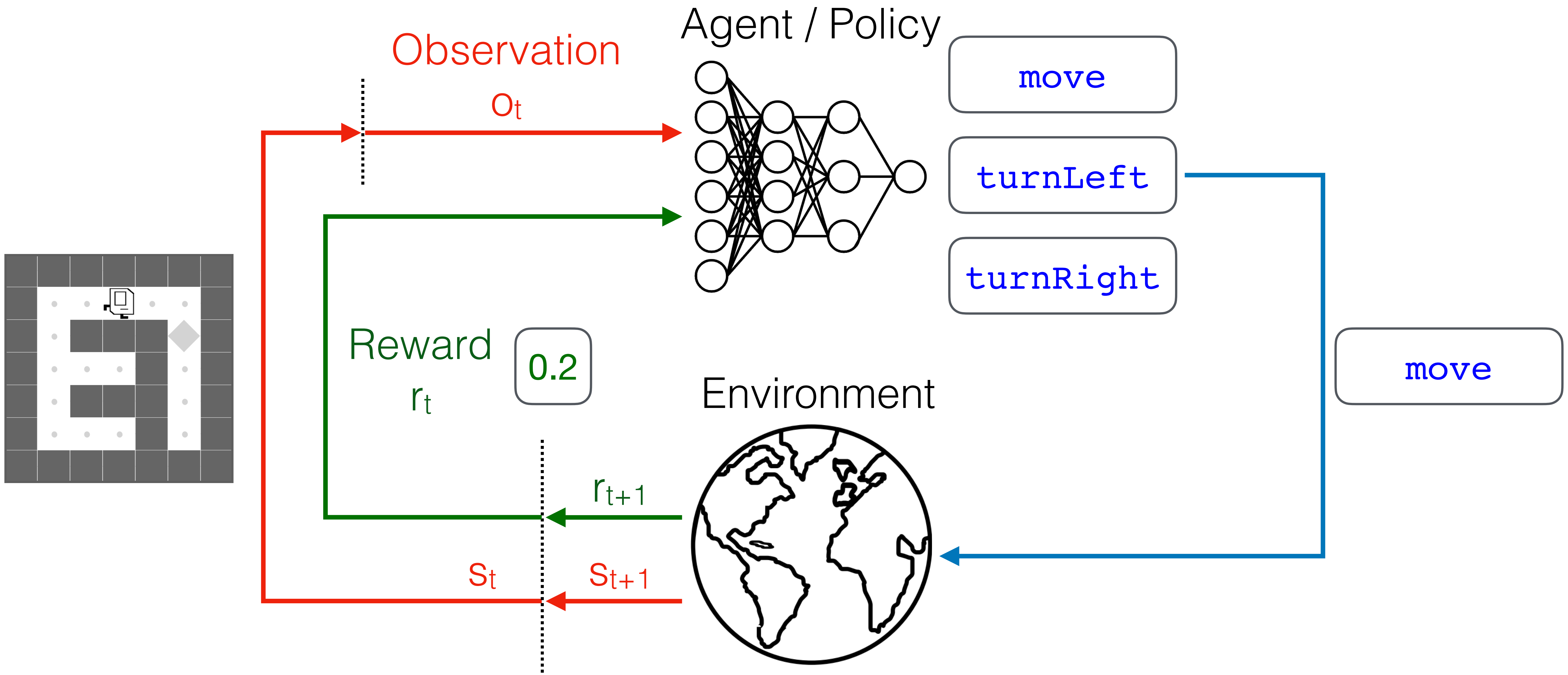

En-Pei Hu*


Pu-Jen Cheng


Hung-Yi Lee


Shao-Hua Sun

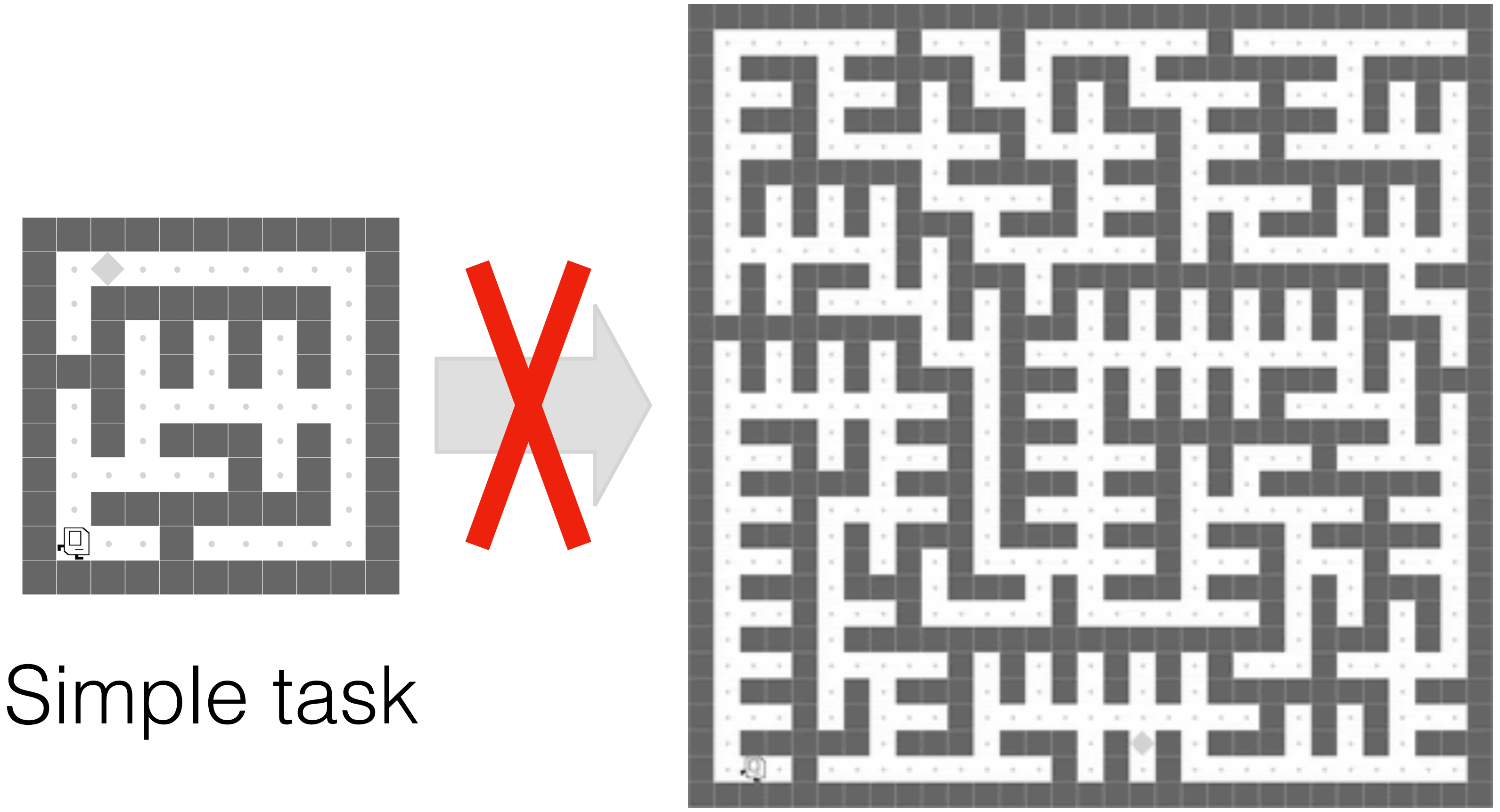
Robot Learning via Deep Reinforcement Learning



Goal: maximize $\sum_{t=0}^{t=H} \gamma^t R_t(s_t, a_t)$

Robot Learning via Deep Reinforcement Learning - Issues

Generalization

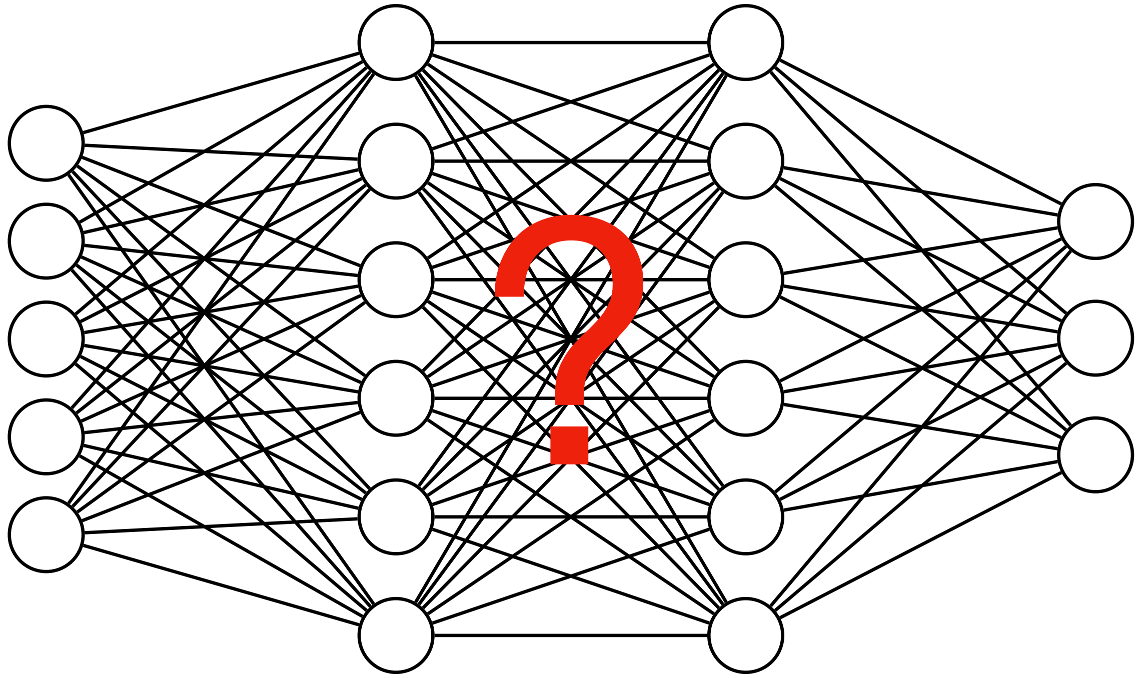


Simple task

Complex task

Interpretability

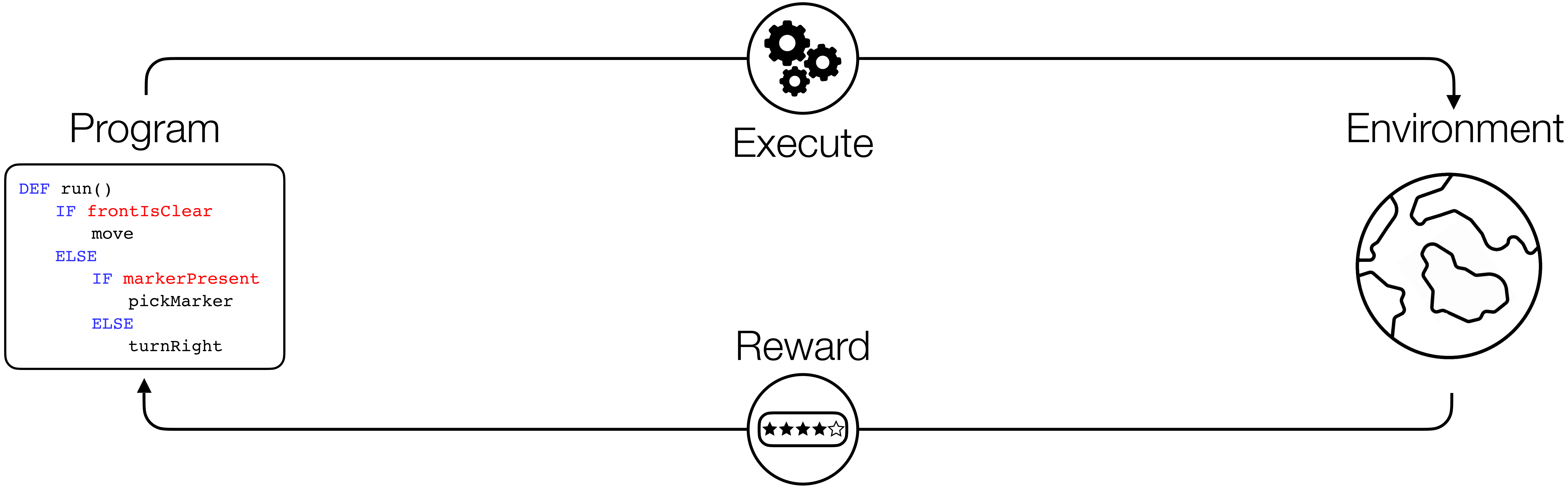
Trust, Safety, and Contestability



Deep neural network policy

Learning to Synthesize Programs as Interpretable and Generalizable Policies

NeurIPS 2021



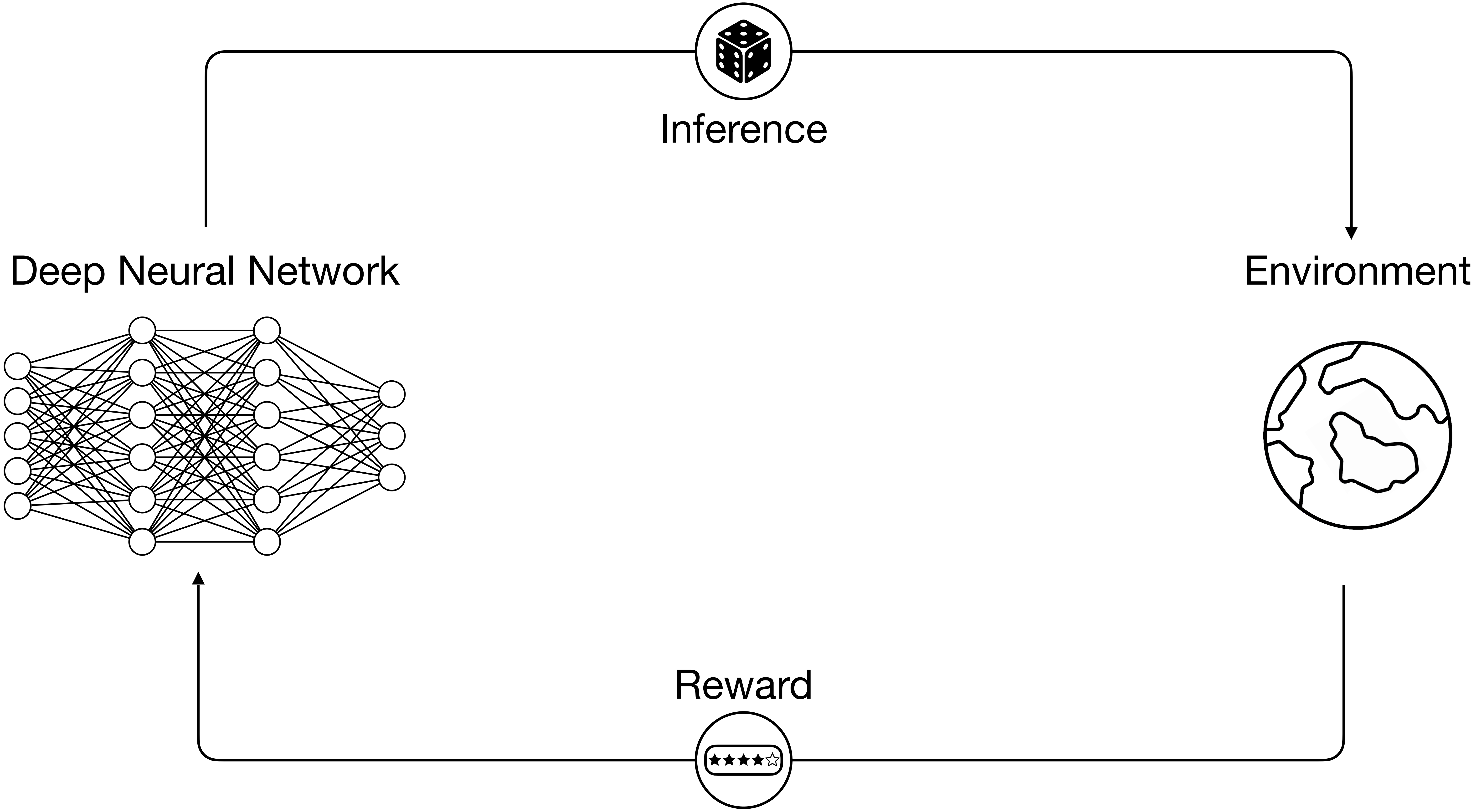

Dweep Trivedi*


Jesse Zhang*

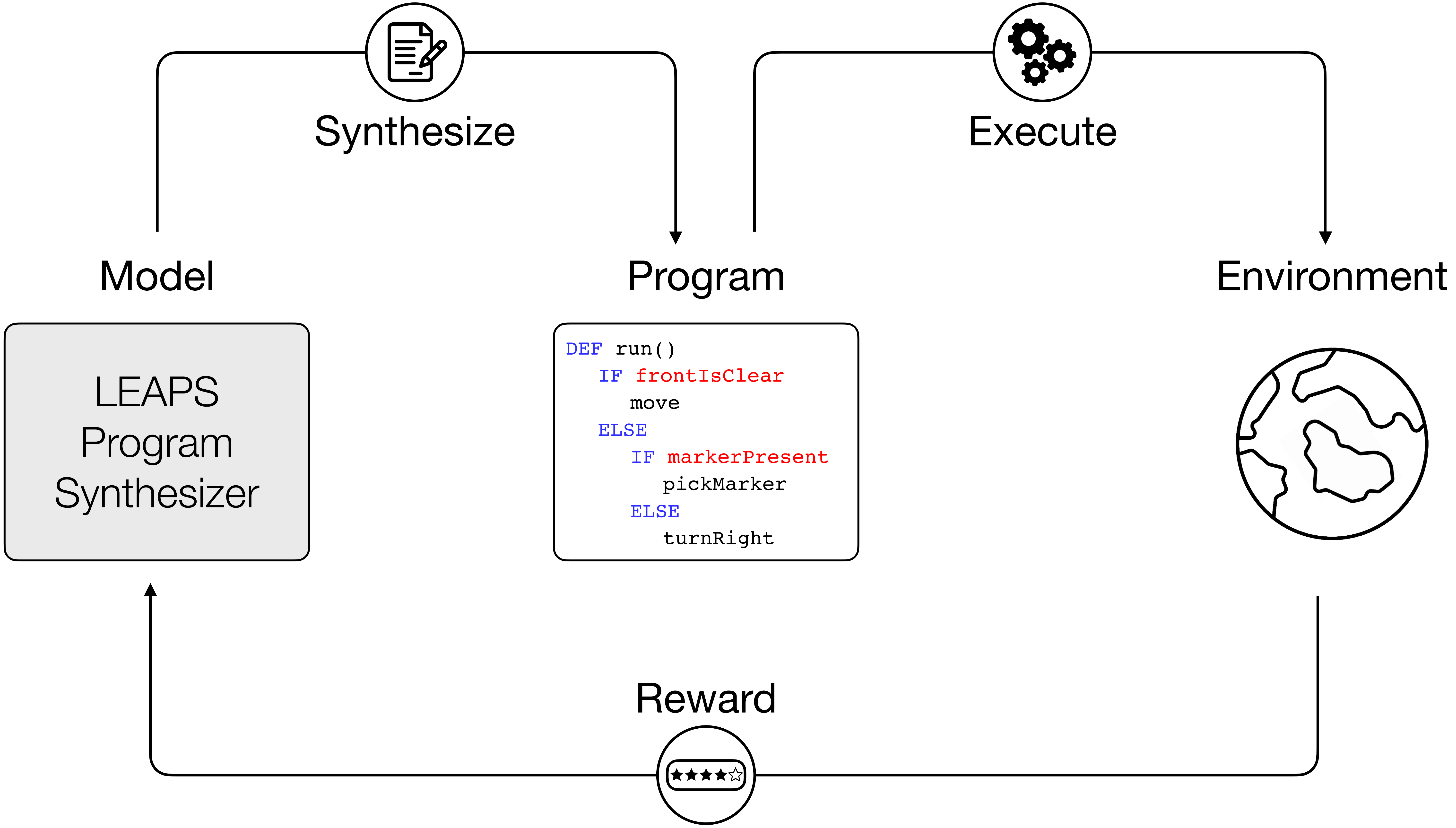

Shao-Hua Sun


Joseph J. Lim

Deep Reinforcement Learning



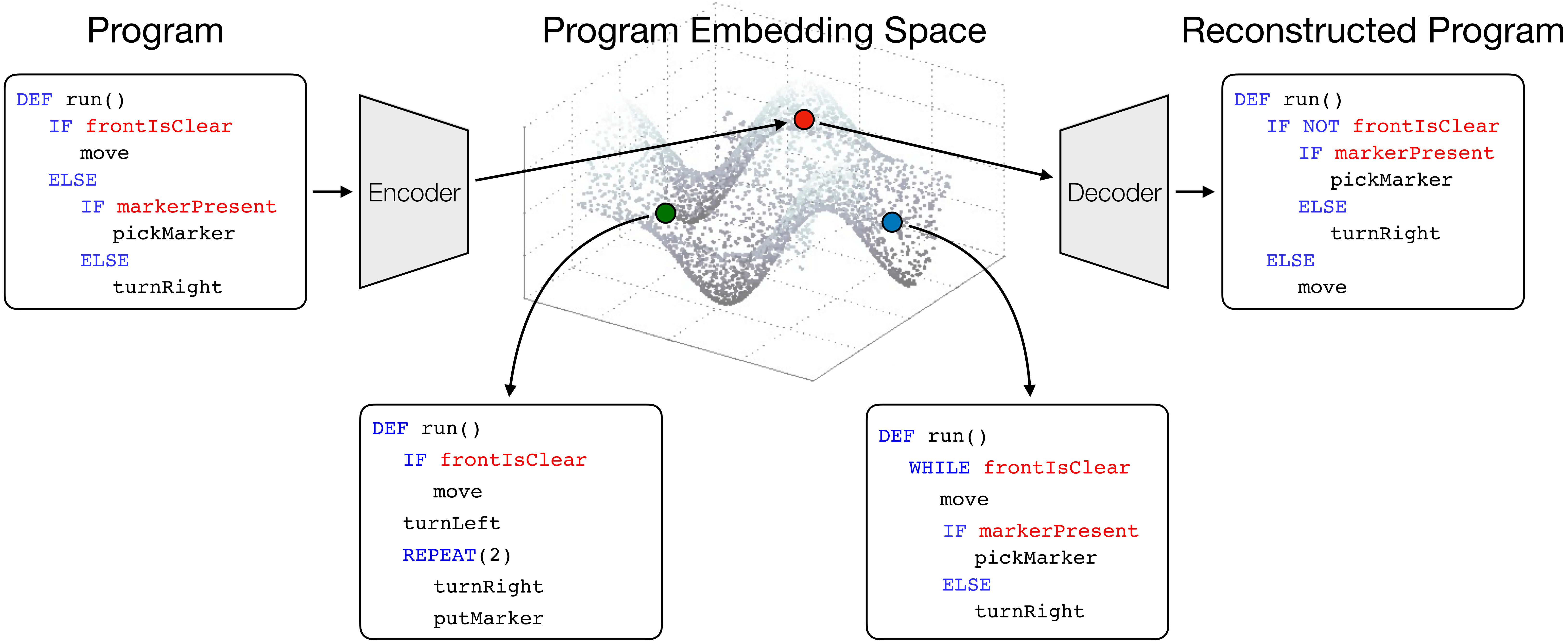
Reinforcement Learning via Synthesizing Programs



LEAPS: Learning Embeddings for lAtent Program Synthesis

Stage 1 Learn a **program embedding space** from randomly generated programs

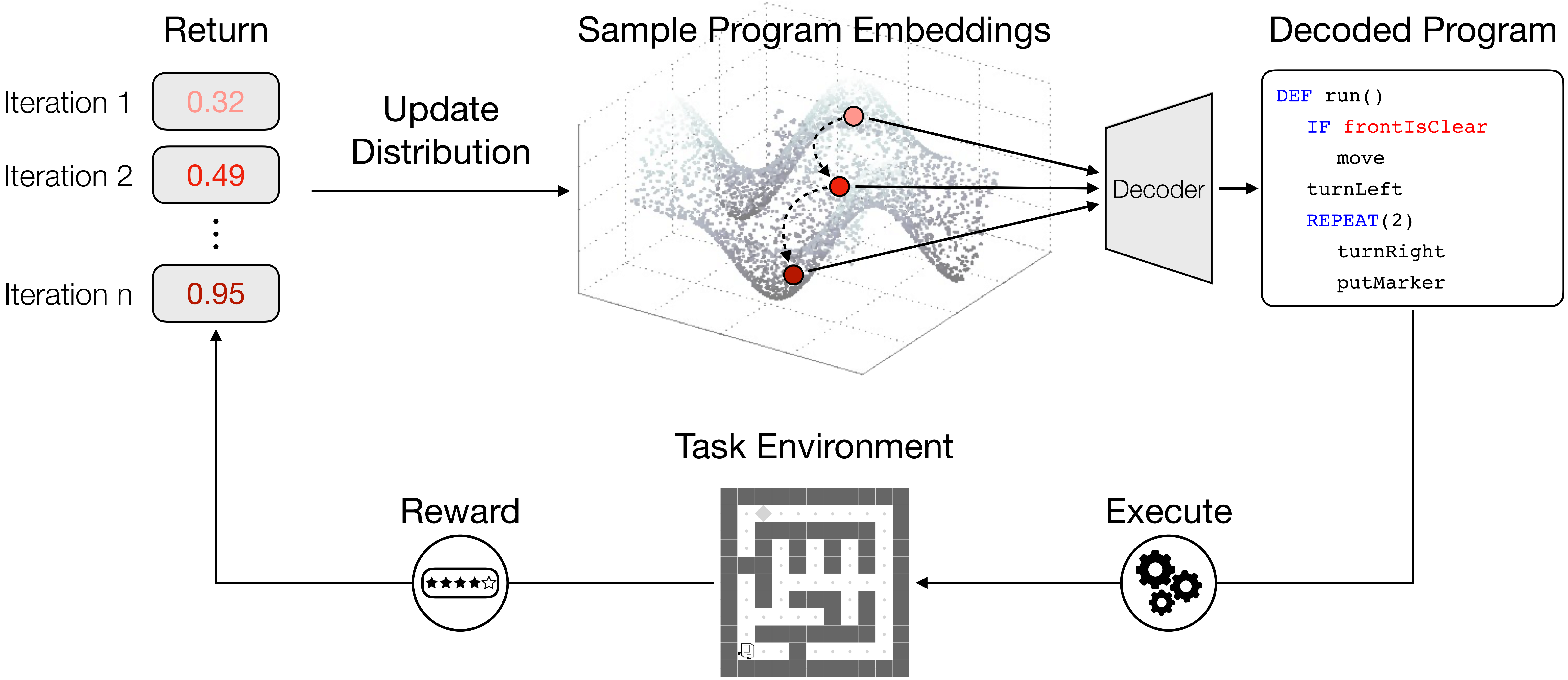
Goal Learn the grammar and the environment dynamics



LEAPS: Learning Embeddings for lAtent Program Synthesis

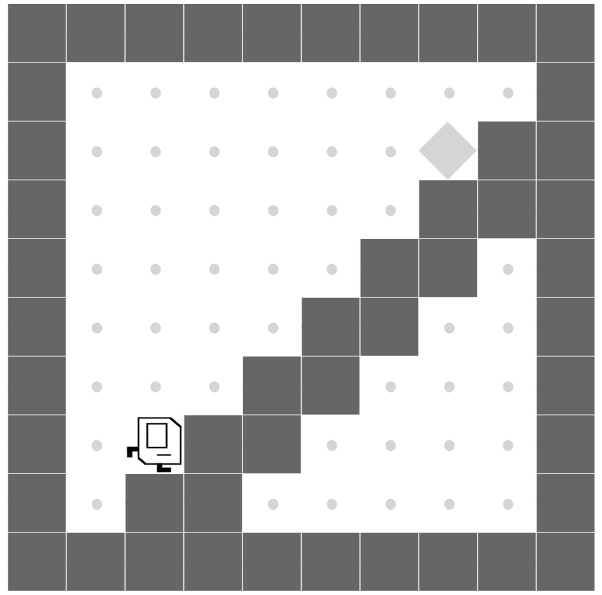
Stage 2 Search for a **task-solving program** using the **cross-entropy method (CEM)**

Goal Optimize the task performance

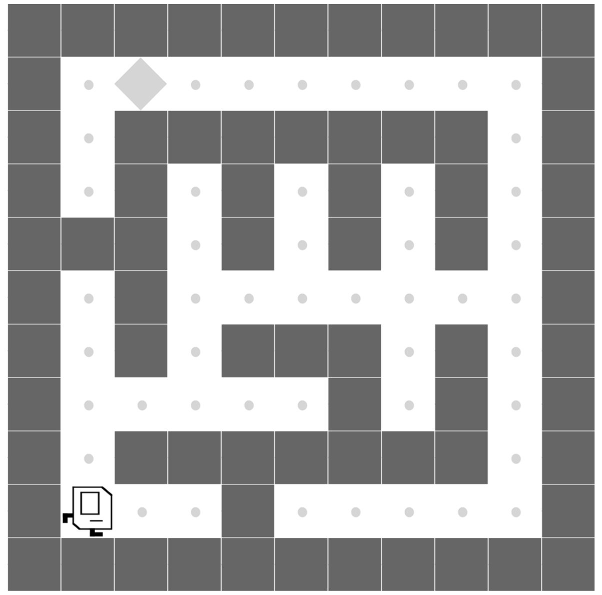


Karel Tasks

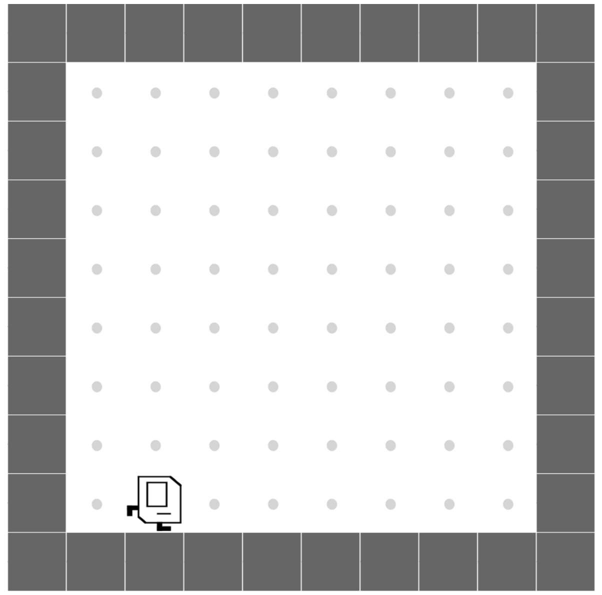
StairClimber



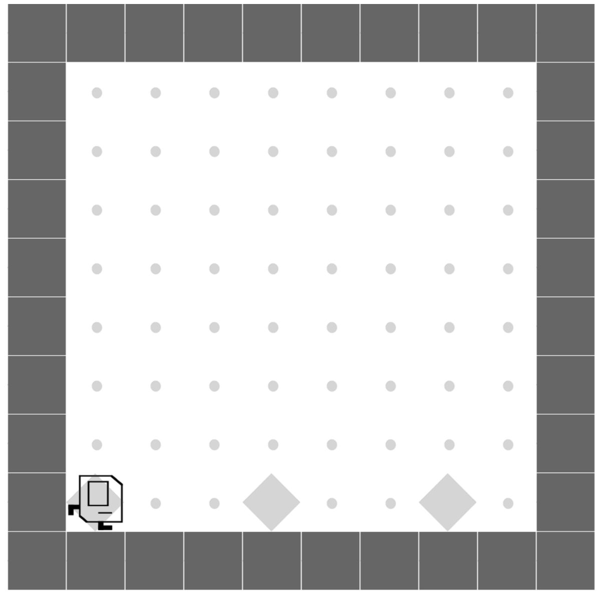
Maze



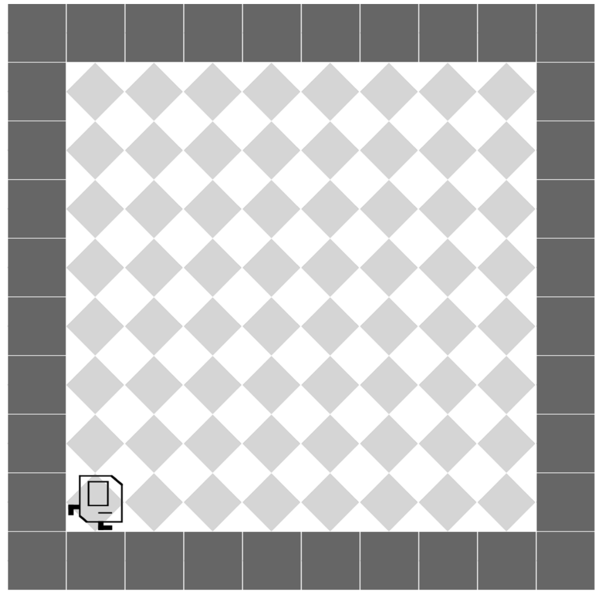
FourCorners



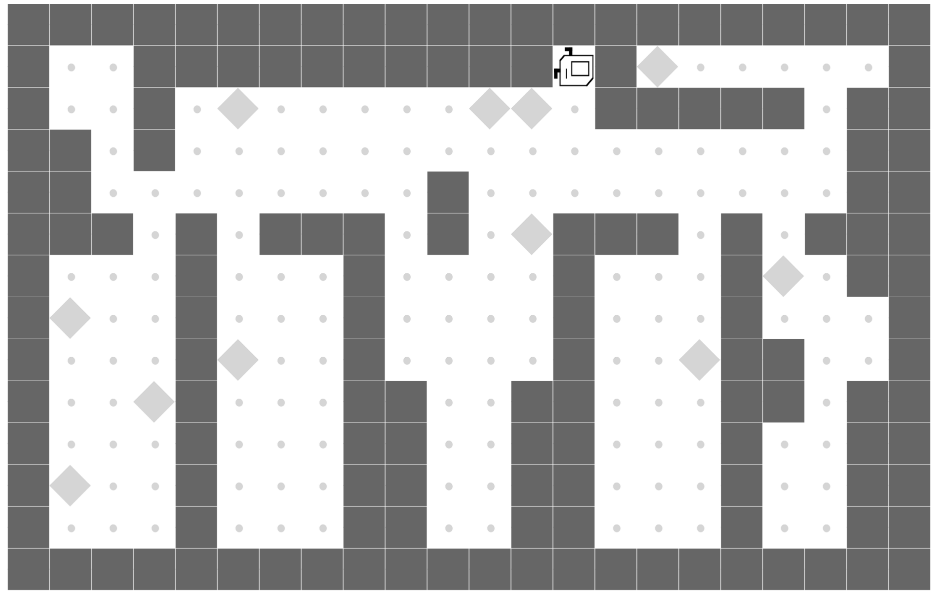
TopOff



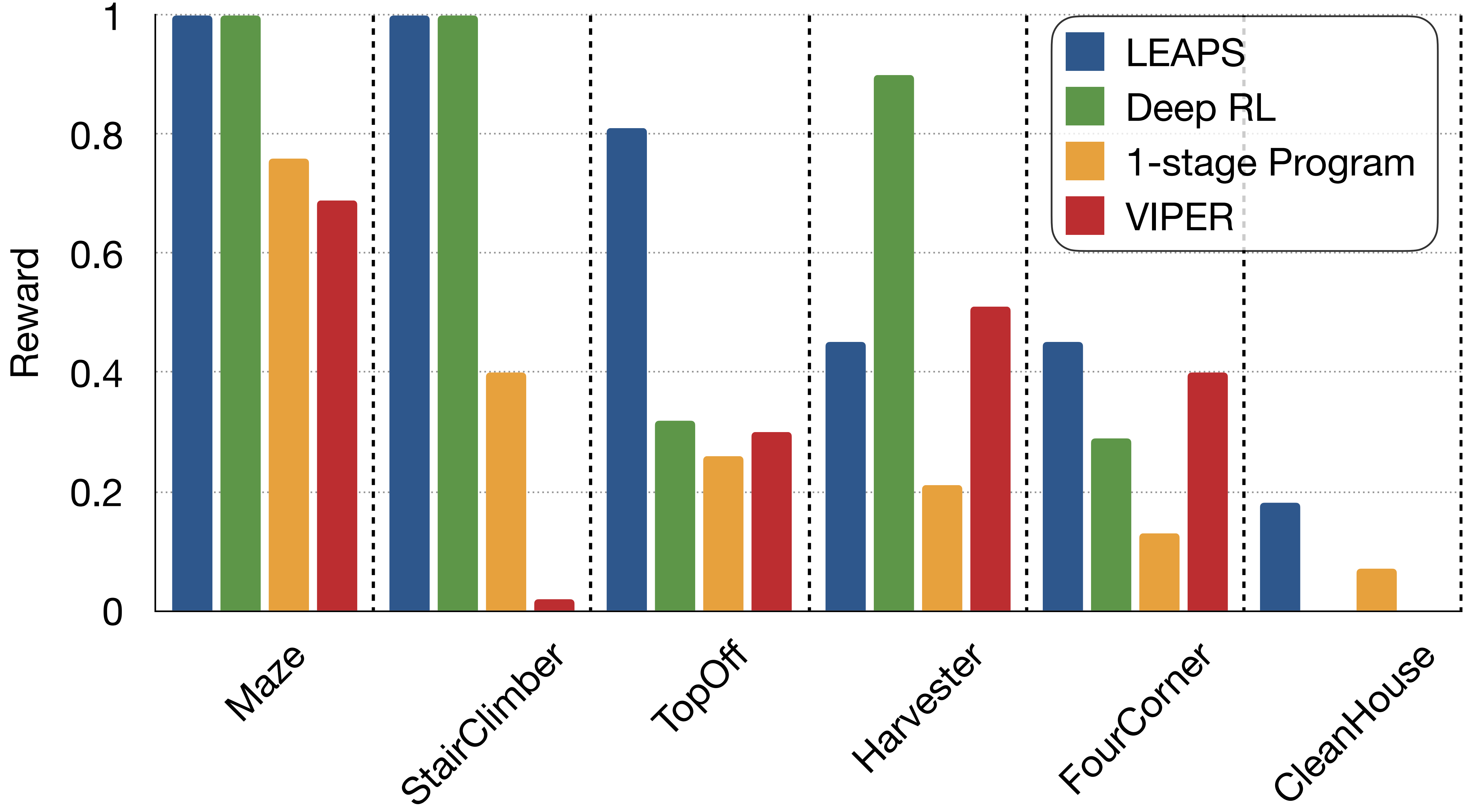
Harvester



CleanHouse

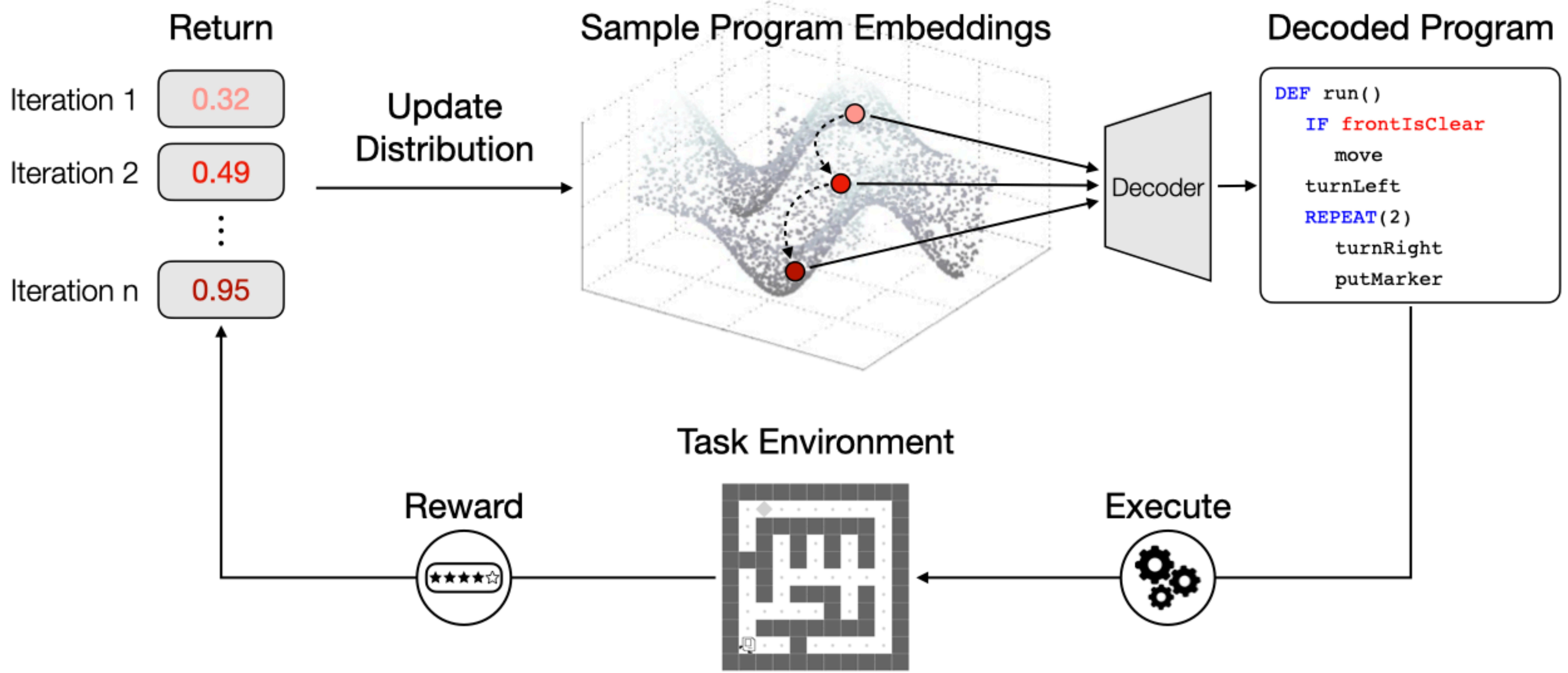


Quantitative Results



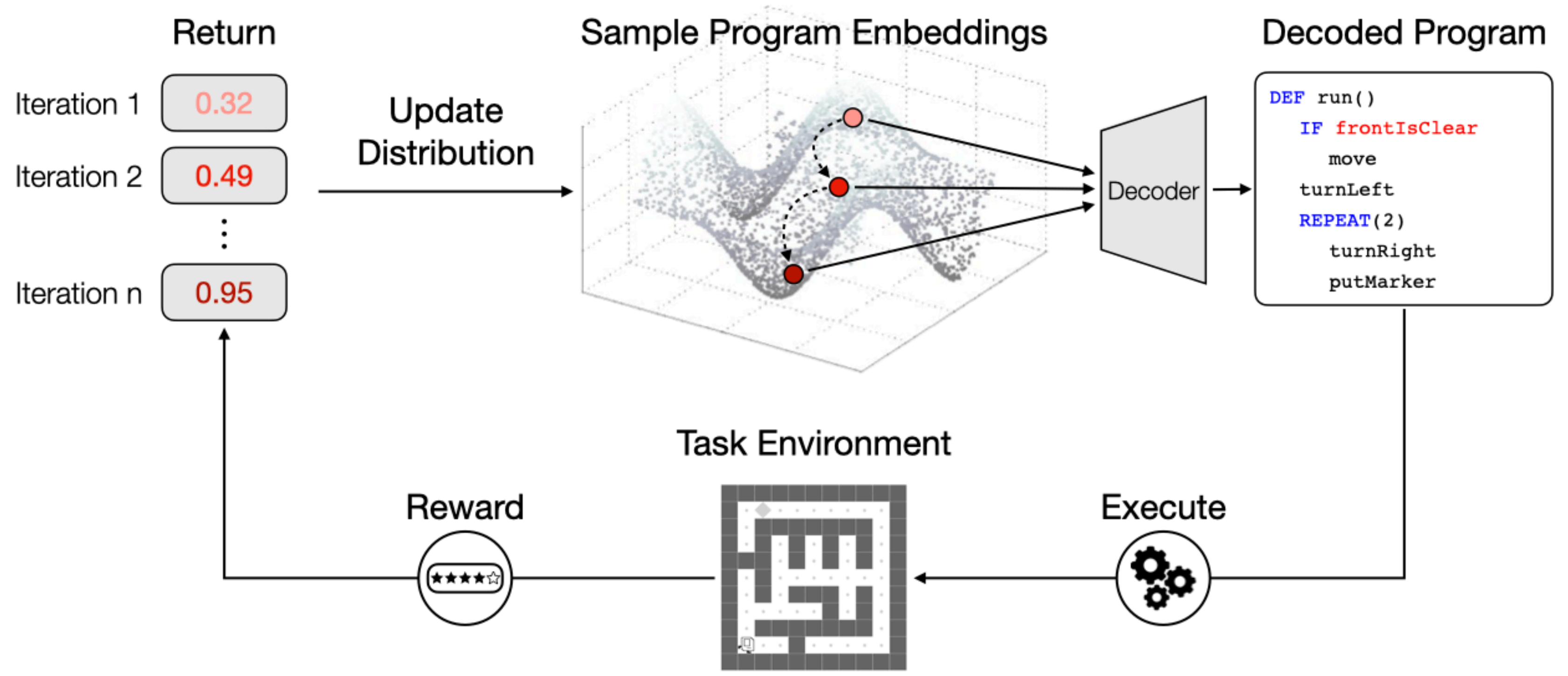
LEAPS: Learning Embeddings for lAtent Program Synthesis

Stage 2 Searching for a **task-solving program** using the **cross-entropy method**



LEAPS: Learning Embeddings for lAtent Program Synthesis

Stage 2 Searching for a **task-solving program** using the **cross-entropy method**



Limited program distribution

Search in the program embedding space spanned by **the dataset programs**

Cannot **synthesize longer or more complex** programs

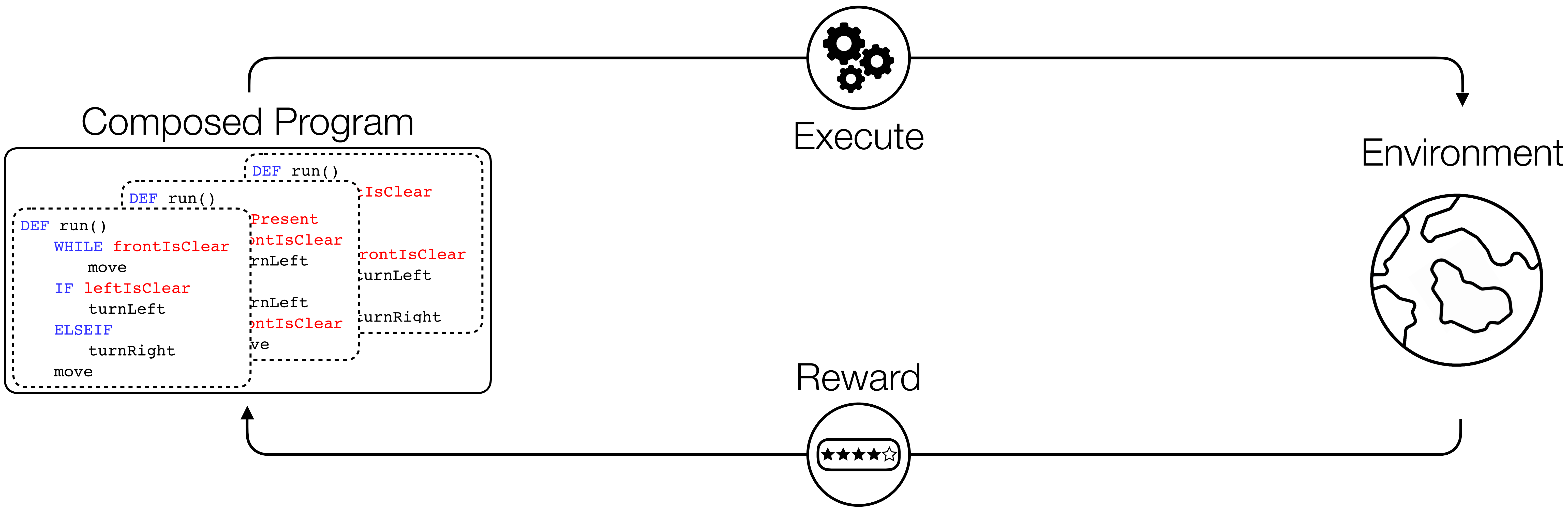
Poor credit assignment

Evaluate each candidate program solely based on the **cumulative return** of its execution trace

Cannot accurately **attribute rewards** to corresponding program parts

Hierarchical Programmatic Reinforcement Learning via Learning to Compose Programs

ICML 2023




Guan-Ting Liu*


En-Pei Hu*


Pu-Jen Cheng


Hung-Yi Lee

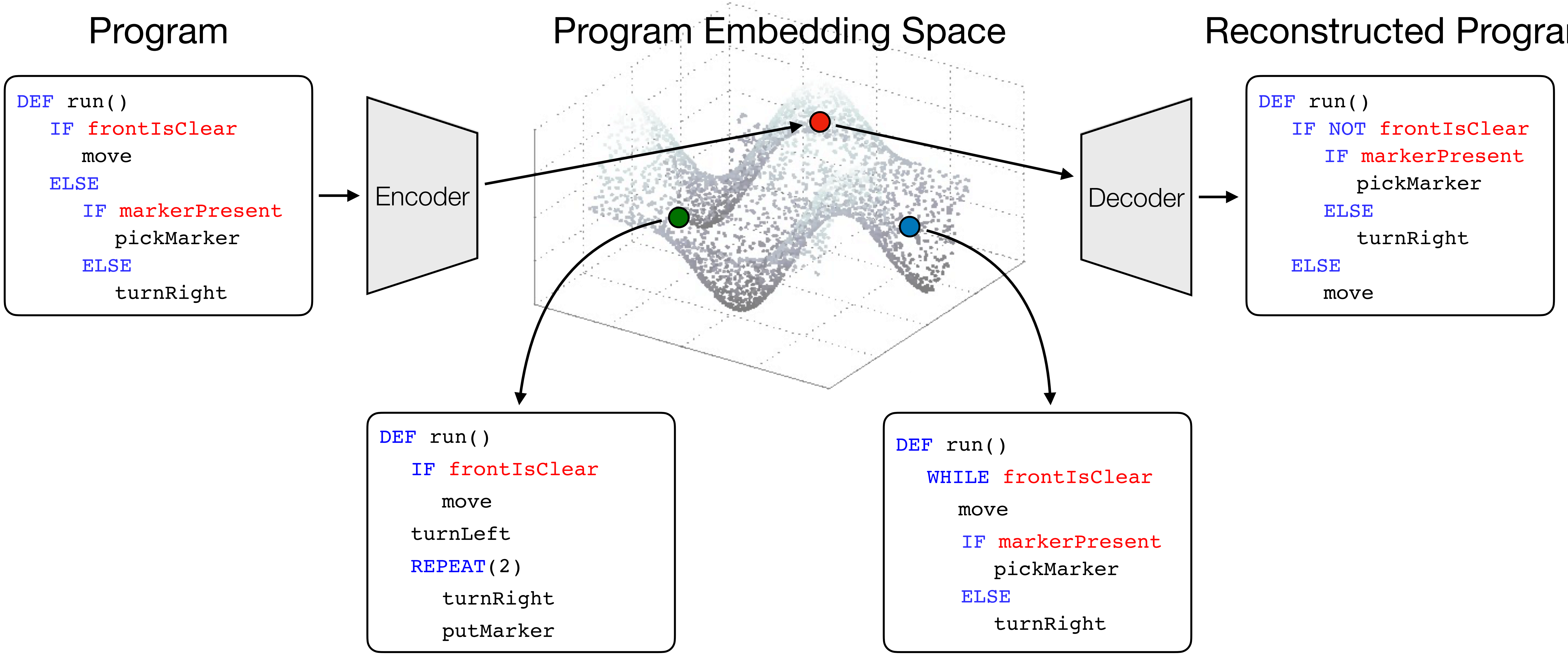

Shao-Hua Sun

ICML International Conference On Machine Learning 40 Years

National Taiwan University 國立臺灣大學

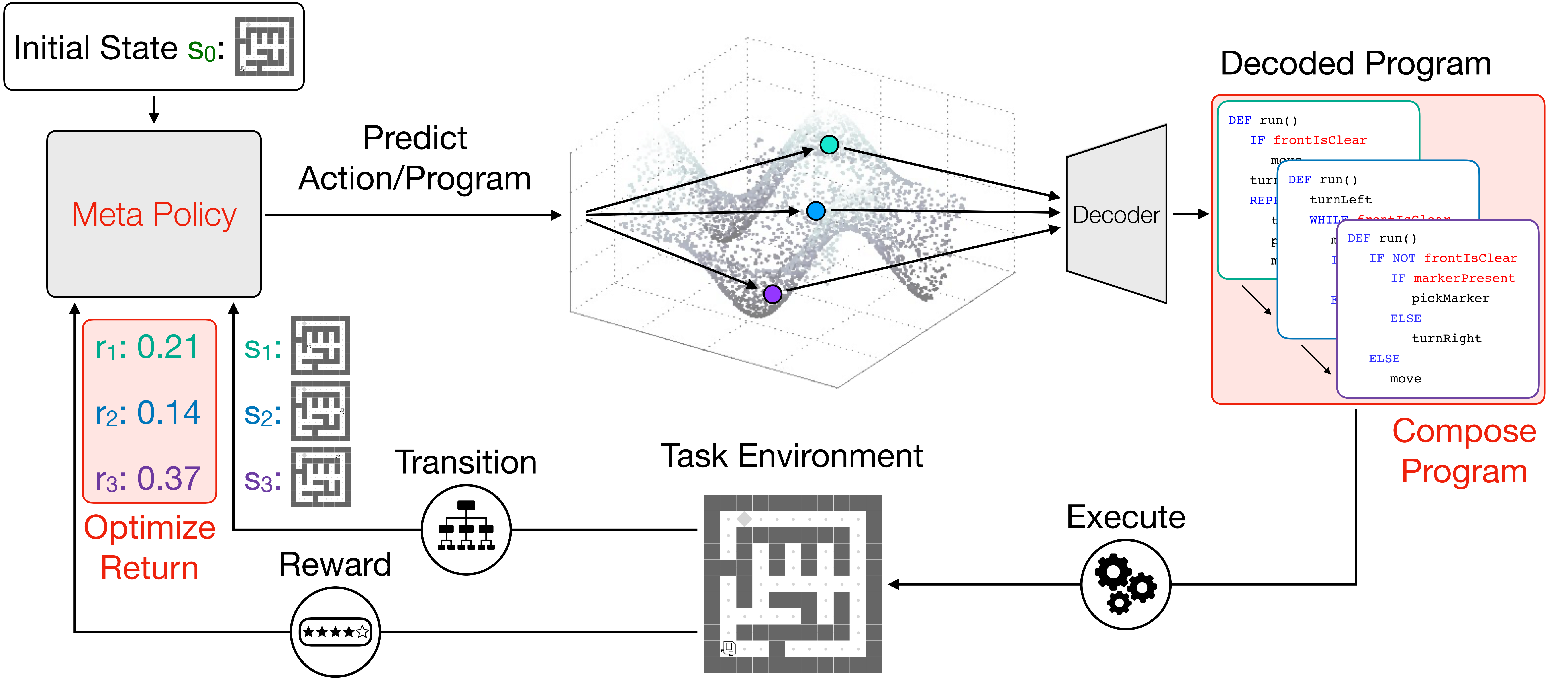
HPRL: Hierarchical Programmatic Reinforcement Learning

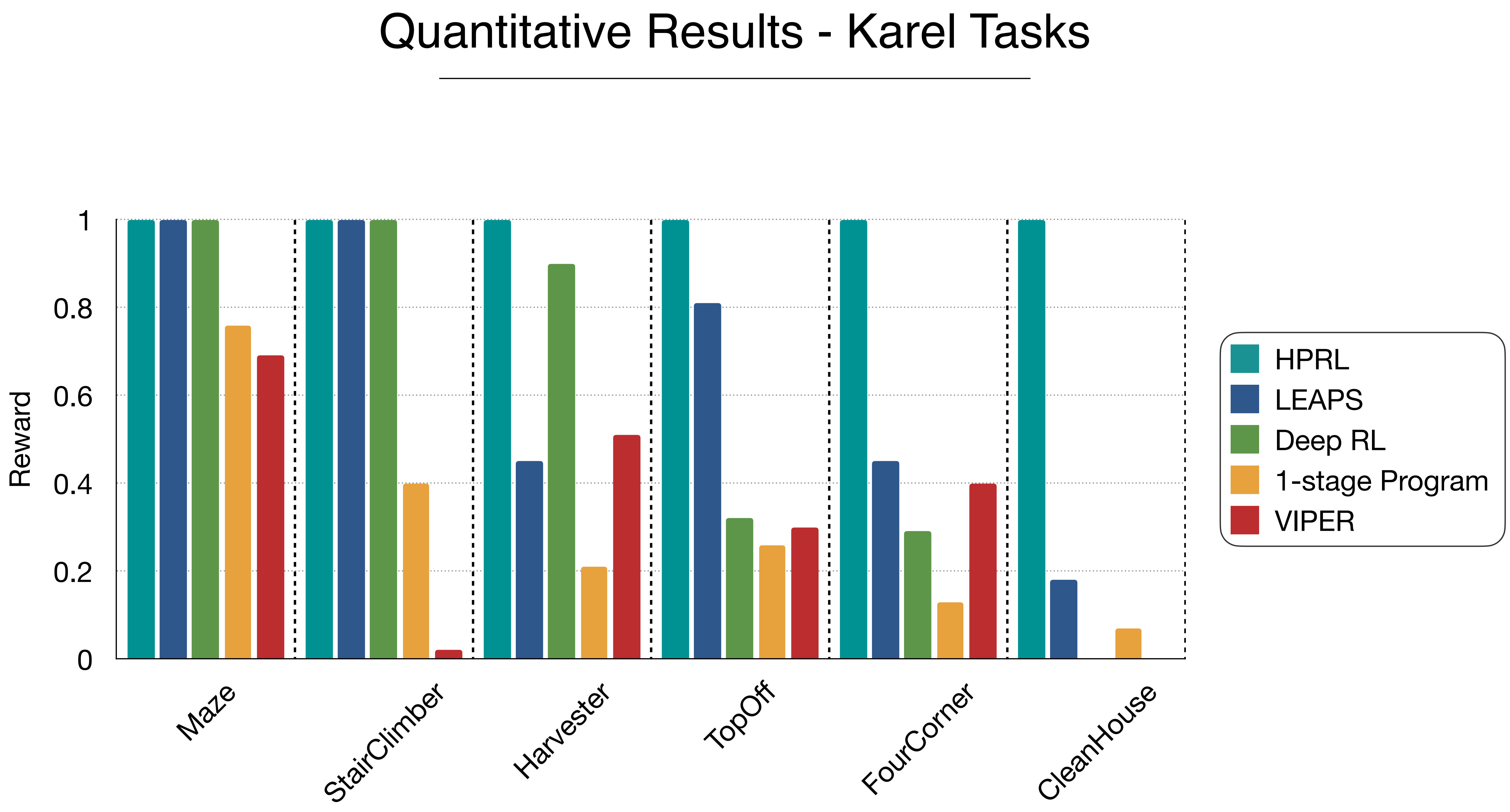
Stage 1 Learning a compressed **program embedding space** from randomly generated programs



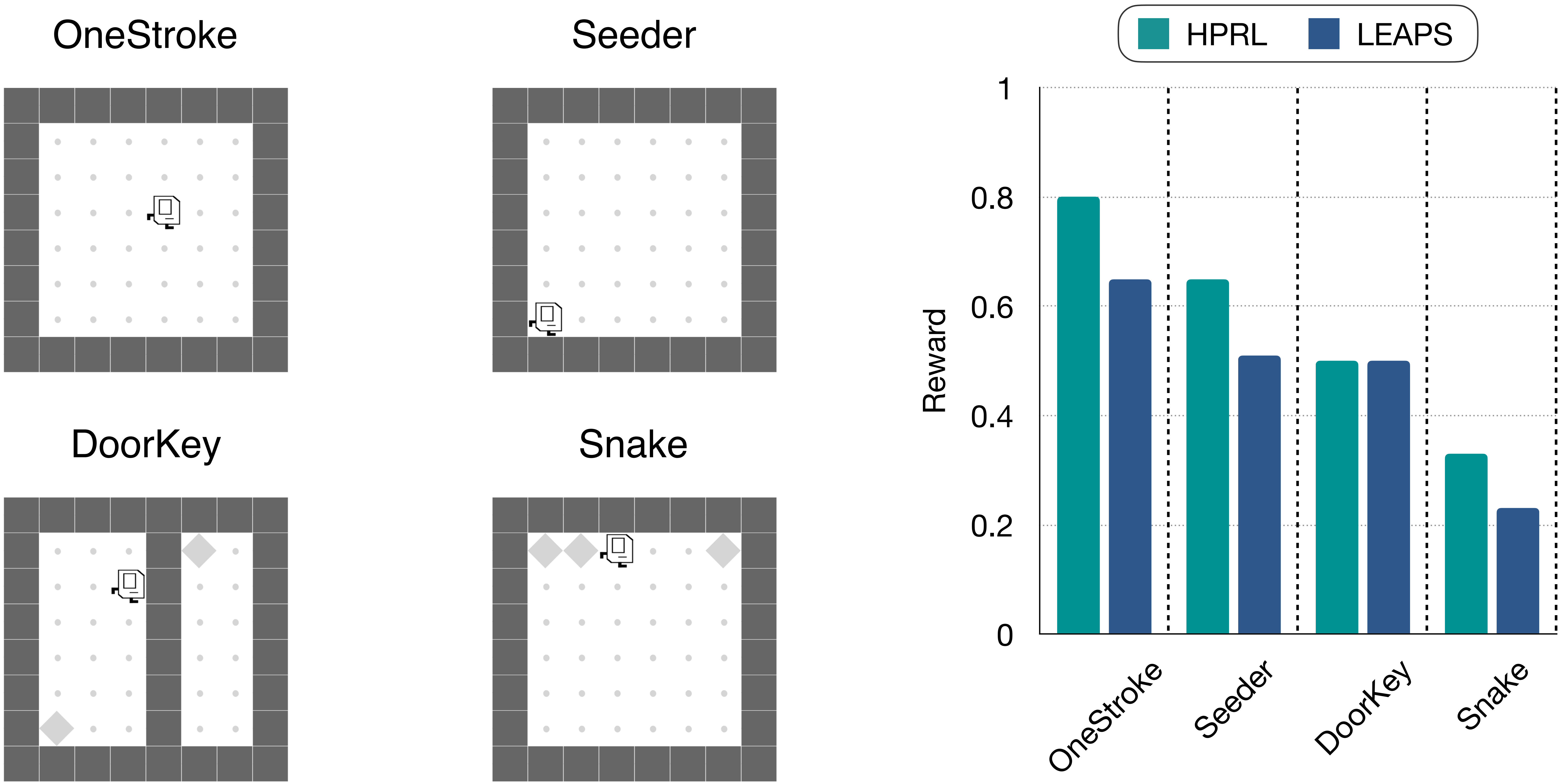
HPRL: Hierarchical Programmatic Reinforcement Learning

Stage 2 Learning a **meta policy** to produce **a series of programs** (*i.e.*, predict a series of **actions**) to yield a **composed task-solving program**




Quantitative Results - Karel Tasks
Reward
1
0.8
0.6
0.4
0.2
0
Maze
StairClimber
Harvester
TopOff
FourCorner
CleanHouse
HPRL
LEAPS
Deep RL
1-stage Program
VIPER

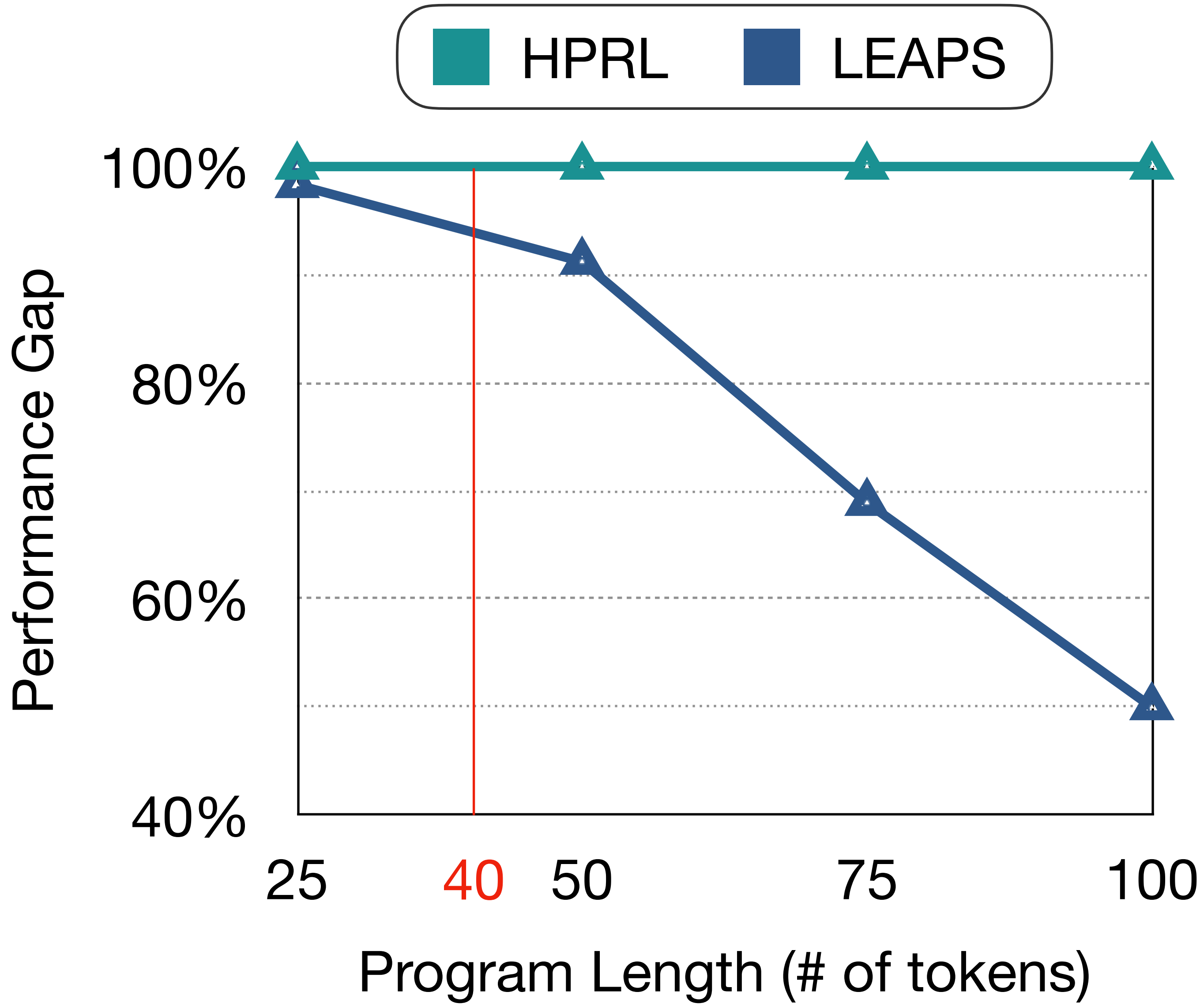
Karel-Hard Tasks



Additional Experiments

Limited program distribution

Synthesize out-of-distributionally long programs

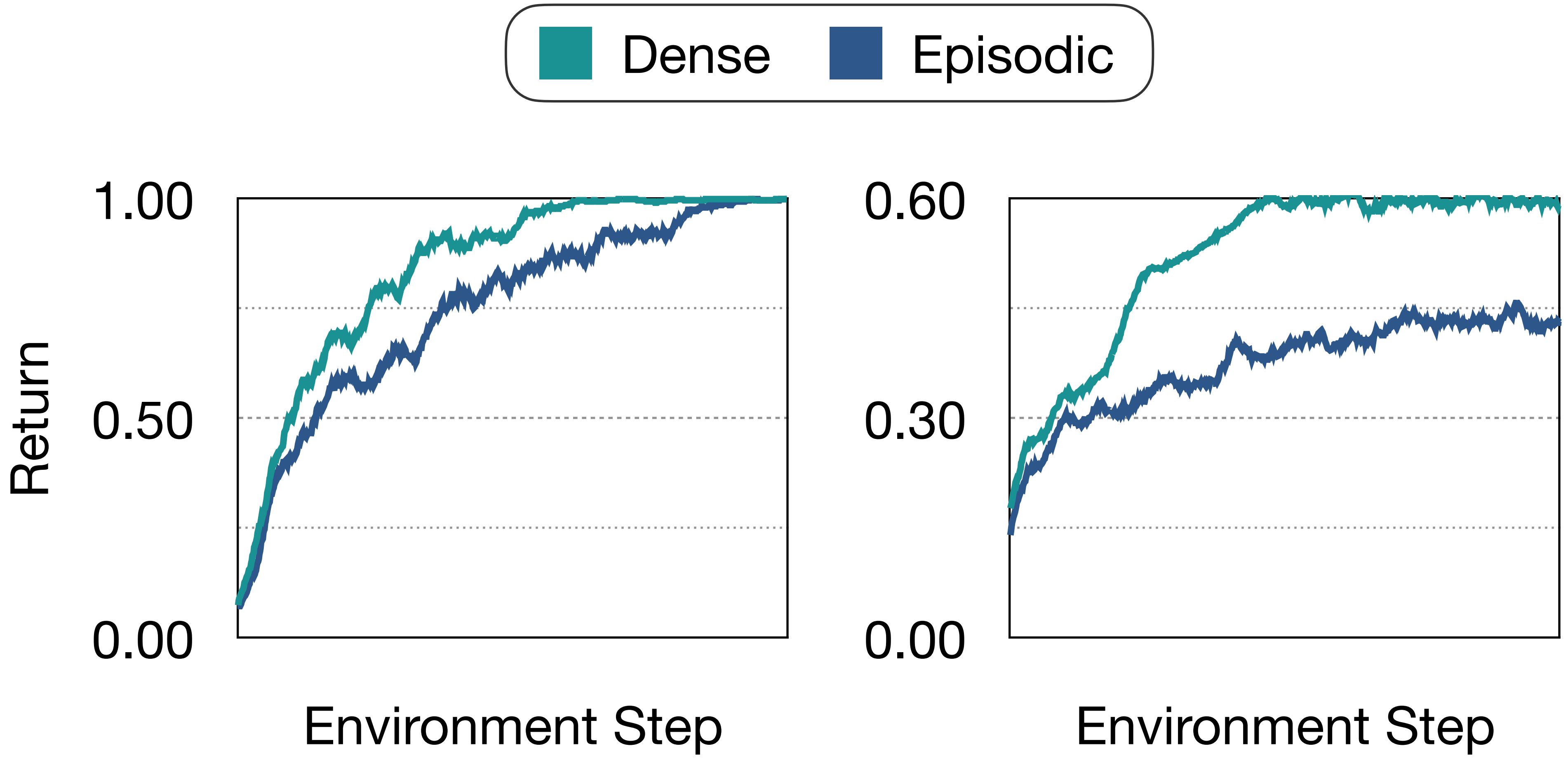


- HPRL can synthesize programs **longer** than the dataset programs (< 40 tokens) better than LEAPS

Poor credit assignment

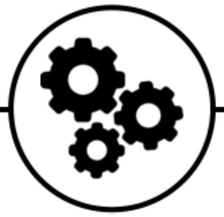
Learning from episodic reward

- **Dense:** Reward each subprogram based on its execution trace
- **Episodic:** Reward the entire composed program at the end

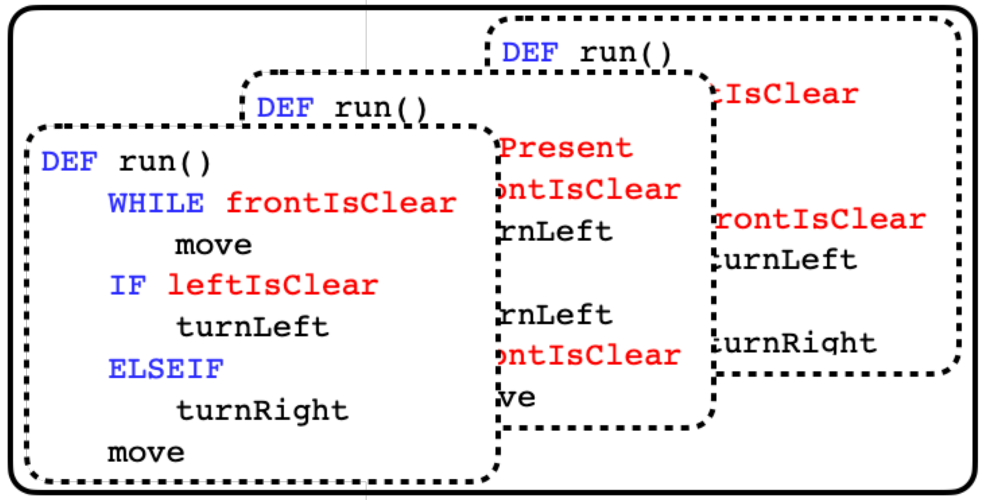


- **The hierarchical design** of HPRL allows for better **credit assignment** with dense rewards, facilitating the learning progress

Execute



Composed Program



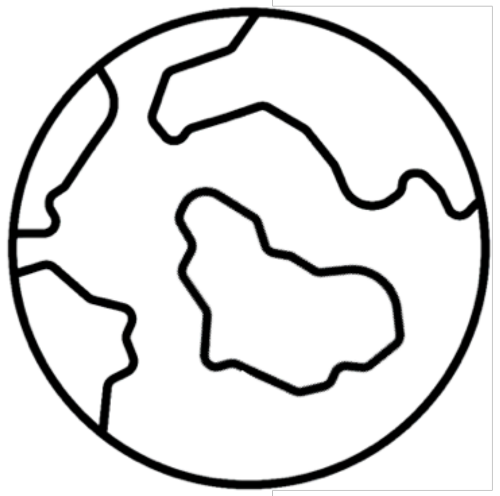
Thank You

Poster Session #4

Jul 26 Wed 2 p.m. — 3:30 p.m.

@ Exhibit Hall 1

Environment



Reward