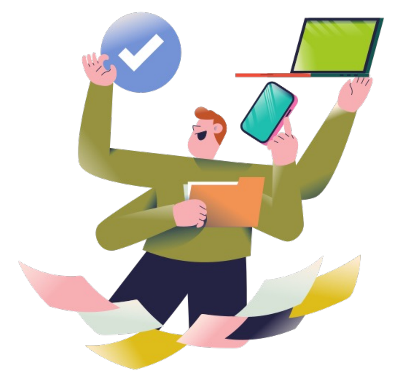
HypRL: Reinforcement Learning of Control Policies for Hyperproperties

Tzu-Han Hsu*, Arshia Rafieioskouei*, Borzoo Bonakdarpour
TART LAB, Michigan State University

Paper

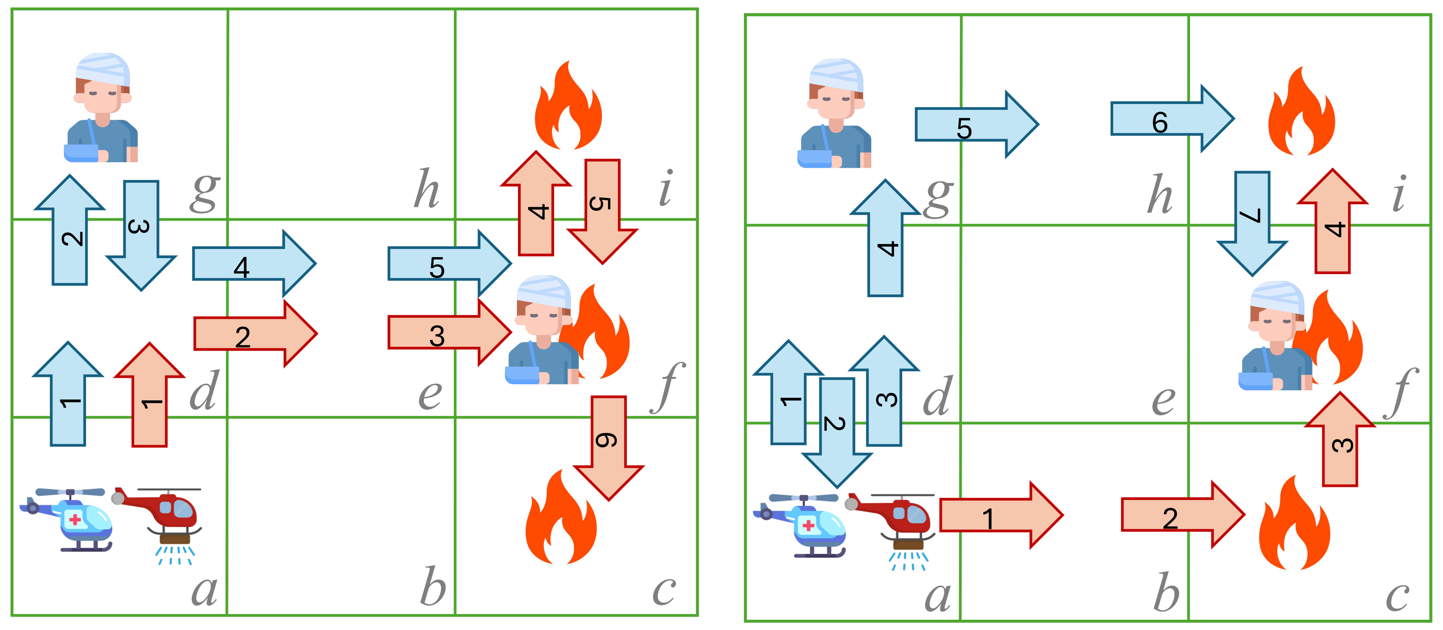
Research Question



Multi-Objective → **Multi-Agent**

How to Shape Optimal **Reward Functions**?

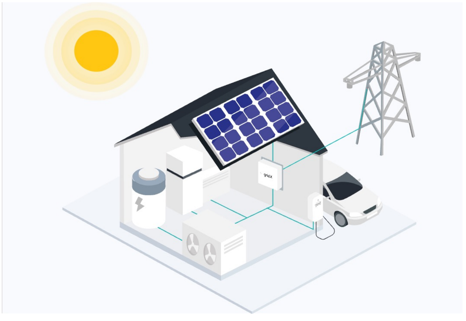
Motivating Example



Real-World Applications

swarm drones

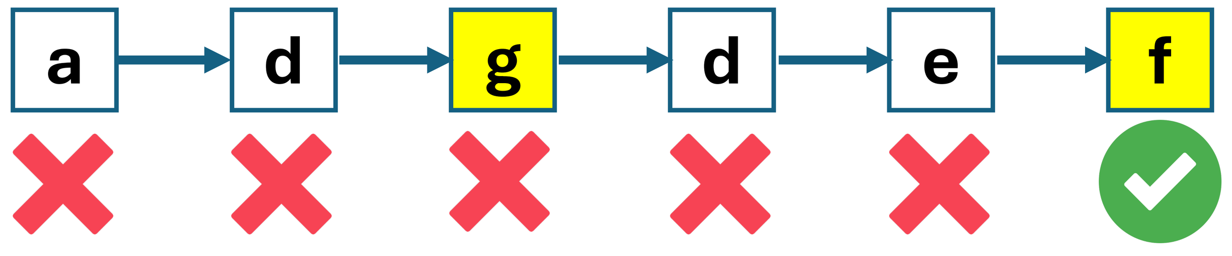
large-language models



energy management system

Linear Temporal Logic (LTL)

- LTL formalizes evolving, multi-objective tasks
- $(\Diamond g \land \Diamond f)$ means eventually reaches both zones to save victims (i.e., goal of 🚁)



- LTL can encode objectives in single-agent setting by recording each episode as a trace and checking it against some logical formula.
- For multi-agent settings, relations among traces are captured through hyperproperties.

What are Hyperproperties?

- Hyperproperties characterize requirements over sets of execution traces, allowing the specification of behaviors that involve multiple agents.
- A HyperLTL formula that captures all objectives and constraints of the example:

$$\varphi_{\text{rescue}} = \forall \tau_1 . \exists \tau_2 . \; (\psi_{\text{dist}} \land \psi_{\text{fire}} \land \psi_{\text{save}})$$

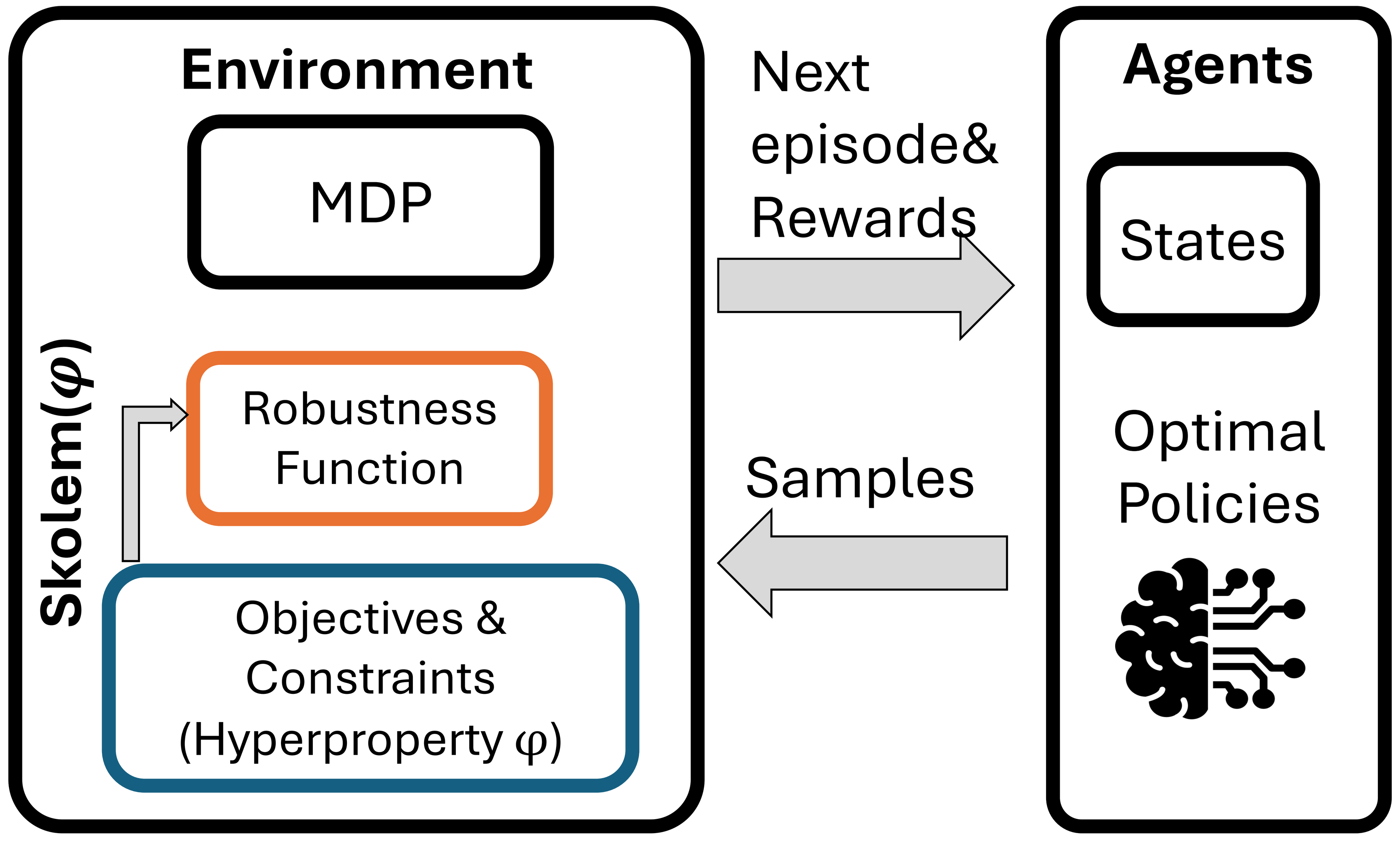
Challenges

- The quantifier alternations introduce interaction or dependencies between multiple traces.
- We use Skolemization technique:

$$\textbf{Skolem}(\varphi_{\text{rescue}}) = \exists \mathbf{f}_2(\tau_1) . \forall \tau_1$$

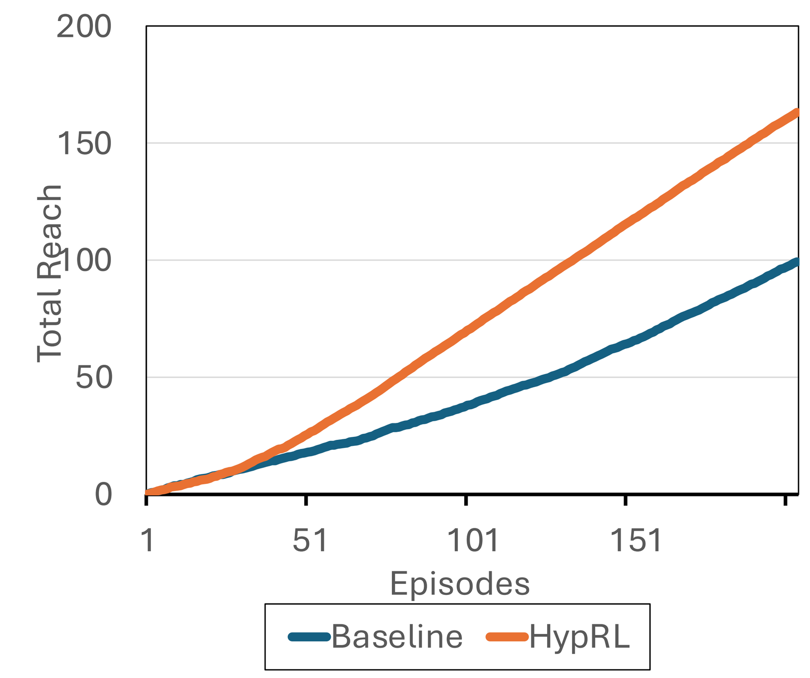
- We define robustness functions that assigns a robustness value to a trace for a formula, serving as quantifiable rewards in RL.

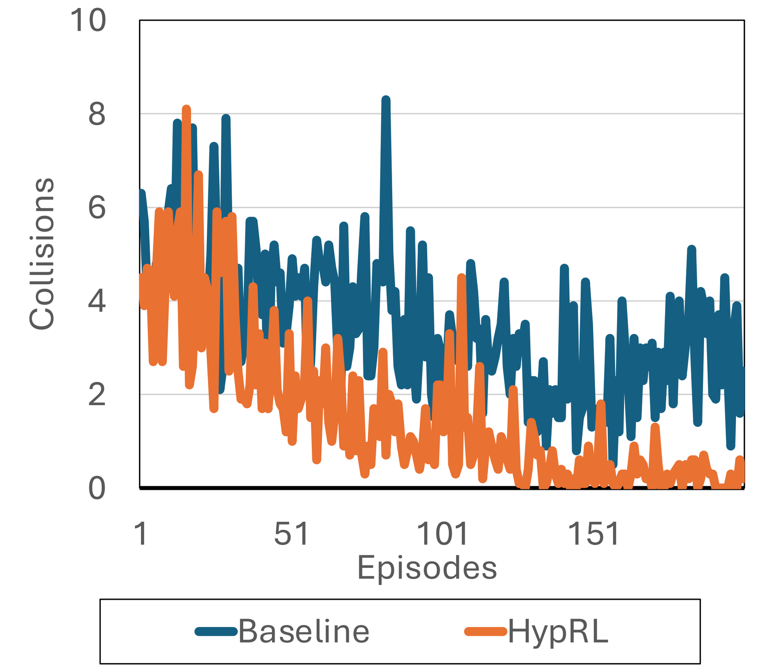
HypRL



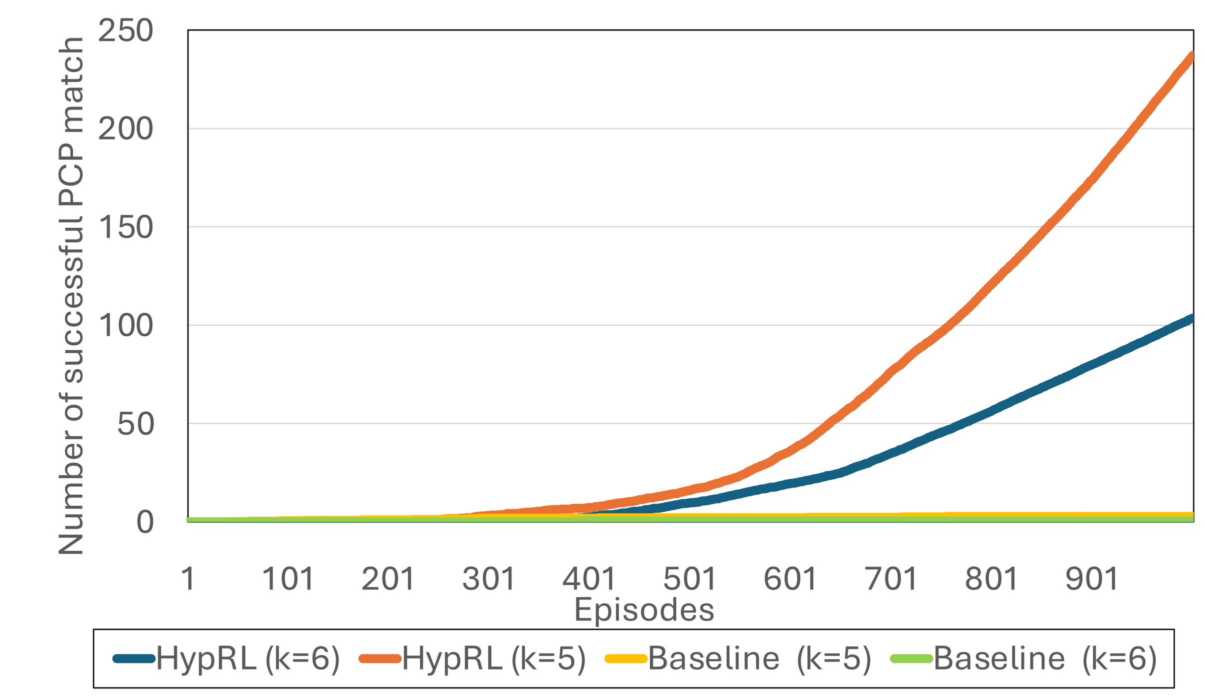
How efficient is HypRL?

- **Safe RL**





- **Post Correspondence Problem**



- **Deep Sea Treasures**

Method	ξ	Measurements $\sum$🧰	Measurements $\sum$🧰$/\beta$	Method	ξ	Measurements $\sum$🧰	Measurements $\sum$🧰$/\beta$
PPO	500	8.89	0.35	DQN	500	1.39	0.05
PPO + HypRL	500	21.80	0.87	DQN + HypRL	500	4.12	0.16
PPO	1000	8.19	0.352	DQN	1000	0.69	0.02
PPO + HypRL	1000	18.16	0.72	DQN + HypRL	1000	4.43	0.17

Take Aways!

- Hyperproperties capture objectives and constraints for multi-agent RL.
- HypRL turns any HyperLTL spec into a reward function, by Skolemizing formulas with quantifier alternations.
- HypRL outperforms baseline reward functions, used in traditional RL frameworks.