Deployment-time Selection of Prompts for LLM-informed Object Search in Partially-Known Environments

GEORGE MASON UNIVERSITY

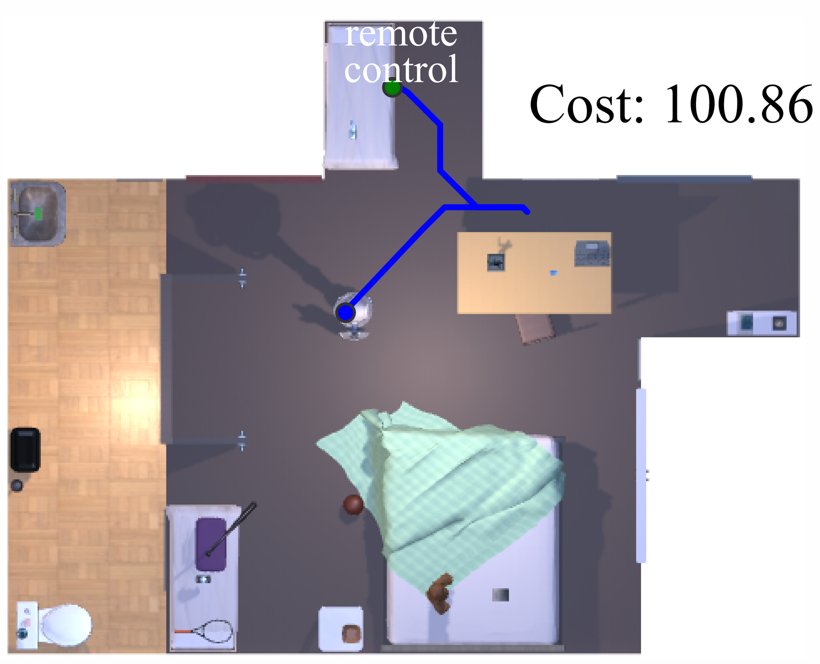
Abhishek Paudel and Gregory J. Stein

RAIL GROUP @ GMU

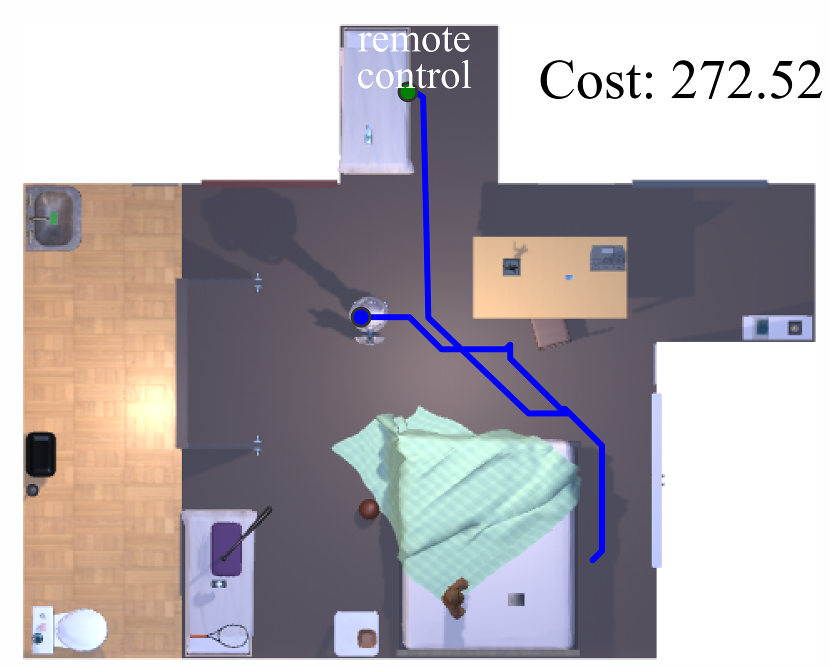
Problem Statement

We consider the problem of selecting best prompts and LLMs during deployment when a robot is deployed for LLM-informed object search tasks in partially-known environments.

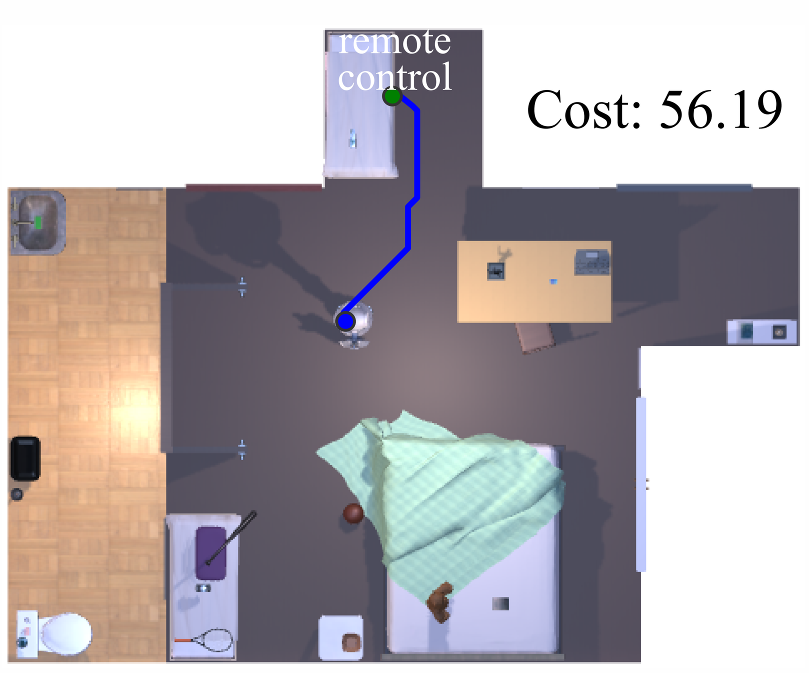
Different prompts lead to different behavior during deployment for object search tasks.



Prompt 1: P-CONTEXT-A



Prompt 2: P-CONTEXT-B



Prompt 3: P-MINIMAL

Prompt 4: P-DIRECT

Target Object: *remote control*

Applying black-box model selection for selecting prompts during deployment is too slow.

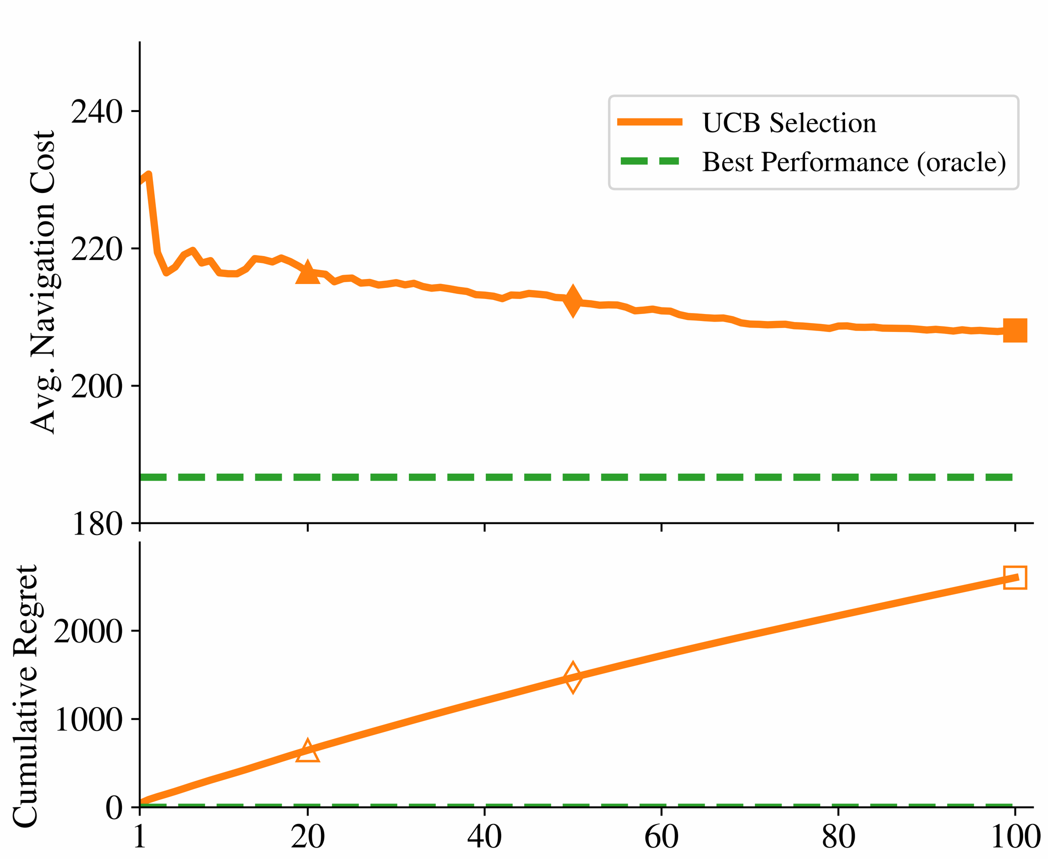
Upper Confidence Bound (UCB) bandit selection[1] trades off between exploitation and exploration to select among prompts and LLMs and takes many trials before converging towards the best prompt-LLM pair θ accumulating large regret.

UCB Bandit based prompt-LLM selection in trial k+1:

$$\pi_\theta^{(k+1)} = \underset{\pi_\theta \in \mathcal{P}}{\operatorname{argmin}} \left[\bar{C}_k(\pi_\theta) - c\sqrt{\frac{\ln k}{n_k(\pi_\theta)}} \right]$$

Exploit: Pick to minimize mean cost so far.

Explore: Pick to potentially improve performance.



Leveraging prior work [2], we seek to enable white-box selection of prompts and LLMs during deployment.

In trial k+1, policy π_θ with prompt-LLM pair θ to be selected for deployment is given by:

$$\pi_\theta^{(k+1)} = \underset{\pi_\theta \in \mathcal{P}}{\operatorname{argmin}} \left[\max\left(C_k^{\text{lb}}(\pi_\theta), \; \bar{C}_k(\pi_\theta) - c\sqrt{\frac{\ln k}{n_k(\pi_\theta)}} \right) \right]$$

Lower bound cost based on offline replay of prompt-LLM pair θ

Lower bound cost based on UCB algorithm

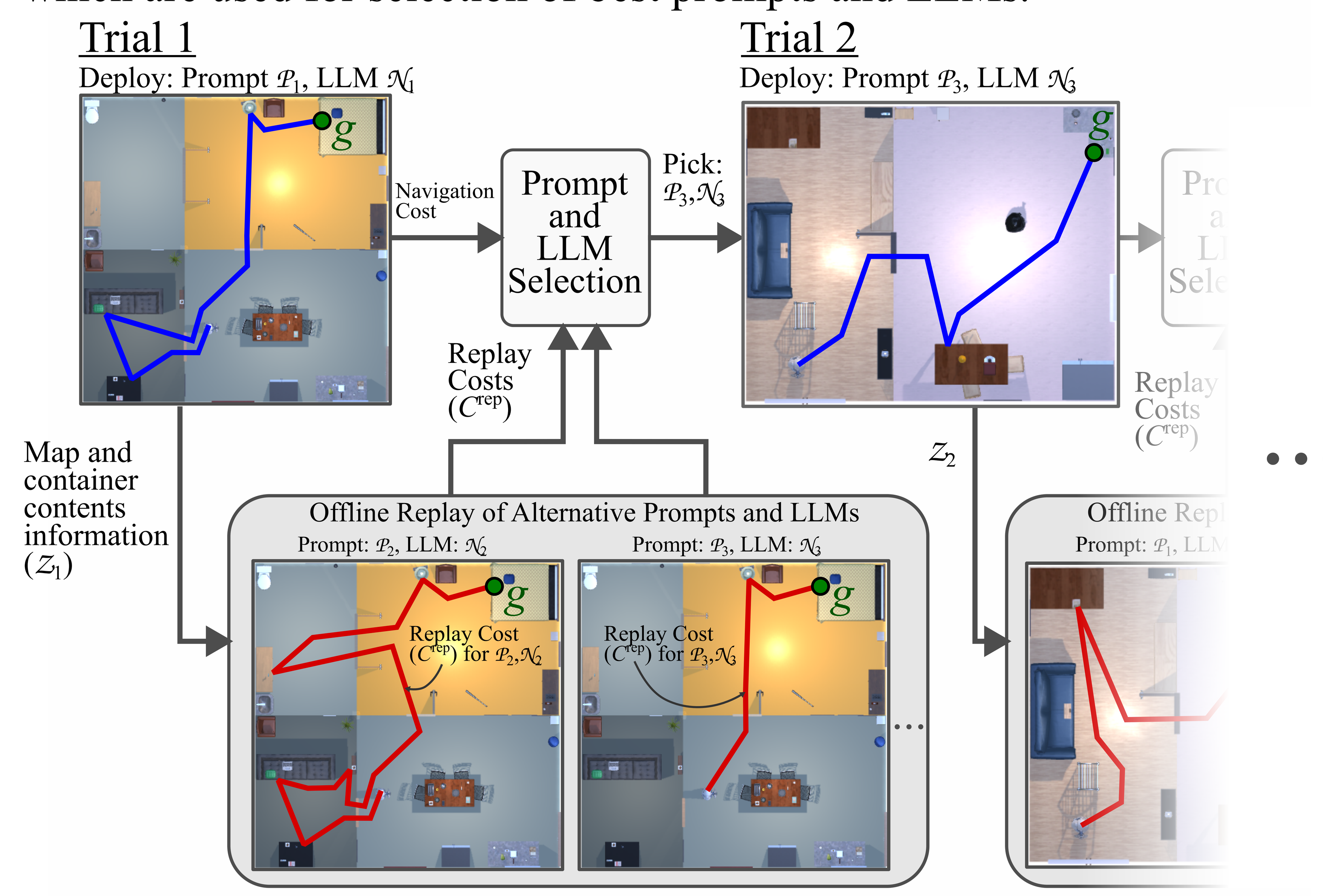
We use high-level action abstration and model-based planning [3] for object search in partially known environments.

$$Q_\pi(b_t, a_t \in \mathcal{A}(b_t)) = D(b_t, a_t) + R_{\text{search}}(b_t, a_t) + (1 - P_S(a_t))Q_\pi(b_t', \pi(b_t'))$$

Informed by LLM

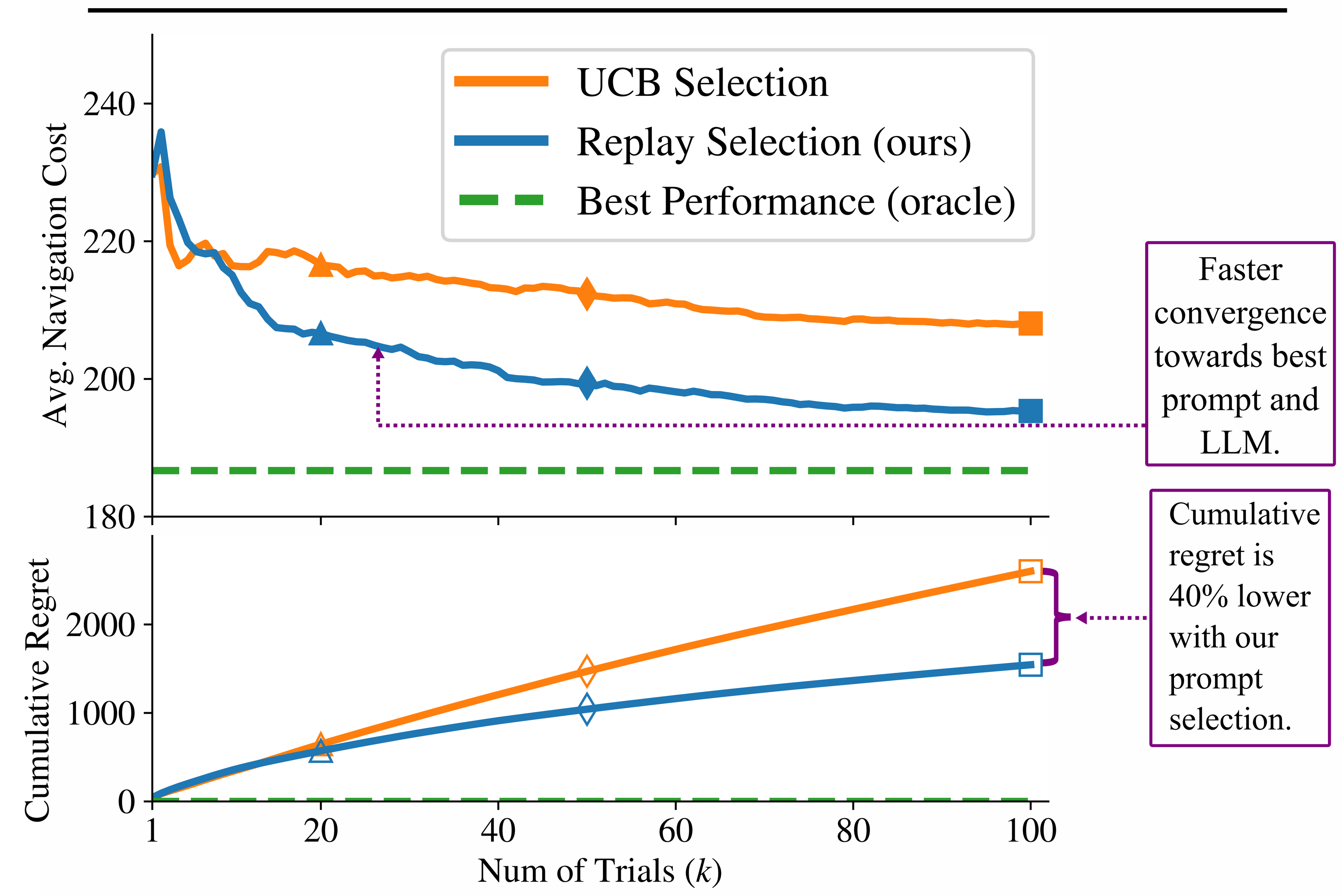
Our approach enables quick selection of best prompts and LLMs during deployment.

Leveraging prior work by Paudel and Stein [2] on white-box selection, our approach uses information collected during trial (e.g. objects found in explored containers) and the map to *replay* the robot behavior informed by all other alternative prompts and LLMs, the outcomes of which are used for selection of best prompts and LLMs.



Prompt Selection Results

Metric	Selection Approach	Num of Trials (k)		
		$k = 20$	$k = 50$	$k = 100$
Avg. Cost	UCB Selection	216.58	212.33	208.06
	Replay Selection (ours)	**206.63**	**199.38**	**195.35**
Cumul. Regret	UCB Selection	646.6	1470.5	2601.5
	Replay Selection (ours)	**572.1**	**1042.3**	**1544.7**



Performance when deploying only one policy/prompt/LLM combination:

Policy / Prompt / LLM	Avg. Cost
LLM+MODEL / P-CONTEXT-A / GPT-4o	227.66
LLM+MODEL / P-CONTEXT-B / GPT-4o	192.25
LLM+MODEL / P-MINIMAL / GPT-4o	205.55
LLM-DIRECT / P-DIRECT / GPT-4o	250.42
LLM+MODEL / P-CONTEXT-A / Gemini	**186.69**
LLM+MODEL / P-CONTEXT-B / Gemini	188.11
LLM+MODEL / P-MINIMAL / Gemini	225.49
LLM-DIRECT / P-DIRECT / Gemini	201.50
OPTIMISTIC+GREEDY / – / –	298.19

References

[1] Lai, T. L., and Robbins, H.. Asymptotically Efficient Adaptive Allocation Rules. Advances in Applied Mathematics. 1985.
[2] Paudel, A., and Stein, G. J.. Data-Efficient Policy Selection for Navigation in Partial Maps via Subgoal-Based Abstraction. International Conference on Intelligent Robots and Systems (IROS). 2023.
[3] Hossain S., Paudel, A., and Stein, G. J.. Enhancing Object Search by Augmenting Planning with Predictions from Large Language Models. CoRL Workshop on Learning Effective Abstractions for Planning (LEAP). 2024.

Paper

Acknowledgements

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 2232733.