NEURAL INFORMATION PROCESSING SYSTEMS

WALL-E: World Alignment by NeuroSymbolic Learning improves World Model-based LLM Agents

Siyu Zhou[1], Tianyi Zhou[2], Yijun Yang[3], Guodong Long[1], Deheng Ye[3], Jing Jiang[1], Chengqi Zhang[4]

[1]Australian AI Institute, University of Technology Sydney,

[2]University of Maryland, College Park, [3]Tencent, [4]Hong Kong Polytechnic University

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

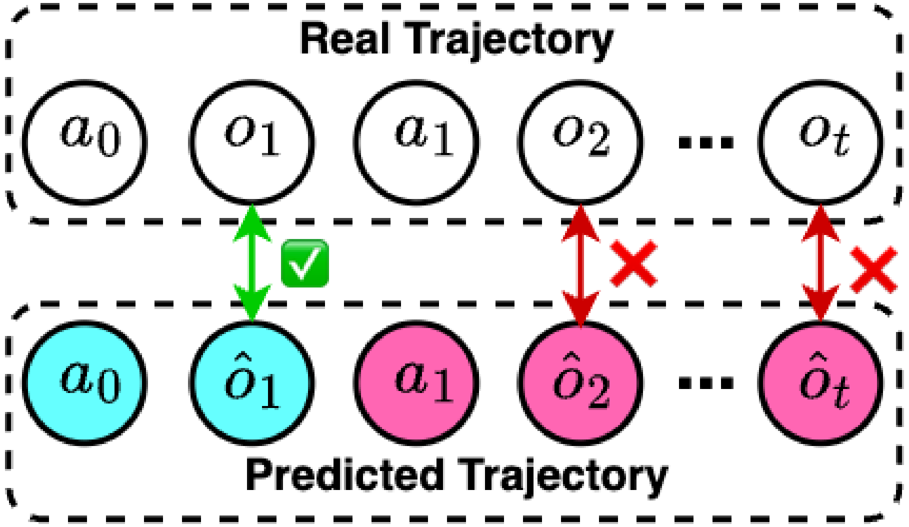
UNIVERSITY OF MARYLAND

Tencent 腾讯

Motivation

- Can LLMs serve as accurate world models?
- How can world models benefit LLM agents?
- Gap between LLMs' commonsense priors and environment dynamics.

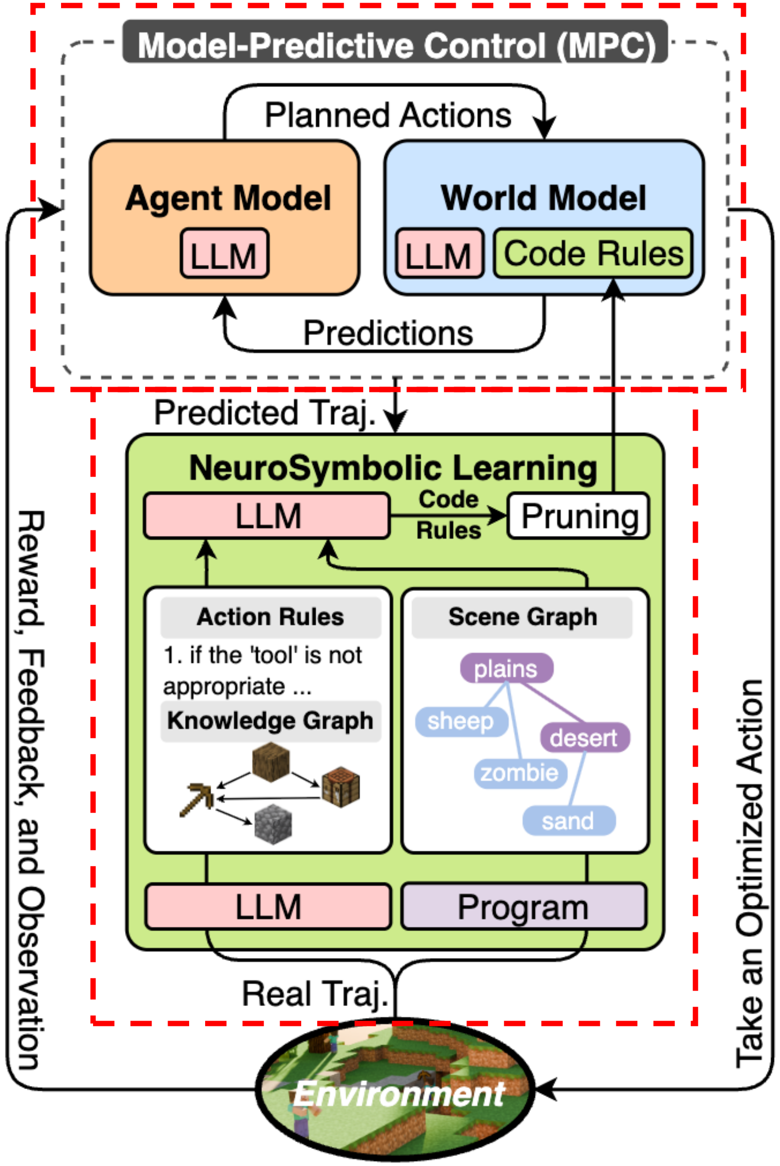
WALL-E mining diamond



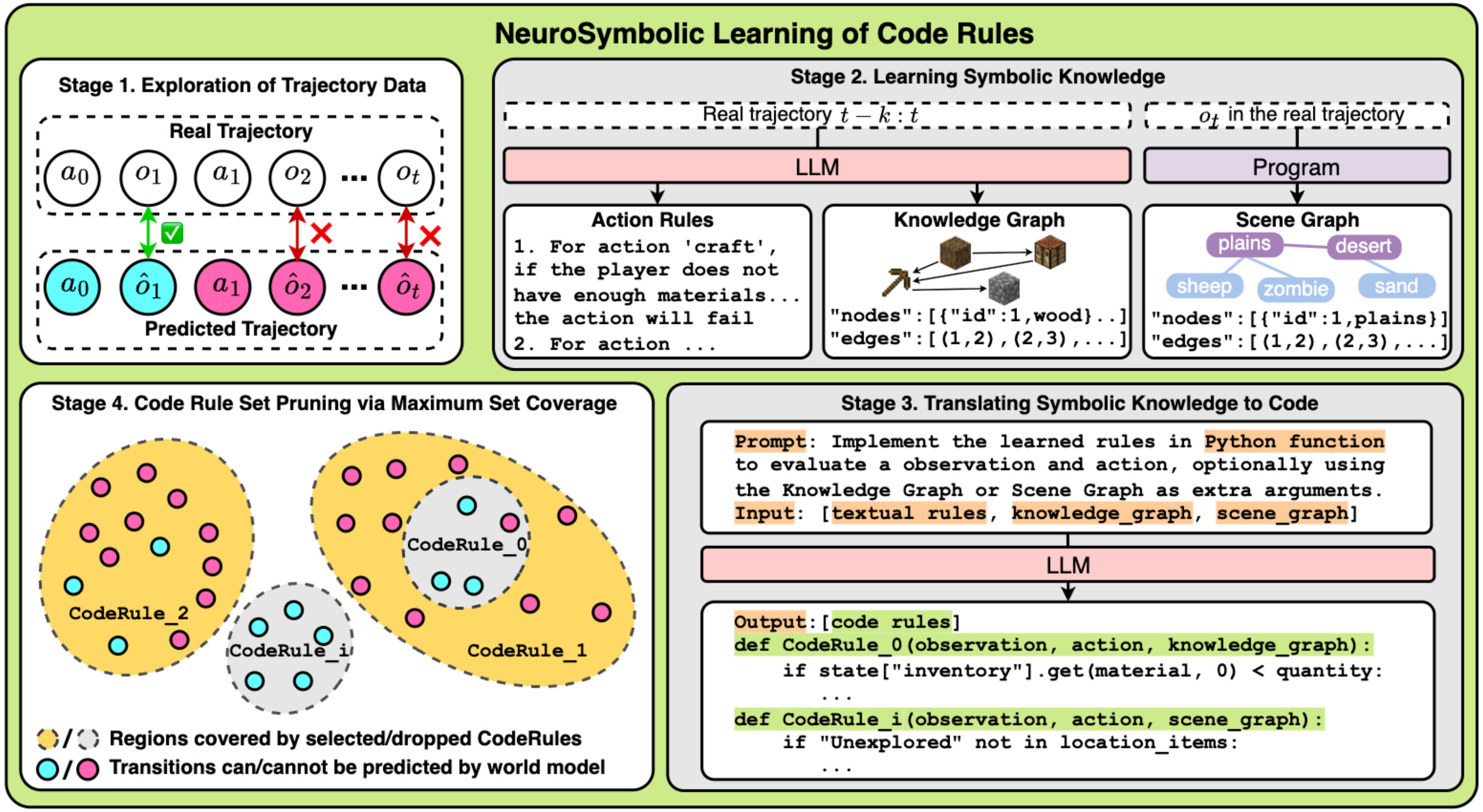
"LLM priors" vs "real environment transitions"

Overview of WALL-E

- We propose WALL-E, a training-free method for world alignment.
- Two alternating stages:
  - NeuroSymbolic Learning (NSLearning) extract & encode symbolic knowledge.
  - Model-Predictive Control (MPC) use aligned world model for look-ahead planning.
- Enables reliable, sample-efficient model-based LLM agents.
- Training-free; no gradients, no fine-tuning.



NeuroSymbolic Learning of Code Rules



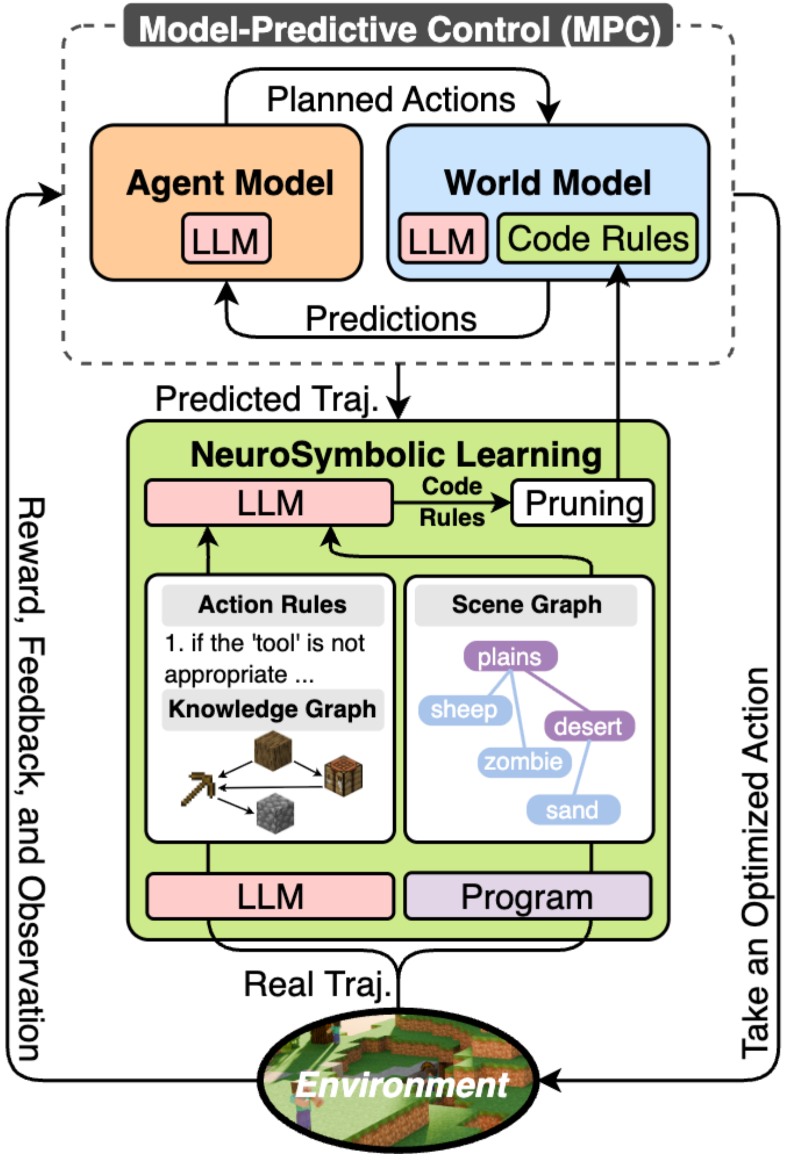
1. Compare predicted vs real trajectories.
2. Extract Action Rules, Knowledge Graph, Scene Graph.
3. Translate symbolic knowledge → Python code rules.
4. Prune rules via verification + maximum coverage.

Model-Predictive Control (MPC)

- Agent queries world model: “What happens if I do a_t in o_t?”
- World model predicts: next state, success flag, feedback, suggestion.
- Code rules verify and correct predictions.
- LLM agent replans until predictions pass all checks.

Algorithm 1 Model-Predictive Control (MPC)

Input: o_t, $\mathcal{R}^{\text{code}}$
Initialize: $fb \leftarrow []$, $sug \leftarrow []$, $count \leftarrow 0$
repeat
  $a_t \leftarrow \text{LLMAGENT}(o_t, fb, sug)$ /*eq.(10)*/
  $\hat{o}_{t+1}, fb, sug, flag \leftarrow \text{ALIGN}(\mathcal{R}^{\text{code}}, f_{\text{LLMwm}}, o_t, a_t)$ /*eq.(8)*/
  $count \leftarrow count + 1$
  if $flag$ **then**
    break /*Action accepted*/
  end if
until $count \geq \text{REPLANLIMIT}$
Output: a_t, $\hat{o}_{t+1}$



Experiments & Results

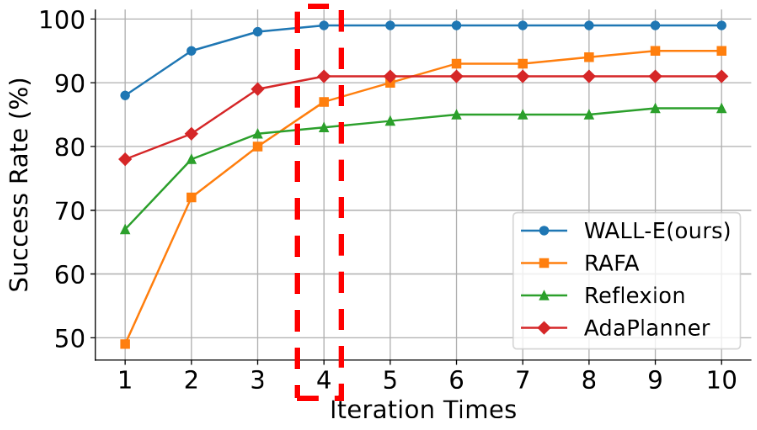
Mars (Minecraft-like)

Metrics	Mod. Type	RL-based methods		LLM-based methods				
		PPO*	DreamerV3*	ReAct*	Reflexion*	Skill Library*	IfR*	WALL-E
Reward ↑	Default	1.9 ± 1.4	11.5 ± 1.6	7.7 ± 1.6	6.0 ± 1.7	8.0 ± 2.1	9.0 ± 2.3	$\mathbf{9.5 \pm 2.1}$
	Terrain	-0.1 ± 0.6	9.3 ± 2.2	7.4 ± 2.7	6.4 ± 3.0	9.5 ± 2.9	8.0 ± 3.7	$\mathbf{10.7 \pm 2.6}$
	Survival	-0.6 ± 0.5	8.6 ± 4.1	6.4 ± 3.7	4.6 ± 3.9	7.9 ± 2.9	7.7 ± 3.7	$\mathbf{13.8 \pm 4.4}$
	Task. Dep.	2.1 ± 1.2	8.8 ± 2.8	5.0 ± 2.1	3.2 ± 1.6	1.5 ± 1.9	5.6 ± 2.9	$\mathbf{6.4 \pm 2.9}$
	Terr. Surv.	0.0 ± 0.7	7.1 ± 2.1	6.7 ± 2.5	4.9 ± 2.5	3.0 ± 2.5	$\mathbf{6.8 \pm 1.9}$	5.5 ± 2.7
	Terr. Task.	-0.7 ± 0.3	6.6 ± 0.7	4.8 ± 2.0	5.3 ± 2.5	5.5 ± 1.5	$\mathbf{6.9 \pm 1.8}$	5.8 ± 2.2
	Surv. Task.	-0.6 ± 0.4	9.6 ± 3.4	1.5 ± 1.3	1.0 ± 1.6	2.3 ± 1.5	$\mathbf{3.3 \pm 1.4}$	3.2 ± 1.4
	All three	0.1 ± 0.8	5.1 ± 1.8	0.7 ± 1.6	-0.4 ± 0.7	-0.5 ± 0.5	0.1 ± 0.5	$\mathbf{1.3 \pm 1.6}$
	Average	0.0	7.9	4.6	3.6	4.2	5.5	**6.7**
Score (%) ↑	Default	1.3 ± 1.7	14.2 ± 1.3	8.0 ± 1.5	5.3 ± 0.9	8.3 ± 1.3	13.0 ± 2.1	$\mathbf{20.3 \pm 1.8}$
	Terrain	0.3 ± 0.1	13.0 ± 1.6	7.6 ± 2.6	7.4 ± 1.6	11.9 ± 3.4	11.8 ± 2.9	$\mathbf{27.8 \pm 1.7}$
	Survival	0.2 ± 0.0	10.8 ± 2.8	8.0 ± 0.6	5.5 ± 1.7	9.7 ± 2.0	11.0 ± 3.7	$\mathbf{50.8 \pm 1.1}$
	Task. Dep.	1.7 ± 0.6	12.1 ± 1.9	4.6 ± 1.6	2.2 ± 0.8	1.5 ± 0.6	6.9 ± 2.5	$\mathbf{9.3 \pm 2.0}$
	Terr. Surv.	0.4 ± 0.1	7.9 ± 1.3	7.1 ± 3.0	4.7 ± 1.6	2.8 ± 0.6	6.7 ± 0.8	$\mathbf{8.6 \pm 1.9}$
	Terr. Task.	0.1 ± 0.1	4.2 ± 0.1	3.8 ± 0.3	5.5 ± 1.7	4.1 ± 0.7	$\mathbf{7.1 \pm 2.5}$	4.7 ± 2.0
	Surv. Task.	0.1 ± 0.1	15.9 ± 2.6	1.3 ± 0.2	1.1 ± 0.1	1.9 ± 0.1	2.1 ± 0.4	$\mathbf{3.3 \pm 1.9}$
	All three	0.6 ± 0.2	4.0 ± 0.3	1.0 ± 0.3	0.2 ± 0.1	0.2 ± 0.0	0.6 ± 0.0	$\mathbf{2.2 \pm 1.6}$
	Average	0.6	10.3	5.2	4.0	5.0	7.4	**15.3**

+16 – 52% reward gain in Mars.

ALFWorld

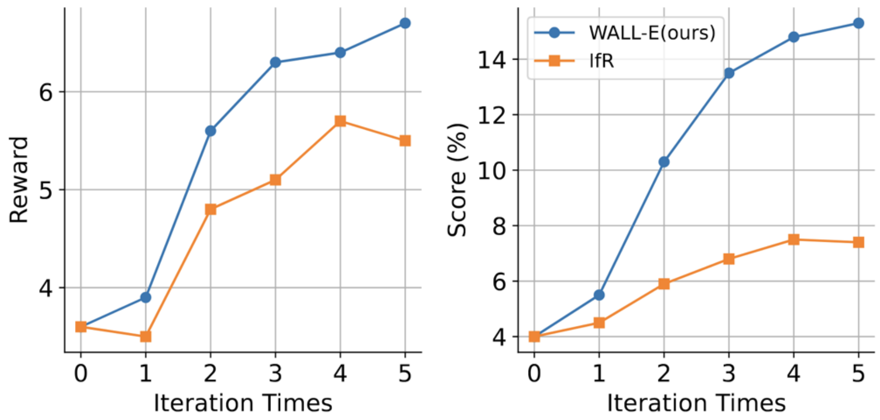
Method	Success Rate (%) ↑						
	Avg.	Pick	Clean	Heat	Cool	Examine	Picktwo
BUTLER	26	31	41	60	27	12	29
GPT-BUTLER	69	62	81	85	78	50	47
DEPS	76	93	50	80	**100**	**100**	0
AutoGen	77	92	74	78	86	83	41
ReAct	74	79	54	96	85	83	51
AdaPlanner	91	**100**	**100**	89	**100**	97	47
Reflexion	86	92	94	70	81	90	88
RAFA	95	**100**	97	91	95	**100**	82
WALL-E (ours)	**98**	**100**	**100**	**96**	**100**	**100**	**94**
Human Performance	91	-	-	-	-	-	-



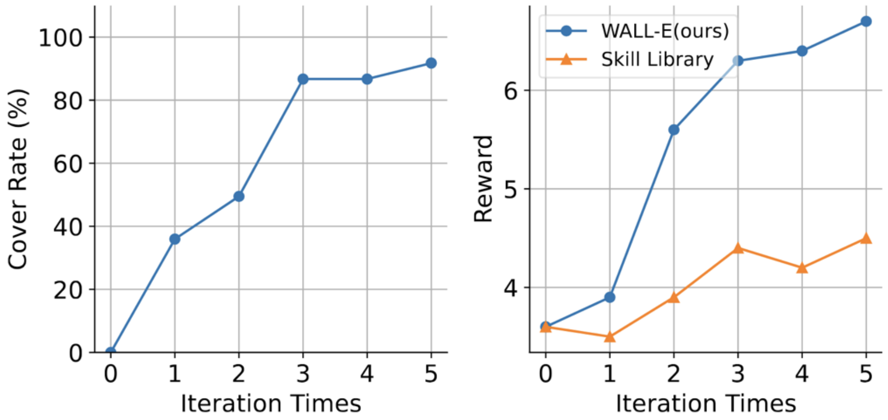
98% success in ALFWorld after 4 iterations.

Analysis & Ablations

Comparison between WALL-E and induction from reflection (IfR)



Comparison between WALL-E and LLM agents with skill library



Ablation study of WALL-E with different symbolic knowledge (KNWL) types

Symbolic KNWL		Mars		ALFWorld
Action Rules	KG/SG	**Reward** ↑	**Score (%)** ↑	**Success Rate (%)** ↑
✓		5.1	8.3	95
	✓	4.4	5.2	88
✓	✓	6.7	15.3	98

Ablation study on the code rule set pruning stage in Mars

Metrics	**w/o Pruning**	**w/ Pruning - Rule Set Limit**				
		1	3	5	7	**no limit (9)**
Reward ↑	1.5	4.3	5.7	6.2	6.5	**6.7**
Score (%) ↑	1.6	8.1	9.5	12.8	14.5	**15.3**

Conclusion & Takeaways

- **World Alignment**: LLMs become accurate **world models** when aligned with **symbolic knowledge**.
- **WALL-E**: A **model-based neurosymbolic agent** powered by **MPC** and **training-free** rule learning.
- **Reliable & Adaptive:** Better prediction, safer decisions, rapid adaptation in **open-world** tasks.
- **Scalable Path Forward**: Toward trustworthy **LLM-based embodied agents**.

Thanks for listening!

Siyu Zhou
Siyu.Zhou-2@student.uts.edu.au