NEURAL INFORMATION PROCESSING SYSTEMS

ZHEJIANG UNIVERSITY

NUS National University of Singapore

Alibaba

Agent Planning with World Knowledge Model

Shuofei Qiao♠*, Runnan Fang♠*, Ningyu Zhang♠†, Yuqi Zhu♠, Xiang Chen♠, Shumin Deng♣, Yong Jiang♢, Pengjun Xie♢, Fei Huang♢, Huajun Chen♠†

♠Zhejiang University
♣National University of Singapore ♢Alibaba Group

https://github.com/zjunlp/WKM

CONTENT

CONTENT

A Motivation Example

Task: clean the soapbar and put it in cabinet.

Task Instruction: You are in the middle of a room. Looking quickly around you, you see a cabinet 4, a cabinet 3, a cabinet 2, a cabinet 1, a countertop 1, a garbagecan 1, a handtowelholder 2, a handtowelholder 1, a sinkbasin 2, a sinkbasin 1, a toilet 1, a toiletpaperhanger 1, and a towelholder 1. Your task is to: clean some soapbar and put it in cabinet.

 # Blind trial-and-error begins.

Agent: go to countertop 1
Observation: On the countertop 1, you see a candle 3, a candle 1, a soapbottle 1, a spraybottle 2, and a toiletpaper 2.
Agent: go to cabinet 1
Observation: The cabinet 1 is closed.
Agent: open cabinet 1
Observation: You open the cabinet 1. The cabinet 1 is open. In it, you see a soapbottle 2.
Agent: close cabinet 1
Observation: You close the cabinet 1.

 # Blind trial-and-error ends.

Agent: go to toilet 1
Observation: On the toilet 1, you see a soapbar 1, and a spraybottle 3.
Agent: take soapbar 1 from toilet 1
Observation: You pick up the soapbar 1 from the toilet 1.
Agent: go to sinkbasin 1
Observation: On the sinkbasin 1, you see nothing.
Agent: clean soapbar 1 with sinkbasin 1
Observation: You clean the soapbar 1 using the sinkbasin 1.
Agent: go to cabinet 1
Observation: The cabinet 1 is closed.
Agent: open cabinet 1
Observation: You open the cabinet 1. The cabinet 1 is open. In it, you see a soapbottle 2.
Agent: close cabinet 1
Observation: You close the cabinet 1.

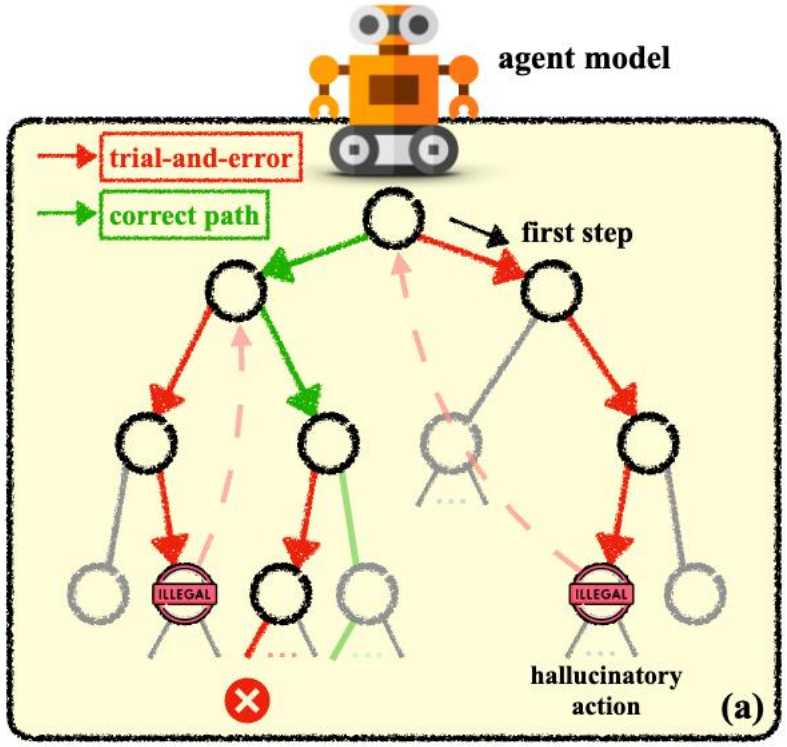
 # Hallucinatory action.

Agent: put soapbar 1 in/on cabinet 1
Observation: Nothing happens.
Agent: check cabinet 1
Observation: Nothing happens.
Agent: go to exit
Observation: Nothing happens.
Agent: exit
Observation: Nothing happens.
.....
Reward: 0.0

The soapbar is most likely in the toilet, but the agent first randomly searches the countertop and cabinet.

The agent wants to put the soapbar into the cabinet after closing the cabinet.

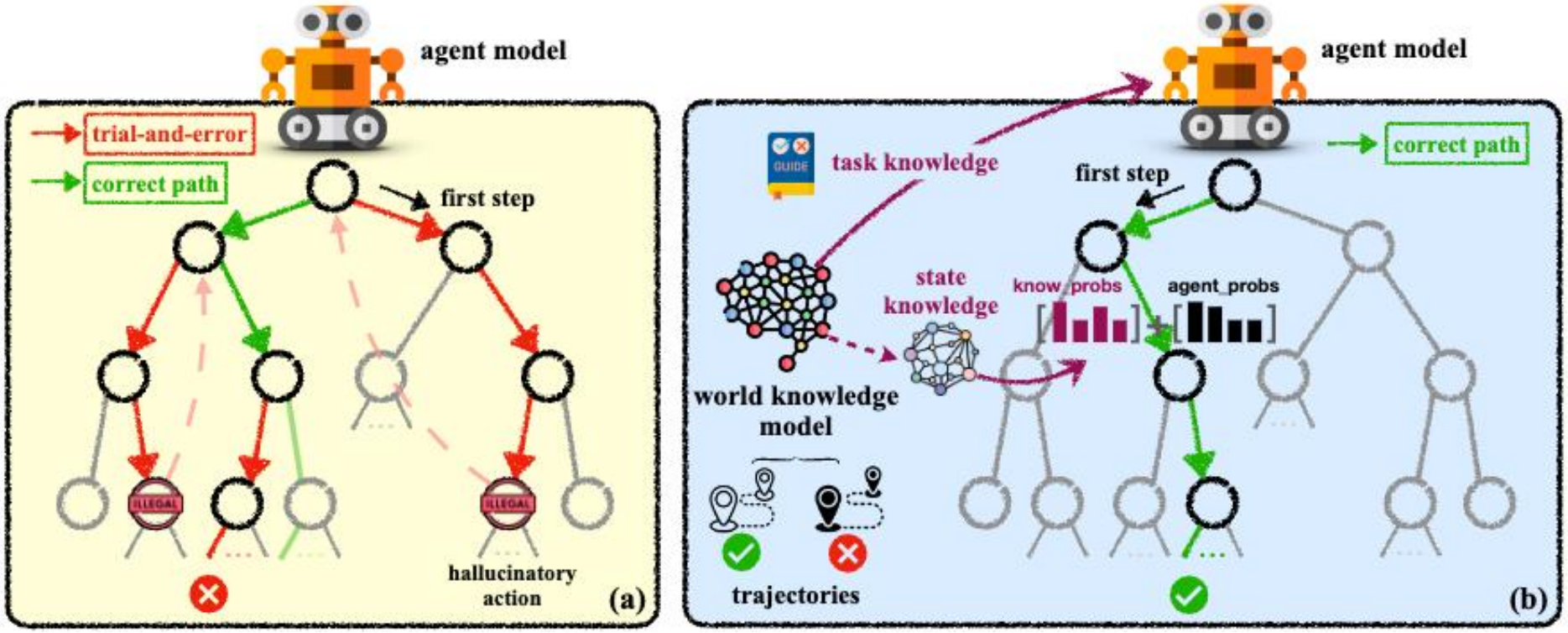
Background



blind trial-and-error → need → static prior knowledge

hallucinated actions → need → dynamic state knowledge

Introduction



CONTENT

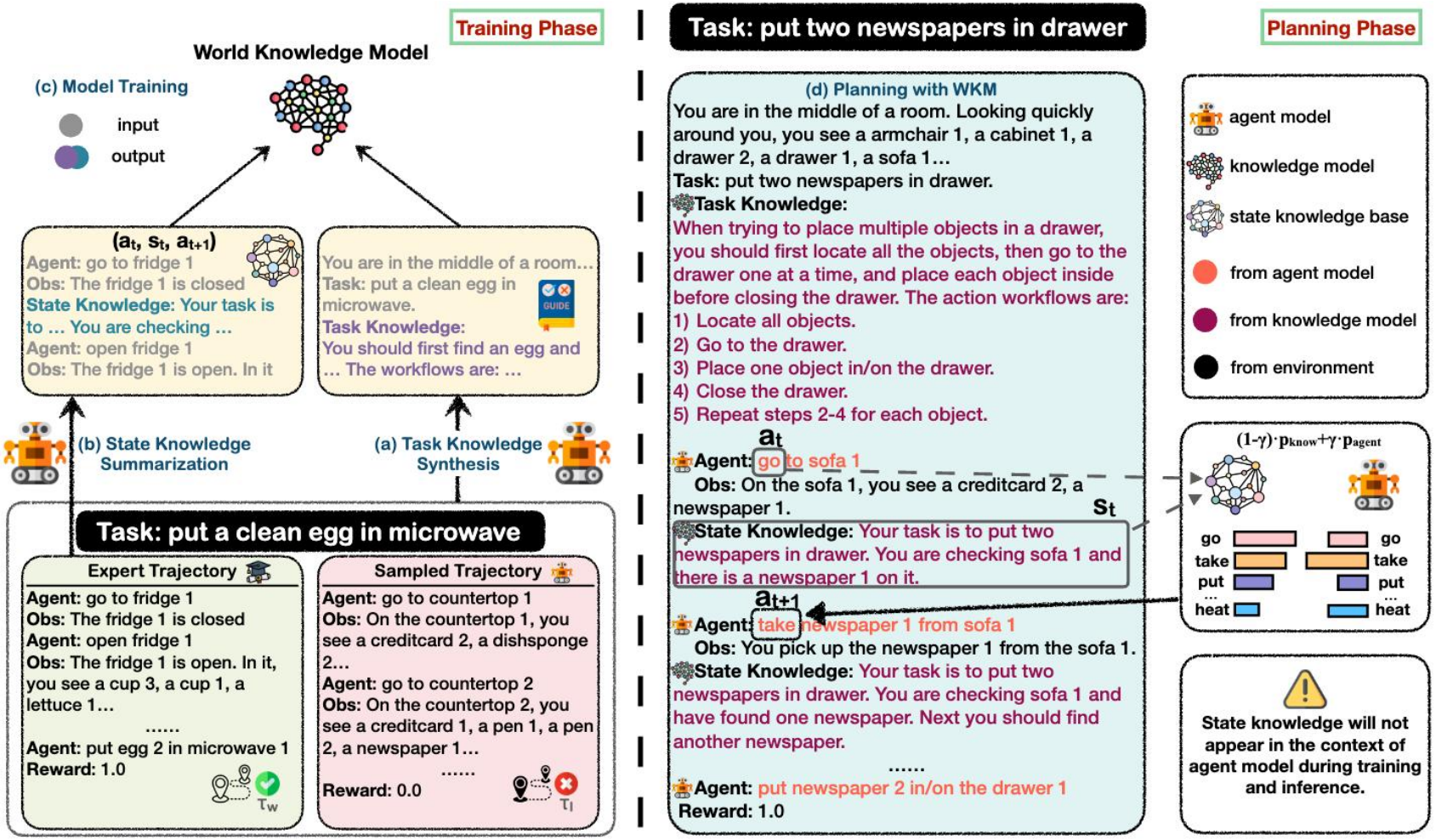
Introduction & Background

WKM: Agent Planning with World Knowledge Model

Experimental Analysis

Discussion & Take Away

Framework of WKM



- We train a world knowledge model on the knowledge synthesized by the agent model itself from both expert and explored trajectories, providing prior task knowledge to guide global planning and dynamic state knowledge to assist local planning.

CONTENT

Introduction & Background

WKM: Agent Planning with World Knowledge Model

Experimental Analysis

Discussion & Take Away

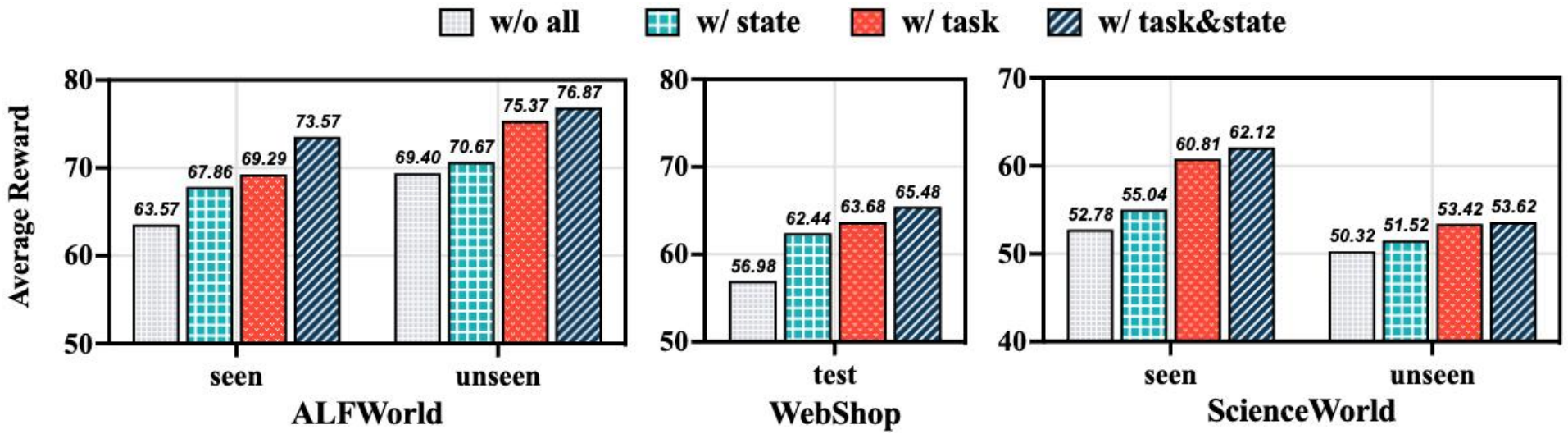
Experiments: Main Results

Backbone	Method	ALFWorld		WebShop	ScienceWorld	
		Seen	Unseen		Seen	Unseen
GPT-3.5-Turbo	REACT	8.57	5.97	44.37	15.41	13.99
GPT-4	REACT	44.29	38.05	62.76	67.32	65.09
Mistral-7B	REACT	7.86	5.22	14.63	20.72	17.65
	Reflexion	11.56	6.00	16.64	21.07	18.11
	NAT	64.43	68.96	61.01	57.12	50.79
	ETO	66.84	71.43	64.09	58.17	51.85
	KNOWAGENT	70.44	70.72	61.28	59.32	47.24
	WKM	**73.57** +3.13	**76.87** +5.44	**65.48** +1.39	**62.12** +2.80	**53.62** +1.77
Gemma-7B	REACT	6.43	2.24	5.93	3.58	3.51
	Reflexion	7.14	2.99	7.71	4.94	3.93
	NAT	67.86	65.88	55.82	47.63	44.98
	ETO	66.43	68.66	62.67	50.44	47.84
	KNOWAGENT	69.29	67.60	58.80	48.55	45.28
	WKM	**70.71** +1.42	**70.40** +1.74	**63.75** +1.08	**53.68** +3.24	**49.24** +1.40
Llama-3-8B	REACT	2.86	3.73	19.32	24.76	22.66
	Reflexion	4.29	4.48	22.73	27.23	25.41
	NAT	60.71	59.70	61.60	55.24	48.76
	ETO	64.29	64.18	64.57	57.90	52.33
	KNOWAGENT	66.71	62.69	64.40	58.67	49.18
	WKM	**68.57** +1.86	**65.93** +1.75	**66.64** +2.07	**60.12** +1.55	**54.75** +2.42

WKM performs well compared to various baselines on different models and datasets, outperforming GPT-4 on ALFWorld and WebShop

Experiments: Ablations

- We analyze the effectiveness of task knowledge and state knowledge respectively.



Findings:

- The improvement through task knowledge is more pronounced than that through state knowledge.
- The impact of state knowledge is more significant on seen tasks compared to unseen tasks, while the influence of task knowledge is sustainable across seen and unseen tasks.

Experiments: Analysis of average planning steps and hallucinated action rates

➢ Average planning steps

Method	ALFWorld		WebShop	ScienceWorld	
	Seen	Unseen		Seen	Unseen
NAT	23.27	23.42	4.08	20.18	21.21
ETO	19.82	22.29	3.99	24.13	26.35
KNOWAGENT	18.51	24.56	4.01	21.06	24.74
WKM	**17.66**	**17.92**	**3.97**	**18.74**	**19.59**

➢ Hallucinated action rates

Method	ALFWorld	
	Seen	Unseen
NAT	45.71%	50.00%
ETO	34.29%	36.57%
KNOWAGENT	33.57%	44.78%
WKM	**32.86%**	**29.85%**

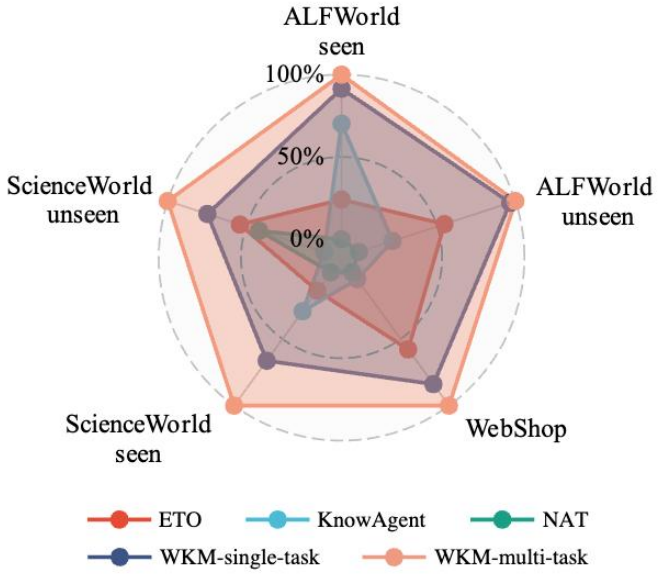
WKM can mitigate blind trial-and-error and reduce hallucinatory actions. And it can maintain stability or even reduce the ratio on unseen tasks.

Experiments: Other Interesting Analysis

- Weak-guide-strong (Mistral-7B guide GPT-3.5/4)

Backbone	Method	ALFWorld	
		Seen	Unseen
GPT-3.5-Turbo	REACT	8.57	5.97
	WKM w/o state	**12.86**	**8.96**
GPT-4	REACT	44.29	38.05
	WKM w/o state	**50.71**	**47.01**

- Unified WKM



1) The knowledge generated by Mistral-7B can effectively guide the planning of GPT-3.5/4;
2) The unified parametric knowledge model jointly trained through multi-task performs better than single-task training.

CONTENT

Discussion & Take Away

- 1) Our primary intention behind designing the WKM is to compensate for the lack of world knowledge in the agent model. However, determining **what a language model knows and doesn't know** has been an ongoing challenge that remains unresolved.
- 2) It is widely acknowledged that world knowledge extends beyond textual representations. While our world knowledge is currently limited to textual information, exploring **multi-modal world knowledge models** is indeed one of our important future tasks.
- 3) Our world knowledge model **cannot dynamically update** with the changes of the world and feedback from the agent.
- 4) Generating world knowledge can introduce **additional inference overhead**.

浙江大學
ZHEJIANG UNIVERSITY

Thanks for listening!

https://github.com/zjunlp/WKM

ACCEPT MY ENDLESS GRATITUDE