



清华大学
Tsinghua University

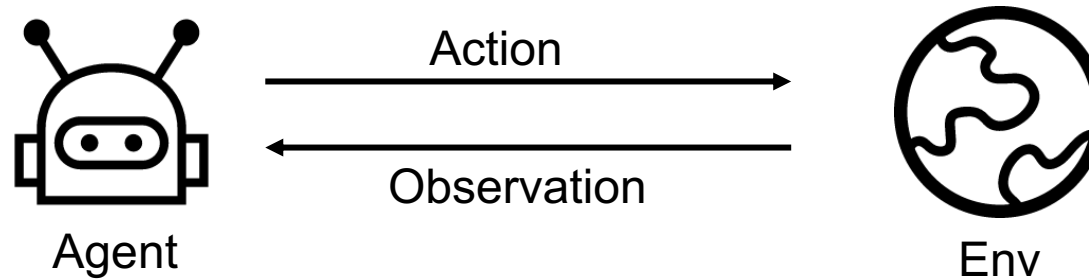
Agent-Environment Alignment via Automated Interface Generation

Kaiming Liu

2025.8.8

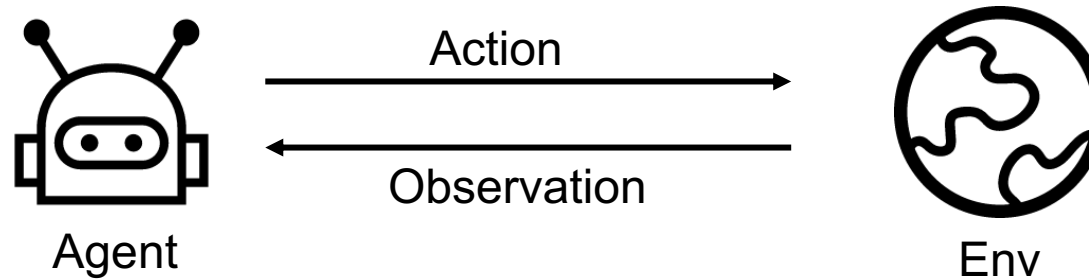
Background: Agent-Environment Misalignment

- Agents interact with environment:



Background: Agent-Environment Misalignment

- Agents interact with environment:



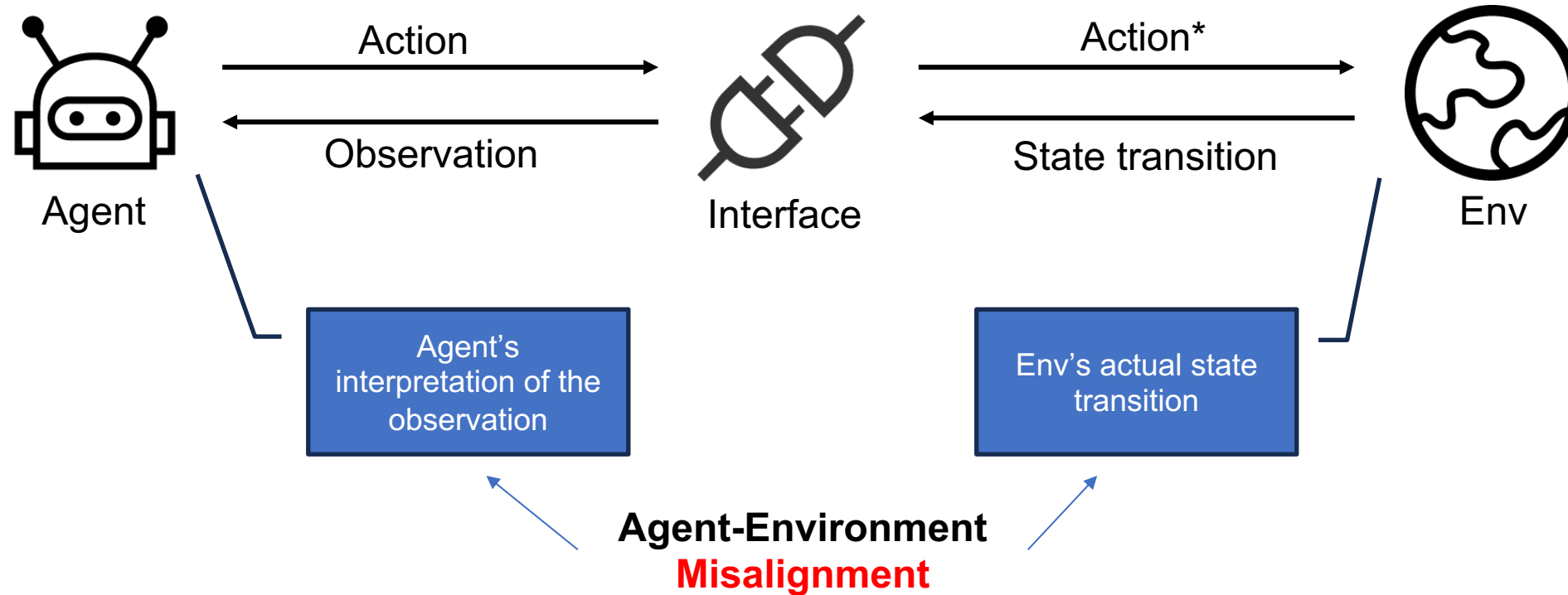
- In the complete setting, agents interact with environment through intermediate **interfaces**:



Such as predefined action spaces. Often given by environment designers.

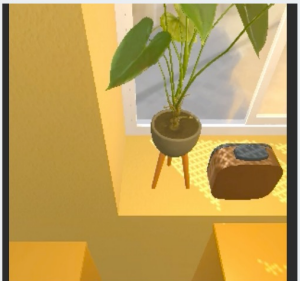
Background: Agent-Environment Misalignment

- Mismatches happen between the **internal expectations of the agent** regarding the influence of its issued **actions** and the **actual state transitions** in the environment.



Background: Agent-Environment Misalignment

- Mismatches happen between the **internal expectations of the agent** regarding the influence of its issued **actions** and the **actual state transitions in the environment**.



You are in the middle of a room. Looking quickly around you, you see a safe 1, a shelf 4, a drawer 2, a bed 1, a drawer 1, a shelf 5, a shelf 2, a sidetable 2, a shelf 3, a drawer 3, a shelf 1, a sidetable 1, a desk 1, and a garbagecan 1.

Your task is to: *examine an alarmclock with the desklamp.*

> go to desk 1

You arrive at loc 8. On the desk 1, you see a pen 1, a bowl 1, a alarmclock 2, a pencil 2, a pencil 3, a creditcard 3, a book 1, a alarmclock 3, a keychain 3, and a book 2.

> take alarmclock 2 from desk 1

You pick up the alarmclock 2 from the desk 1.

> go to sidetable 2

You arrive at loc 1. On the sidetable 2, you see a desklamp 1, and an alarmclock 1.

> use desklamp 1

You won!

Next Example

Agents need first `go to receptacle` then execute `examine receptacle`

Agent-Environment Interaction



THOUGHT: The book may be on the shelf. First I need to examine the shelves.

ACTION: examine shelf 1



OBSERVATION: Nothing happens.

Agent's Inner Monologue



THOUGHT: The environment returned "Nothing happens." This means the book not on shelf 1.

Environment Logic



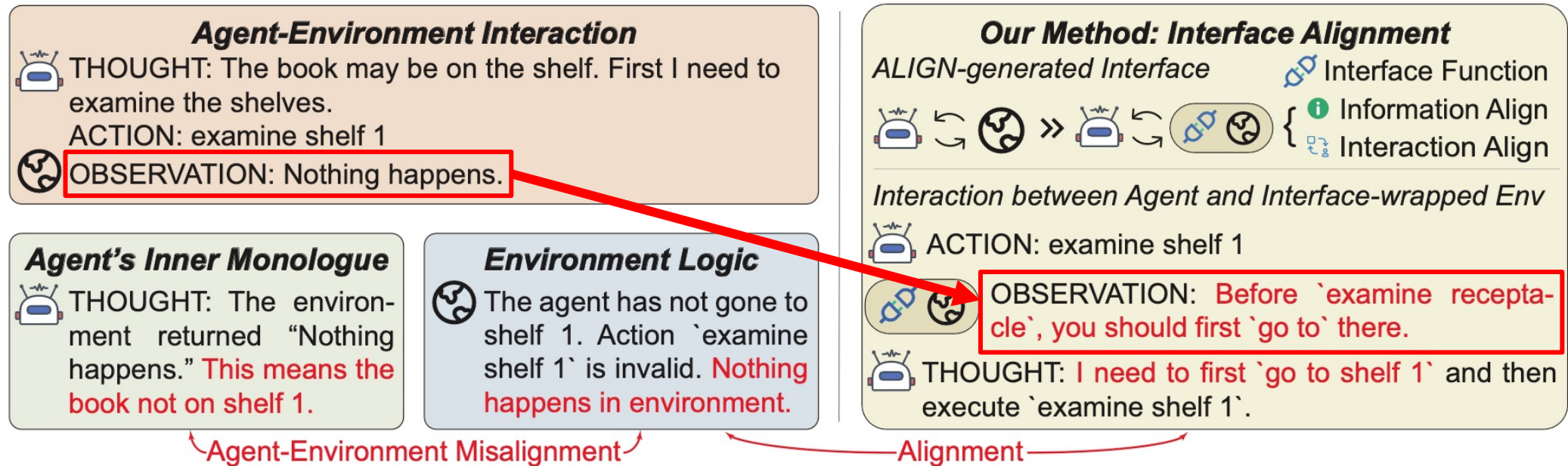
The agent has not gone to shelf 1. Action `examine shelf 1` is invalid. Nothing happens in environment.

Agent-Environment Misalignment

Preliminary Experiment: Observation Adjustment

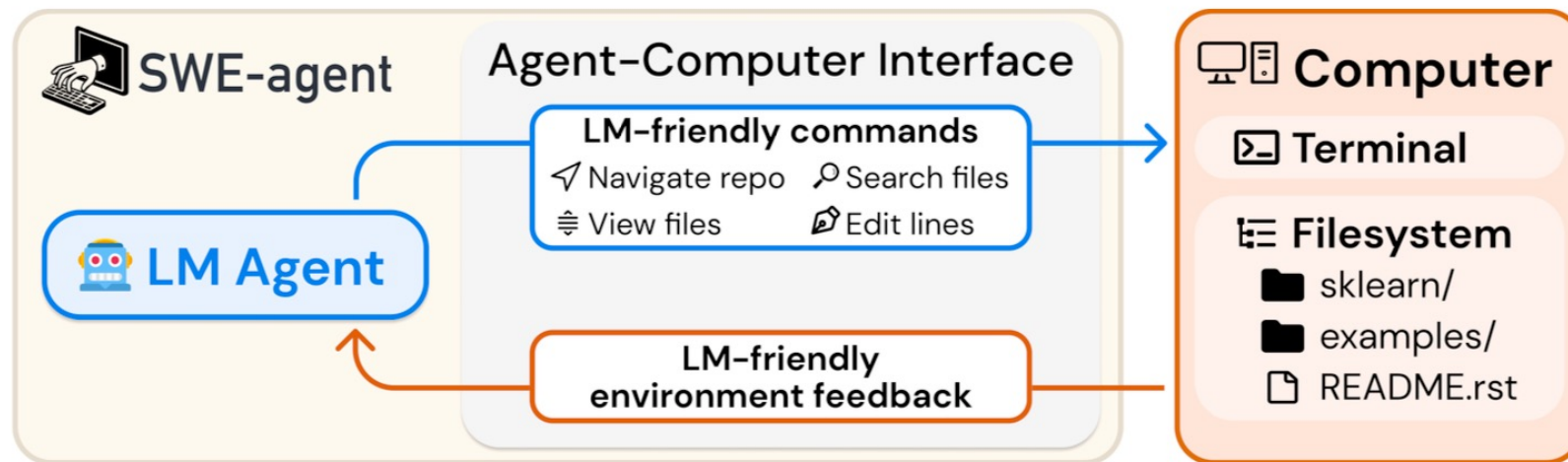
Preliminary Experiment: Correcting observations only when "Examine" is invalid improved task accuracy **from 13.4% to 31.3%**:

- The importance of addressing agent-environment misalignment in improving agent performance;
- Designing more robust interfaces may be an effective way to resolve agent-environment misalignments.



Representative Method: Human-designed Interface

- Since each environment and even each method requires a separately designed interface, this results in very **high labor costs**.
- Whether manually designed interfaces are optimal still requires further investigation.



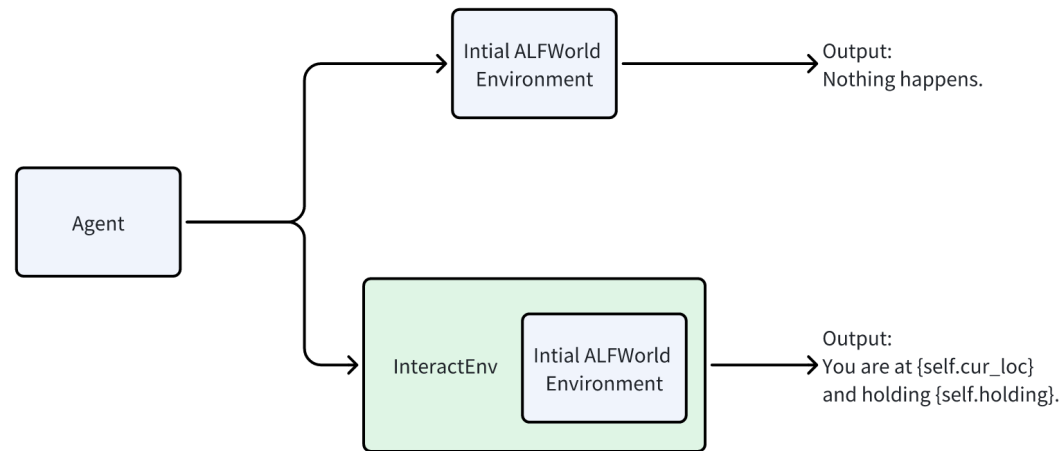
Representative Method: Method-customized Interfaces

Agent methods that do not explicitly specify interface design often still involve customized interface design:

- **WALL-E** manually maintains environment state information in JSON format [1];
- **AgentBoard** adds a new action, *check valid actions*, allowing the agent to obtain a list of valid actions [2];
- **AutoManual** wraps a new class, `InteractEnv`, to reimplement the interaction mechanism of ALFWorld [3].

New challenges:

- It becomes **difficult to directly compare results across methods** — are performance differences due to the methods themselves or the customized interfaces?



[1] Zhou, Zhou, et al. WALL-E: world alignment by rule learning improves world model-based LLM agents. CoRR 2024

[2] Ma, Zhang, et al. AgentBoard: An analytical evaluation board of multi-turn LLM agents. NeurIPS 2024.

[3] Chen, Li, et al. AutoManual: Constructing instruction manuals by LLM agents via interactive environmental learning. NeurIPS 2024

Motivation

Aligning agents with environments requires a strategy for
automated interface generation.

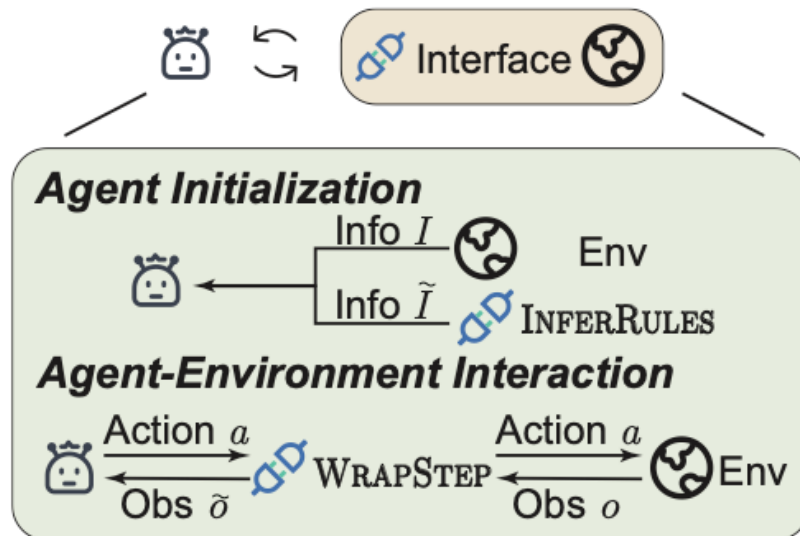
Auto-Aligned Interface Generation (ALIGN)

Interface design:

- **InferRules:** Provides the agent with more descriptive information about the environment during initialization, such as environment rules, potential limitations, etc.
- **WrapStep:** Enhances the observation returned to the agent for each environment step, when necessary, to deliver information in a format that is easier for the agent to interpret.

Implementation:

- Environment wrapper;
- No need to modify the agent logic or environment code.



Interface Example

Information Align

```
def InferRules():  
    return """1. Before examining or interacting  
    with any receptacle, you must first go to that  
    receptacle."""
```

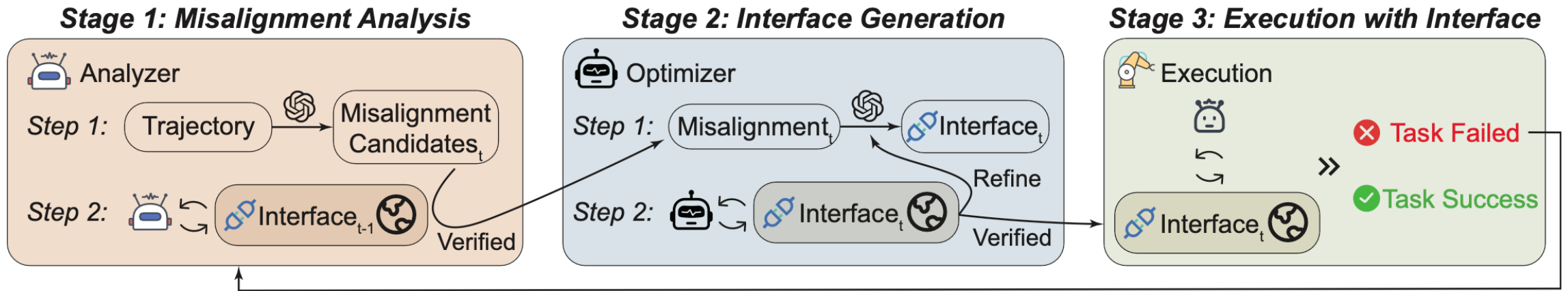
Interaction Align

```
def WrapStep():  
    ...  
    if target not in current_location:  
        obs = f"You need to go to {target} before  
        examining it. You must first navigate to a  
        receptacle before you can examine it."
```

Auto-Aligned Interface Generation (ALIGN)

ALIGN iteratively optimizes by automating the analysis of agent-environment misalignments between the agent and the environment, as well as automatically generating more robust interfaces. Each iteration consists of three stages:

- Stage 1: Misalignment Analysis
- Stage 2: Interface Generation
- Stage 3: Execution with Interface



Auto-Aligned Interface Generation (ALIGN)

Stage1: Misalignment Analysis:

- The *Analyzer* identifies agent-environment misalignments from previous erroneous interaction trajectories.

Misalignment Example

Analysis Result 1

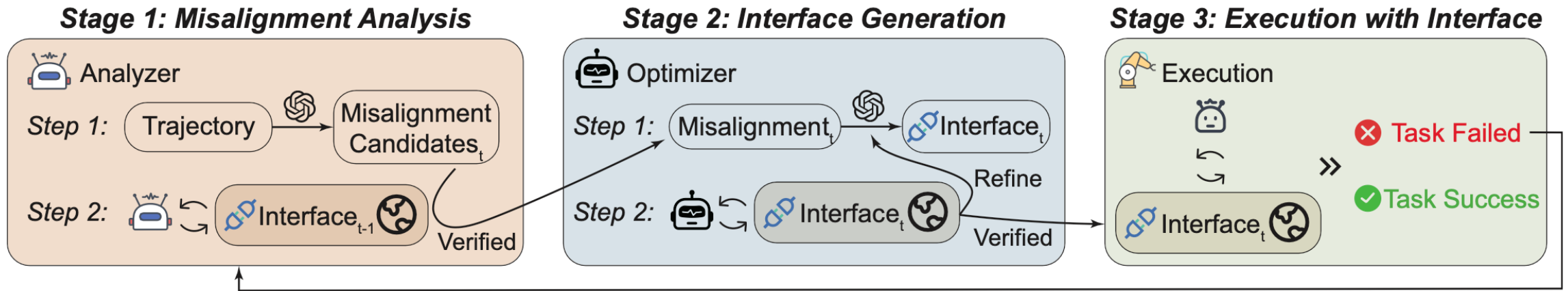
Agent Action Type: examine

Agent Action Case: examine drawer 1

Agent High-Level Reasoning Intent: The Agent is attempting to locate the box and desk lamp by examining potential receptacles.

Environment Rule: The Environment may require the Agent to first "go to" a receptacle before performing actions like "examine" on it.

Sufficient Observation: The environment should provide observation such as "You need to go to drawer 1 before examining it" when the Agent attempts to examine a receptacle without first moving to it.



Auto-Aligned Interface Generation (ALIGN)

Stage2: Interface Generation:

- The *Optimizer* generates new interface.

Interface Example

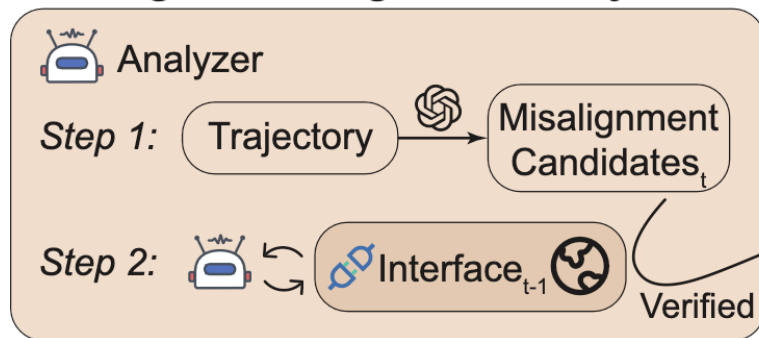
Information Align

```
def InferRules():  
    return """1. Before examining or interacting  
    with any receptacle, you must first go to that  
    receptacle."""
```

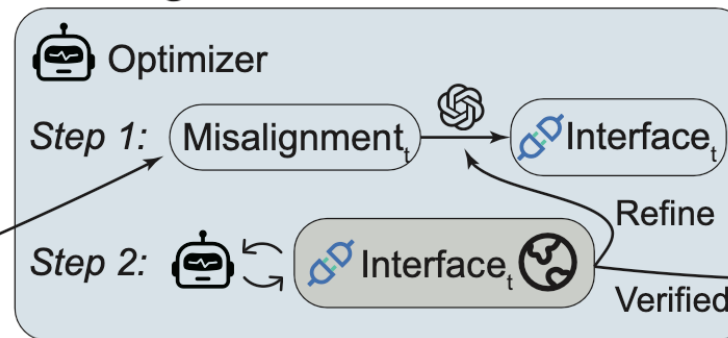
Interaction Align

```
def WrapStep():  
    ...  
    if target not in current_location:  
        obs = f"You need to go to {target} before  
        examining it. You must first navigate to a  
        receptacle before you can examine it."
```

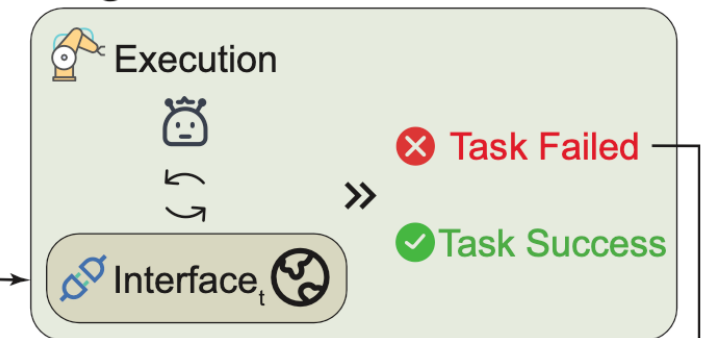
Stage 1: Misalignment Analysis



Stage 2: Interface Generation



Stage 3: Execution with Interface



Auto-Aligned Interface Generation (ALIGN)

Additional Design: Experimental Verification

- To mitigate the hallucination problem of LLMs, the *Analyzer* and *Optimizer* conduct experiments to validate the misalignments identified and the interfaces generated.

Experiment Verification Example

Optimizer: <thought>...</thought>

<action>init_simulator(task_id="4-293")</action>

Experiment: ...

Optimizer: <thought>Now I will simulate an invalid "examine" action where ...</thought>

<action>exec_agent_action("examine drawer 1")</action>

Experiment: ... Observation: Nothing happens. ...

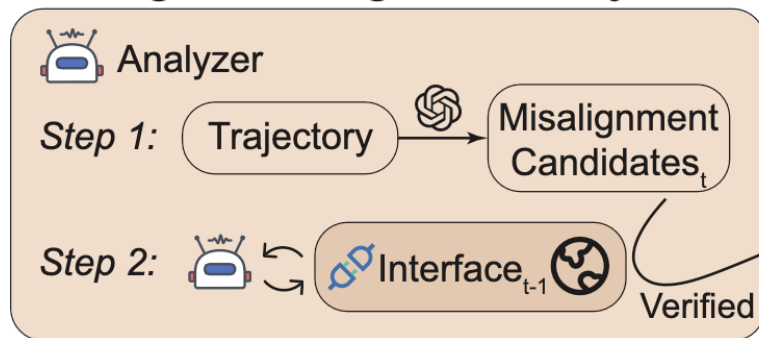
...

Optimizer: <thought>The "examine drawer 1" action was executed, but the environment did not provide sufficient feedback ...</thought>

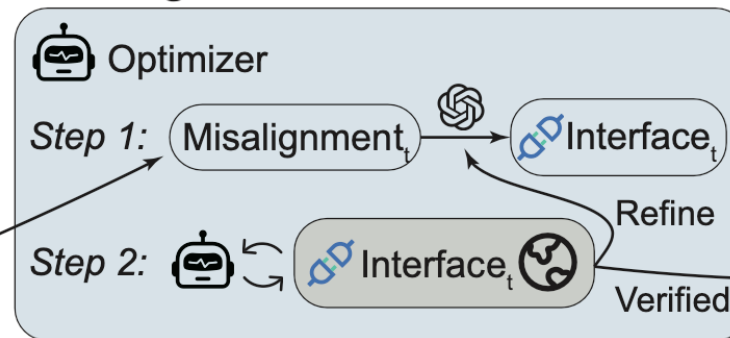
<if_need_refine>True</if_need_refine>

<refine_strategy>...</refine_strategy>

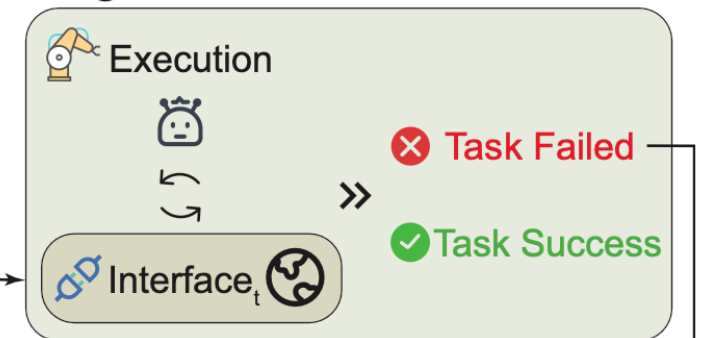
Stage 1: Misalignment Analysis



Stage 2: Interface Generation



Stage 3: Execution with Interface



Experimental Settings

Benchmarks:

- Embodied tasks: ALFWorld [1], ScienceWorld [2]
- Web navigation: WebShop [3]
- Tool-use: M^3 ToolEval [4]

Agent Methods:

- Vanilla: A basic implementation with no specific prompting strategies.
- ReAct [5]: Leverages the reasoning potential of LLMs through interactive reasoning and action.
- Self-Consistency [6]: Enhances the stability of the agent's decisions through multiple samples and a voting mechanism.
- Self-Refine [7]: The agent reflects on and revises its previous decisions to improve decision quality.
- Planning: Inspired by RAP [8], utilizes LLM's planning capability to decompose complex tasks into subtasks.

[1] Shridhar, Yuan, et al. Aligning text and embodied environments for interactive learning. ICLR 2021

[2] Wang, Jansen, et al. ScienceWorld: Is your agent smarter than a 5th grader? EMNLP 2022

Agent Base Model: Qwen2.5-7B-Instruct [3] Yao, Chen, et al. WebShop: Towards scalable real-world web interaction with grounded language agents.

[4] Wang, Chen, et al. Executable code actions elicit better LLM agents. ICML 2024

[5] Yao, Zhao, et al. ReAct: Synergizing reasoning and acting in language models. ICLR 2023

Model Selection for Analyzer and Optimizer: [6] Wang, Wei, et al. Self-Consistency improves chain of thought reasoning in language models. NeurIPS 2022

• Model for interface generation: Gemini2.5 Pro

[7] Madaan, Tandon, et al. Self-Refine: Iterative refinement with self-feedback. ICLR 2023

• Others: GPT-4.1

[8] Hao, Gu, et al. Reasoning with language model is planning with world model. EMNLP 2023

Main Results

Principle findings:

- ALIGN consistently enhances performance across different domains;
- Agent-environment misalignment is a pervasive phenomenon impeding the agent performance;

Method	Interface	Embodied		Web	Tool-use
		ALFWorld	ScienceWorld	WebShop	M ³ ToolEval
Vanilla	w/o ALIGN	13.43	14.94	54.10	11.11
	w/ ALIGN	60.45 (+47.02)	27.69 (+12.75)	61.23 (+7.13)	20.83 (+9.72)
ReAct	w/o ALIGN	19.40	20.03	37.20	9.72
	w/ ALIGN	63.43 (+44.03)	28.97 (+8.94)	42.93 (+5.73)	18.06 (+8.34)
Self-Consistency	w/o ALIGN	11.94	14.07	56.23	11.11
	w/ ALIGN	69.40 (+57.46)	25.41 (+11.34)	61.10 (+4.87)	16.67 (+5.56)
Self-Refine	w/o ALIGN	3.73	14.87	44.80	5.55
	w/ ALIGN	40.30 (+36.57)	22.99 (+8.12)	52.30 (+7.50)	6.94 (+1.39)
Planning	w/o ALIGN	9.70	17.13	46.95	11.11
	w/ ALIGN	52.99 (+43.29)	26.34 (+9.21)	54.67 (+7.72)	18.06 (+6.95)

Main Results

Principle findings:

- ALIGN consistently enhances performance across different domains;
- Agent-environment misalignment is a pervasive phenomenon impeding the agent performance;
- Alignment between agent and environment facilitates identification of additional performance-influencing factors.

Method	Interface	Embodied		Web	Tool-use
		ALFWorld	ScienceWorld	WebShop	M ³ ToolEval
Vanilla	w/o ALIGN	13.43	14.94	54.10	11.11
	w/ ALIGN	60.45 (+47.02)	27.69 (+12.75)	61.23 (+7.13)	20.83 (+9.72)
ReAct	w/o ALIGN	19.40	20.03	37.20	9.72
	w/ ALIGN	63.43 (+44.03)	28.97 (+8.94)	42.93 (+5.73)	18.06 (+8.34)
Self-Consistency	w/o ALIGN	11.94	14.07	56.23	11.11
	w/ ALIGN	69.40 (+57.46)	25.41 (+11.34)	61.10 (+4.87)	16.67 (+5.56)
Self-Refine	w/o ALIGN	3.73	14.87	44.80	5.55
	w/ ALIGN	40.30 (+36.57)	22.99 (+8.12)	52.30 (+7.50)	6.94 (+1.39)
Planning	w/o ALIGN	9.70	17.13	46.95	11.11
	w/ ALIGN	52.99 (+43.29)	26.34 (+9.21)	54.67 (+7.72)	18.06 (+6.95)

Indicate potential deficiencies in the critic and self-refinement capabilities of the Qwen2.5-7B-Instruct model.

Main Results

Principle findings:

- ALIGN consistently enhances performance across different domains;
- Agent-environment misalignment is a pervasive phenomenon impeding the agent performance;
- Alignment between agent and environment facilitates identification of additional performance-influencing factors.

Method	Interface	Embodied		Web	Tool-use
		ALFWorld	ScienceWorld	WebShop	M ³ ToolEval
Vanilla	w/o ALIGN	13.43	14.94	54.10	11.11
	w/ ALIGN	60.45 (+47.02)	27.69 (+12.75)	61.23 (+7.13)	20.83 (+9.72)
ReAct	w/o ALIGN	19.40	20.03	37.20	9.72
	w/ ALIGN	63.43 (+44.03)	28.97 (+8.94)	42.93 (+5.73)	18.06 (+8.34)
Self-Consistency	w/o ALIGN	11.94	14.07	56.23	11.11
	w/ ALIGN	69.40 (+57.46)	25.41 (+11.34)	61.10 (+4.87)	16.67 (+5.56)
Self-Refine	w/o ALIGN	3.73	14.87	44.80	5.55
	w/ ALIGN	40.30 (+36.57)	22.99 (+8.12)	52.30 (+7.50)	6.94 (+1.39)
Planning	w/o ALIGN	9.70	17.13	46.95	11.11
	w/ ALIGN	52.99 (+43.29)	26.34 (+9.21)	54.67 (+7.72)	18.06 (+6.95)

Indicate potential deficiencies in the long-reasoning and scientific causal reasoning capabilities of the Qwen2.5-7B-Instruct model.

Interface Quality Analysis

- Measure *the frequency of consecutive invalid actions* (the proportion of the actions that occur within sequences of two or more consecutive invalid steps).
- Provide evidence that ALIGN effectively renders latent constraints explicit, thereby preventing agents from entering repetitive error cycles.

Method	ALFWorld			ScienceWorld		
	w/o ALIGN	w/ ALIGN	Δ	w/o ALIGN	w/ ALIGN	Δ
Vanilla	77.91	26.59	66%	49.12	24.47	50%
ReAct	82.23	38.63	53%	46.61	29.99	36%
Self-Consistency	77.71	15.08	81%	51.10	31.51	38%
Self-Refine	90.38	45.84	49%	58.02	29.48	49%
Planning	74.09	19.14	74%	68.67	20.94	70%
Average	80.46	28.51	65%	54.70	27.28	49%

Generalization Study

Principle findings:

- ALIGN can generalize to different agent architectures;
- ALIGN can generalize to larger and heterogenous LLMs.

(a) Interface source: Vanilla agent				
Target method	ALFWorld	ScienceWorld	WebShop	M ³ ToolEval
ReAct	+39.56	+12.29	+7.87	+5.56
Self-Consistency	+51.49	+15.30	+3.00	+8.33
Self-Refine	+34.33	+14.11	+6.17	+4.17
Planning	+41.05	+9.66	+3.26	+11.11
(b) Interface source: Qwen2.5-7B-Instruct agent				
Target LLM	ALFWorld	ScienceWorld	WebShop	M ³ ToolEval
Qwen2.5-14B-Instruct	+17.46	+4.61	+4.66	+6.11
Llama3.1-8B-Instruct	+5.97	+10.27	+0.33	+0.83
Llama3.3-70B-Instruct	+5.82	+3.99	+5.68	+1.67

Ablation Study

Ablation on interface components:

- Each component of the interface contributes meaningfully.
- The critical role of fine-grained, enriched observation during interaction.

Method	w/o INFERRULES		w/o WRAPSTEP	
	ALFWorld	ScienceWorld	ALFWorld	ScienceWorld
Vanilla	-8.96	-3.35	-33.58	-4.72
ReAct	-5.22	-2.08	-17.91	-6.44
Self-Consistency	-1.49	-2.30	-37.27	-10.59
Self-Refine	-7.46	-1.72	-34.33	-7.59
Planning	-10.45	-0.78	-26.87	-9.86
<i>Mean</i>	-6.72	-2.05	-31.79	-7.84

Ablation on Experimental Verification:

- Experimental setting: Multi-sample (n=6) and select the best by Analyzer or Optimizer itself.
- Underscore the necessity of Experimental Verification.

Temp.	Turn0	Turn1	Turn2	Turn3
0.2	13.43	22.39	0.00	0.00
0.5	13.43	23.88	1.49	0.75

Discussion: Model Selection for Analyzer & Optimizer

Model Selection:

- Analyzer: gpt-4.1-mini
 - Optimizer: gpt-4.1-mini for Experimental Verification, gemini2.5 pro for interface generation
- Using weaker LLMs as the Analyzer can also achieve good performance.

Base Model	Interface	pick and place	pick clean and place	pick heat and place	pick cool and place	look at or examine in light	pick two obj and place	Average Tasks-success Rate (%)
Qwen2.5-7B-Instruct	w/o ALIGN	20.83	25.81	17.39	0	0	5.88	13.43
	w/ ALIGN	79.17	58.06	78.26	66.67	11.11	94.12	64.93
gpt-4.1-mini-2025-04-14	w/o ALIGN	58.33	22.58	8.70	9.52	22.22	52.94	28.36
	w/ ALIGN	95.83	87.10	26.09	80.95	27.78	52.94	64.93
gpt-4.1-2025-04-14	w/o ALIGN	100	93.55	13.04	71.43	61.11	100	73.88
	w/ Interface-GPT4_1-mini	100	100	78.26	100	77.78	100	93.28

Discussion: ALIGN for SOTA Model

- Closed-source LLMs: gpt-4.1-mini & gpt-4.1

Base Model	Interface	pick and place	pick clean and place	pick heat and place	pick cool and place	look at or examine in light	pick two obj and place	Average Tasks-success Rate (%)
gpt-4.1-mini-2025-04-14	w/o ALIGN	58.33	22.58	8.70	9.52	22.22	52.94	28.36
	w/ ALIGN	95.83	87.10	26.09	80.95	27.78	52.94	64.93
gpt-4.1-2025-04-14	w/o ALIGN	100	93.55	13.04	71.43	61.11	100	73.88
	w/ ALIGN	100	100	78.26	100	77.78	100	93.28

- RL-trained LLMs: GiGPO-Qwen2.5-7B-Instruct-ALFWorld

Agent method setting	Interface	Average Tasks-success Rate (%)
Vanilla Agent	w/o ALIGN	35.04
	w/ ALIGN	55.97
Same as Training Setting	w/o ALIGN	89.55
	w/ ALIGN	92.54

Discussion: ALIGN for SOTA Agent Framework

Agent Framework: AgentSquare

- Planning module: OPENAGI
- Reasoning module: Self-Refine
- Memory module: Generative, DiLu, TP and VOYAGER

Base Model	Agent FrameWork	Interface	Memory Module	pick and place	pick clean and place	pick heat and place	pick cool and place	look at or examine in light	pick two obj and place	Average Tasks-success Rate (%)
gpt-4.1-2025-04-14	AgentSquare	/	Generative	95.83	87.10	69.57	95.24	83.33	88.24	86.57
	AgentSquare	/	DiLu	91.67	87.10	52.17	95.24	83.33	70.59	80.60
	AgentSquare	/	TP	87.50	51.61	4.35	61.90	27.78	47.06	47.76
	AgentSquare	/	VOYAGER	95.83	83.87	52.17	90.48	83.33	64.71	79.10
	Vanilla Agent	w/o ALIGN	/	100	93.55	13.04	71.43	61.11	100	73.88
	Vanilla Agent	w/ ALIGN	/	100	100	78.26	100	77.78	100	93.28

- This suggests that in the future, using base models equipped with automatically generated aligned interfaces for specific environments could achieve LLM-based agent adaptation in particular tasks and achieve strong performance.

Discussion: Token Consumption

- LLM hallucination issues decreases → cost reduce;
- Only code generation needs SOTA model;
- The cost of interface generation is a one-time expense.

		ALFWorld	ScienceWorld	WebShop	M ³ ToolEval
Analyzer	Input Token (M)	0.2770	0.4333	0.1783	0.1094
	Output Token (M)	0.0040	0.0036	0.0048	0.0016
	Total Token (M)	0.2809	0.4370	0.1831	0.1109
Optimizer	Input Token (M)	0.2619	0.2288	0.0669	0.1100
	Output Token (M)	0.0087	0.0172	0.0040	0.0118
	Total Token (M)	0.2706	0.2460	0.0709	0.1217
Total	Total Token (M)	0.5515	0.6830	0.2540	0.2326

Future...

Validation on more difficult environments and advanced agent methods.

Perspectives from environment designers: detecting design issues through ALIGN.

Interface design: detecting agent internal expectations and transforming actions accordingly.



Evolving and diversifying interfaces & RL Agent.

Thanks!