

Life Cycle of Skills

Construction, Update, Selection, Utilization & Internalization

Meituan LongCat Team

General Agent 团队负责人：顾奇

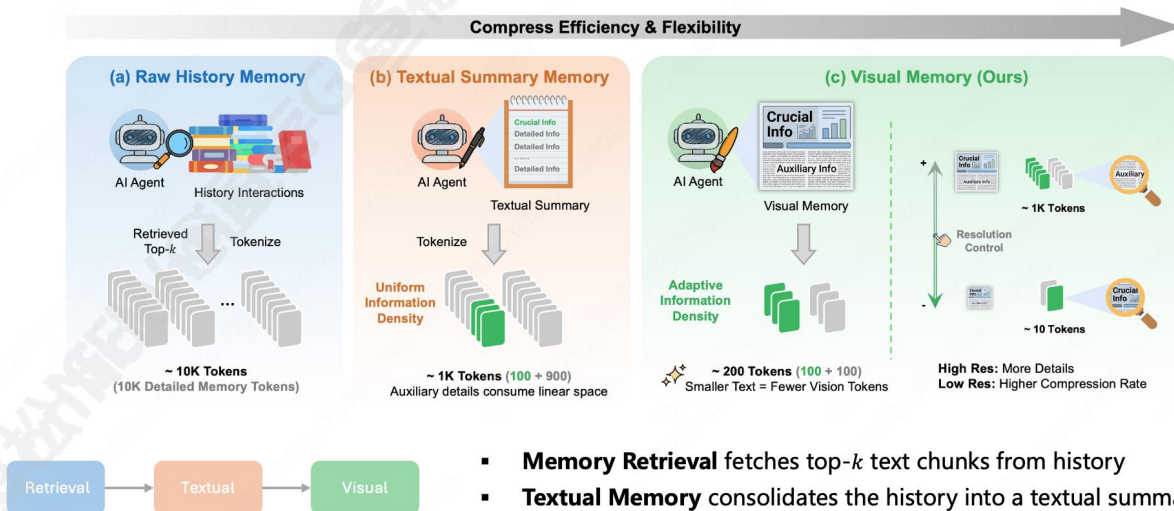
2026年7月



Information: Format, Structure & Topology

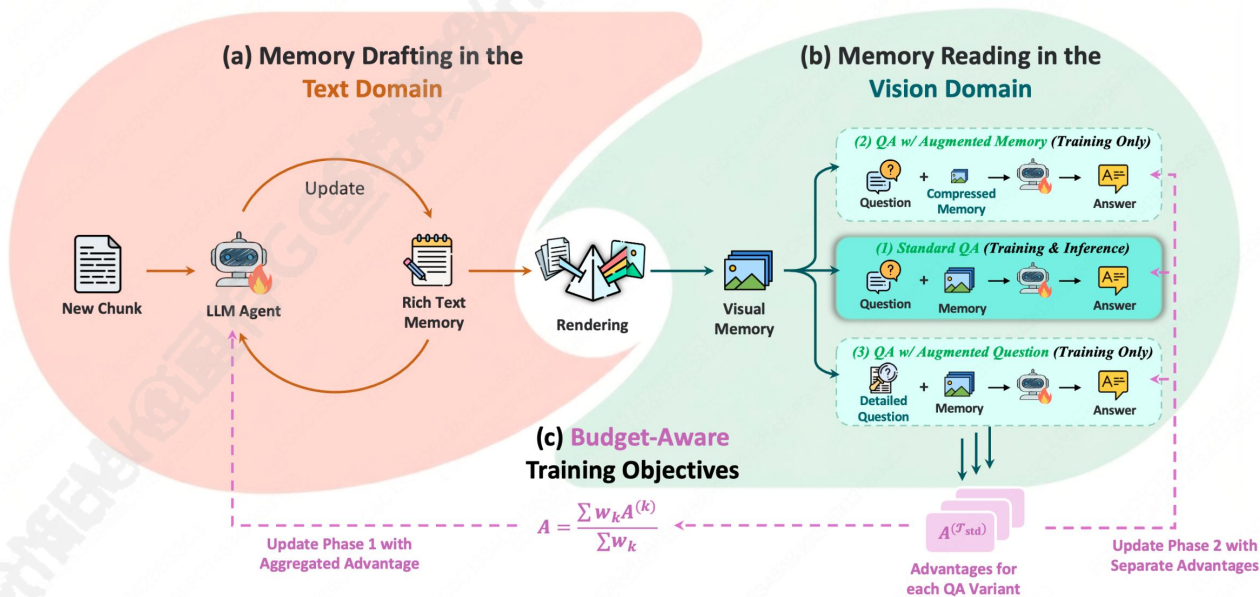
Format, Structure & Topology

- Information Density — Diverse / Native Information Carrier
- Easy / Possible to Use at least with Specifically Tuned Models



- Parametrized?
- Multiple Skills : Hierarchical Topology

Format & Structure——MemOCR (ICML 26)

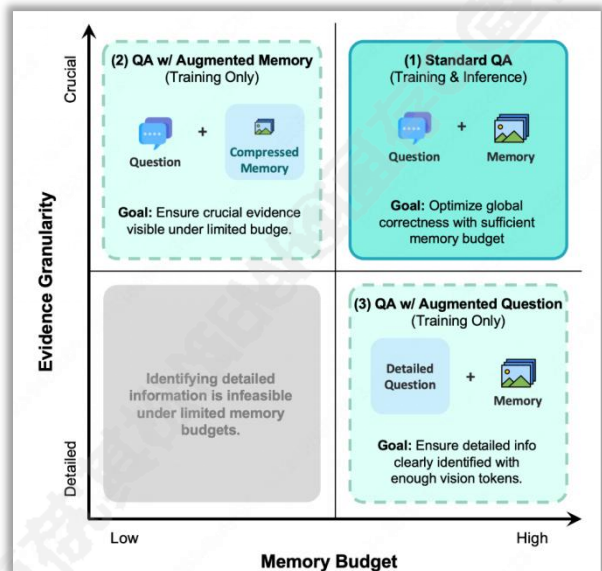


1. In the **text domain**, agent drafts memory with **layout control**
2. In the **vision domain**, agent reads the memory under **different visibility**
3. Two stages **jointly optimized** via budget-aware objectives.

Format & Structure——MemOCR (ICML 26)



- Method: Aligning evidence importance with visibility via layout control



To prevent the shortcut of uniform text rendering, MemOCR employs three complementary training tasks optimized via GRPO (Group Relative Policy Optimization):

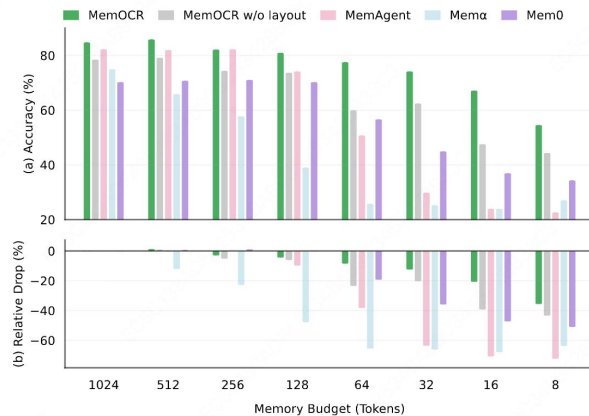
- ◆ **Standard QA ($\mathcal{T}_{std}$):** 512-token budget ensures global correctness
- ◆ **QA w/ Augmented Memory ($\mathcal{T}_{augM}$):** 4× per dimension downsampling (16× fewer pixels) forces crucial evidence to remain readable under compression
- ◆ **QA w/ Augmented Question ($\mathcal{T}_{augQ}$):** Detail-oriented questions on uncompressed memory preserve auxiliary information

Format & Structure——MemOCR (ICML 26)



Table 1. Comparison of accuracy (%) across different context lengths (10K, 30K, and 100K tokens). Best results are highlighted in **bold** and second best in underline. The percentages in the Average column represent the performance drop relative to the 1024-token budget.

Method	Memory Budget		HotpotQA			2Wiki			NQ			TriviaQA			Average		
	(Tokens)		10K	30K	100K	10K	30K	100K	10K	30K	100K	10K	30K	100K	10K	30K	100K
Raw History Memory																	
Qwen2.5	100K	70.3	-	-	53.9	61.1	-	50.8	-	-	71.1	-	-	61.5	-	-	-
R1-Distill Qwen	100K	50.8	12.5	0.8	50.8	26.6	2.3	35.2	3.9	1.6	54.7	11.7	0.8	47.9	13.7	1.4	1.4
Qwen2.5-1M	100K	75.0	66.4	68.8	67.2	65.6	54.7	53.1	43.8	50.8	71.1	77.3	59.4	66.6	63.3	58.4	58.4
Textual Summary Memory																	
Mem0	1024	70.3	66.1	64.1	60.4	47.7	40.9	47.7	48.8	48.2	69.5	70.6	61.4	62.0	58.3	53.6	53.6
	256	71.1	65.6	63.8	57.8	44.8	43.5	46.9	51.1	49.7	69.5	70.6	62.8	61.3 (-1.1%)	58.0 (-0.5%)	55.0 (+2.5%)	55.0
	64	56.7	60.9	56.2	53.9	42.2	43.3	46.9	43.8	50.0	66.4	65.9	62.5	56.0 (-9.7%)	53.2 (-8.8%)	53.0 (-1.2%)	53.0
	16	37.0	43.3	34.9	28.4	31.5	27.6	41.4	34.1	43.0	65.3	62.8	58.9	43.0 (-30.6%)	42.9 (-26.4%)	41.1 (-23.4%)	41.1
Mem- α	1024	75.0	77.9	79.4	47.7	40.1	50.0	45.6	44.8	43.0	21.1	23.2	19.3	47.3	46.5	47.9	47.9
	256	57.8	57.3	57.0	38.3	32.8	41.7	47.9	33.6	38.0	17.2	24.2	18.3	40.3 (-44.9%)	37.0 (-20.5%)	38.7 (-19.2%)	38.7
	64	25.8	29.7	28.9	34.9	31.5	39.1	27.1	25.0	24.0	13.0	19.5	13.8	25.2 (-46.8%)	26.4 (-43.2%)	26.4 (-44.9%)	26.4
	16	24.0	30.2	26.3	35.4	32.3	40.1	23.7	26.3	24.2	11.7	19.0	14.3	23.7 (-50.0%)	27.0 (-42.0%)	26.2 (-45.3%)	26.2
MemAgent	1024	82.3	78.9	79.4	65.4	63.8	61.2	53.4	46.1	55.5	70.1	78.1	66.7	67.8	66.7	65.7	65.7
	256	82.3	76.8	76.6	66.1	62.5	58.1	51.3	44.3	53.6	69.5	77.6	67.2	67.3 (-0.7%)	65.3 (-2.1%)	63.9 (-2.8%)	63.9
	64	50.8	52.3	51.0	38.5	42.7	36.2	48.7	36.5	47.9	64.6	69.5	59.6	50.7 (-25.3%)	50.3 (-24.7%)	48.7 (-25.9%)	48.7
	16	24.0	26.6	23.2	28.4	26.3	16.7	24.5	20.1	27.3	49.7	56.3	41.7	31.6 (-53.3%)	32.3 (-51.6%)	27.2 (-58.6%)	27.2
Visual Summary																	
MemOCR (Ours)	1024	84.8	75.1	78.3	72.2	73.7	62.7	61.8	49.2	55.4	79.6	81.3	69.8	74.6	69.8	66.6	66.6
	256	82.2	75.4	75.7	71.2	72.8	65.5	57.3	48.8	58.3	79.7	80.9	68.9	72.6 (-2.7%)	69.5 (-0.5%)	67.1 (+0.8%)	67.1
	64	77.6	68.1	67.2	62.9	66.2	62.4	51.0	43.6	47.5	77.6	80.7	70.2	67.3 (-9.8%)	64.7 (-7.4%)	60.7 (-8.8%)	60.7
	16	67.2	57.9	52.4	57.9	56.0	45.7	42.8	35.9	45.9	80.8	72.2	67.2	62.2 (-16.6%)	55.5 (-20.5%)	52.8 (-20.8%)	52.8



Better Performance under **limited memory budget**.
Memory Budgets: tokens for memory at inference time.

Actions: Construction, Update, Selection, Utilization & Internalization

Construction, Update, Selection, Utilization & Internalization



Information by Model

Exploration
Construction

Information by Others

Standardize

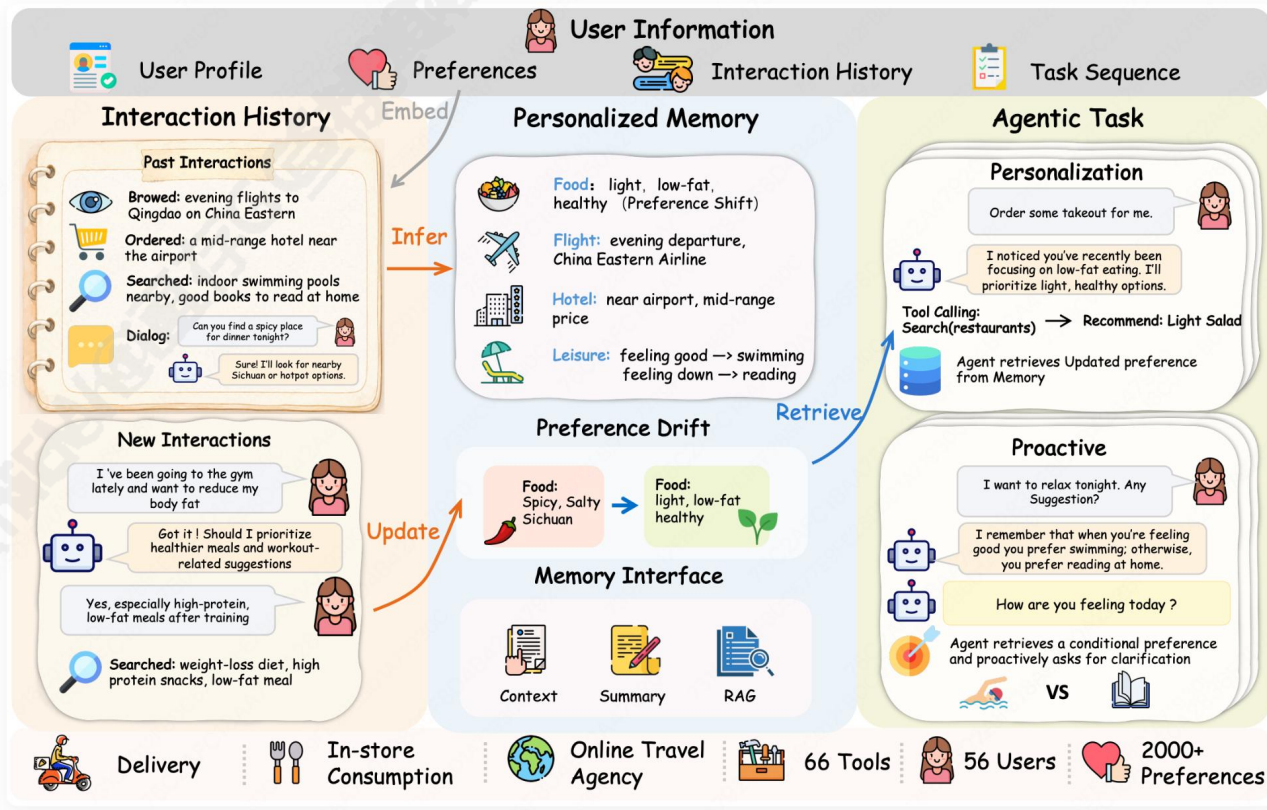
Update

Select 0-n & Utilize

Internalize

Why Internalize?

Construction & Update——VitaBench 2.0



Construction & Update——VitaBench 2.0



🏆 Leaderboard

Performance under three memory settings, sorted by **Avg@4 on Full Context**. **Bold** = best in column.

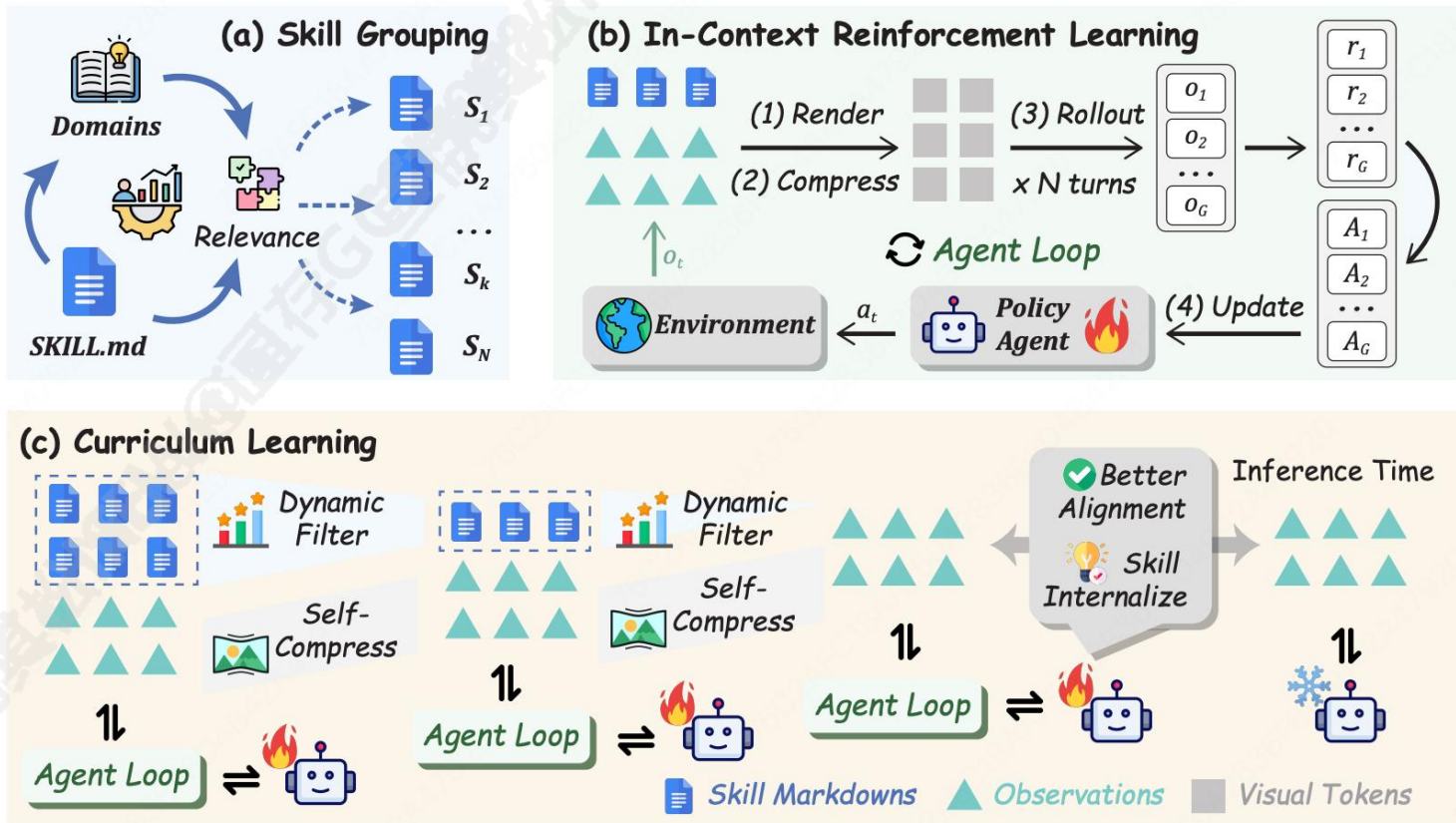
Thinking Models

Non-thinking Models

#	Model	Full Context			Agentic Memory			RAG Memory		
		Avg@4	Pass@4	Pass^4	Avg@4	Pass@4	Pass^4	Avg@4	Pass@4	Pass^4
1	Claude-Opus-4.6	0.503	0.664	0.337	0.454	0.645	0.259	0.430	0.566	0.299
2	Doubao-Seed-2.0-pro	0.474	0.683	0.270	0.428	0.650	0.225	0.339	0.496	0.205
3	DeepSeek-V4-Pro	0.472	0.649	0.295	0.449	0.656	0.255	0.430	0.584	0.271
4	GPT-5	0.441	0.658	0.226	0.421	0.647	0.204	0.410	0.591	0.236
5	Claude-4.5-Sonnet	0.417	0.658	0.197	0.397	0.642	0.178	0.374	0.573	0.186
6	o3	0.403	0.653	0.169	0.401	0.669	0.154	0.362	0.587	0.158
7	DeepSeek-R1-0528	0.396	0.691	0.131	0.412	0.712	0.118	0.390	0.643	0.153
8	GLM-5.1	0.394	0.587	0.213	0.352	0.556	0.150	0.328	0.485	0.185
9	Doubao-Seed-1.6	0.373	0.599	0.176	0.383	0.646	0.123	0.375	0.591	0.179
10	GLM-4.5	0.364	0.623	0.156	0.311	0.596	0.106	0.336	0.555	0.147
11	GLM-4.6	0.359	0.612	0.116	0.351	0.625	0.107	0.336	0.574	0.135
12	MiniMax-M2.7	0.345	0.584	0.145	0.351	0.609	0.124	0.314	0.518	0.143
13	Gemini-2.5-Pro	0.331	0.605	0.109	0.378	0.638	0.138	0.320	0.579	0.109
14	Kimi-K2.6	0.293	0.533	0.099	0.280	0.508	0.088	0.303	0.511	0.118
15	Qwen3-Max	0.284	0.499	0.105	0.324	0.599	0.091	0.315	0.519	0.134
16	Gemini-2.5-Flash	0.282	0.556	0.063	0.312	0.567	0.098	0.309	0.544	0.107

<https://vitabench2.github.io/>

Internalization——Skill0



Internalization——Skill0



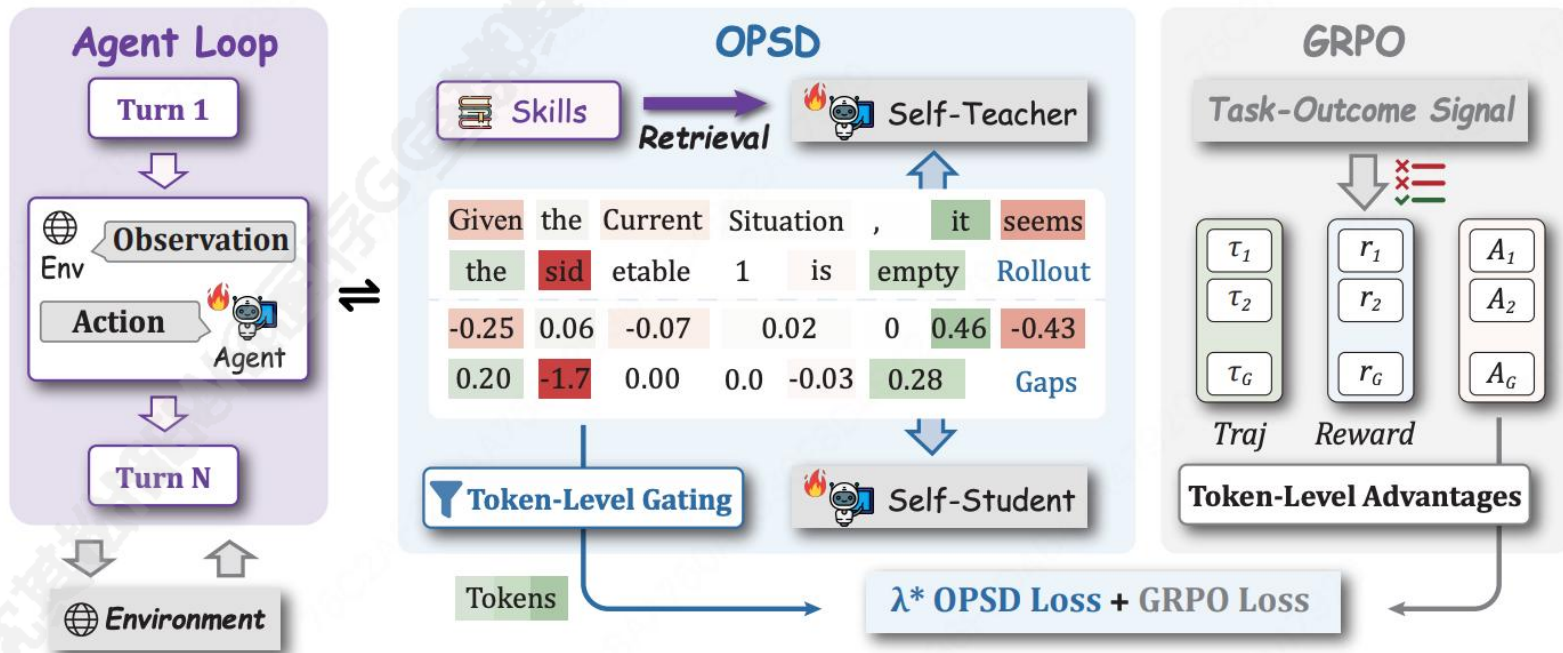
Table 1: **Performance on ALFWorld and Search-QA tasks.** We report the success rate (%) and the average context token cost (k) per step. [†] denotes models validated with skill augmentation; * denotes methods that encodes visual context with reduced token overhead. We simply reproduce results of SkillRL-3B without cold start and skill evolution. **Best** and **second-best** are highlighted.

Method	ALFWorld								Search-QA								
	Pick	Look	Clean	Heat	Cool	Pick2	Avg [†]	Cost [↓]	NQ	Triv	Pop	Hotp	2Wk	MuS	Bam	Avg [†]	Cost [↓]
Qwen2.5-(VL)-3B-Instruct																	
Zero-Shot	27.0	24.3	4.5	20.5	10.2	0.0	15.2	1.21	9.4	31.3	19.8	15.0	14.8	4.7	16.8	15.9	0.48
Few-Shot [†]	44.1	45.1	30.5	44.1	9.2	3.3	29.3	2.30	11.8	30.9	20.2	13.7	18.4	4.5	25.6	17.9	0.86
Zero-Shot*	44.3	27.6	8.6	3.1	5.7	3.1	17.6	0.48	10.2	27.7	10.9	9.1	12.2	3.7	15.2	12.7	0.15
Few-Shot* [†]	57.1	25.3	4.5	5.5	10.2	9.4	23.8	0.88	11.1	26.2	14.2	15.4	13.4	3.0	19.2	14.6	0.36
GRPO	92.6	85.7	70.6	86.6	79.3	65.0	79.9	1.02	39.3	60.6	41.1	37.4	34.6	15.4	26.4	36.4	0.61
AgentOCR*	91.9	81.8	76.0	73.3	76.1	70.0	78.2	0.38	38.6	56.5	41.7	33.6	30.7	14.6	24.0	34.2	0.26
EvolveR	77.3	24.5	47.9	41.7	24.6	22.5	44.1	1.89	43.4	58.4	43.4	37.3	38.1	13.7	32.8	38.2	—
SkillIRL [†]	91.9	100	82.9	87.4	78.7	70.0	82.4	2.21	38.6	57.6	40.3	33.6	31.1	13.3	58.1	38.9	0.87
SKILL0*	95.6	80.4	100	86.7	78.7	75.2	87.9	0.38	39.8	57.5	42.3	35.1	33.7	13.3	63.7	40.8	0.18
Qwen2.5-(VL)-7B-Instruct																	
Zero-Shot	67.6	35.4	19.3	31.3	30.1	4.4	31.3	1.08	10.4	32.4	22.3	15.8	15.4	7.2	19.2	17.5	0.70
Few-Shot [†]	75.4	64.9	67.5	26.7	19.4	8.9	48.4	2.12	12.3	36.8	24.5	17.7	18.2	6.5	24.8	20.1	0.97
Zero-Shot*	46.0	35.6	19.1	7.1	5.5	5.4	21.1	0.52	6.9	30.4	12.0	10.5	9.1	5.5	24.0	14.0	0.26
Few-Shot* [†]	44.3	55.4	52.9	0.0	11.2	5.4	28.9	1.79	10.5	31.9	18.7	14.2	14.4	6.9	24.8	17.3	0.41
GRPO	92.6	93.8	85.2	80.0	82.7	56.5	81.8	0.95	45.1	63.7	44.0	43.6	43.2	16.8	37.6	41.9	0.73
AgentOCR*	95.6	96.2	78.1	73.2	72.4	72.0	81.2	0.43	43.1	61.0	45.4	40.8	38.3	15.7	36.8	40.1	0.36
EvolveR	64.9	33.3	46.4	13.3	33.3	33.3	43.8	—	43.5	63.4	45.9	38.2	42.0	15.6	54.4	43.1	—
SkillIRL [†]	97.9	71.4	90.0	90.0	95.5	87.5	89.9	—	45.9	63.3	45.9	43.2	40.3	20.2	73.8	47.1	—
SKILL0*	100	85.8	94.6	81.9	85.7	80.1	89.8	0.41	42.7	61.1	45.3	40.0	38.3	16.4	66.9	44.4	0.34

Table 2: **Performance on WebShop (128 tasks).** We report the Score (%), Acc (%) and the average context token cost (k) per step. All methods encode visual context with reduced token overhead.

Method	Inference	Qwen2.5-VL-3B			Qwen2.5-VL-7B		
	w/ Skills	Score↑	Accuracy↑	Tokens↓	Score↑	Accuracy↑	Tokens↓
Baselines							
Zero-Shot	✓	23.4	6.3	0.46k	26.8	7.8	0.48k
Few-Shot		45.8 [+22.4]	18.4 [+12.1]	0.78k [+0.52k]	51.2 [+24.4]	24.2 [+16.4]	0.91k [+0.43k]
AgentOCR		75.2 [+51.8]	56.3 [+50.0]	0.42k [-0.04k]	78.6 [+51.8]	59.3 [+51.5]	0.40k [-0.08k]
Ours: RL w/ Skills							
SKILL0	✓	80.9 [+57.5]	64.1 [+57.8]	0.94k [+0.48k]	85.3 [+58.5]	71.9 [+64.1]	0.89k [+0.41k]
SKILL0		78.6 [+55.2]	66.4 [+60.1]	0.49k [+0.03k]	85.1 [+58.3]	74.2 [+66.4]	0.46k [-0.02k]

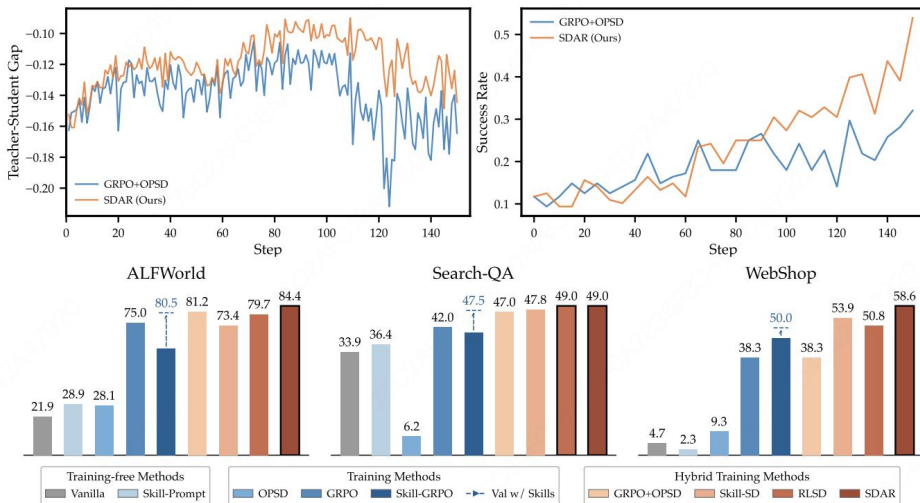
Internalization——SDAR



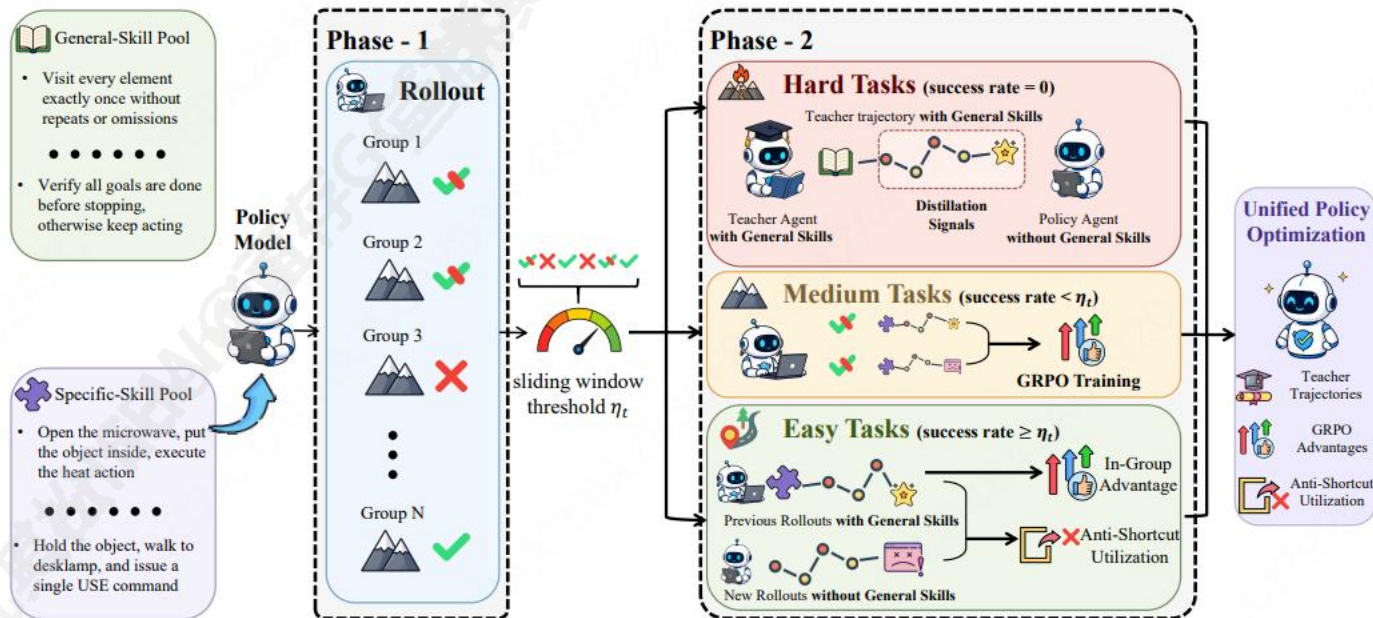
Internalization——SDAR



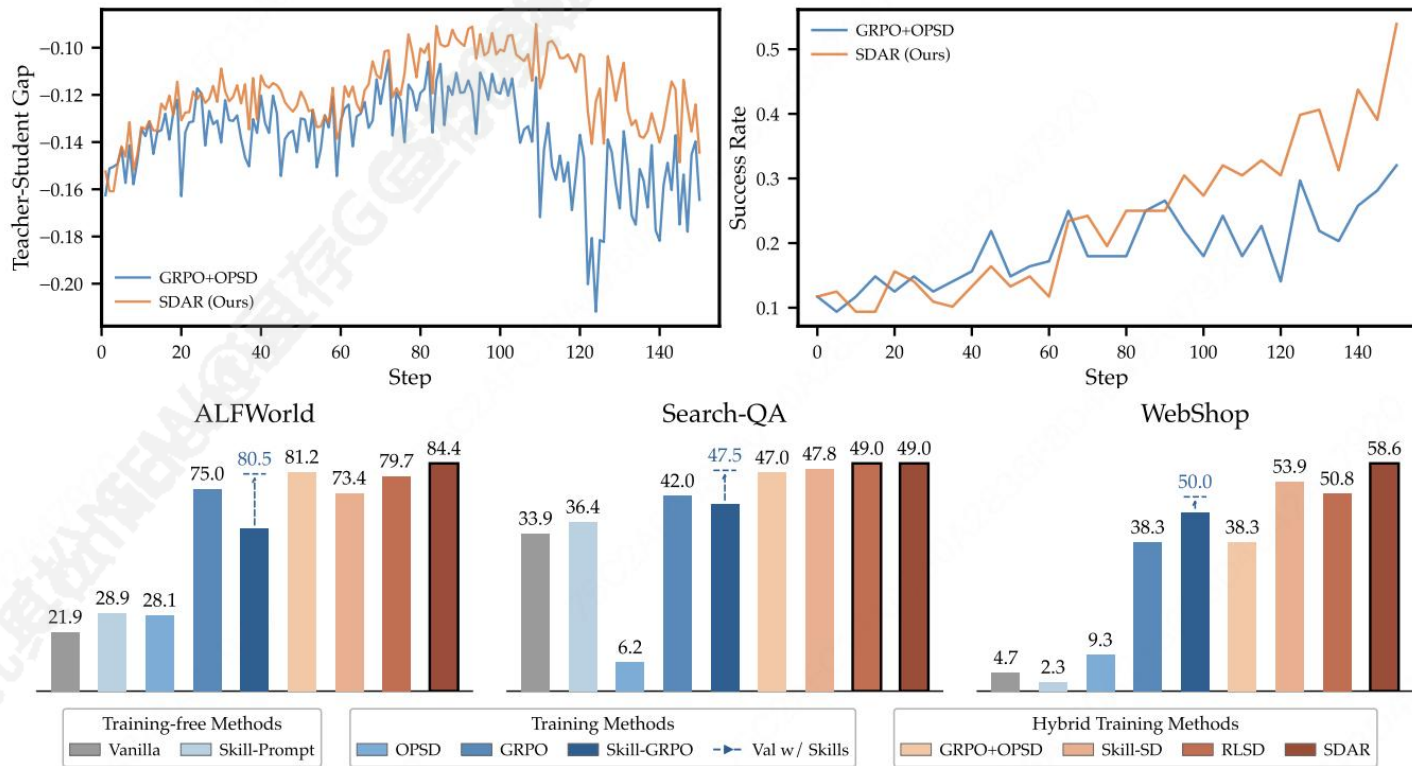
Method	ALFWorld					Search-QA										WebShop	
	Pick	Look	CleanHeat	Cool	Pick2 Avg	NQ	Triv	Pop	Hotp	2Wk	MuS	Bam	Avg	Score	Acc		
<i>Qwen2.5-3B-Instruct</i>																	
Vanilla	44.4	11.1	6.2	15.4	28.6	12.5	21.9	24.6	48.1	31.0	26.3	25.3	7.2	59.7	31.7	6.7	0.8
Skill-Prompt*	51.7	66.7	48.4	0.0	4.3	10.0	28.9	23.7	46.2	30.6	24.4	22.1	7.5	12.5	23.9	0.2	0.8
OPSD	48.8	41.7	16.7	0.0	15.8	16.7	28.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	11.3	3.1
GRPO	91.2	62.5	96.2	61.9	65.0	47.4	75.0	39.3	60.6	41.1	37.4	34.6	15.4	26.4	36.4	79.8	63.3
Skill-GRPO	88.9	71.4	58.8	70.6	40.7	29.2	60.2	43.5	58.8	43.0	36.8	32.2	11.7	12.5	34.1	77.3	60.9
Skill-GRPO*	94.3	57.1	100	66.7	73.1	57.1	80.5	44.3	59.6	44.3	39.0	36.1	14.5	14.9	36.1	76.3	66.4
GRPO+OPSD	100	82.4	85.7	75.0	70.0	60.0	81.2	44.9	61.2	45.2	40.4	38.5	16.0	66.1	44.6	77.8	66.4
Skill-SD	88.2	50.0	96.2	52.4	65.0	57.9	73.4	44.4	60.4	44.0	39.5	40.4	15.4	64.9	44.1	75.9	64.0
RLSD	87.9	75.0	90.9	75.0	73.1	68.4	79.7	41.5	58.6	42.3	40.4	40.2	16.8	66.9	43.8	84.4	66.4
SDAR	97.1	62.5	100	61.9	75.0	84.2	84.4	44.8	58.1	44.3	38.6	36.2	15.7	66.1	43.4	85.0	68.0
<i>Qwen2.5-7B-Instruct</i>																	
Vanilla	36.1	22.2	3.1	0.0	0.0	12.5	25.2	50.8	29.5	29.0	29.0	10.4	63.7	33.9	5.9	1.6	
Skill-Prompt*	51.7	50.0	32.3	5.3	4.3	0.0	23.4	30.9	52.1	32.7	32.7	27.9	12.7	66.1	36.4	1.7	0.8
OPSD	50.0	60.0	22.7	21.4	17.6	9.5	32.8	8.8	8.6	17.5	2.5	4.2	0.5	1.2	6.2	4.5	2.3
GRPO	91.2	87.5	96.2	81.0	65.0	57.9	81.2	45.1	63.7	44.0	43.6	43.2	16.8	37.6	42.0	80.9	72.6
Skill-GRPO	88.5	66.7	65.2	61.1	57.7	73.1	69.5	45.2	63.7	45.7	43.1	43.3	19.6	21.4	40.3	80.4	71.9
Skill-GRPO*	100	83.3	96.4	83.3	75.0	78.9	88.3	44.8	63.0	45.1	43.7	43.7	20.5	71.4	47.5	87.0	81.2
GRPO+OPSD	91.4	61.5	100	87.5	76.5	52.2	80.4	47.3	64.5	46.9	43.8	39.3	18.0	69.4	47.0	86.8	76.5
Skill-SD	93.9	93.8	90.9	100	69.2	68.4	85.1	47.1	64.5	47.8	44.2	42.1	20.2	69.0	47.8	86.1	76.5
RLSD	100	87.5	92.3	58.8	80.0	65.2	82.0	46.8	63.0	44.4	45.5	48.9	21.5	73.0	49.0	87.4	77.3
SDAR	94.7	75.0	100	86.7	68.2	78.9	85.9	46.3	63.5	48.2	43.8	48.4	19.6	73.0	49.0	89.4	82.8
<i>Qwen3-1.7B-Instruct</i>																	
Vanilla	25.0	22.2	3.1	0.0	21.4	4.2	12.5	29.4	46.9	37.0	23.5	19.6	6.4	10.5	24.8	46.5	4.7
Skill-Prompt*	10.3	50.0	16.1	0.0	0.0	5.0	9.4	29.4	46.5	36.2	22.9	20.8	4.3	10.1	24.3	23.0	2.3
OPSD	26.3	33.3	9.1	0.0	4.5	5.3	14.1	4.2	8.3	4.6	6.6	15.3	0.7	1.2	5.8	47.4	9.3
GRPO	71.1	41.7	36.4	40.0	31.8	31.6	46.1	40.0	58.9	43.5	35.4	30.3	12.0	65.7	40.8	67.3	38.3
Skill-GRPO	27.6	54.5	22.7	27.3	0.0	19.2	21.1	39.2	58.6	43.9	35.2	28.2	11.5	66.1	40.4	73.4	46.1
Skill-GRPO*	31.4	42.9	51.9	8.3	11.5	7.1	28.1	38.0	58.4	43.9	36.3	29.0	12.5	66.9	40.7	80.4	50.0
GRPO+OPSD	38.2	50.0	30.8	28.6	30.0	21.1	32.0	40.7	58.9	45.0	37.0	34.6	13.3	65.7	42.2	70.7	38.3
Skill-SD	52.9	37.5	69.2	42.9	60.0	36.8	52.3	39.1	57.5	45.4	34.8	34.1	10.7	64.1	40.8	81.8	53.9
RLSD	50.0	37.5	61.5	19.0	50.0	21.1	42.2	38.6	57.3	43.0	34.5	34.1	11.5	65.3	40.6	74.0	50.8
SDAR	73.5	25.0	76.9	33.3	40.0	36.8	53.9	39.7	58.9	45.3	35.9	35.5	12.6	65.3	41.9	76.8	58.6



Utilization & Internalization——Skill0.5



Utilization & Internalization——Skill0.5



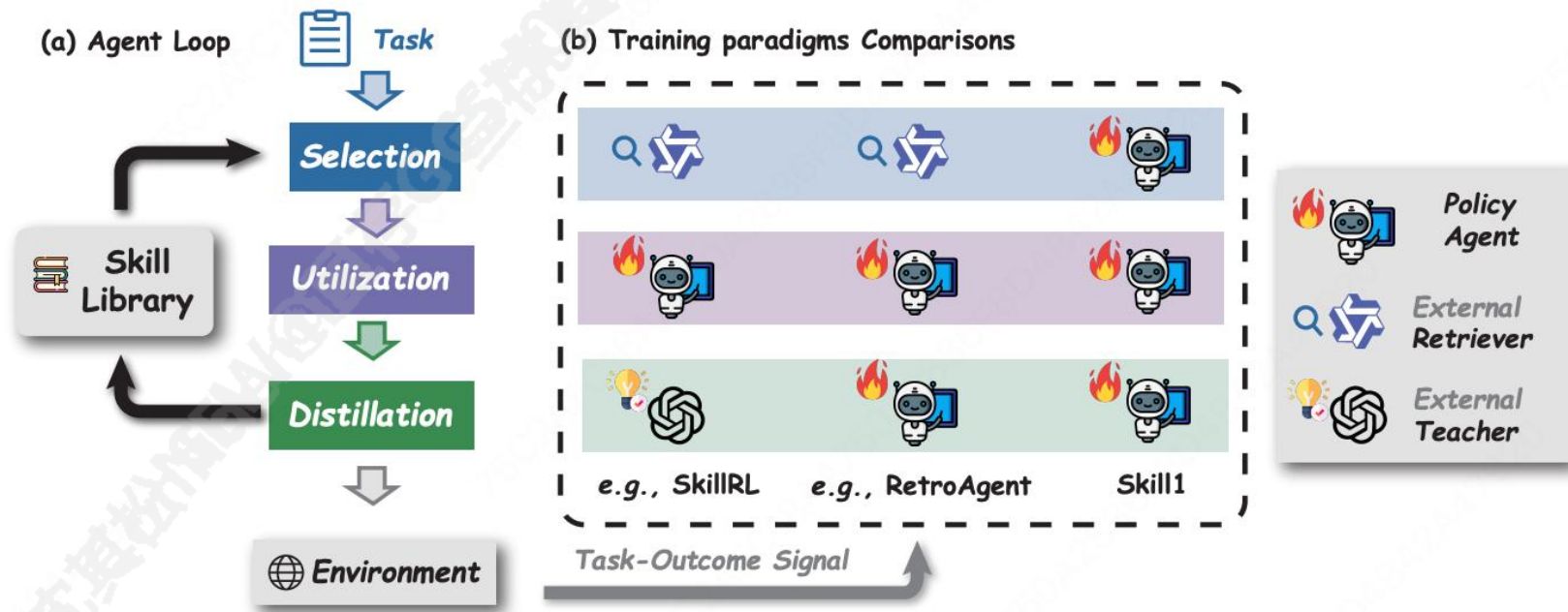
Utilization & Internalization——Skill0.5



Method	ID					OOD				
	Pick	Cool	Clean	Avg.	Rank ↓	Look	Heat	Pick2	Avg.	Rank ↓
<i>Prompt-based Methods</i>										
Zero-shot	28.6	12.0	18.5	20.7	17.2	38.5	12.5	12.5	18.9	15.5
Few-shot	62.9	44.0	63.0	57.5	12.0	46.2	31.2	8.3	24.5	12.8
ReAct	71.4	28.0	33.3	47.1	13.3	46.2	18.8	12.5	22.6	12.8
Reflexion	85.7	44.0	44.4	60.9	10.3	46.2	31.3	29.2	34.0	8.7
Mem0	54.3	4.0	18.5	28.7	17.5	38.5	18.8	4.2	17.0	15.7
ExpeL	80.0	44.0	66.7	65.5	9.5	46.2	18.8	20.8	18.9	11.2
MemP	65.7	12.0	33.3	40.2	15.3	46.2	37.5	12.5	28.3	11.0
SimpleMem	71.4	16.0	44.4	47.1	13.3	53.8	18.8	20.8	28.3	9.5
<i>RL-based Methods</i>										
RLOO	91.4	80.0	81.5	85.1	4.3	61.5	56.3	20.8	41.5	5.8
GRPO	80.0	72.0	88.9	80.5	6.2	76.9	56.3	16.7	43.4	5.5
<i>Memory-Augmented RL Methods</i>										
MemRL	74.3	12.0	55.6	50.6	12.7	46.2	12.5	45.8	35.8	9.7
EvolveR	88.6	52.0	81.5	75.9	6.2	46.2	6.2	50.0	35.8	10.0
Mem0+GRPO	65.7	20.0	51.9	48.3	13.2	23.1	6.2	20.8	17.0	15.0
SimpleMem+GRPO	85.7	52.0	70.3	71.3	7.7	61.5	43.8	41.7	47.2	4.5
<i>Skill-Augmented RL Methods</i>										
SkillRL	91.4	84.0	96.3	90.8	2.7	69.2	75.0	12.5	45.3	6.3
SKILL0	94.3	76.0	81.5	85.1	3.8	46.2	50.0	29.2	39.6	7.5
SLIM	91.4	84.0	70.4	82.8	4.5	53.8	31.3	29.2	35.8	7.0
Skill0.5	94.3	88.0	96.3	93.1	1.3	69.2	87.5	33.3	58.5	2.5

Table 1: Performance comparison on ALFWorld under ID and OOD task settings. **Best** and second-best results in each column are highlighted, respectively. Average Rank is computed across all evaluated settings.

Selection & Utilization & Internalization——Skill1



Selection & Utilization & Internalization——Skill1

Stage 1: Selection



? Write a one-sentence search query to find...



Query: Tips for heating an object with microwave then placing..

Retrieve

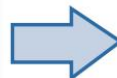
Skill Library

- ☐ Skill-01
- ☐ Skill-05
- ☐ Skill-14

Strategy: "The desk lamp 1 was used on alarmclock 2 instead..."

Description: Useful when the agent needs to target a specific object under a desk lamp before using the desk lamp to view...

Re-Rank



Skill Library

- ☒ Skill-14 (Selected)
- ☐ Skill-09
- ☐ Skill-01
- ☐ Skill-08

Stage 2: Utilization

Skill-14



Environment

Skill-Conditioned Rollout



τ_1

τ_2

τ_3

τ_6

G Rollouts

Query Generator

Re-Ranker

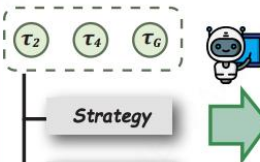
Policy Agent

Skill Generator



$$J = J_{util} + \lambda_1 J_{rerank} + \lambda_2 J_{distill}$$

Stage 3: Distillation



Successful Trajectories
 τ_6

Skill Library

- ☐ Skill-14
- ☐ Skill-09
- ☐ Skill-01
- ☐ Skill-08
- ☒ Skill-16 (Added)

Selection & Utilization & Internalization——Skill1

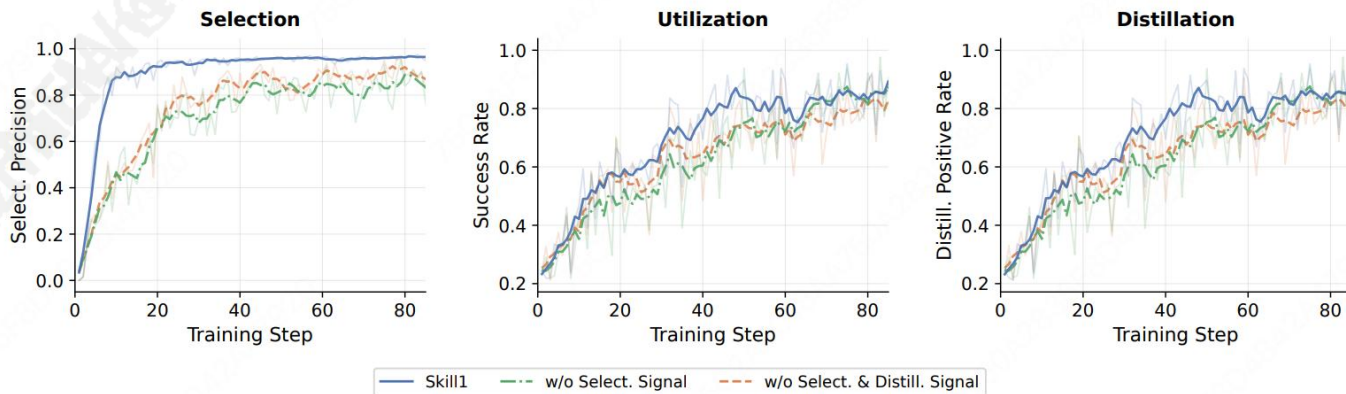


Method	ALFWorld (Success %)							WebShop		
	Pick	Look	Clean	Heat	Cool	Pick2	Avg.	Score	Succ.	
w/o Training										
Zero-Shot	33.4	21.6	19.3	6.9	2.8	3.2	14.8	26.4	7.8	
ReAct (Yao et al., 2022b)	48.5	35.4	34.3	13.2	18.2	17.6	31.2	46.2	19.5	
Reflexion (Shinn et al., 2023)	62.0	41.6	44.9	30.9	36.3	23.8	42.7	58.1	28.8	
Mem0 (Chhikara et al., 2025)	54.0	55.0	26.9	36.4	20.8	7.7	33.6	23.9	2.0	
ExpeL (Zhao et al., 2024)	21.0	67.0	55.0	52.0	71.0	6.0	46.3	30.9	11.2	
RL-Trained w/o Skills										
PPO (Schulman et al., 2017)	92.3	64.0	92.5	89.5	80.3	68.8	80.4	81.4	68.7	
RLOO (Ahmadian et al., 2024)	87.6	78.2	87.3	81.3	71.9	48.9	75.5	80.3	65.7	
GRPO (Shao et al., 2024)	90.8	66.1	89.3	74.7	72.5	64.7	77.6	79.3	66.1	
GiGPO (Feng et al., 2025)	97.7	82.7	98.8	83.7	89.3	79.2	90.8	84.4	72.8	
RL-Trained w/ Skills										
EvolveR (Wu et al., 2025)	64.9	33.3	46.4	13.3	33.3	33.3	43.8	42.5	17.6	
Mem0 (Chhikara et al., 2025) w/ GRPO	78.1	54.8	56.1	31.0	65.0	26.9	54.7	58.1	37.5	
SimpleMem (Liu et al., 2026a) w/ GRPO	89.5	36.3	60.0	50.0	64.9	26.3	62.5	67.8	46.9	
SkillRL (Xia et al., 2026)	97.9	71.4	90.0	90.0	95.5	87.5	89.9	85.2	72.7	
RetroAgent (Zhang et al., 2026a)	97.9	90.9	99.2	92.9	85.3	91.0	94.9	88.9	82.3	
Skill1 (Ours)	100.0 _{↑2.1}	98.6 _{↑7.7}	97.3	99.2 _{↑6.3}	96.1 _{↑0.6}	96.0 _{↑5.0}	97.5 _{↑2.6}	89.7	82.9	

Selection & Utilization & Internalization——Skill1



	Pick	Look	Clean	Heat	Cool	Pick2	Avg.
Skill1	100.0	98.6	97.3	99.2	96.1	96.0	97.5
w/o Select.	96.9	90.3	98.0	90.4	86.5	85.3	91.8
w/o Distill.	97.4	88.5	98.1	96.1	87.6	89.5	92.4
w/o Library	96.7	71.5	94.9	70.7	71.5	65.5	80.9
w/ $\lambda_1=0$	99.5	80.5	98.8	100.0	90.6	84.9	94.0
w/ $\lambda_2=0$	100.0	85.4	95.5	96.4	91.0	96.2	94.9
w/ $\lambda_1=\lambda_2=0$	98.1	74.9	95.6	95.6	79.5	87.2	90.2



Discussion