CRADLE: Empowering Foundation Agents Towards General Computer Control

Weihao Tan[3 *], Wentao Zhang[3 *], Xinrun Xu[5 *], Haochong Xia[3 †], Ziluo Ding[2 †], Boyu Li[3 †], Bohan Zhou[4 †], Junpeng Yue[4 †], Jiechuan Jiang[4 †], Yewen Li[3 †], Ruyi An[3 †], Molei Qin[3 †], Chuqiao Zong[3 †], Longtao Zheng[3 †], YuJie Wu[1 †], Xiaoqiang Chai[1 †], Yifei Bi[2], Tianbao Xie[6], Pengjie Gu[3], Xiyun Li[2], Ceyao Zhang[7], Long Tian[1], Chaojie Wang[1], Xinrun Wang[3 ‡], Börje F. Karlsson[2 ‡], Bo An[3, 1 §], Shuicheng Yan[1 §], Zongqing Lu[4, 2 §]

[1] Skywork AI [2] Beijing Academy of Artificial Intelligence [3] Nanyang Technological University, Singapore
[4] Peking University [5] Institute of Software, Chinese Academy of Sciences [6] The University of Hong Kong
[7] The Chinese University of Hong Kong, Shenzhen

[*] Equal contribution [†] Core contribution [‡] Equal advising [§] Corresponding authors

General Computer Control (GCC)

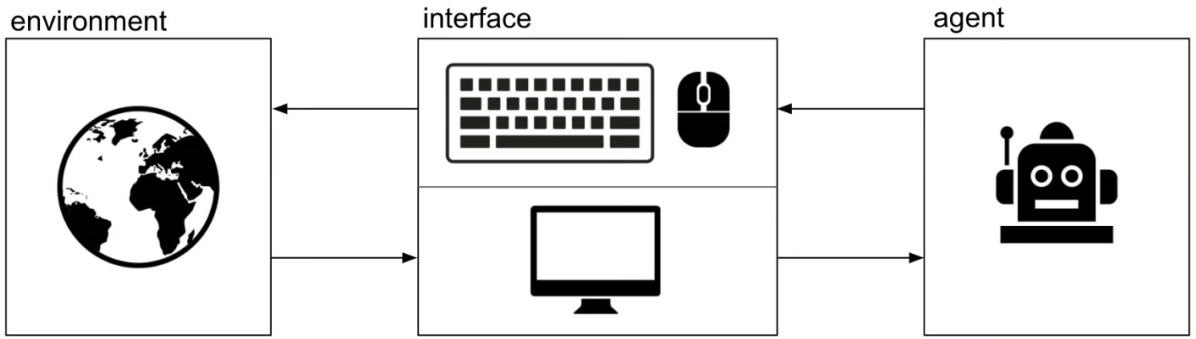
Advantages

NO built-in APIs needed

Generalization to (unseen) tasks

Data collection with the same granularity for **self-improvement**

Building foundation agents that can master **ANY** computer task via the universal human-style interface by receiving input from **screens** and **audio** and outputting **keyboard** and **mouse** actions.



Challenges

- **Good Alignment** across multimodal inputs
- **Precise control** of keyboard & mouse
- **Long-term memorizing & reasoning**
- **Efficient exploration** and **self-improving**

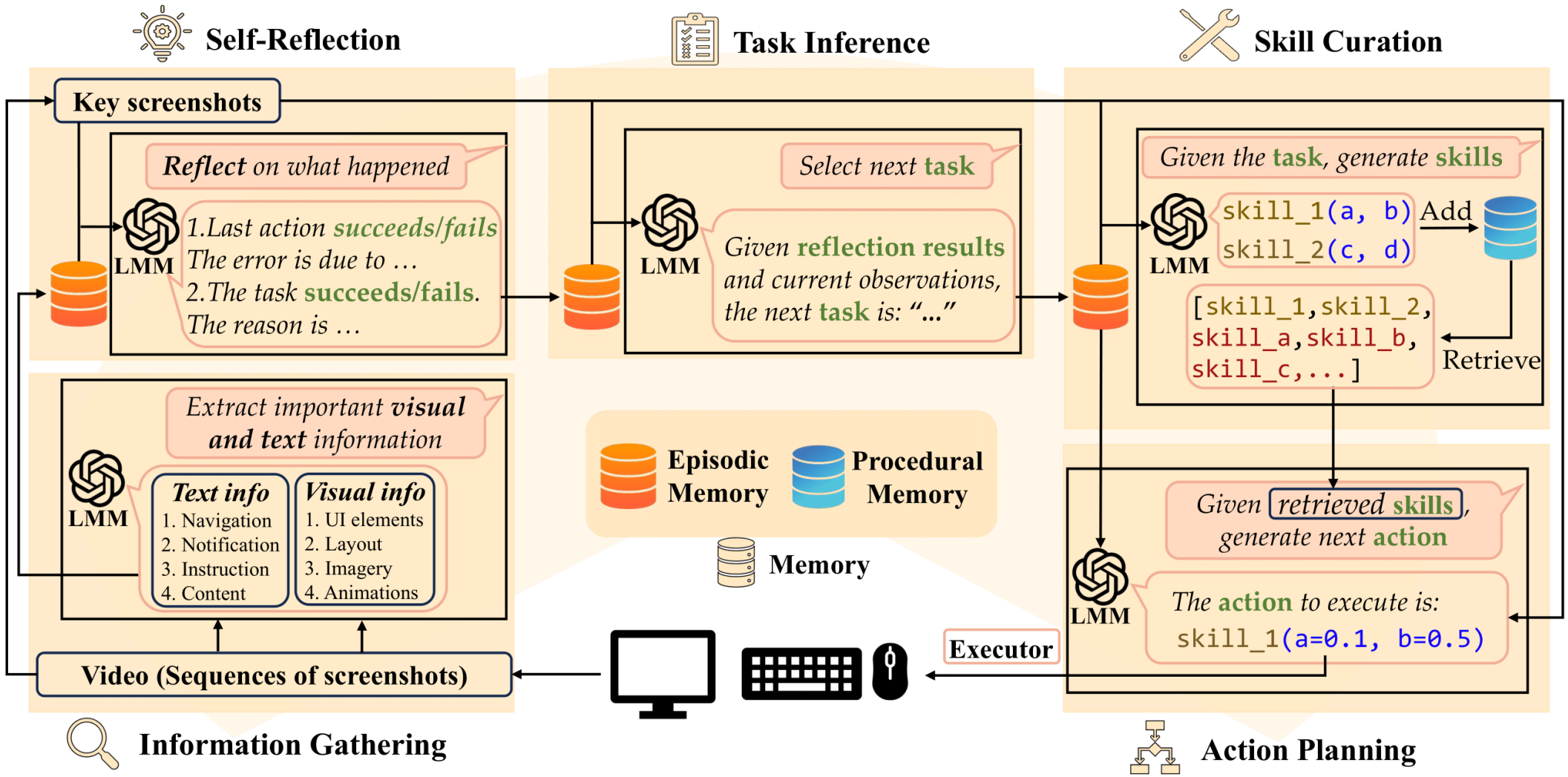
https://openai.com/research/universe

Cradle Framework

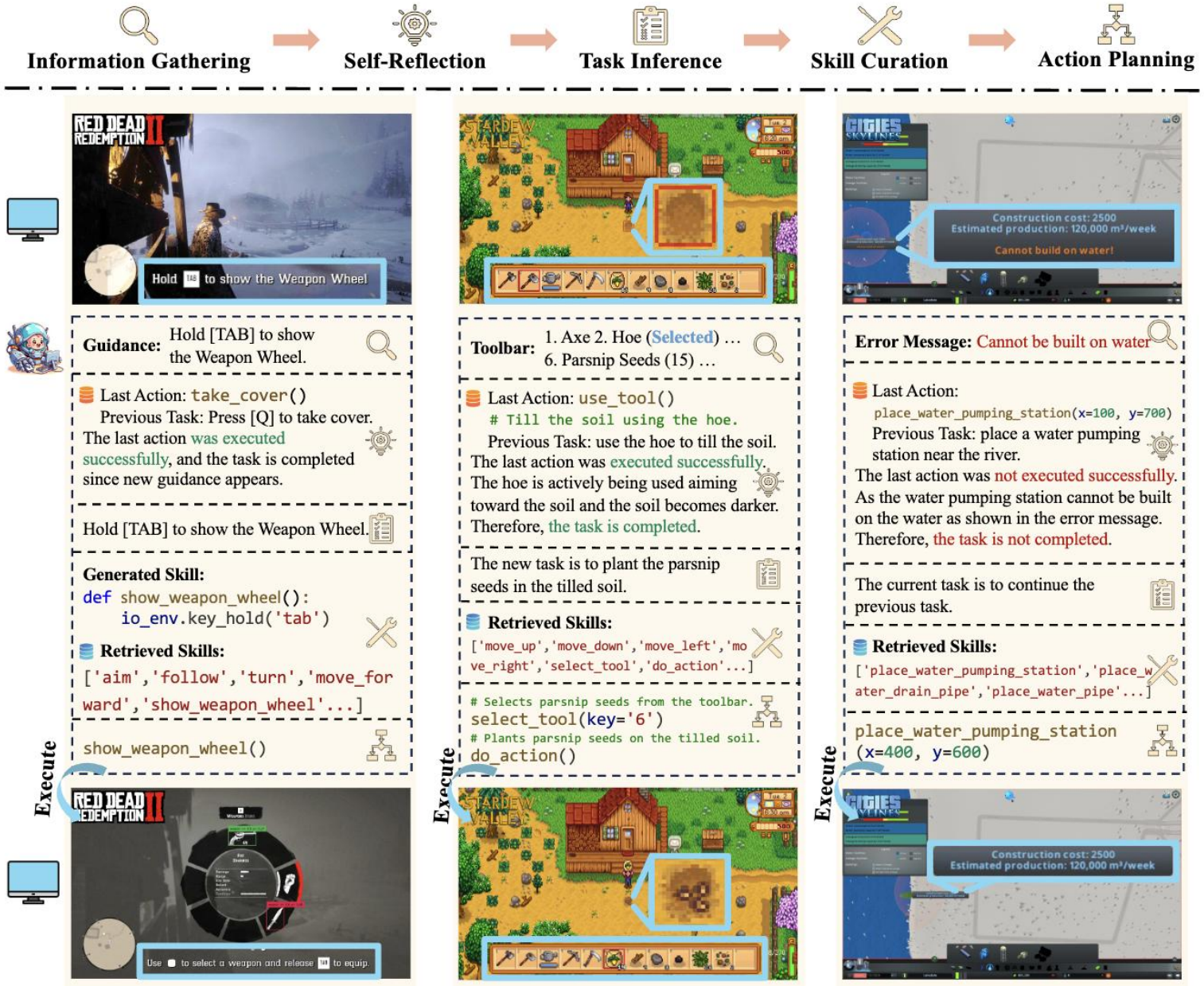
- Cradle: A Novel Agent Framework under GCC Setting

Cradle Framework

- Six key modules

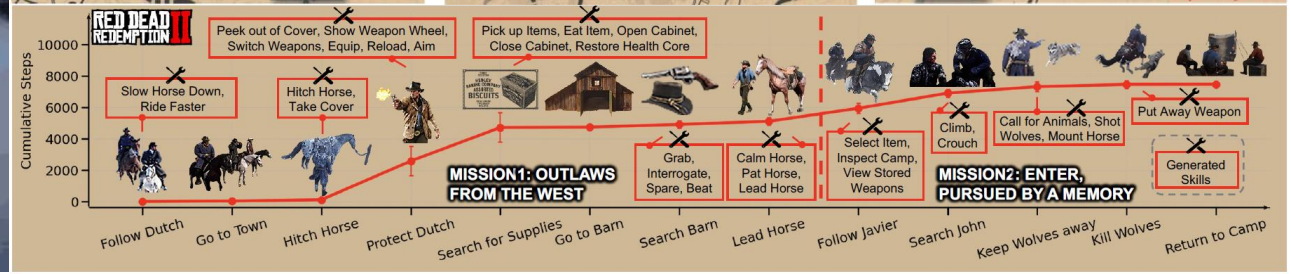


Complete Workflow



Cradle in RDR2

- 60 minutes main storyline missions
- Long-horizon open-ended tasks



Cradle in Other Games

- **Stardew Valley**
  - Farm clearup, cultivation and shopping
- **Dealer's Life**
  - Run a pawn shop for a week and maximize profits
- **Cities: Skylines**
  - Build a reasonable city and maximize the population with initial budgets

Stardew Valley		
Task	CRADLE	Human
Farm Clearup Grids	14.8 ± 5.0	**35.2** ± 14.5
Cultivation	4/5	5/5
Shopping	1/5	5/5

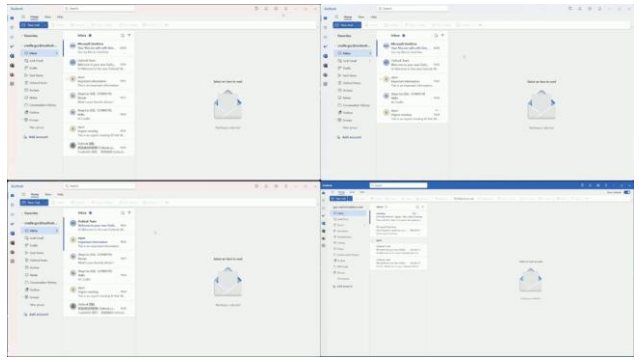
Dealer's Life 2		
Metrics	CRADLE	Human
Avg. Haggling Count	**1.95** ± 0.43	1.63 ± 0.53
Turnover Rate (%)	**93.6** ± 6.9	68.4 ± 22.2
Item Profit Rate (%)	**37.8** ± 19.1	21.1 ± 13.6
Total Profit Rate (%)	**39.6** ± 27.3	17.3 ± 15.1

Cities: Skylines		
Metrics	CRADLE	Human
Smooth Closed-loop Road	4/5	5/5
Sufficient Water Supply	1/5	3/5
Sufficient Power Supply	5/5	5/5
High Zoning Coverage	4/5	4/5
Population	**450** ±224	415 ±416

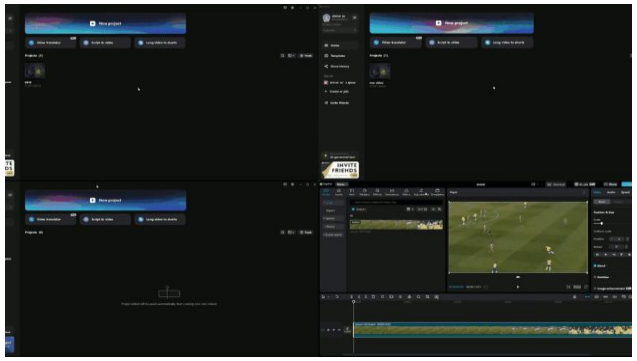
Cradle in Software Applications

Chrome

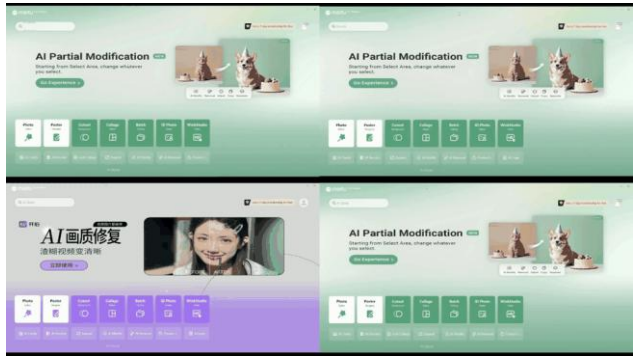
Outlook



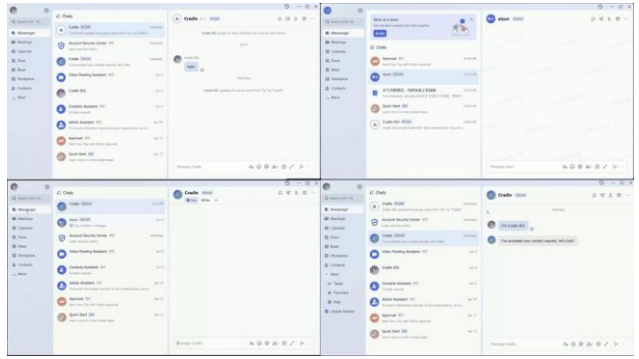
CapCut



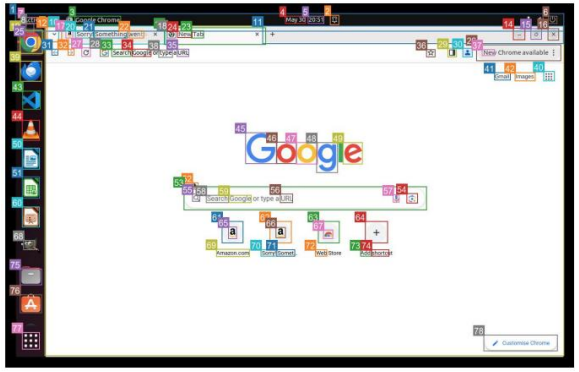
Meitu Xiuxiu

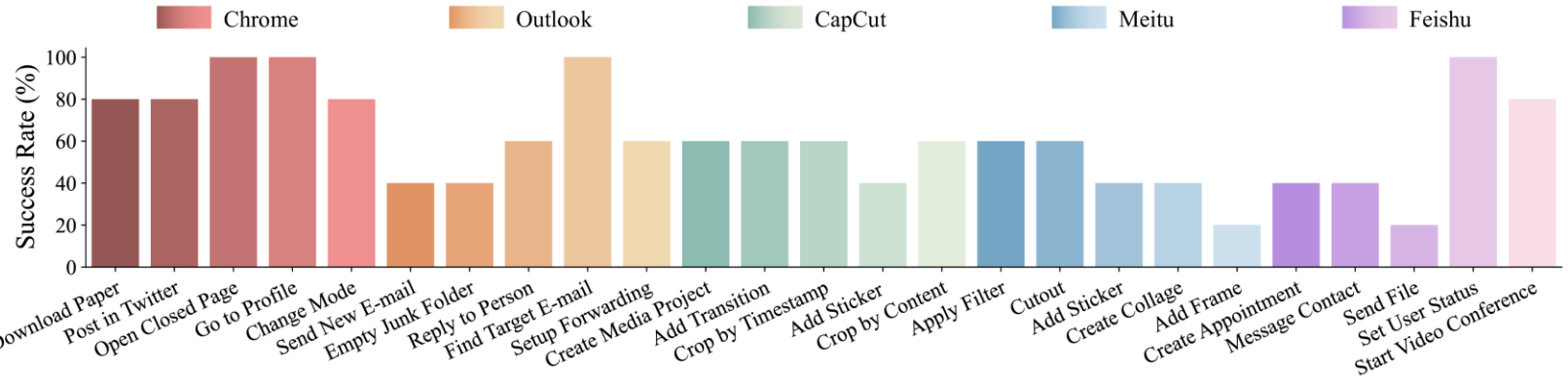


Feishu



SAM2SOM Labels



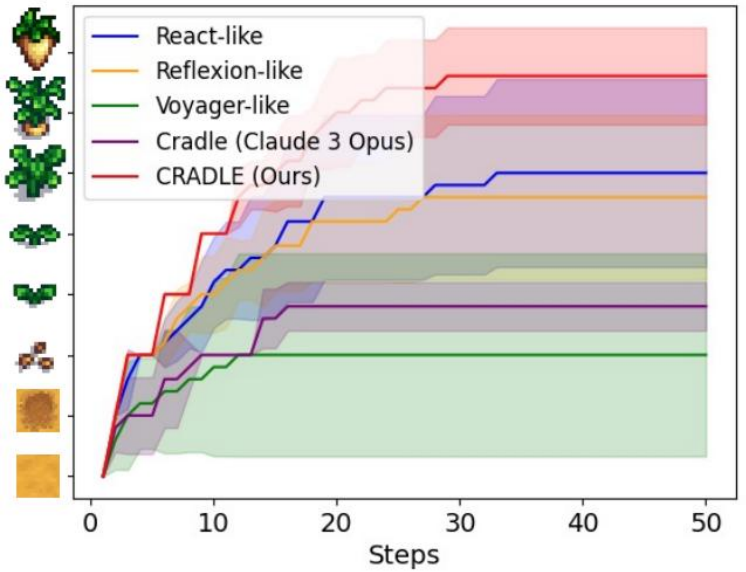


Cradle in OSWorld

Method	Office (117)	OS (24)	Daily (78)	Workfl-ow(101)	Professi-onal (49)	All (369)
GPT-4o	**3.58**	8.33	6.07	**5.58**	4.08	5.03
GPT-4o+SoM	**3.58**	**20.83**	3.99	3.60	2.04	4.59
CRADLE	**3.58**	16.67	**6.55**	5.48	**20.41**	**7.81**

Ablation Studies & Baselines Comparison

Method	Follow Dutch	Follow Micah	Hitch Horse	Protect Dutch	Search for Supplies	Cultivation
React [72]-like (GPT-4o)	15 ± 2 ($5/5$)	74 ± 0 ($1/5$)	N/A	N/A	N/A	N/A
Reflextion [53]-like (GPT-4o)	19 ± 4 ($5/5$)	58 ± 14 ($2/5$)	N/A	N/A	N/A	N/A
Voyager [57]-like (GPT-4o)	32 ± 12 ($3/5$)	N/A	N/A	N/A	N/A	N/A
CRADLE (Claude 3 Opus)	30 ± 7 ($5/5$)	52 ± 17 ($4/5$)	N/A	N/A	N/A	N/A
CRADLE (GPT-4o) (Ours)	**13 ± 3** ($5/5$)	**33 ± 3** ($5/5$)	**26 ± 5** ($4/5$)	**461 ± 0** ($1/5$)	**134 ± 0** ($1/5$)	**24 ± 4** ($4/5$)



Subtask	w/o Information Gathering	w/o Self-Reflection	w/o Task Inference	w/o Episodic Memory	CRADLE
Follow Micah	0%	0%	40%	80%	**100%**
Hitch Horse	0%	**100%**	**100%**	**100%**	**100%**
Go to Shed	0%	20%	40%	20%	**80%**
Peek out of Cover	60%	**100%**	80%	**100%**	**100%**
Switch Weapon	0%	80%	60%	80%	**100%**
Combat	0%	0%	0%	0%	**20%**

Looking Ahead

- Modular growth
  - As base model's capabilities improves, Cradle improves accordingly!
  - Some or even all of modules can be combined into a single request.
- Self-improvement
  - Greatly extends the reach of AI agents. **Every software is potential out-of-box testbed and benchmark.**
  - Cradle collects both high-level semantic actions and low-level keyboard and mouse control for self-improvement.