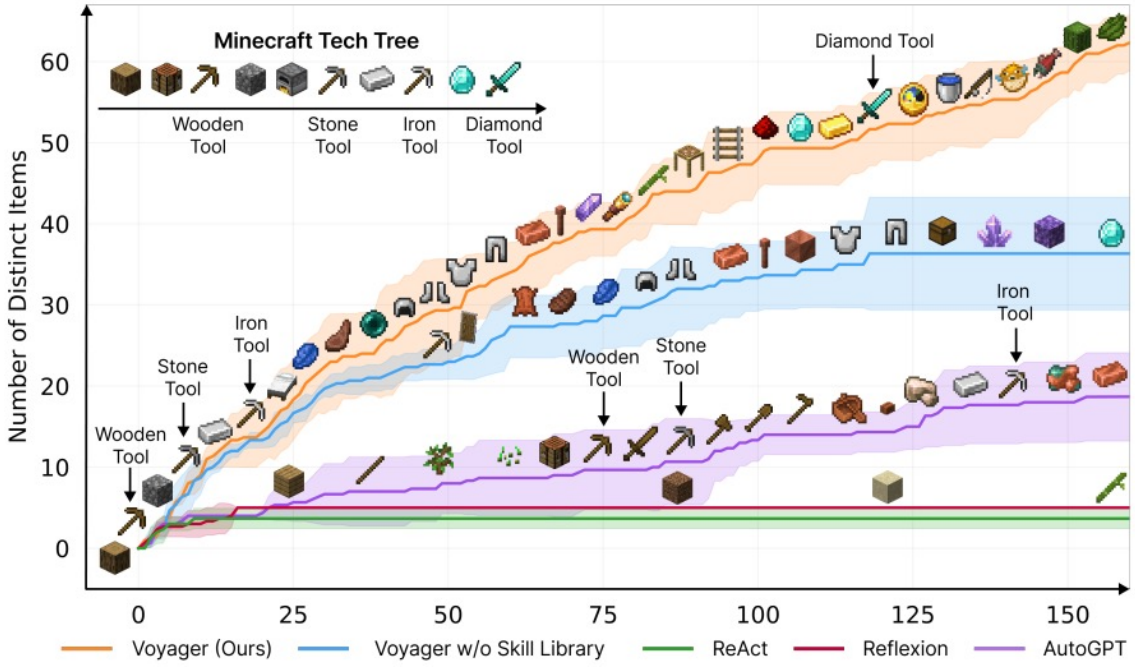
Towards Unified Alignment Between Agents, Humans, and Environment

Zonghan Yang
Tsinghua University

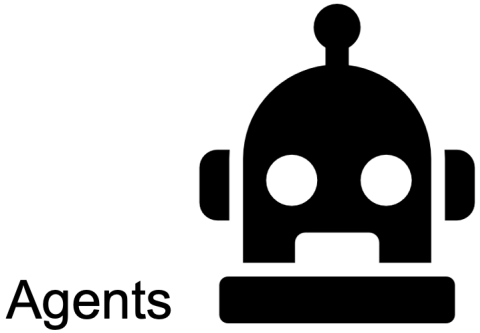
The Prosperity of Agents



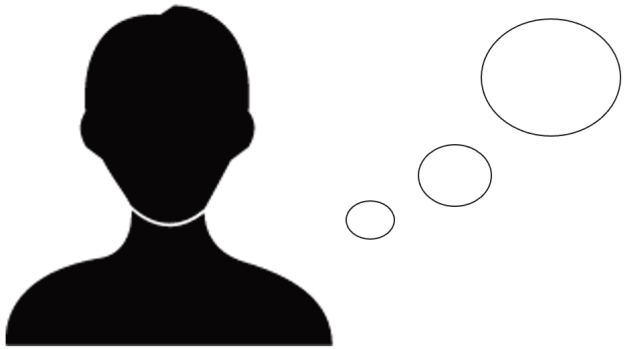
Task Solving (Wang et al., 2023)

Behavior Simulation (Park et al., 2023)

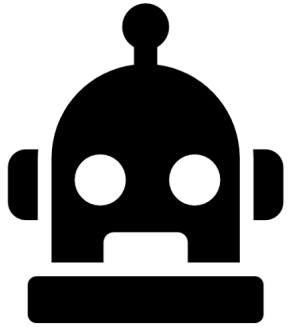
A Working System of Agents



A Working System of Agents

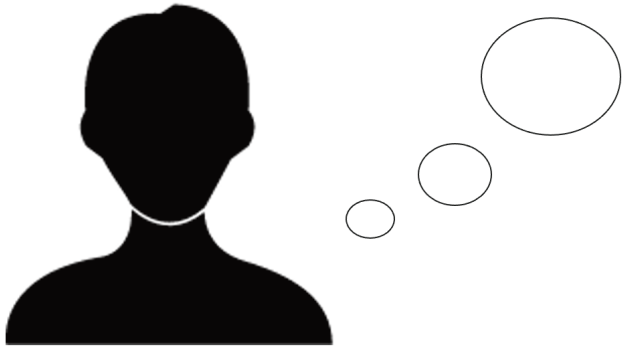


Humans



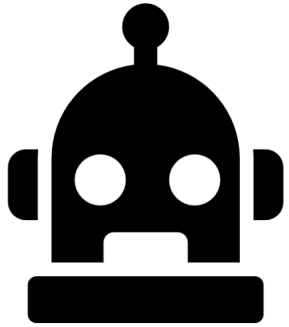
Agents

A Working System of Agents



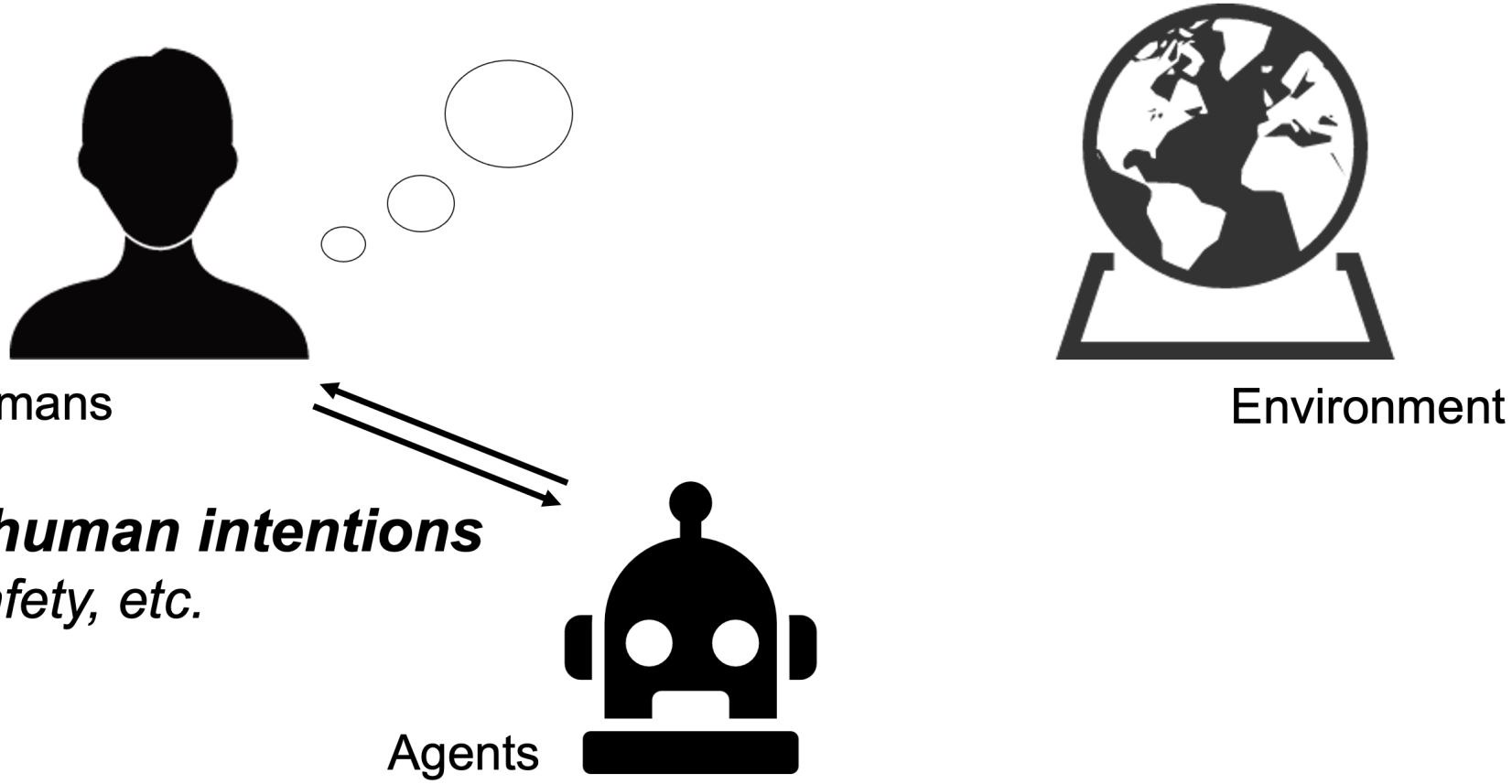
Humans

Environment



Agents

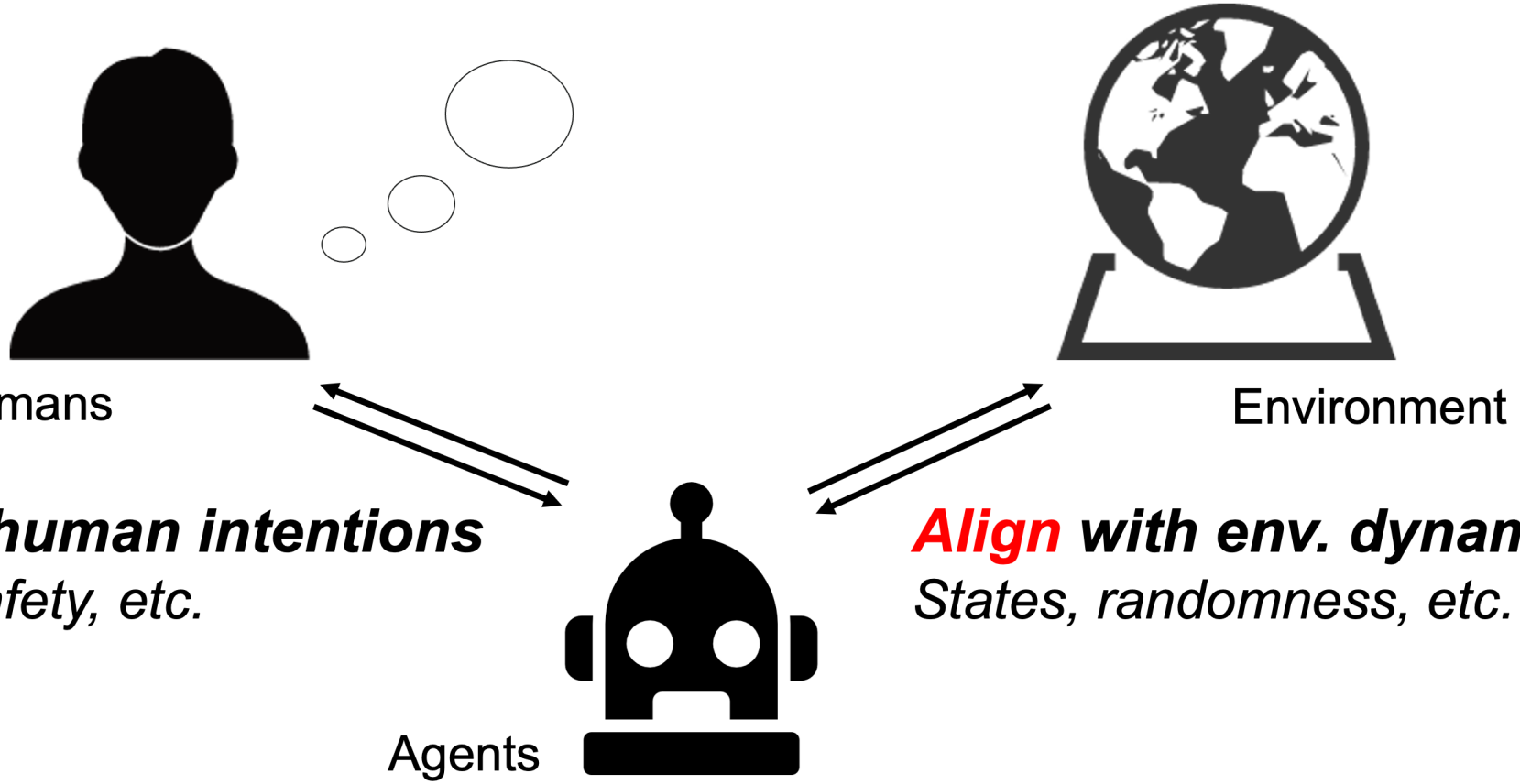
Unified Alignment for Agents (UA^2)



***Align* with human intentions**

Ambiguity, safety, etc.

Unified Alignment for Agents (UA^2)



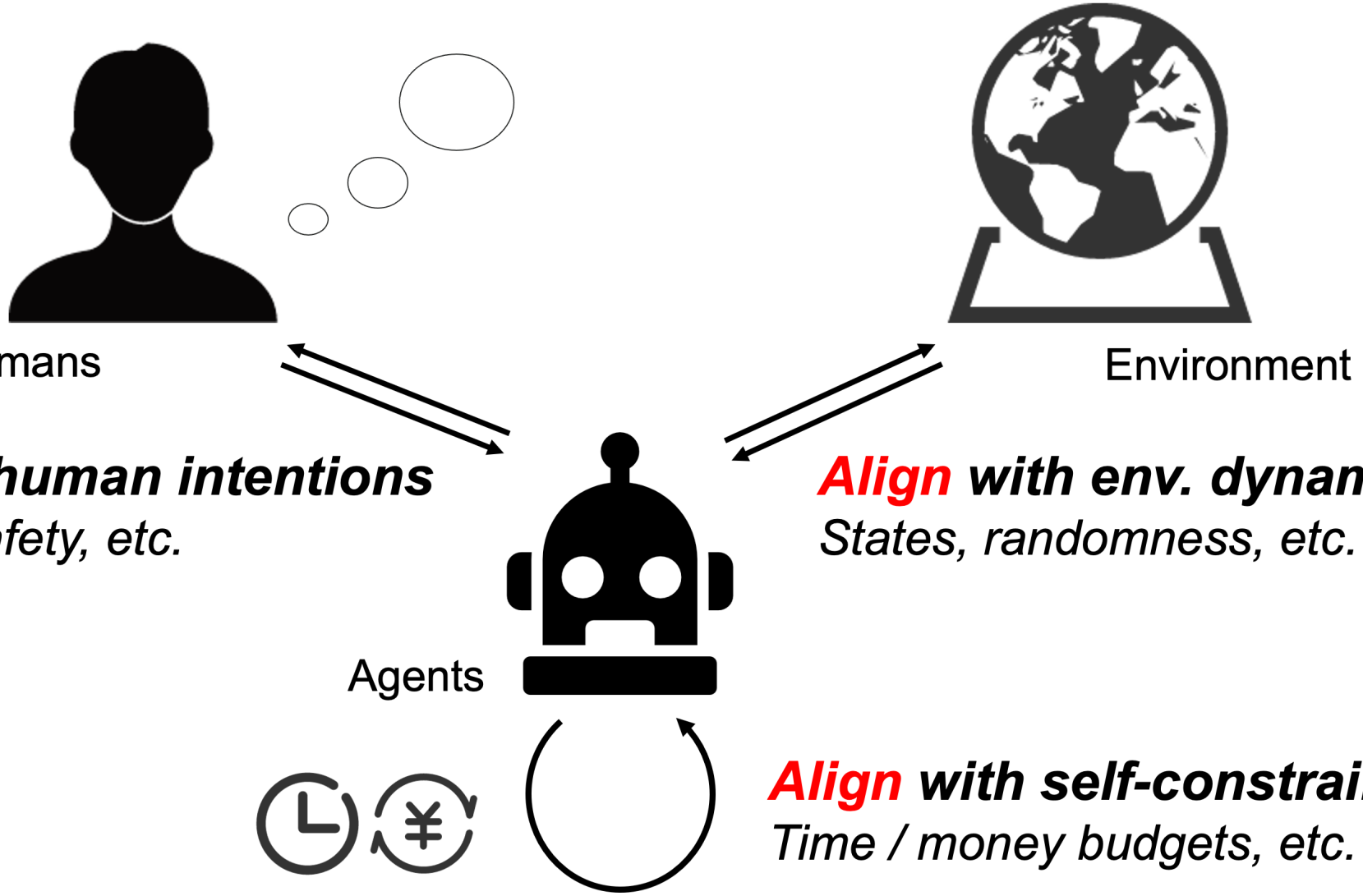
***Align* with human intentions**

Ambiguity, safety, etc.

***Align* with env. dynamics**

States, randomness, etc.

Unified Alignment for Agents (UA^2)

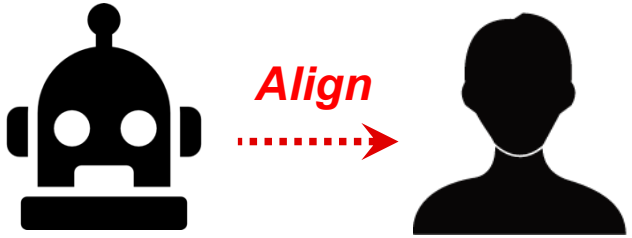


***Align* with human intentions**
Ambiguity, safety, etc.

***Align* with env. dynamics**
States, randomness, etc.

***Align* with self-constraints**
Time / money budgets, etc.

UA²: Aligning with Human Intentions



User profiles

I love the cute creatures in the world. I prefer ____ over ____. And I do not like ____ . In fact, My special likes are ____.

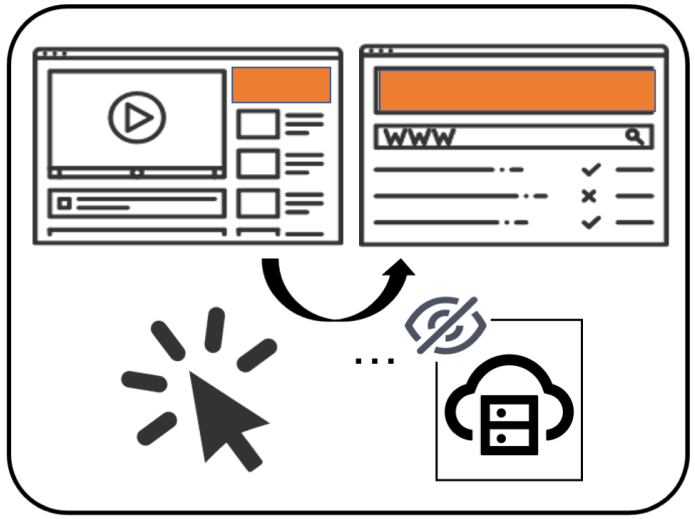
Ambiguity

Based on my purchase preference from history, help me buy some good face towels that cost as little as possible.

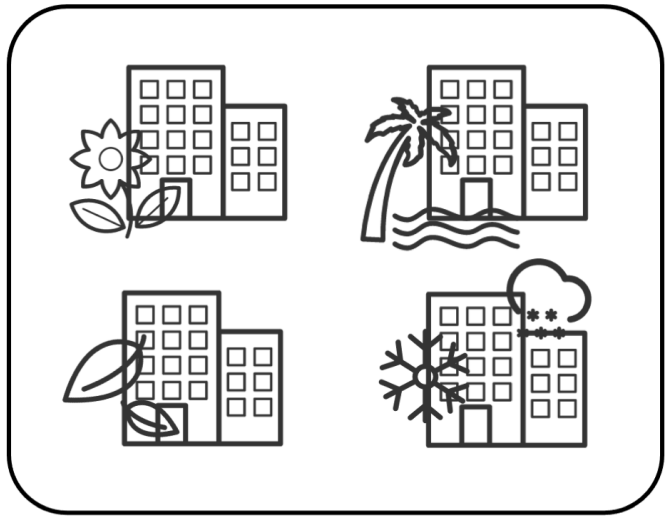
Safety concerns

DO NOT take malicious or destructive actions in the environment. Treat all the action categories with fairness and equity.

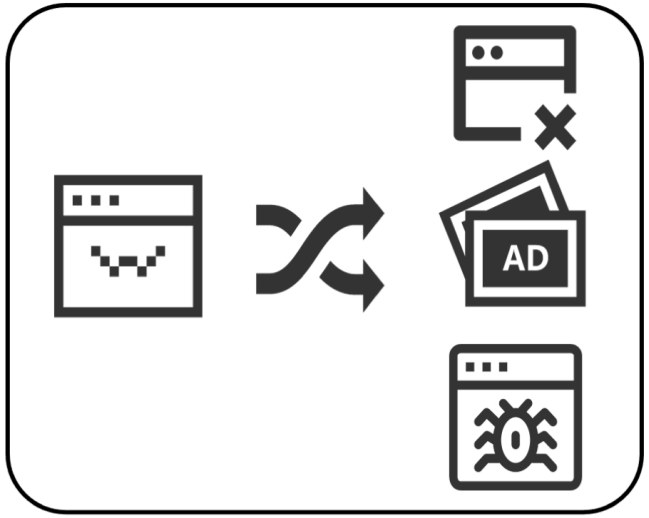
UA^2: Aligning with Env. Dynamics



Partial observability

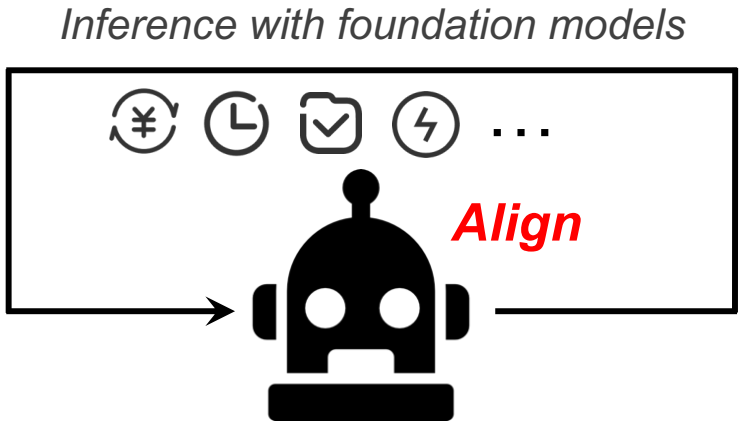


Temporality



Stochasticity

UA^2: Aligning with Self-Constraints



Money cost

Time limit

Storage capacity

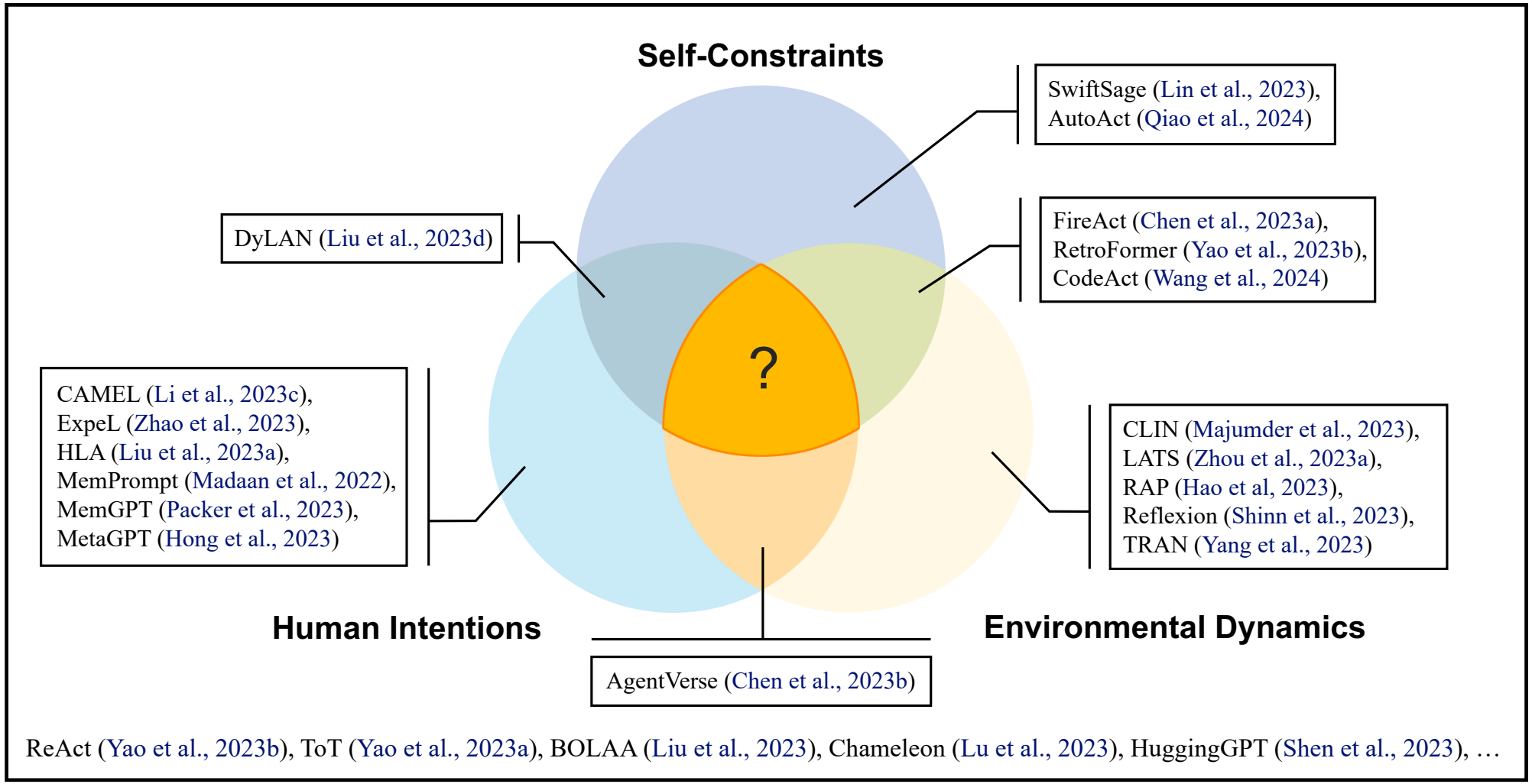
Battery cost

...

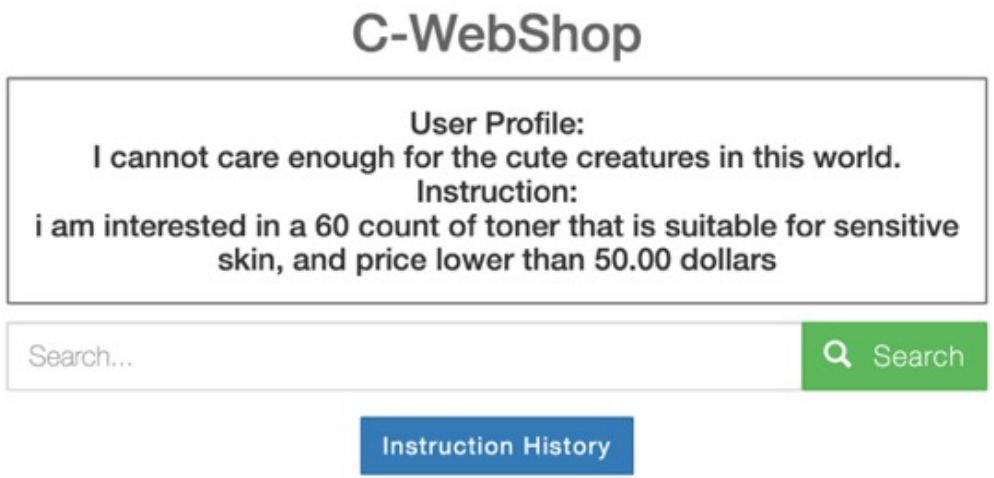
Benchmarks Review from UA^2

Type	Benchmarks	Human Intentions	Environmental Dynamics	Self-Constraints
Digital	Androidenv (Toyama et al., 2021)	None	Partial Obs.	None
	WebShop (Yao et al., 2022a)	None	Full Obs.†	None
	Mind2Web (Deng et al., 2023)	None	Partial Obs.	None
	ToolBench (Qin et al., 2023)	None	Full Obs. & Temp. & Stoch.	None
	WebArena (Zhou et al., 2023b)	Fixed and Given	Partial Obs.	None
Embodied	VirtualHome (Puig et al., 2018)	None	Partial Obs.	None
	BabyAI (Chevalier-Boisvert et al., 2019)	None	Partial Obs.	None
	ALFWorld (Shridhar et al., 2020)	None	Partial Obs.	None
	MineDojo (Fan et al., 2022)	None	Partial Obs. & Stoch.	None
	ScienceWorld (Wang et al., 2022a)	None	Partial Obs.	None
	Interactive Gibson (Xia et al., 2020)	None	Partial Obs.	#Actions
	AGENT (Shu et al., 2021)	None	Partial Obs.	#Actions
	RFUniverse (Fu et al., 2022)	Fixed and Given	Partial Obs.	#Actions
	BEHAVIOR-1K (Li et al., 2023b)	None	Full Obs.	#Actions
	HAZARD (Zhou et al., 2024)	None	Partial Obs. & Temp.	#Actions
Mixed	MINT (Wang et al., 2023b)	None	Partial Obs.	#Actions
	SmartPlay (Wu et al., 2023)	None	Partial Obs. & Stoch.	None
	AgentBench (Liu et al., 2023c)	None	Partial Obs.	None
	AgentBoard (Ma et al., 2024)	None	Partial Obs. & Temp. & Stoch.	None

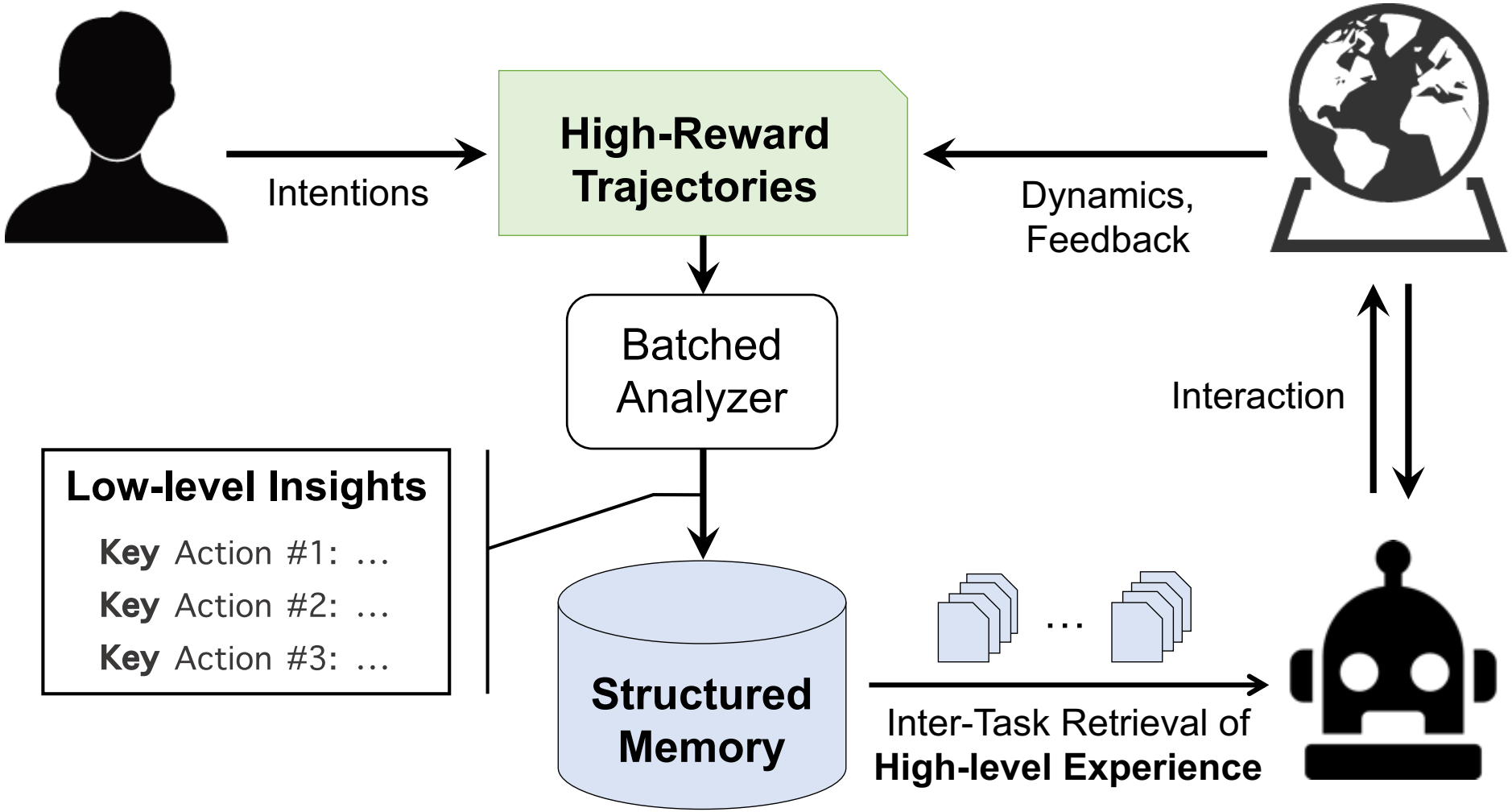
Techniques Review from UA^2



Our Benchmark with UA^2



Our Framework with UA^2



Our Evaluation with UA²

Method	Reward ↑	SR (%) ↑	G_{HI} (%) ↓	G_{ED} (%) ↓	Time (s) ↓	Money ($) ↓
ReAct	50.3	8.0	11.7	14.9	**1.716**	**0.013**
ReAct-SC	49.9	7.4	14.4	14.6	1.720	0.039
Reflexion	44.4	**13.8**	22.5	25.7	5.539	0.045
LATS*	**52.4**	10.0	18.5	**14.3**	125.935	5.508
Ours	51.9	9.6	**6.7**	14.8	1.779	0.014

$$\mathbf{G}_{\mathrm{HI}} = (R_{\mathrm{full}} - R_{\mathrm{HI}})/R_{\mathrm{full}} \times 100\% \qquad \mathbf{G}_{\mathrm{ED}} = (R_{\mathrm{full}} - R_{\mathrm{ED}})/R_{\mathrm{full}} \times 100\%$$

R_{full} : reward in the fully-retrofitted environment
R_{HI} : reward in the environment **w/o** the computation of human intentions
R_{ED} : reward in the environment **w/o** personalized reranking mechanisms

Actionable Insights

- Synergizing agents with alignment research
- Constructing realistic agent benchmarks
- Holistic evaluations for agent frameworks
- On agent frameworks that self-evolve
- Details in https://arxiv.org/abs/2402.07744

Scan For More!