From History to State: Constant-Context Skill Learning for LLM Agents

NeurIPS 2026 Main Track

Haoyang Xie, Xinyuan Wang, Yancheng Wang, Puda Zhao, Feng Ju
School of Computing and Augmented Intelligence, ASU

Arizona State University

Copyright © 2026 Arizona Board of Regents

Personal agents face a three-way tension

1

Cloud models

Strong on multi-step workflows, but emails, files, and tool outputs leave the device.

Capable, not private

2

Local models

Keep data on the device, but are less reliable on long, multi-step tasks.

Private, less capable

3

Both

Reread long skill prompts and a growing history at every step.

Costly at every step

Goal: make small local models reliable on recurring workflows, with a constant-size input.

Copyright © 2026 Arizona Board of Regents

From history to state: a constant-size input

Recurring workflows repeat one procedure with different details, e.g. sorting email every morning.

Standard agent: the prompt grows with every step

Step 1	Goal	Skill text	Obs.	
Step 5	Goal	Skill text	History so far	
Step 10	Goal	Skill text	Full execution history	

Our agent: the prompt stays short at every step

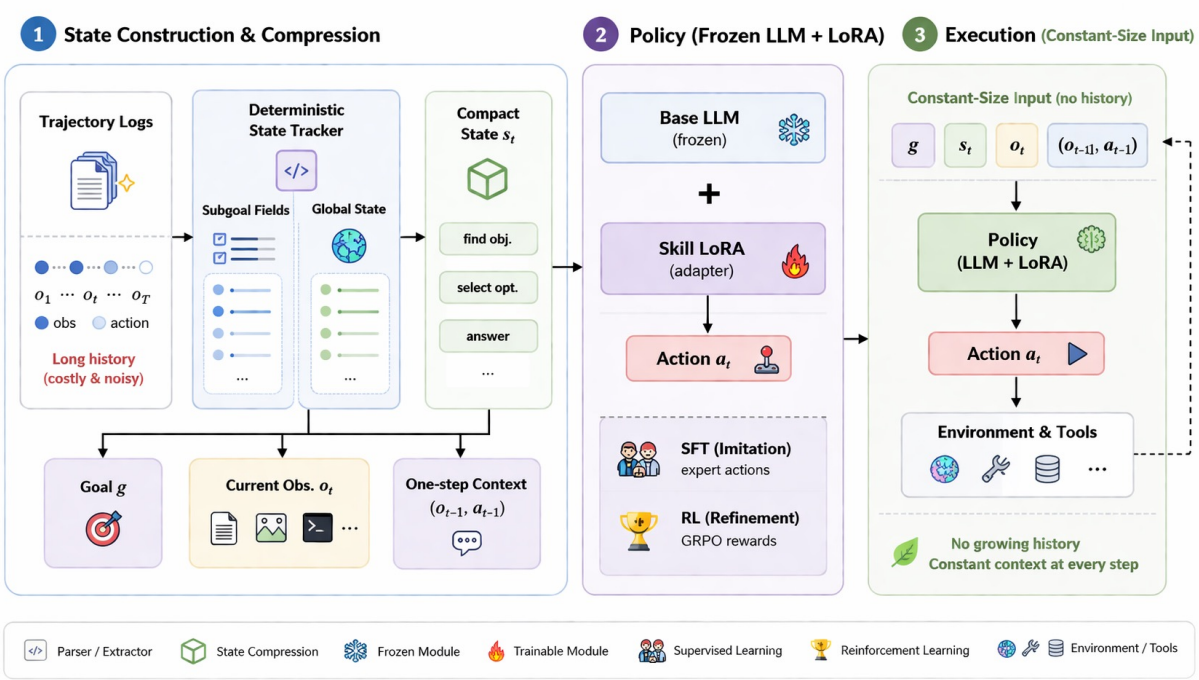
Step 1	Goal	Obs.	Previous step	**State block**
Step 5	Goal	Obs.	Previous step	**State block**
Step 10	Goal	Obs.	Previous step	**State block**

Procedure → LoRA weights

Progress → short state block

Copyright © 2026 Arizona Board of Regents

Pipeline: one tracker for state and reward



1 Deterministic tracker

Plain-code rules, no LLM calls, render the state block.

2 Step-level SFT

Expert runs give (state, action) pairs. One LoRA per task family; base frozen.

3 Online RL

GRPO-style updates with subgoal rewards from the same tracker.

Copyright © 2026 Arizona Board of Regents

Each step gets a short state, not a log

Step 1 find_object | **Step 2** take_object | **Step 3 reach_dest** | **Step 4** place_object

Step 3: Go to desk 1



ALFWorld: the full input at this step.
No skill text, no history.

Goal	put a mug in desk
Previous step	**Saw:** You arrive at shelf 1. On the shelf 1, you see a mug 2. **Did:** take mug 2 from shelf 1
Current observation	You pick up the mug 2 from the shelf 1.
State block	Current subgoal: reach_dest Target object: mug Holding: mug 2 Destination: desk At destination: no
Next action	**go to desk 1**

Copyright © 2026 Arizona Board of Regents

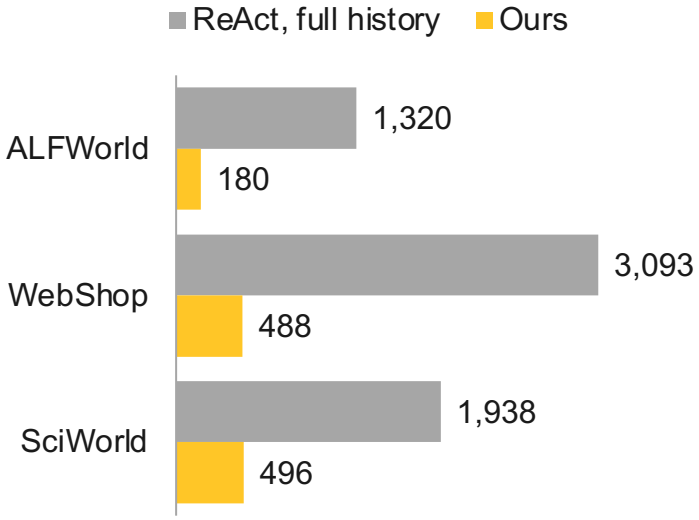
Near state of the art with 4–7× fewer tokens

Task performance vs. published methods

	Ours	Best published	ReAct
ALFWorld	89.6%	**90.8% (GiGPO)**	46.3%
WebShop	76.8%	**76.9% (EMPO2)**	3.8%
SciWorld	**79.7**	75.9 (EMPO2)	−21.5

Success rate (%), SciWorld: unseen task score. Ours: Qwen3-8B SFT+RL, trained on one A100. ReAct: same model, full history. Bold: best in row.

Prompt tokens per turn

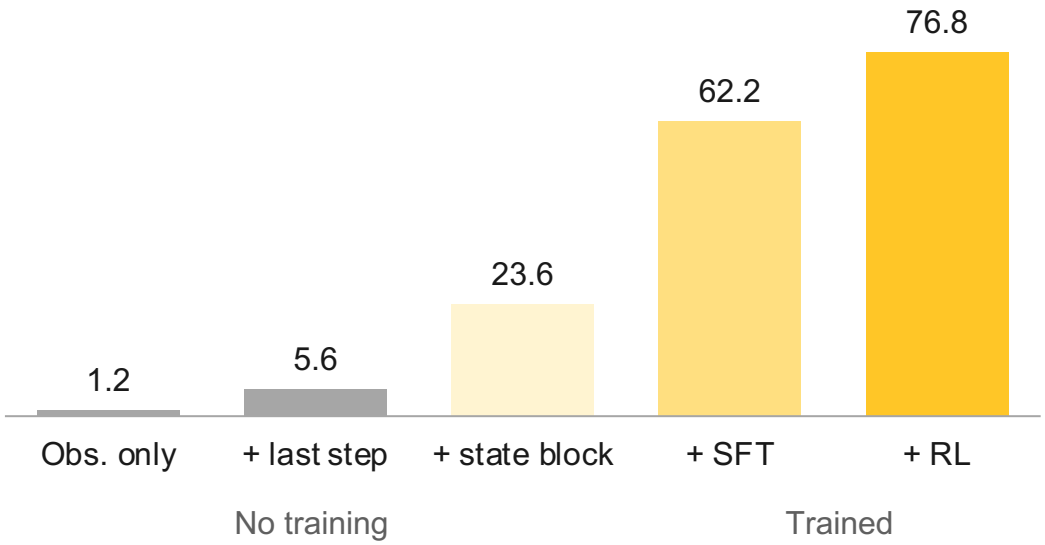


4–7× fewer than ReAct with full history

Copyright © 2026 Arizona Board of Regents

The state block and training both matter

WebShop success rate (%), Qwen3-8B



- **−15.6 pts** removing the state block after training
- **−4.6 pts** using only the final task reward in RL
- **4 backbones:** Qwen3-4B, 8B, 14B and Llama-3.1-8B all gain from SFT+RL
- **Robust:** paraphrased ALFWorld goals 89.1% vs. 89.6%

Copyright © 2026 Arizona Board of Regents



Thank you

From history to state

Haoyang Xie, hxie40@asu.edu

Paper

Code

Copyright © 2026 Arizona Board of Re