MobileIPL: Enhancing Mobile Agents' Thinking Process via Iterative Preference Learning

Kun Huang, Weikai Xu, Yuxuan Liu, Quandong Wang, Pengzhi Gao, Wei Liu, Jian Luan, Bin Wang, Bo An

Xiaomi Inc. · Nanyang Technological University · Renmin University of China

ICLR 2026

Overview

- **A Typical Mobile GUI Task**

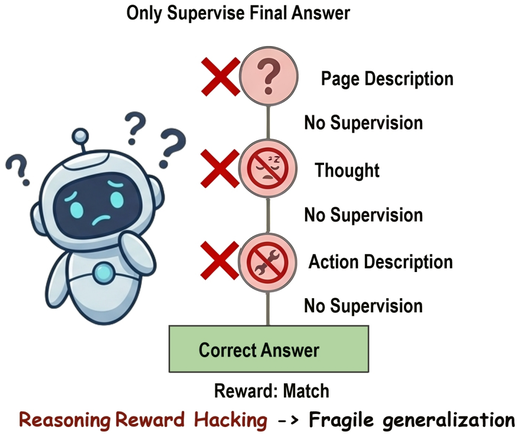
  - **Task:** Automating user tasks on mobile apps through GUI interaction.

  - **Example:** Find tomorrow's Beijing AQI in Caiyun Weather.

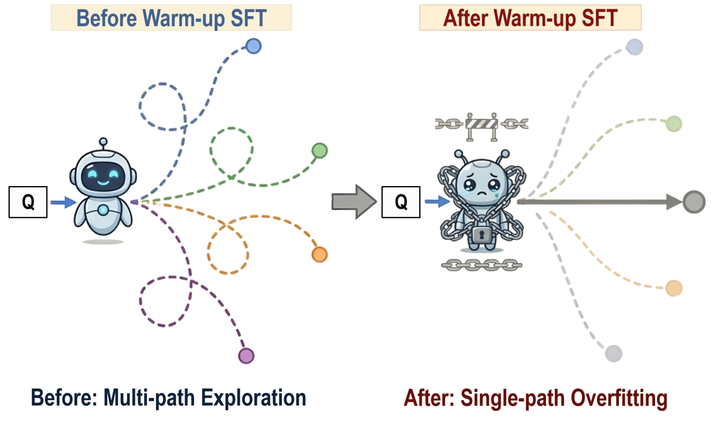
- **Two Key Challenges in GUI-Agent Self-training**

  - **Challenge 1:** Existing self-training supervises actions or outcomes, but not step-level thought quality.

  - **Challenge 2:** Warm-up SFT may overfit easy downstream patterns and reduce reasoning diversity.

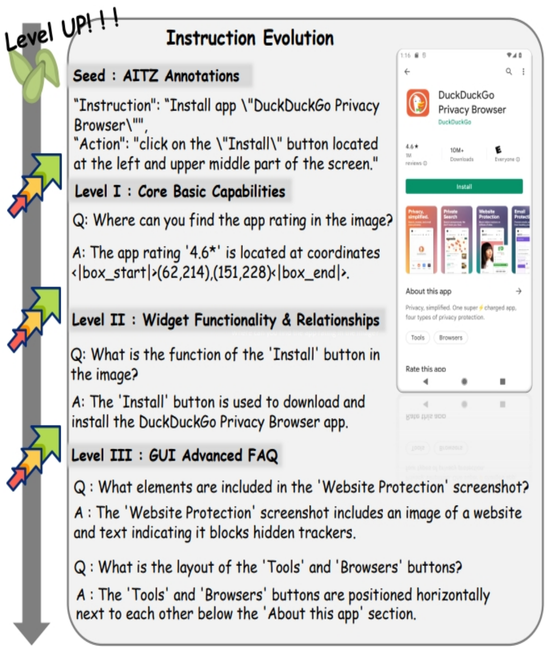


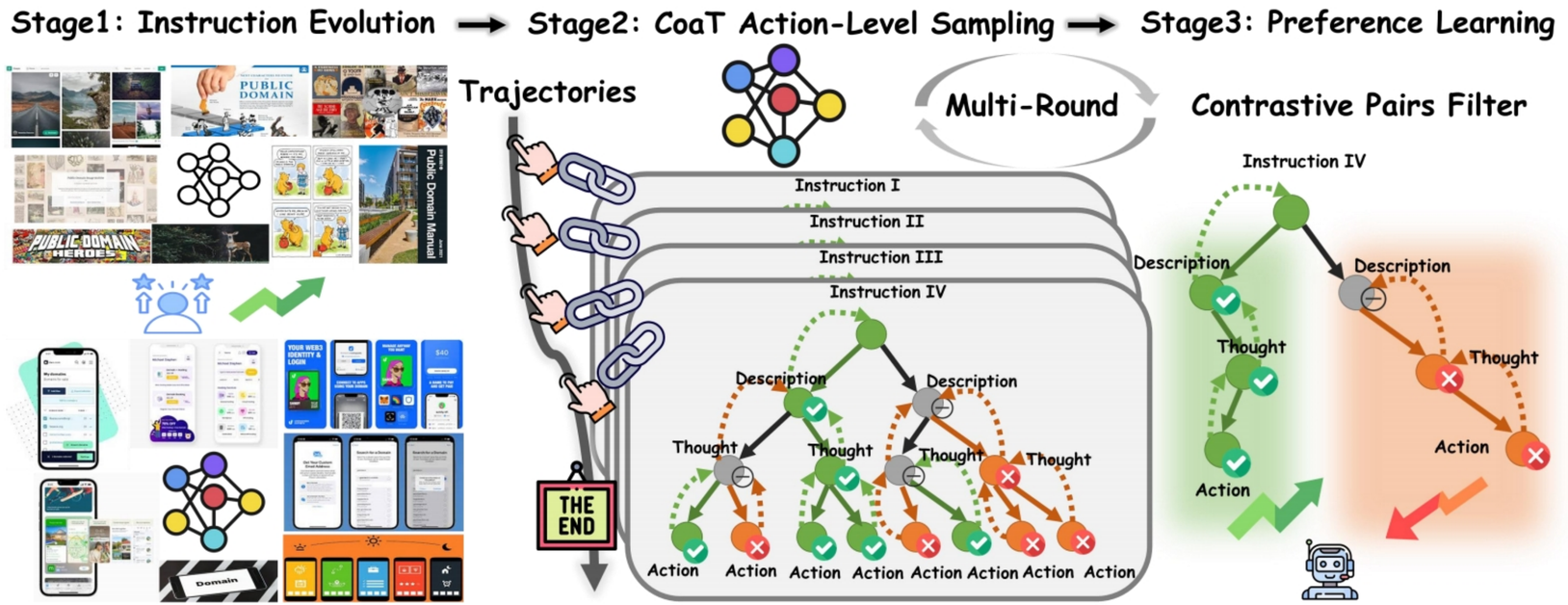
Challenge 1



Challenge 2

Method





- **Instruction Evolution:** warm up SFT with diverse Q&A pairs based on real mobile UI screenshots to improve GUI understanding and reduce overfitting.
- **CoaT Tree Sampling:** build a multi-turn reasoning tree for each action step to explore diverse step-level reasoning paths.
- **T-DPO Training:** score leaf actions and backpropagate rewards to optimize step-level thought quality.

Main Experiments

Model	Mode	Atomic								
		SCROLL	CLICK		TYPE		PRESS	STOP	Total	
			type	match	type	match			type	match
CogAgent (CoaT)	ZS	70.22	88.23	66.15	45.80	21.80	45.95	24.60	72.59	53.28
AUTO-GUI (CoaT)	FT	61.40	74.56	32.20	87.20	81.40	57.70	74.40	82.98	47.69
AriaUI-MoE	FT	53.73	85.51	60.20	84.20	80.80	63.70	76.38	78.53	63.56
Seeclick-7B	PF	11.14	69.92	52.96	53.80	53.00	67.88	55.36	62.93	49.11
UGround-7B	PF	58.22	80.94	58.48	82.56	73.85	58.22	68.78	74.54	60.19
OS-Atlas-7B	PF	76.12	75.82	54.83	87.80	81.60	68.67	**81.75**	77.83	65.11
UI-Tars-7B	PF	52.50	83.03	64.27	**89.97**	82.76	61.87	74.35	77.59	65.61
Falcon-UI-7B	PF	-	-	-	-	-	-	-	**84.70**	69.10
Qwen2-VL-7B (CoaT)	FT	47.50	81.53	59.72	81.96	73.85	**58.22**	67.39	74.26	60.36
AITZ-Seed	FT	42.83	82.48	53.16	82.56	75.29	56.65	61.82	73.14	55.40
MobileIPL	IPL	51.08	**91.73**	**71.45**	88.20	**83.40**	51.69	78.17	81.90	**69.15**

AITZ

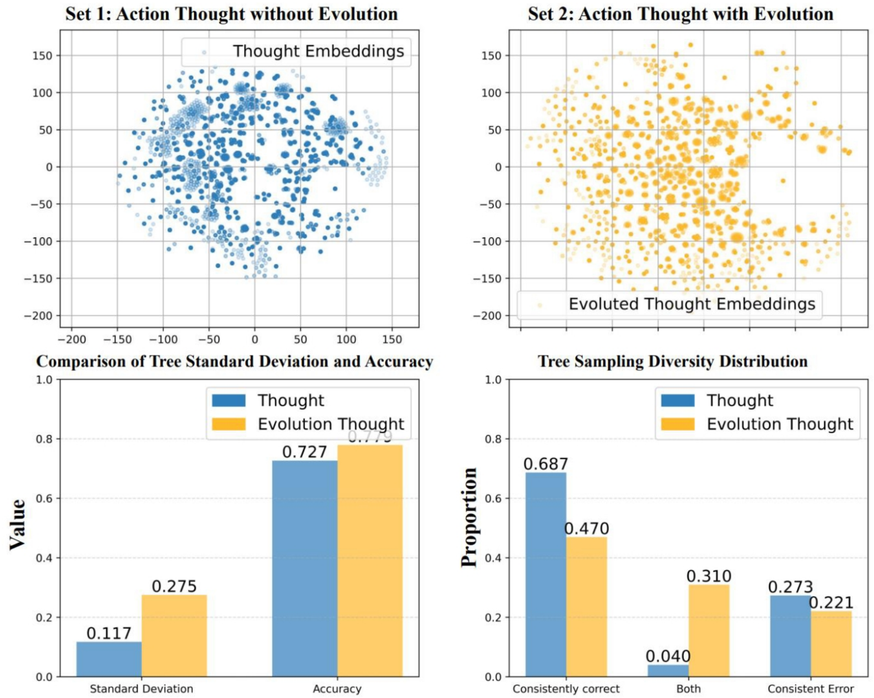
- MobileIPL significantly improves atomic action prediction on AITZ, with clear gains in overall type and match accuracy.
- This shows better fine-grained grounding at the action level.

Model	Training Data	Gmail	Booking	Music	SHEIN	News	CM	ToDo	Signal	Yelp	Overall
SeeClick-7B	AITW+External	28.2	29.4	18.1	20.0	30.0	53.1	30.7	37.1	27.4	30.44
SphAgent-7B	AITW	32.1	45.9	46.1	35.1	48.3	61.1	55.9	43.3	42.9	45.63
SphAgent-7B	AMEX	61.7	68.2	77.7	**72.0**	71.9	64.6	**79.6**	71.3	69.6	70.71
AriaUI-MoE	AMEX	63.1	62.3	68.5	58.9	83.0	54.7	62.5	83.3	66.9	64.10
UGround-7B	AMEX	70.9	68.8	72.7	63.7	77.7	67.7	63.7	80.1	67.6	69.12
SphAgent-7B	AITW + AMEX	62.4	68.1	76.3	71.9	68.6	67.3	77.6	66.0	64.1	69.14
OS-Atlas-7B	AMEX	61.1	**73.5**	77.9	61.6	75.2	66.4	71.0	75.9	72.0	70.33
UI-Tars-7B	AMEX	67.7	70.0	71.8	63.8	71.5	67.7	77.0	**86.4**	72.8	70.33
Qwen2-VL-7B	AMEX	58.0	70.1	76.6	63.8	79.4	66.8	67.8	80.2	**76.6**	69.01
	+ CoaT	75.9	68.1	77.7	66.2	76.8	66.4	77.5	79.6	65.6	70.93
MobileIPL-7B	AMEX (Seed)	57.0	60.2	68.8	63.1	75.0	50.2	65.6	77.7	62.6	62.19
	MobileIPL	**77.3**	71.8	**80.0**	68.4	**85.3**	**71.3**	73.5	82.1	71.8	**74.29**

AMEX

- MobileIPL achieves the best overall result on AMEX, outperforming both the seed model baseline and CoaT-based training.
- This demonstrates stronger generalization across realistic mobile apps and tasks.

Analysis Experiments



Diversity Analysis

- Instruction evolution increases thought diversity and improves tree-level accuracy.
- This helps the model explore more reasonable alternatives during sampling.

Model	Scroll	Click	Type	Press	Total
MobileIPL-R1	45.8	71.1	81.2	23.5	65.4
- IPL	46.9	59.4	78.6	55.4	60.4 (-5.0)
- Evo (R1)	44.8	67.7	78.8	24.0	62.9 (-2.5)
- IPL Negative (R1)	46.9	61.1	74.2	56.6	61.4 (-4.0)
- IPL + Naive DPO (R1)	47.5	59.7	73.8	58.2	60.3 (-5.1)
- 1/2 training data (R1)	42.9	68.3	79.0	43.8	64.8 (-0.6)
- 4/5 training data (R2)	30.8	67.1	77.6	33.2	60.6 (-4.8)

Ablation Study

Model	Accuracy	Rollouts per sentence
MobileIPL (Ours)	69.15	∼27
SPO-Chain	68.03	∼54
GRPO	66.29	8

Compared with other RL method

- Ablation study show that both instruction evolution and step-level IPL are essential, and replacing our design with naive DPO consistently hurts performance.
- Compared with other RL baselines, MobileIPL also achieves the best accuracy with a better trade-off between rollout cost and effectiveness.


Thanks !