

OS-ATLAS: A Foundation Action Model for Generalist GUI Agents



Zhiyong Wu^{1*}, Zhenyu Wu^{1,2*}, Fangzhi Xu^{1*}, Yian Wang^{2*}, Qiushi Sun³, Chengyou Jia¹, Kanzhi Cheng¹, Zichen Ding¹, Liheng Chen³, Paul Pu Liang⁴, Yu Qiao¹

¹Shanghai AI Lab, ²Shanghai Jiao Tong University, ³The University of Hong Kong, ⁴MIT
*Equal contribution



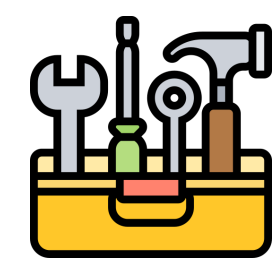
Paper



What We Provide -- Contributions

Foundation Action Model **OS-Atlas** --- “Hands” of AI Agents

- 4B / 7B sizes
- Planner + Action Model (**OS-Atlas**) = Minimal Computer Agent



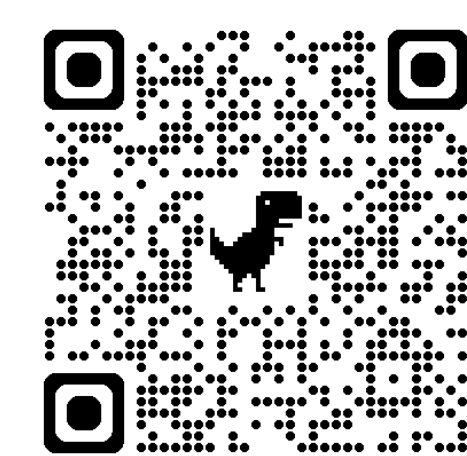
Complete Infrastructure for GUI Data Synthesis

- Across 5 Platforms (Windows / MacOS / Linux / Android / Web)
- **2.3 M** distinct screenshots and more than **13 M** GUI elements



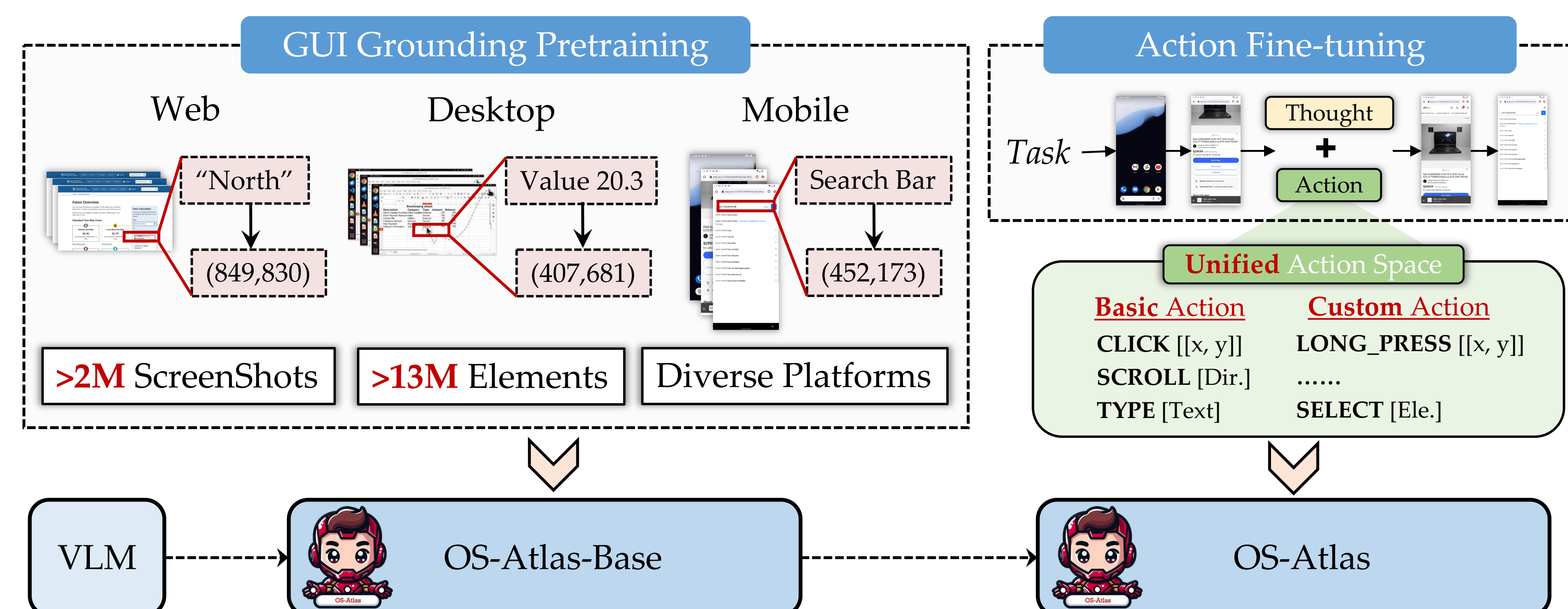
Open-Source Everything !

- Model weights, Pretraining Data, Code, Toolkit ...



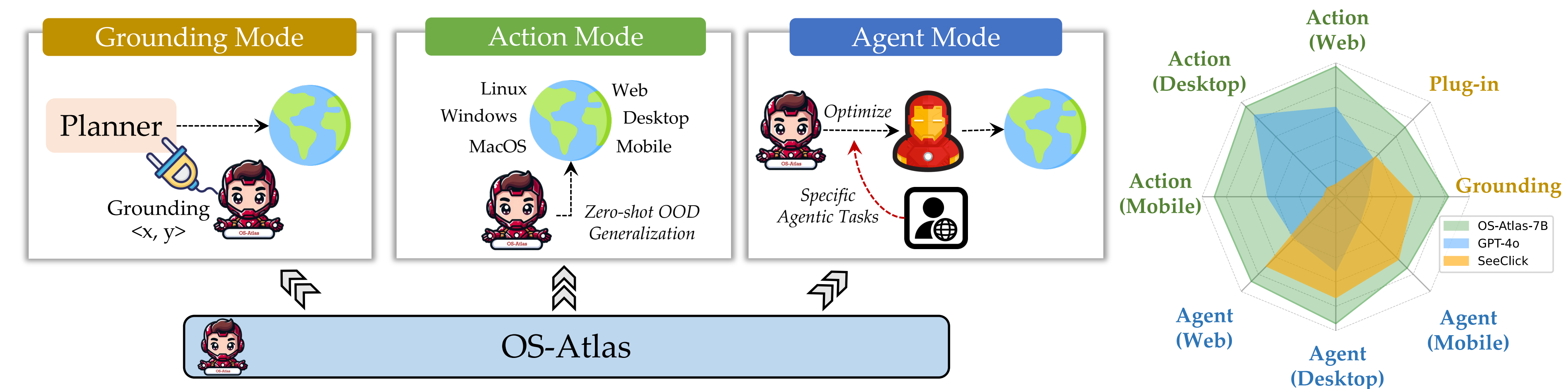
How We Do That -- Approaches

- GUI Grounding Pre-training --- **OS-Atlas-Base**
- Action Finetuning --- **OS-Atlas**



How Does OS-Atlas Perform -- Experiments

Three Critical Capabilities --- (1) GUI Grounding; (2) OOD Action Performance; (3) DIY Agents



- **SOTA** GUI Grounding on ScreenSpot !

Planner	Grounding Models	Mobile		Desktop		Web		Avg.
		Text	Icon/Widget	Text	Icon/Widget	Text	Icon/Widget	
	Fuyu	41.00	1.30	33.00	3.60	33.90	4.40	19.50
	CogAgent	67.00	24.00	74.20	20.00	70.40	28.60	47.40
	SeeClick	78.00	52.00	72.20	30.00	55.70	32.50	53.40
	InternVL-2-4B	9.16	4.80	4.64	4.29	0.87	0.10	4.32
	Qwen2-VL-7B	61.34	39.29	52.01	44.98	33.04	21.84	42.89
	UGround-7B	82.80	60.30	82.50	63.60	80.40	70.40	73.30
	OS-Atlas-Base-4B	85.71	58.52	72.16	45.71	82.61	63.11	70.13
	OS-Atlas-Base-7B	93.04	72.93	91.75	62.86	90.87	74.27	82.47
GPT-4o	SeeClick	83.52	59.39	82.47	35.00	66.96	35.44	62.89
	UGround-7B	93.40	76.90	92.80	67.90	88.70	68.90	81.40
	OS-Atlas-Base-4B	94.14	73.80	77.84	47.14	86.52	65.53	76.81
	OS-Atlas-Base-7B	93.77	79.91	90.21	66.43	92.61	79.13	85.14

- GPT-4o: 5% on OSWorld
- Latest Claude-3.5-Sonnet: 14.9%
- GPT-4o + **OS-Atlas**: **14.6%**

Models	Successful Rate										Avg.
	OS	Calc	Impress	Writer	VLC	TB	Chrome	VSC	GIMP	WF	
GPT-4o + SoM	20.83	0.00	6.77	4.35	6.53	0.00	4.35	4.35	0.00	3.60	4.59
GPT-4o	8.33	0.00	6.77	4.35	16.10	0.00	4.35	4.35	3.85	5.58	5.03
+ SeeClick	16.67	0.00	12.76	4.35	23.52	6.67	10.86	8.70	11.54	7.92	9.21
+ OS-Atlas-Base-4B	20.83	2.23	14.89	8.70	23.52	13.33	15.22	13.04	15.38	7.92	11.65
+ OS-Atlas-Base-7B	25.00	4.26	17.02	8.70	29.41	26.67	19.57	17.39	19.23	8.91	14.63
Human	75.00	61.70	80.85	73.91	70.59	46.67	78.26	73.91	73.08	73.27	72.36

- **SOTA** Performances on 6 Benchmarks (both **Out-of-distribution** and After SFT)

Models	GUI-Act-Web			OmniAct-Web			OmniAct-Desktop		
	Type	Grounding	SR	Type	Grounding	SR	Type	Grounding	SR
Zero-shot OOD Setting									
GPT-4o	77.09	45.02	41.84	79.33	42.79	34.06	79.97	63.25	50.67
OS-Atlas-4B	79.22	58.57	42.62	46.74	49.24	22.99	63.30	42.55	26.94
OS-Atlas-7B	86.95	75.61	57.02	85.63	69.35	59.15	90.24	62.87	56.73
Supervised Fine-tuning Setting									
InternVL-2-4B	81.42	47.03	36.17	47.51	51.34	24.39	67.00	44.47	29.80
Qwen2-VL-7B	89.36	90.66	82.27	89.22	85.94	78.58	96.27	94.52	91.77
SeeClick	88.79	78.59	72.34	86.98	75.48	68.59	96.79	70.22	72.69
OS-Atlas-4B	89.36	89.16	81.06	88.56	82.00	73.91	96.51	85.53	84.78
OS-Atlas-7B	89.08	91.60	82.70	97.15	95.41	93.56	97.15	95.85	94.05

Models	AndroidControl-Low			AndroidControl-High			GUI-Odyssey		
	Type	Grounding	SR	Type	Grounding	SR	Type	Grounding	SR
Zero-shot OOD Setting									
GPT-4o	74.33	38.67	28.39	63.06	30.90	21.17	37.50	14.17	5.36
OS-Atlas-4B	64.58	71.19	40.62	49.01	49.51	22.77	49.63	34.63	20.25
OS-Atlas-7B	73.00	73.37	50.94	57.44	54.90	29.83	60.42	39.74	26.96
Supervised Fine-tuning Setting									
InternVL-2-4B	90.94	84.05	80.10	84.09	72.73	66.72	82.13	55.53	51.45
Qwen2-VL-7B	91.94	86.50	82.56	83.83	77.68	69.72	83.54	65.89	60.23
SeeClick	93.00	73.42	75.00	82.94	62.87	59.11	70.99	52.44	53.92
OS-Atlas-4B	91.92	83.76	80.64	84.69	73.79	67.54	83.47	61.37	56.39
OS-Atlas-7B	93.61	87.97	85.22	85.22	78.48	71.17	84.47	67.80	61.98

Web & Desktop

Mobile

OS-Atlas --- ICLR 2025 (Spotlight)