Model	Mind2Web	MiniWoB++	WebArena
	Single step Reference-based	Short Execution-based	Long Execution-based
API-based Models			
`GPT-3.5`	12.79	39.57	6.16
`GPT-4`	29.09	53.04	14.41
Open-source Instructed Models			
`CodeLlama-instruct-7b`	6.62	23.04	0.00
`Llama3-chat-8b`	11.50	31.74	3.32
`Llama3-chat-70b`	22.27	48.70	7.02
Open-source Interactive Data Finetuned Models			
`FireAct-7b` [3]	-	-	0.25*
`AgentLM-7b` [51]	2.99	15.65	0.86
`CodeActAgent-7b` [41]	3.13	9.78	2.34
`AutoWebGLM-7b(S1)` [19]	-	-	2.50*
`AgentFlan-7b` [5]	3.80	20.87	0.62
`Lemur-chat-70b` [46]	14.28	21.30	2.95
`AgentLM-70b` [51]	10.61	36.52	3.07
`Synatra-CodeLlama-7b`	15.85	38.20	6.28