

SWE ATLAS - CODEBASE QNA

Evaluating deep code comprehension and reasoning

#	MODEL	PROVIDER	SCORE
1	GPT 5.5 (Codex)	openai	45.43
2	Gpt 5.4 xHigh (Codex)	openai	40.8
2	Gpt 5.4 xHigh (Mini-SWE-Agent)	openai	36.3
2	Opus 4.6 (Claude Code)	anthropic	33.3
2	Gpt 5.3 (Codex)	openai	32.6
2	Sonnet 4.6 (Claude Code)	anthropic	31.2
7	Opus 4.6 (Mini-SWE-Agent)	anthropic	30
8	Muse Spark	meta	24.2
9	Glm 5 (Mini-SWE-Agent)	zai	20.5
10	Gemini 3.1 Pro (Mini-SWE-Agent)	google	13.5
10	Kimi K2.5 (Mini-SWE-Agent)	moonshot	13.1
10	Minimax M2.5 (Mini-SWE-Agent)	minimax	10.3
13	Gemini 3 Flash (Mini-SWE-Agent)	google	8.2