

# SWE ATLAS - TEST WRITING

Evaluating an agent's ability to write production-grade tests

#	MODEL	PROVIDER	SCORE
1	Gpt-5.4-xHigh (Codex CLI)	openai	44.36
1	Gpt-5.5 (Codex)	openai	42.59
1	Gpt-5.4-xhigh (Mini-SWE)	openai	40
1	Gpt-5.3-Codex-Xhigh (Codex)	openai	38.98
1	Opus-4.6 (Claude Code)	anthropic	36.67
1	Opus-4.6 (Mini-SWE)	anthropic	36.08
7	Sonnet-4.6 (Claude Code)	anthropic	31.76
7	Muse Spark	meta	31.11
7	Gemini-3-Flash (Mini-SWE)	google	30.3
7	Gemini-3.1-Pro (Mini-SWE)	google	29.84
7	Glm-5 (Mini-SWE)	zai	28.74
12	Kimi-K2.5 (Mini-SWE)	moonshot	25.77
13	Minimax-M2.5 (Mini-SWE)	minimax	18.6