MO§ES™ × Artificial Analysis Coding Agent Benchmarks

Operator-augmented **Claude Code + Opus 4.7** vs. 13 published combinations · Field: **artificialanalysis.ai/agents/coding-agents** · 2026-05-14

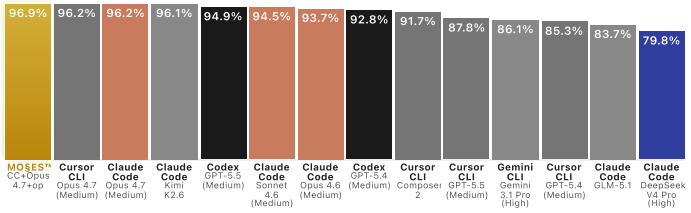
MO§ES™ leads all 5 measured economic categories — 7-day window (2026-05-08 → 2026-05-14)

CACHE HIT RATE	OUTPUT : INPUT	TOKENS / TASK	TIME / TASK	COST / LOC
96.88%	**31.7×**	**767K**	**1.84 min**	**$0.0007**
#1 — > field's 96.2%	*30d 42.5× · 90d 22.1×*	#1 — 3.6× more efficient	#1 — 3.2× faster	*plan · API: $0.044 · ccusage: $0.018*
SRC: Token Dashboard 7d	#1 — 83× field leader (0.38)	SRC: 1.12B ÷ 1,465 tasks	SRC: ~45 hr ÷ 1,465 tasks	#1 — < 1¢ per line
	SRC: 3.90M out ÷ 123K in			SRC: $23.33 ÷ 35,242 LOC

Testing methodology — not the same thing being measured: AA field: per-task isolated runs on SWE-Bench-Pro-Hard-AA, Terminal-Bench v2, SWE-Atlas-QnA. Each task = one bug/issue. **MO§ES™:** sustained operator work on App Hub (real product) over 7-day measurement window (2026-05-08 → 2026-05-14), 21 sessions, 7,327 turns. App Hub = 5 build days within the window (5/10–5/14), 35,242 LOC shipped. Cost/task ($0.017 sub · $1.05 API) computed similarly but not on cards. $/LOC: AA = cost_per_task ÷ 20 LOC industry convention. MO§ES = $23.33/wk sub ÷ 35,242 actual LOC.

1. Cache Hit Rate

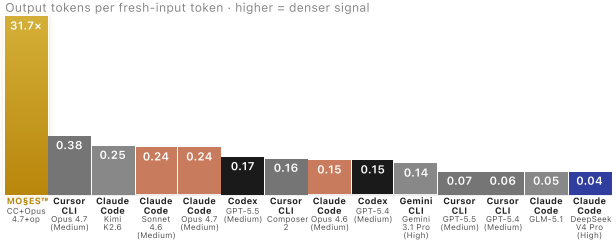
Cache reuse % · higher better · MO§ES sustained multi-project



SRC: cache_read 1.084B ÷ (1.084B + 34.83M cache_create + 123K input) = 96.88%

2. Output : Fresh Input (log scale)

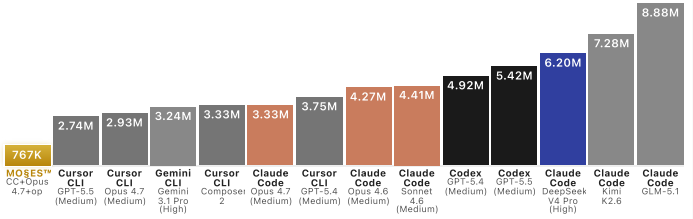
Output tokens per fresh-input token · higher = denser signal



SRC: 3,902,803 output ÷ 123,246 fresh input = 31.7× · 30d 42.5× · 90d/all-time 22.1×

3. Tokens per Task

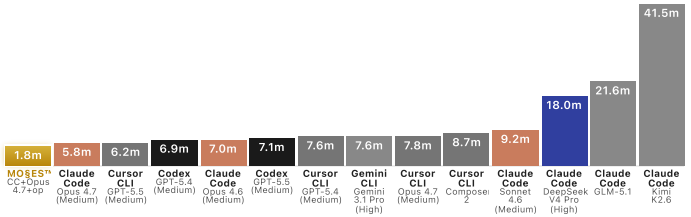
Total tokens · lower better



SRC: 1.123B 7d total ÷ 1,465 task-equivs (7,327 turns ÷ 5) = 767K

4. Time per Task

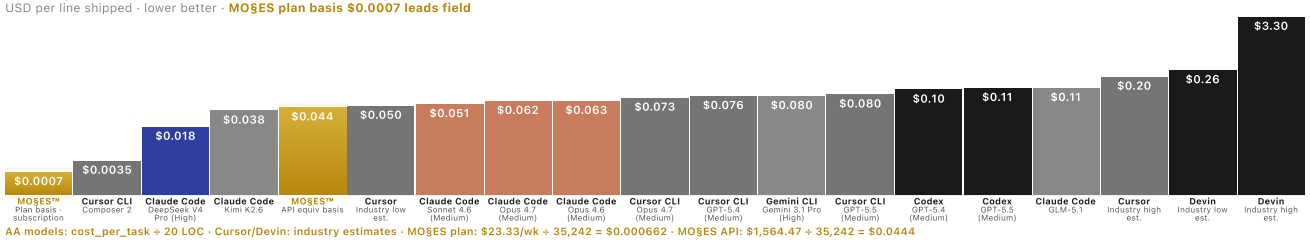
Wall time · lower better



SRC: ~45 hr active ÷ 1,465 = 1.84 min

5. Cost per LOC — all field models + MO§ES™ (log scale)

USD per line shipped · lower better · **MO§ES plan basis $0.0007 leads field**



AA models: cost_per_task ÷ 20 LOC · Cursor/Devin: industry estimates · MO§ES plan: $23.33/wk ÷ 35,242 = $0.000662 · MO§ES API: $1,564.47 ÷ 35,242 = $0.0444

RAW DATA — 7-DAY MEASUREMENT WINDOW (2026-05-08 → 2026-05-14)

INPUT	OUTPUT	CACHE CREATE	CACHE READ	TOTAL TOKENS
123,246	**3.90 M**	**34.83 M**	**1.084 B**	**1.123 B**
FRESH TOKENS IN	3,902,803	34,826,779	1,084,399,183	1,123,252,011
drives **cache hit** · **out:in**	drives **out:in**	drives **cache hit**	drives **cache hit**	drives **tokens/task** · **time/task**

ancillary inputs › **21** sessions · **7,327** turns · **1,465** tasks · **~45 hr** active · **35,242** LOC · **$23.33/wk** plan — **$1,564.47** API equiv

MO§ES™ = operator-augmented Claude Code, concurrent sessions, $100/mo Max plan · 7d window (2026-05-08 → 2026-05-14): 21 sessions · 7,327 turns · 1.12B tokens · $1,564.47 API-equivalent · App Hub: 35,242 LOC in 5 days · Sources: Token Dashboard API, ccusage report, App Hub wc -l count, artificialanalysis.ai/agents/coding-agents