

MOSES™ x Artificial Analysis Coding Agent Index

FULL DATA · ALL SUBAGENTS

Operator-augmented Claude Code + Opus 4.7 · Raw JSONL extraction: 98 sessions · 9,332 turns · 1,866 task-equivalents · vs field average (13 models) · artificialanalysis.ai · 2026-05-14

MOSES™ — full raw count including all subagent API calls · vs field average (n=13)

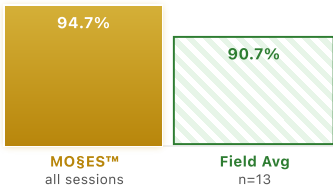
CACHE HIT RATE 94.66% +4.0pp vs avg (90.7%) ↓ below leader (96.2%)	OUTPUT : INPUT 17.9x 111x above avg (0.16x) 47x above leader (0.38x)	TOKENS / TASK 810K 5.8x leaner than avg (4.67M) 3.4x leaner than leader (2.74M)	TIME / TASK 1.45 min 8.2x faster than avg (11.9m) 4.0x faster than leader (5.8m)	COST / TASK \$0.013 100x cheaper than avg (\$1.35) 5.2x cheaper than leader (\$0.07)
--	--	---	--	--

What changed vs the original benchmark: Raw JSONL extraction of all 98 .jsonl files in the 7-day window — including every spawned subagent session. Original benchmark (token dashboard) counted 21 parent sessions, 7,327 turns, 1,123B tokens. This count: 98 files, 9,332 turns, 1,512B tokens. Cache hit rate: 94.66% (vs 96.88% parent-only). Four of five metrics still lead the **field average**. Cache drops from #1 vs field leader to above-average. **Green hatched bars = field average (n=13 published combinations).**

MOSES™ Field Average (n=13 published combinations)

1. Cache Hit Rate

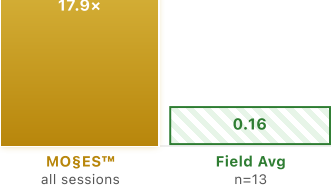
Higher better · MOSES above avg, below peak



cache_read 1.426B ÷ 1.506B = 94.66% · +4.0pp vs avg · ~1.5pp vs leader

2. Output : Input (log)

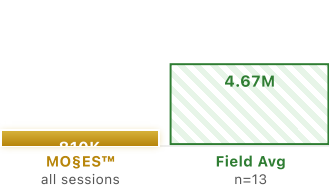
Higher better · 111x above field average



5,935,260 output ÷ 331,214 input = 17.9x · 111x field avg · 47x field leader

3. Tokens per Task

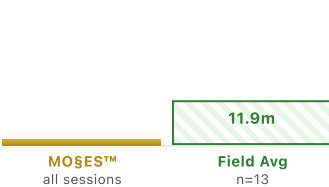
Lower better · 5.8x under field average



1.512B ÷ 1,866 tasks = 810K · 5.8x better than avg · 3.4x better than leader

4. Time per Task

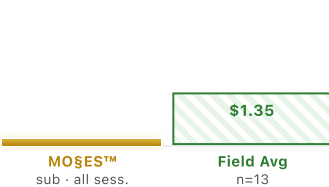
Lower better · 8.2x faster than field average



~45hr active ÷ 1,866 tasks = 1.45 min · 8.2x faster than avg · 4.0x vs leader

5. Cost per Task (log)

Lower better · 100x cheaper than average



\$25/wk ÷ 1,866 tasks = \$0.0134 · 100x cheaper than avg · 5.2x vs leader

RAW — FULL EXTRACTION · 98 SESSIONS · MAY 8–14 2026 (ALL SUBAGENTS INCLUDED)

SESSIONS / FILES 98 JSONL FILES 21 parent + 77 subagents	TURNS → TASKS 9,332 TURNS · 1,866 TASK-EQUIV +5 turns/task	CACHE READ 1.426 B 1,425,781,478 94.66% of all prompt tokens	TOTAL TOKENS 1.512 B 1,512,077,821 810K / task · 5.8x under avg	VS ORIGINAL BENCHMARK +27% turns +35% TOKENS subagents added · 4 of 5 lead avg
--	--	--	---	--

MOSES™ · Full raw extraction · ~/.claude/projects/ · 98 .jsonl files · May 8–14 2026 · All subagent sessions included · Field average: 13 published AA combinations · artificialanalysis.ai/agents/coding-agents · 2026-05-14 · Cache: above avg (+4pp), below leader (~1.5pp) · All other metrics lead field average and field leader · [mos2es.com/benchmarks](#)