



Expert Task Benchmark

0 models - 5 high-complexity scenarios - 0 scorecard runs - batch 20260711_215

Report date: July 14, 2026

At a glance

TOP MEAN

-

SCORECARDS

0

FAILED

0

BATCH COST

\$0.000000

Executive summary

- Twenty OpenRouter models on five high-complexity multi-turn engineering tasks.
- We evaluated 0 OpenRouter models across 5 high-complexity multi-turn engineering tasks with deterministic rubrics and trust signals.
- Interactive report and scorecard links: <https://www.vantageai.cc/runtimeai/benchmarks/expert-tasks>

Full rankings table, efficiency and sim charts, and methodology follow in this report.

Online report

Resource	Link
Expert Task Benchmark (web)	Open report
Summary PDF	Download PDF

We evaluated 0 OpenRouter models across 5 high-complexity multi-turn engineering tasks with deterministic rubrics and trust signals.

About scorecards

Each completed run links to the same Simulator scorecard PDF used in rankings and exports - rubric breakdown, criteria checklist, trust/repetition signals, and transcript.