

Codex and Claude Sonnet CLIs — text and structured, ten attempts each

PUBLIC RESULT · CAMPAIGN SUMMARY

40 passes

Across 40 graded attempts; no cross-exam quality score

COVERAGE

2 exams · 2 lanes

COMPLETED ATTEMPTS

40

MODEL FAILURES

0

LANE ERRORS

0

WITHHELD / INVALID

0 / 0

Results by AI exam

Exam	Model	Passed / attempts	Errors
Record normalization and duplicate QA	cli/claude-sonnet	10/10	0
Intent and entity annotation	cli/claude-sonnet	10/10	0
Record normalization and duplicate QA	cli/codex	10/10	0
Intent and entity annotation	cli/codex	10/10	0

Showing all 4 model/exam results. Technical errors are counted in attempts but are not model failures.

2 verified records. Compare models within the same exam and input condition.

What was evaluated

INPUT CONDITION
Conditions differ by exam and model lane; the detailed matrix retains each cell without merging unlike tasks.
EXAM AND REVISION
2 exam revisions; each fingerprint is retained in the detailed report

Limits on the claim

- This summary is accounting, not a leaderboard or a broad capability score.
- Low sample cells support counts only; no ranking is inferred.
- Lane errors, withheld and invalid attempts are retained outside the model-failure denominator.
- Per-exam conditions and quality axes remain in the detailed report.

Evidence

CAMPAIGN	CANONICAL DIGEST
fc-6d6679dea47a	2 sealed record digests in the detailed report

Open the full evidence matrix and audit report →

<https://vvdexops.com/reports/fc-6d6679dea47a/>

Public disclosure excludes submissions, raw answers, private references, exact geometry, frame timestamps, and hidden-test material.