

Five exams, five families — ten rollouts per lane

PUBLIC RESULT · CAMPAIGN SUMMARY

93 passes

Across 148 graded attempts; no cross-exam quality score

COVERAGE

5 exams · 3 lanes

COMPLETED ATTEMPTS

150

MODEL FAILURES

55

LANE ERRORS

2

WITHHELD / INVALID

0 / 0

Results by AI exam

Exam	Model	Passed / attempts	Errors
Streaming items must equal what the full parse would have built	cli/codex	10/10	0
Answering questions from a document collection with outdated and distractor sources, and refusing when the evidence is insufficient	cli/codex	10/10	0
Updating a cross-session fact: answer from the newer authoritative roster, not stale carried memory	cli/codex	10/10	0
Harness recovery under a faulted tool loop	cli/codex	10/10	0
Correct one order in an admin application, through a real browser	cli/codex	10/10	0
Streaming items must equal what the full parse would have built	cli/claude-sonnet	9/10	0
Answering questions from a document collection with outdated and distractor sources, and refusing when the evidence is insufficient	cli/claude-sonnet	1/10	0
Updating a cross-session fact: answer from the newer authoritative roster, not stale carried memory	cli/claude-sonnet	10/10	0

First 8 of 15 model/exam results in report order. The complete table is linked on page 2. Technical errors are counted in attempts but are not model failures.

5 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
5 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-d89e429d2781	5 sealed record digests in the detailed report

Open the full evidence matrix and audit report →

<https://vvdexops.com/reports/fc-d89e429d2781/>

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