GPT-5.5 annotation evaluation — ten attempts on each of four tasks

PUBLIC RESULT · CAMPAIGN SUMMARY

25 passes

Across 40 graded attempts; no cross-exam quality score

COVERAGE

4 exams · 1 lanes

COMPLETED ATTEMPTS	MODEL FAILURES	LANE ERRORS	WITHHELD / INVALID
40	15	0	0 / 0

Results by AI exam

Exam	Model	Passed / attempts	Errors
Object geometry annotation	`cli/codex-gpt-5.5`	5/10	0
Robot catch temporal and tracking annotation	`cli/codex-gpt-5.5`	0/10	0
Record normalization and duplicate QA	`cli/codex-gpt-5.5`	10/10	0
Intent and entity annotation	`cli/codex-gpt-5.5`	10/10	0

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

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

CONDITIONS · LIMITS · EVIDENCE

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

4 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

fc-ec21239c00aa

CANONICAL DIGEST

40 sealed record digests in the detailed report

Open the full evidence matrix and audit report →

https://vvdexops.com/reports/fc-ec21239c00aa/

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