

Gemini 3.6 Flash native video with exact frames — 10 FPS requested

PUBLIC RESULT · FIXED EVALUATION

65.65/100

Measured quality, separate from the pass gate

OUTCOME

Did not pass

COMPLETED

1 recorded attempt

ANNOTATIONS · VALID / INVALID

6 / 0

MODEL

gemini-3.6-flash

Box matching F1: 0.50. Required: 1.00. The attempt did not meet the exam pass criteria.

Recorded quality axes and margins

MEASURE	VALUE	MINIMUM · MARGIN
Box F1	0.5	pass gate ≥ 1 · -0.50
Box IoU	0.61	item match ≥ 0.5 · mean +0.11
Classification accuracy	1	pass gate ≥ 1 · met (gate is the maximum)
Event temporal IoU	0.5	item match ≥ 0.5 · mean on the minimum
Segment temporal IoU	0.71	item match ≥ 0.5 · mean +0.21
Track consistency on matched boxes	1	pass gate ≥ 1 · met (gate is the maximum)

Every matched event sits exactly on the item-match minimum (0.5). That part of the pass carries no margin.

A pass gate is the value an axis must reach for the exam to pass. An item-match minimum is the overlap one item needs to count as matched; the margin shown for a mean overlap is the mean's distance from that minimum, not any single item's.

Track consistency is conditional on successful box matching. It does not repair a missed localization.

What was evaluated

INPUT CONDITION
native video with supplied exact-frame images; provider-side processing is not independently observed.
EXAM AND REVISION
Robot catch temporal and tracking annotation · v0.2.0

Limits on the claim

- n=1: this result does not establish general model capacity or a stable success rate.
- Provider-side video or frame processing was not independently observed.
- This record does not include independent human validation of the reference answers.
- Quality is a grader measurement; the pass decision is a separate threshold verdict.

Evidence

CAMPAIGN	CANONICAL DIGEST
fc-403cfc23f9e5	f9e6baef1c1acb4e45e2bd7e027b368b5e594f33037c643dc8ee9edb75daa2e9

Open the full evidence matrix and audit report →

<https://vvdexops.com/reports/fc-403cfc23f9e5/>

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