

# Gemini 3.6 Flash native-video exam — revision 0.2.0

PUBLIC RESULT · FIXED EVALUATION

42.11/100

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: 0.00. Required: 1.00. The attempt did not meet the exam pass criteria.

## Recorded quality axes and margins

| MEASURE                            | VALUE | MINIMUM · MARGIN                          |
|------------------------------------|-------|-------------------------------------------|
| Box F1                             | 0     | pass gate ≥ 1 · -1.00                     |
| Box IoU                            | 0     | item match ≥ 0.5 · mean -0.50             |
| Classification accuracy            | 1     | pass gate ≥ 1 · met (gate is the maximum) |
| Event temporal IoU                 | 0.31  | item match ≥ 0.5 · mean -0.19             |
| Segment temporal IoU               | 0.9   | item match ≥ 0.5 · mean +0.40             |
| Track consistency on matched boxes | 0     | pass gate ≥ 1 · -1.00                     |

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 file; provider-side sampling 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-2917108cf6be | 59f431f7ed9ec5214737424a5c886cbe65f3caf49b1a11807bc27be17428b613 |

Open the full evidence matrix and audit report →

<https://vvdexops.com/reports/fc-2917108cf6be/>

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