

ALI: Action-Language Inference

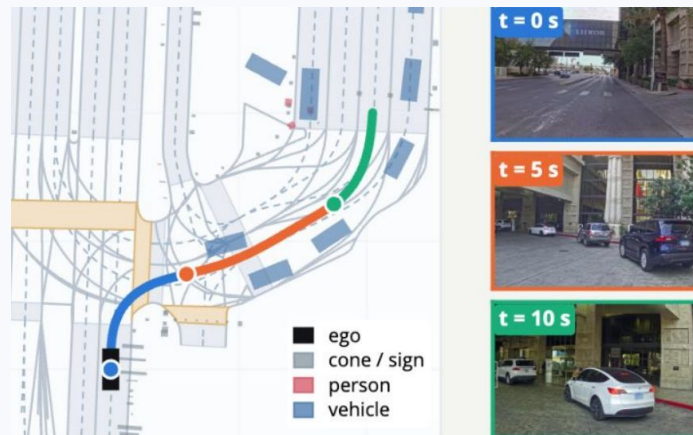
Reasoning-action alignment in autonomous driving Vision-Language-Action models

Oren Sultan

Google Research for Waymo, Perception Team

WITH YOTAM INTRATOR, IDAN KLIGVASSER,
REGEV COHEN, EHUD RIVLIN

Google Research



0-5 s Gently accelerate · Turn right

5-10 s Slow down gently · Left lane change

10-15 s Slow down gently · Turn left

Google Research for Waymo, Perception

Te



Oren Sultan



Yotam Intrator



Moshe Kimhi



Idan Kligvasser



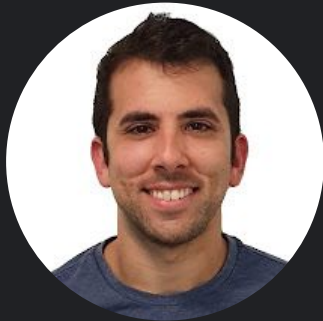
Regev Cohen



Gali Bar-On



Guy Ben-Yosef



Ori Linial



Ariav Kenig

THE PROBLEM

Descriptive drift: the actions does not follow the reasoning

VLA's reason, then act: with *embodied reasoning*, a natural-language trace precedes the trajectory.

THE VLA SAYS

*“Decelerate and move left
to maintain a safe distance.”*

THE VLA DOES

Accelerates and stays in lane.

WHY REASONING MATTERS

- 1. Planning:** reasoning improves VLA plans.
- 2. Interpretability:** a readable account of the plan, valid only if the action follows it.

HOW COMMON IT IS

50%+

of long-tail scenarios.
Semantic Reasoning
Coherence Check (SRAC),
Wagner et al., 2026

38%

of cases: Alpamayo-R1 fails to
stop when its reasoning says it
should (Mayumu et al., 2026)

PROBLEM FORMULATION

ALI: does the trajectory **entail** its description?

Premise P: the ego trajectory

$W = ((x_1, y_1), \dots, (x_n, y_n))$

$T = S \cdot h$ waypoints (S seconds at h Hz), in meters

x: longitudinal (accelerate, decelerate)

y: lateral (turn, lane change)

Hypothesis H: the stated actions

$H = (H_1, \dots, H_n)$

Action phrases extracted from the model's reasoning

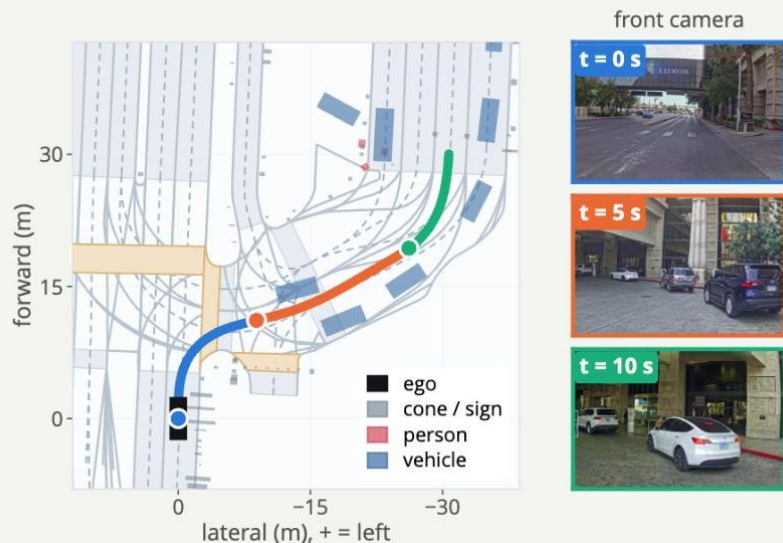
ALI
classifier

$s = p(\text{Entailment} \mid P, H)$

A score in $[0, 1]$. Thresholded:
Entailment or Contradiction.

Built on nuReasoning: real drives with human-verified

(a) **Driving context** *not given to the ALI classifier*



HUMAN-VERIFIED REASONING · EVERY 5 S

0–5 s *Gently accelerate* to execute the **right turn** into the driveway since the path and crosswalk are clear.

5–10 s *Slow down gently* and execute a **left lane change** to bypass the stationary SUV blocking the ego lane, as the right lane is also occupied by a parked car.

10–15 s *Slow down gently* and **turn left** to follow the routing intent and safely navigate around the stationary vehicle ahead in the pick-up zone.

X Rejected: Maintain speed (unsafe)

Driving context

Bird's-eye map with agents, the ego trajectory and the front camera.

Decisions every 5 s

One longitudinal and one lateral action, chosen from a closed vocabulary.

Rejected alternatives

What the annotator ruled out for the same scene, and why.

The ALI benchmark: negatives change **only the text**

The trajectory never changes.

Only the text does.

No cameras, no map.

(b) Given to the ALI classifier

Premise P 149 waypoints at 10 Hz, (forward, lateral) in metres

(0.25, +0.00), (0.49, +0.00), ... (6.72, -1.07), ... (11.15, -8.88), ... (14.76, -18.52), ...
(19.34, -26.18), ... (25.23, -30.06), ... (30.03, -30.72)

Hypothesis H one longitudinal (lon) and one lateral (lat) action per 5 s decision

+ POSITIVE

the annotated decisions

- lon Gently accelerate
lat Turn right
- lon Slow down gently
lat Left lane change
- lon Slow down gently
lat Turn left

→ Entailment

- COUNTERFACTUAL

one action replaced

- lon Gently accelerate
lat Turn right
- lon Slow down gently
lat Left lane change
- lon **Maintain speed**
lat Turn left

longitudinal medium change
safety-critical Unsafe

→ Contradiction

- WRONG ORDER

decisions 2 and 3 swapped

- lon Gently accelerate
lat Turn right
- lon Slow down gently
lat **Turn left**
- lon Slow down gently
lat **Left lane change**

→ Contradiction

5, 10 and 15 s horizons · **59K** positives · **216K** negatives · **no drive shared** between train and test

The ALI classifier: fine-tuning reaches **0.94 ROC AUC**

Zero-shot LLMs $\leq$ **0.79** · ten-label coherence check **0.68** · kinematic MLP **0.89** · fine-tuned Qwen3.5-4B **0.94** · ROC AUC on 22,705 test pairs

Method	Input	All	CF	Order	k=1	k=2	k=3
<i>Zero-shot LLMs</i>							
Gemini 3.7 Flash	Plain	0.78	0.77	0.86	0.81	0.75	0.73
	Kin.	0.78	0.77	0.85	0.81	0.75	0.73
GPT-5.6 terra	Plain	0.78	0.77	0.87	0.80	0.72	0.75
	Kin.	0.79	0.78	0.89	0.79	0.76	0.72
Qwen3.5-4B	Plain	0.65	0.63	0.77	0.59	0.56	0.55
	Kin.	0.72	0.71	0.79	0.71	0.69	0.62
<i>Kinematic classifiers, no language model, fitted on the training set</i>							
Per-claim logistic	–	0.82	0.81	0.91	0.84	0.80	0.76
Per-claim MLP	–	0.89	0.88	0.94	0.89	0.88	0.86
<i>Discretized coherence check, ten shared labels (Wagner et al., 2026), adapted to our vocabulary and refitted to our data</i>							
SRAC	–	0.68	0.68	0.70	0.73	0.64	0.59
<i>Fine-tuned on the training set (LoRA, 3 epochs)</i>							
Qwen3.5-4B	Plain	0.93	0.93	0.99	0.94	0.92	0.91
	Kin.	0.94	0.93	0.99	0.94	0.93	0.92
Floor: no trajectory	Hyp. only	0.72	0.71	0.79	0.67	0.67	0.67

Plain: raw waypoints at 10 Hz.

Kin.: plus per-decision speed, lateral displacement and heading

CF: counterfactual negatives.

Order: wrong-order negatives.

k: decisions per hypothesis.

Diagnosing a driving VLM: a correct decision, **never driven**

1 INPUT



route command **go straight**

current speed **7.0 m/s**

ego motion of **the last 3 s**

2 DRIVING VLM

Qwen3.5-9B, zero-shot

reasoning

*"The car is approaching a junction with traffic lights ahead showing red, and there is a large tanker truck in the right lane and other vehicles ahead. The car must **Slow down gently** to prepare for the stop at the red light, while maintaining **No lateral action** to stay centered in its lane."*

10 waypoints at 2 Hz

holds 7.1 m/s

(3.55, +0.00), (7.10, +0.00), ... (35.50, +0.00)

Diagnosing a driving VLM: a correct decision, **never driven**

1 INPUT



route command **go straight**
current speed **7.0 m/s**
ego motion of **the last 3 s**

2 DRIVING VLM

Qwen3.5-9B, zero-shot

reasoning

*"The car is approaching a junction with traffic lights ahead showing red, and there is a large tanker truck in the right lane and other vehicles ahead. The car must **Slow down gently** to prepare for the stop at the red light, while maintaining **No lateral action** to stay centered in its lane."*

10 waypoints at 2 Hz **holds 7.1 m/s**
(3.55, +0.00), (7.10, +0.00), ... (35.50, +0.00)

3 ALI CLASSIFIER

accepted if $s > 0.5$

stated decision

extracted by Qwen3.8-27B

Slow down gently
No lateral action

human annotator, for reference

"Slow down gently with no lateral action to safely approach the stationary vehicle and red light ahead."

FAITHFULNESS

s = 0.01 ✗ **unfaithful**

does it drive what it says?

premise **the VLM's waypoints**



DECISION CORRECTNESS

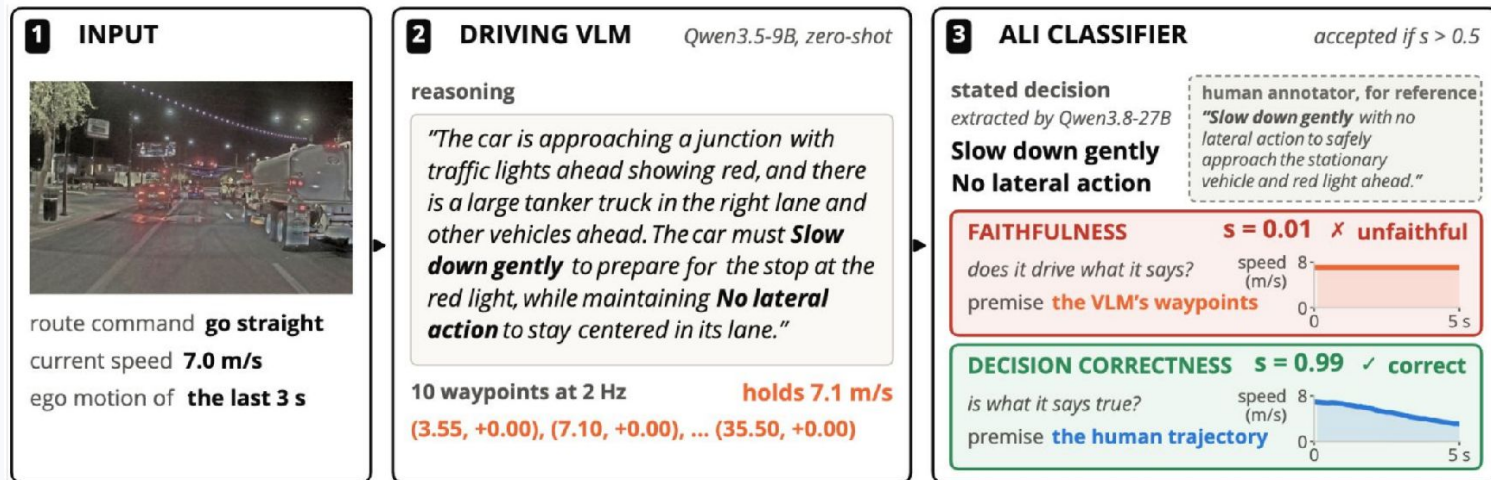
s = 0.99 ✓ **correct**

is what it says true?

premise **the human trajectory**



Diagnosing a driving VLM: a correct decision, **never driven**



Faithfulness (F+)

Premise the model's trajectory

↓ ALI entails?

Hypothesis the model's decision

Does it drive what it says?

Decision correctness (R+)

Premise the human's trajectory

↓ ALI entails?

Hypothesis the model's decision

Is what it says right?

Trajectory correctness (W+)

Premise the model's trajectory

↓ ALI entails?

Hypothesis the human's decision

Does it drive what the human decided?

Fine-tuned VLAs: faithful, correct, better planners

2,396 TEST FRAMES · FINE-TUNED ON 29,818 ALIGNED REASONING-TRAJECTORY PAIRS

Model	Stated (%)			ALI classifier (%)				NPS ↑	ADE ↓
	Maint.	No lat.	Static	F+ ↑	R+ ↑	R+F+ ↑	W+ ↑		
Zero-shot									
Qwen3.5-4B	88	76	66	76	41	38	45	0.50	2.82
Qwen3.5-9B	63	77	53	67	45	36	45	0.51	2.95
Gemma-4-12B	47	86	47	89	46	42	43	0.47	3.22
Gemini 3.7 Flash	36	86	34	97	64	63	53	0.49	2.93
VLAs fine-tuned on the benchmark's positives									
Qwen3.5-4B	35	75	32	91	76	73	69	0.65	1.54
Qwen3.5-9B	32	73	29	93	77	74	70	0.66	1.52
Gemma-4-12B	37	77	35	94	76	74	66	0.59	1.62

$$\text{ADE} = \frac{1}{T} \sum_{t=1}^T \|\hat{\mathbf{p}}_t - \mathbf{p}_t\|_2$$

Mean waypoint error

$$\text{NPS} = \text{NC} \cdot \text{DA} \cdot (0.3 \text{ EP} + 0.2 \text{ CF} + 0.5 \text{ HL})$$

Gates NC no at-fault collision {0, 0.5, 1} · DA drivable area {0, 1}

Weighted EP ego progress [0, 1] · CF comfort {0, 1} · HL human-likeness [0, 1]

Faithfulness helps planning **only when the decision is correct**

ZERO-SHOT PREDICTIONS · ACTIVE MANEUVERS · 2,396 TEST FRAMES

Model	Stated decision	n		NPS ↑		ADE (m) ↓	
		Faithful	Unfaithful	Faithful	Unfaithful	Faithful	Unfaithful
Qwen3.5-9B	correct	245	223	0.79	0.42	0.9	3.4
	incorrect	114	542	0.39	0.46	4.6	3.6
Gemma-4-12B	correct	471	90	0.67	0.22	1.6	3.8
	incorrect	542	171	0.34	0.22	4.7	5.1
Alpamayo 1.5	correct	491	147	0.78	0.48	1.1	3.2
	incorrect	178	767	0.49	0.53	3.1	2.6

Correct decision: driving it pays.

Faithful plans score far higher:

NPS 0.79 vs. 0.42, 0.67 vs. 0.22, 0.78 vs. 0.48.

Wrong decision: faithfulness does not help, and can hurt.

Qwen and Alpamayo plan better when they do *not* drive the wrong decision. Faithfulness by itself is not sufficient.

ALI as a training reward: better aligned, planning intact

FINE-TUNED QWEN3.5-9B · GRPO · 8 SAMPLES PER SCENE · 2,560 SCENES

Reward $\sqrt[3]{(\mathbf{F} \cdot \mathbf{R} \cdot \mathbf{W})}$: the geometric mean of the three ALI scores, zero if any one is zero.

Reward	ALI classifier (%)					NPS ↑	ADE ↓
	Static	F+ ↑	R+ ↑	R+F+ ↑	W+ ↑		
Fine-tuned Qwen3.5-9B (SFT), no reward	29	93	77	74	70	0.65	1.52
+ GRPO, reward $\sqrt[3]{(\mathbf{F} \cdot \mathbf{R} \cdot \mathbf{W})}$	26	97	79	78	72	0.65	1.52
+ GRPO, reward $\frac{1}{2} \sqrt[3]{(\mathbf{F} \cdot \mathbf{R} \cdot \mathbf{W})} + \frac{1}{2} \text{NPS}$	32	96	79	77	70	0.67	1.49

Alignment reward: all three rates rise.

Correct and faithful 74% → 78%,
with NPS and ADE unchanged.

Add the planning score: planning improves too.

NPS 0.65 → 0.67, ADE 1.52 → 1.49 m,
and 77% still correct and faithful.

Alpamayo 1.5 says stop, then accelerates

RELEASED 10B DRIVING VLA • ZERO-SHOT

71%

trajectory correctness, NPS 0.60: the strongest zero-shot planner on our test set.

32–41%

of its predictions are both correct and faithful.

22%

of its stated slowing, stopping or yielding decisions are followed by acceleration.

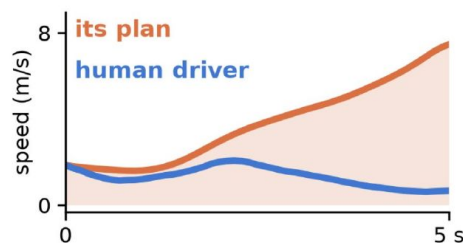


Alpamayo 1.5 writes

“Stop to yield to the truck turning across our lane since it blocks our path”

car moving at 1.9 m/s

human driver: slow down gently



ALI classifier (score)

✓ decision correct 0.73

✗ plan unfaithful 0.05

NPS 0.50 • ADE 3.5 m

ALI checks that a driving model **does what it says**

01

Descriptive drift is severe

Even Alpamayo 1.5, the strongest zero-shot planner, drives what it says on only 44–56% of frames.

02

Fine-tuning on aligned pairs helps

Faithfulness reaches 91–94%, decision correctness 76–77%, and planning improves.

03

Faithfulness needs a correct decision

A correct decision driven: NPS 0.79. Not driven: 0.42. A wrong one driven: no gain.

04

ALI improves VLAs, in training and at inference

As a GRPO reward: faithfulness, decision correctness and planning all improve.

Future work

Richer reward formulations · longer horizons and multiple agents · beyond driving, to robotics VLAs

Thank you.

Questions? Come find me at the poster.

OREN SULTAN · ORENSULTAN@GOOGLE.COM