NVIDIA

The ChatGPT Moment for Autonomous Driving

Marco Pavone, Sr. Director of Autonomous Vehicle Research

Alpamayo Summit @ CVPR 2026 | June 4, 2026

AV represents one of the best examples of Physical AI to date

Progress in AV has been mind-blowing





2005 → **2026**

更多免费行业报告 并购家 www.ipoipo.net

AV represents one of the best examples of Physical AI to date

Progress in AV has been mind-blowing





2005 → **2026**

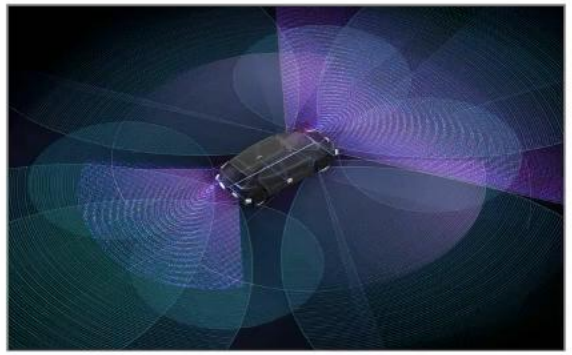
NVIDIA

Key Pillars of NVIDIA Autonomous Vehicle Solutions

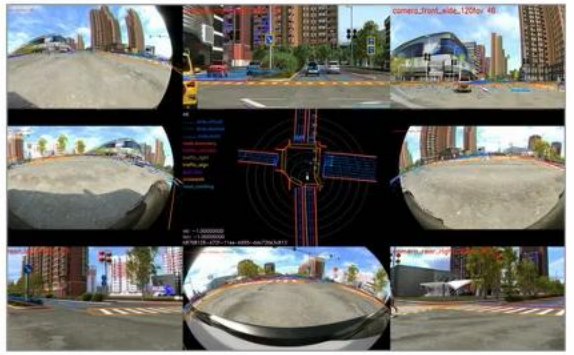
A full stack program targeting L4 autonomy | Billions of investment over 10 years



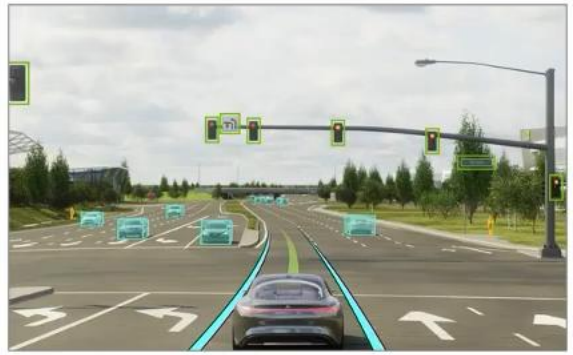
Scalability/One Chip Architecture



Drive SDK – OS, Tools, DriveWorks



DRIVE AV Stack & Fleet



Functional Safety



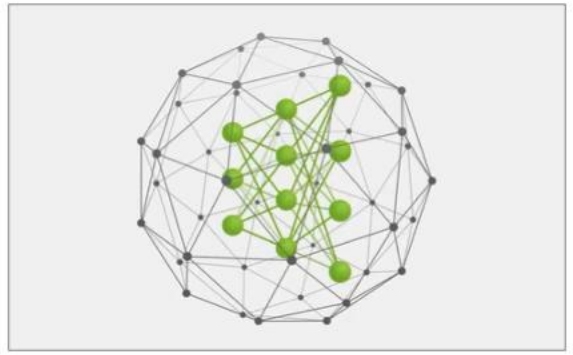
AI / DL Infrastructure to Edge



Simulation



AI Engineering



AV Research

更多免费行业报告 并购家 www.ipoipo.net

Foundation Models for Physical AI

A **foundation model** is any model that is trained on broad data (using self-supervision at internet scale) and that can be adapted to a wide range of downstream tasks

Cosmos World Foundation Models (WFMs)

Embed | Predict | Transfer | Reason

Physical AI Synthesis

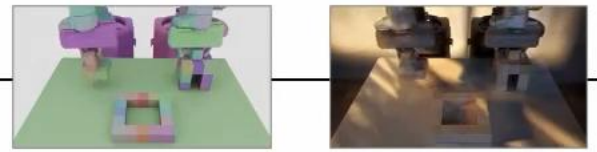
Cosmos Predict



Generate future world states from multimodal inputs

For post-training to develop specialized physical AI models

Cosmos Transfer



Generate Physics-grounded photoreal outputs conditioned on structural inputs or ground truth from NVIDIA Omniverse

For synthetic data generation with NVIDIA Omniverse

Physical AI Understanding

Cosmos Embed



"A car overtaking a truck" 0.15 0.21 -0.07 0.63

A joint video-text embedder tailored for physical AI

For text-to-video retrieval, inverse video search, semantic deduplication, and data curation

Cosmos Reason



Chain-of-thought reasoning to provide the best response

For out-of-the box physical AI reasoning and post-training for embodied AI models

Foundation Models Driving End-to-End Physical AI Development



Cosmos World Foundation Models (WFMs)
Embed | Predict | Transfer | Reason

Long-tail generalization via common-sense reasoning

AV development processes	On-vehicle AV stack
Data generation Data curation Sensor and traffic simulation Rate estimation & safety validation ...	FM-empowered E2E stacks Interactive vehicle interface Explainability: from vision to action Run-time monitoring ...



Large, intelligent models running in the cloud



Small, efficient models running onboard

更多免费行业报告 并购家 www.ipoipo.net

The Emergence of Embodied Reasoning



Background: Alpamayo Vision-Action (VA) Model

Foundation for Alpamayo production (NDAS E2E program)

Action: Autoregressive Trajectory Decoder

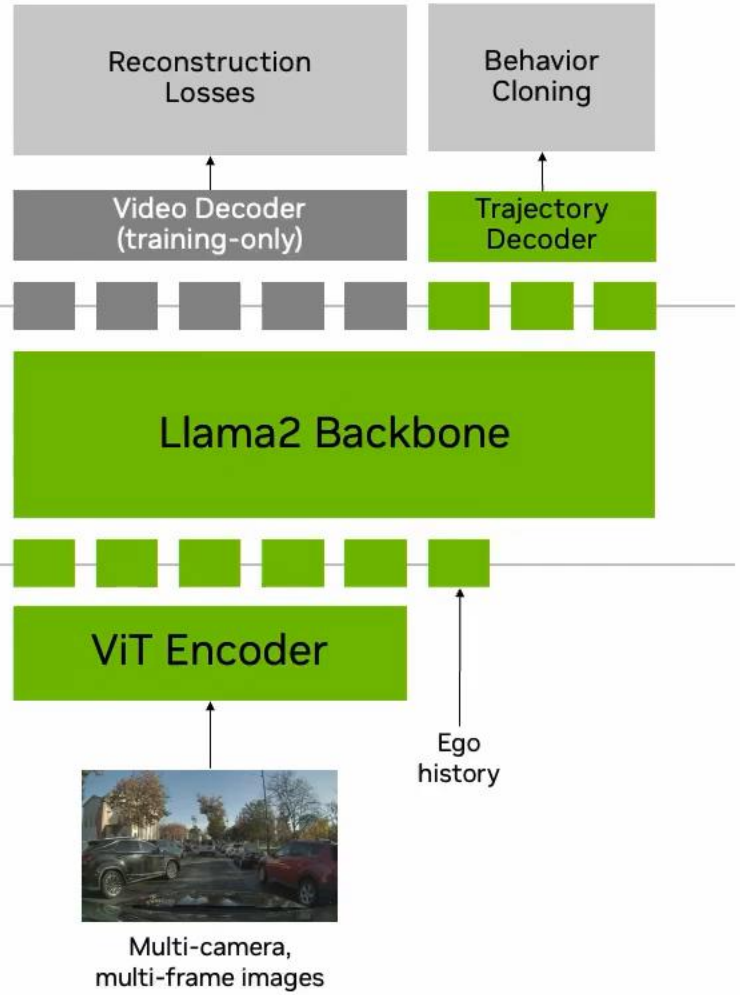
- Next-token-prediction training pipeline on the full multimodal sequence (images, actions)

Token Processing: LLM Backbone

- Trained over large-scale AV data
- Runtime-friendly size for embedded deployment by design

Vision: Spatiotemporal Context Encoder

- Efficient multi-camera, multi-frame tokenization to reduce token sequence lengths



更多免费行业报告 并购家 www.ipoipo.net

Alpamayo VA in Action

Videos 2x speed



Unprotected left, with VRU crossing, then nudging for VRU getting into vehicle with open door



Nudging for DPV, slowing for speed bump, nudging for trash can



Lane change with deadline into stop and go traffic for upcoming construction zone



Creeping and yielding at minor to major occluded left and stop sign



Nudging for a parking car, slowing/waiting for an oncoming bus in our lane, stopping at a stop sign and taking off



Nudge for VRU, yield behind parking truck, and nudge into oncoming lane

From Alpamayo VA to Alpamayo VLA

Assessing the benefits of internet-scale pre-training

Action: Multimodal Trajectory Decoder

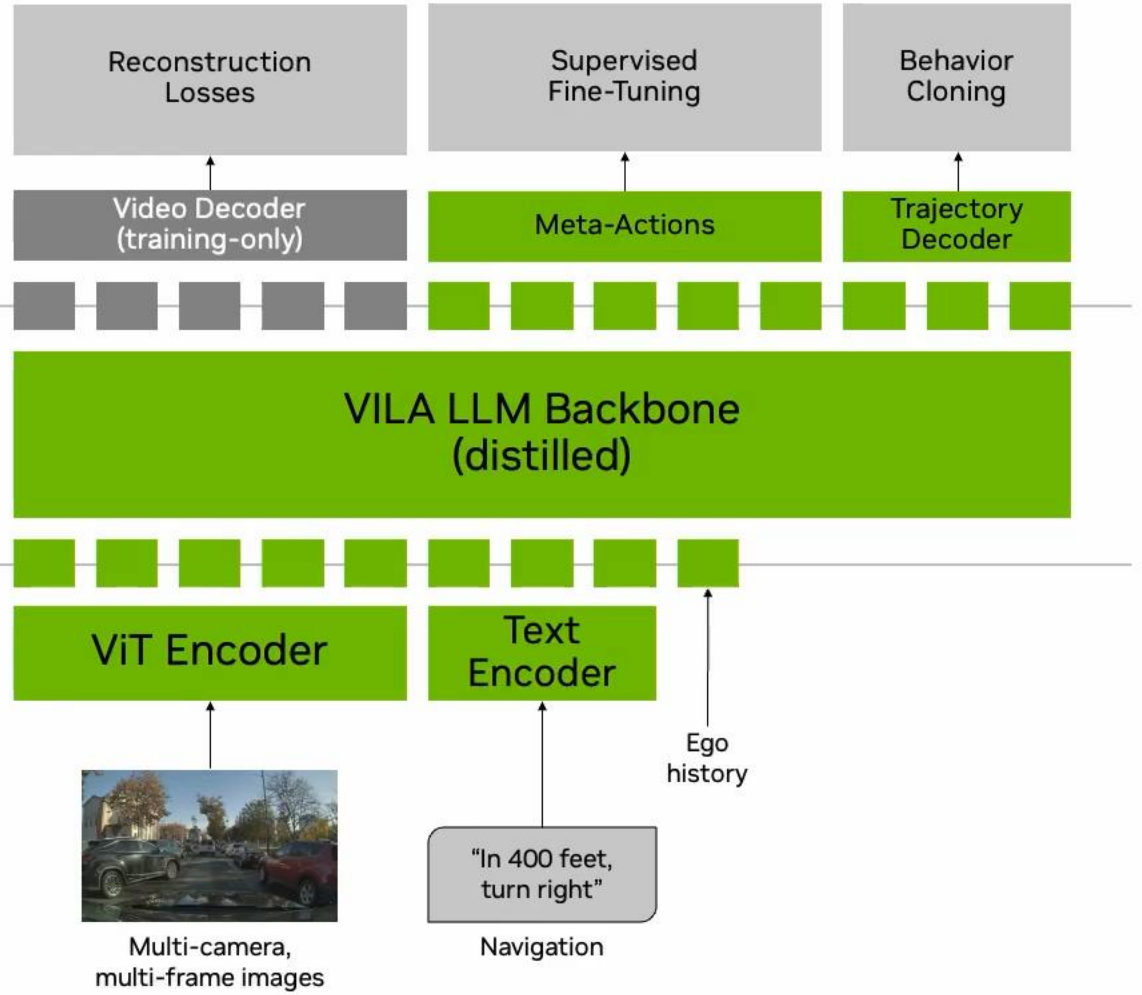
- Training with autolabeled meta-actions to improve trajectory multi-modality and interpretability

Language: Internet-Pretrained LLM Backbone

- Pretrained over internet-scale data
- Distilled into a runtime-friendly size for embedded deployment

Vision: Spatiotemporal Context Encoder

- Efficient multi-sensor tokenization to reduce token sequence lengths
- Native scaling to higher resolutions and sensor counts



更多免费行业报告 并购家 www.ipoipo.net

Alpamayo VLA Converges Faster and Drives Better

Key result: Language pre-training synergizes well with the task of driving

- Language pre-training provides **faster convergence** and leads to a **better steady state** compared to training from scratch, evaluated on 25,600 scenarios across the US and EU*
- Closed-loop evaluation corroborates the results



On the value of language pre-training



**Not the first time we have seen such a result: Mao, Ye, Qian, Pavone, Wang. A Language Agent for Autonomous Driving. COLM 2024*

Alpamayo Reasoning VLA: Scalable Architecture for Reasoning VLAs

Action: Reasoning-informed Trajectory Diffusion

- Efficient action-expert trajectory decoding via lightweight conditional flow-matching
- RL post-training to improve reasoning-action alignment and action quality

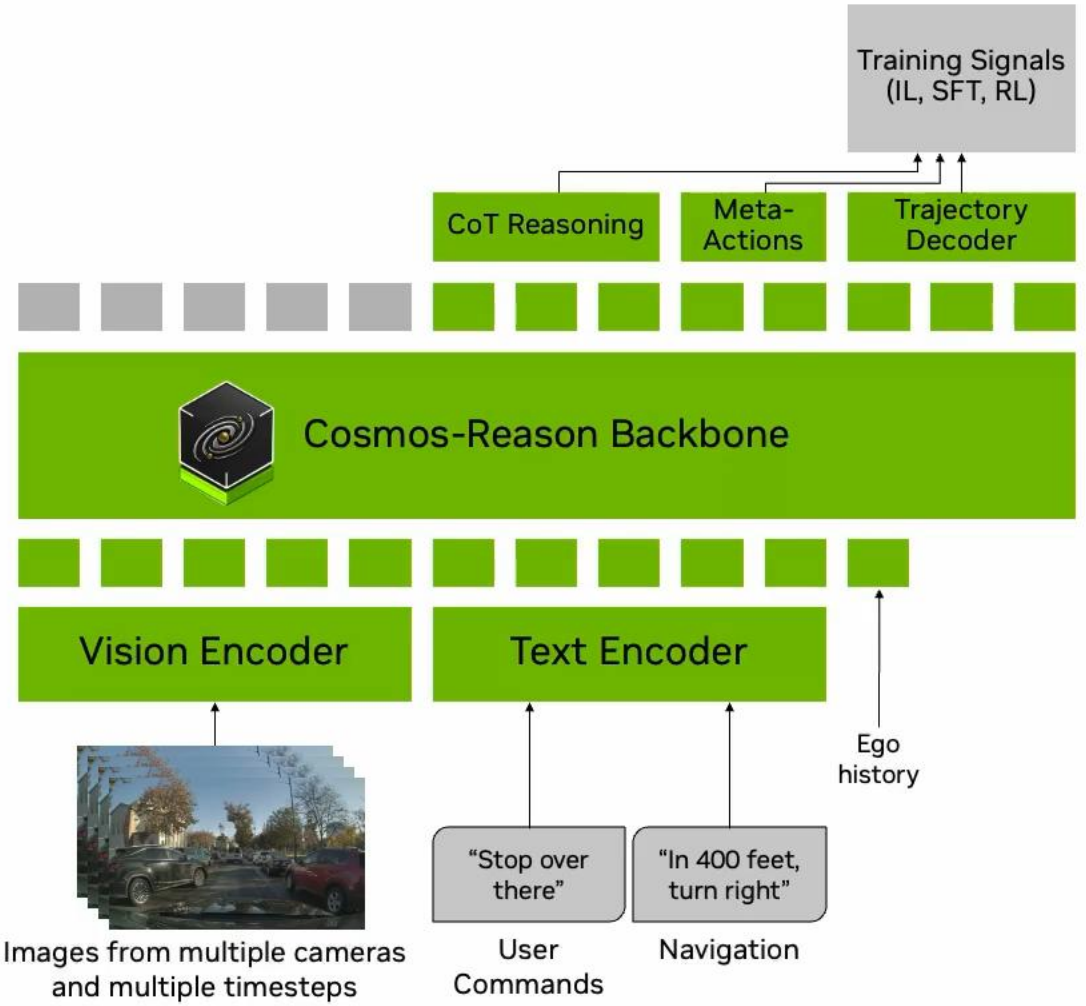
---

Reasoning: Internet-Pretrained Reasoning Backbone

- Pretrained to reason over internet-scale data
- RL with verifiable rewards improve causal reasoning

Vision: Efficient Context Encoder

- Handles multiple input modalities (cameras, text)
- Efficient multi-camera, multi-timestep tokenization to reduce token sequence lengths
- Native scaling to higher resolutions and sensor counts



更多免费行业报告 并购家 www.ipoipo.net

NVIDIA

2025-10-28

Alpamayo-R1: Bridging Reasoning and Action Prediction for Generalizable Autonomous Driving in the Long Tail

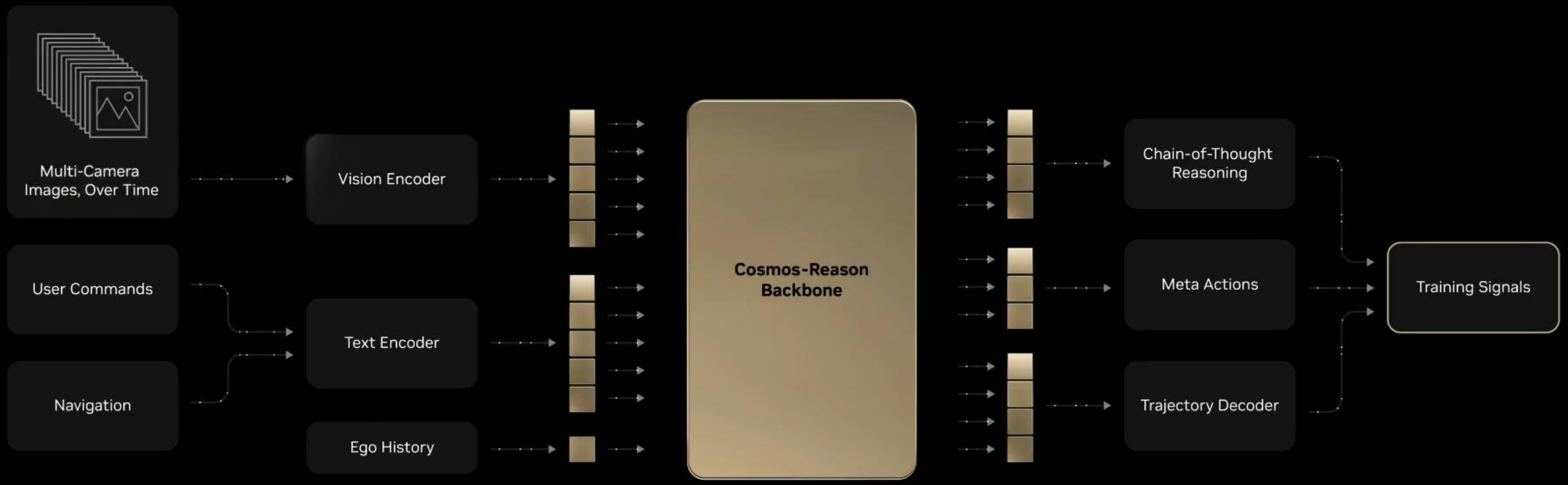
NVIDIA[1]

Abstract

End-to-end architectures trained via imitation learning have advanced autonomous driving by scaling model size and data, yet performance remains brittle in safety-critical long-tail scenarios where supervision is sparse and causal understanding is limited. We introduce Alpamayo-R1 (AR1), a vision–language–action model that integrates Chain of Causation reasoning with trajectory planning to enhance decision-making in complex driving scenarios. Our approach is built on three key innovations:

NVIDIA, *Alpamayo-R1: Bridging Reasoning and Action Prediction for Generalizable Autonomous Driving in the Long Tail*, 2025; https://research.nvidia.com/publication/2025-10_alpamayo-r1

Alpamayo 1



更多免费行业报告 并购家 www.ipoipo.net


Reasoning
Stop due to pedestrians in the crosswalk.
Trajectory Prediction
History
Prediction
Ego Vehicle
Longitudinal (m)
Lateral (m)
80
60
40
20
0
−20
−10
0
10
20
Alpamayo 1 Driving in AlpaSim



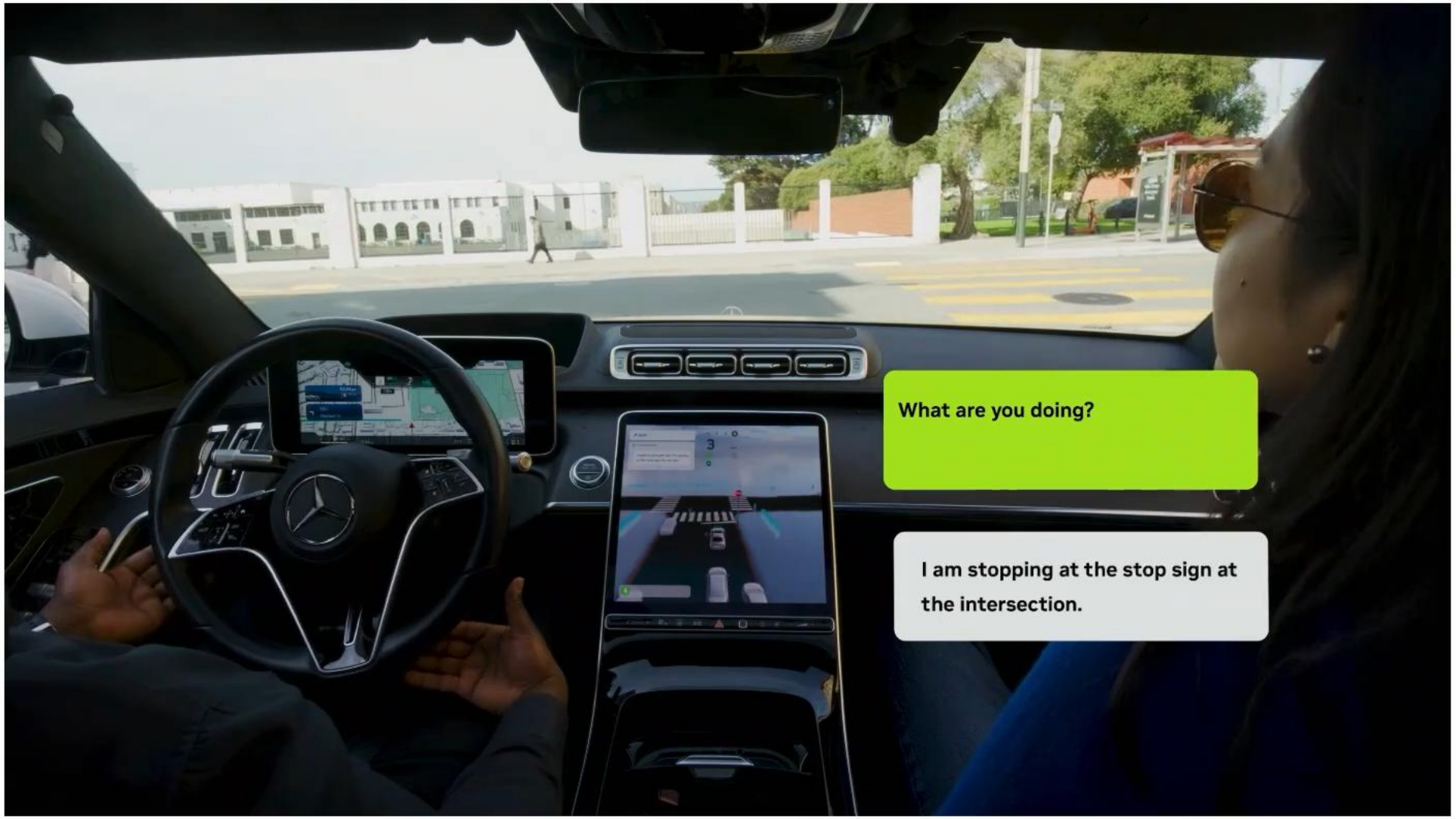




Physical AI AV Dataset

更多免费行业报告 并购家 www.ipoipo.net

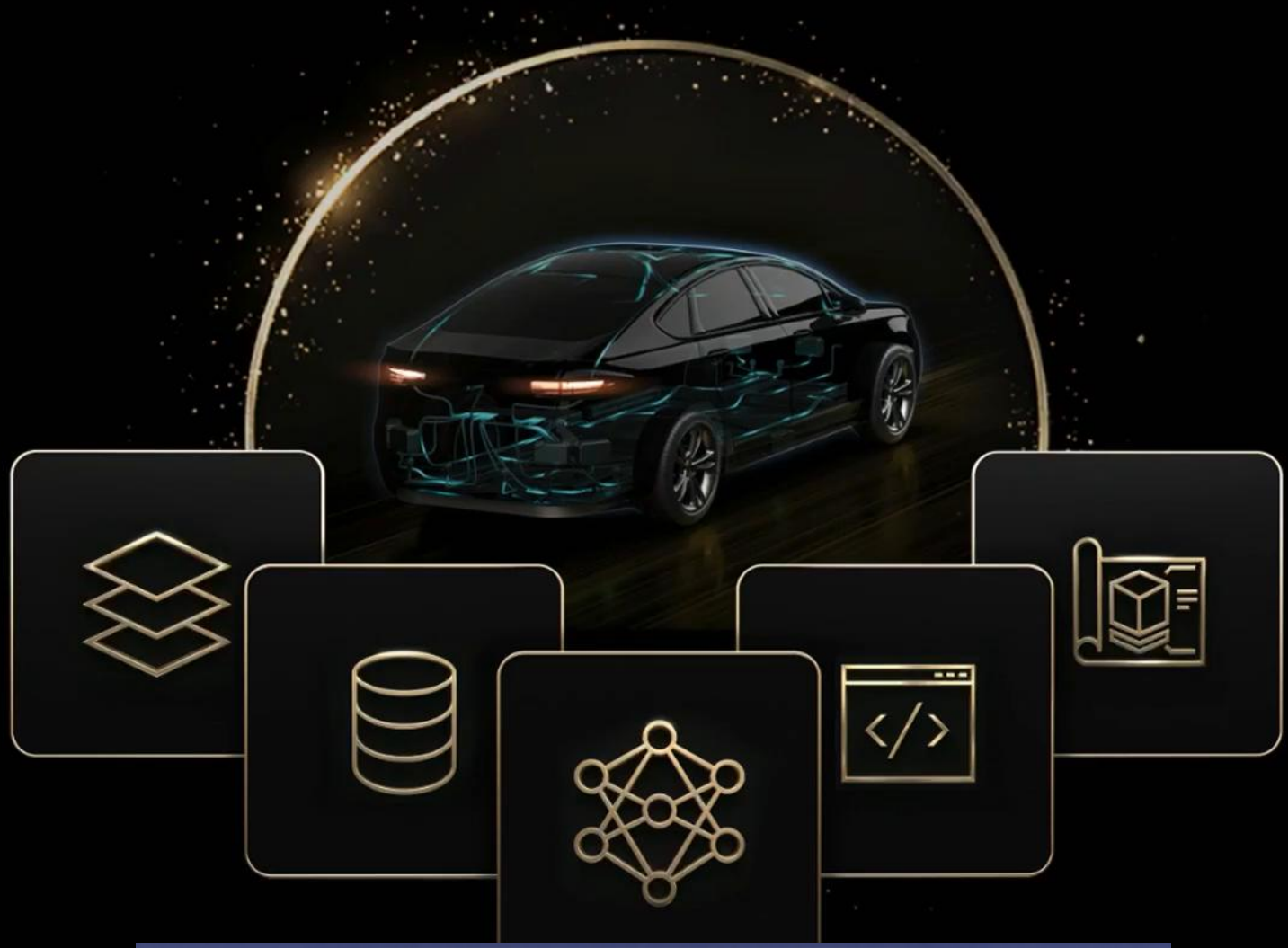
Alpamayo Deployed on Car



Pavone, Liu, *From Research to Production: How Alpamayo Accelerates Autonomous Vehicle Development*, GTC Talk 2026
https://www.nvidia.com/en-us/on-demand/session/gtc26-s81779/

NVIDIA Alpamayo

An Open Ecosystem Designed to Accelerate Reasoning-Based Autonomous Vehicle Development



更多免费行业报告 并购家 www.ipoipo.net

NVIDIA Alpamayo Open Platform



Alpamayo Models

Open 10B-parameter Reasoning VLAs

- Steerable behavior via navigation and text prompts
- Flexible multi-camera support
- Post-training scripts



AlpaSim

Open simulation framework for closed-loop testing and validation

- Extensible plugin system
- Launching an AlpaSim-based Closed-Loop E2E Driving Benchmark



Physical AI Dataset

Most diverse open AV dataset

- ~1700 hours of multi-sensor data from 2500 cities in 25 countries
- Human QA'd reasoning labels
- CoC auto-labeling pipeline
- Launching a Reasoning Benchmark focused on challenging scenarios

Entry points:

- https://huggingface.co/blog/drmapavone/nvidia-alpamayo
- https://huggingface.co/blog/drmapavone/nvidia-alpamayo-1-5

NVIDIA

Ecosystem Recognition

Thank YOU for engaging with Alpamayo!

Computex 2026 Best Choice Award

Alpamayo Open Platform recognized for:

- Empowering manufacturers to build AVs with human-like perception, reasoning, and decision-making capabilities.
- Enabling closed-loop testing across diverse traffic scenarios, weather conditions, and edge cases.
- Fostering a robust open-source AI ecosystem.



Zero-shot application to Japanese data!

(Tier IV)

400,000+ downloads of Alpamayo Models
2,300,000+ downloads of the Physical AI AV Dataset
3,000+ stars across all Alpamayo GitHub repos

2026 Challenge WINNERS | Leaderboard | Submit | About

Method	Team	MMS	MMS (selected)	MMS (heavy rai…	MMS (constructio…
TTVLM-Base Qwen 3.5 2B MD+BV0.25	STLA-MINES	**5.15**	4.52	6.44	5.89
Kinematics V2	Rangers	**4.31**	4.18	4.39	4.45
A1.5	KE:SAI	**4.31**	4.21	4.50	4.84

Zero-shot 3rd place on KITScenes LongTail Challenge!

(eam)

更多免费行业报告 并购家 www.ipoipo.net

NVIDIA Alpamayo Open Platform





Recipes

Each recipe folder contains its own README with installation and training instructions.

Recipe	Description
`recipes/alpamayo1_sft/`	Alpamayo 1 supervised fine-tuning (HuggingFace Trainer + DeepSpeed)
`recipes/alpamayo1_5_sft/`	Alpamayo 1.5 SFT (HuggingFace Trainer + DeepSpeed)
`recipes/alpamayo1_x_rl/`	Alpamayo 1 and 1.5 RL post-training (Cosmos-RL / GRPO)

Utility Scripts

Script	Purpose
`scripts/curate_pai_samples.py`	Curate a subset of PAI samples
`scripts/convert_checkpoint.py`	Convert between Alpamayo 1 and 1.5 checkpoints

Chain-of-Causation Reasoning Labels

100% manually-verified labels to bootstrap reasoning model R&D, evaluation, and benchmarks

- Explore reasoning immediately, from training to evaluation.

Reasoning Auto-Labeling Pipeline

Tools to generate reasoning labels on *your own* data

- Generate meta-actions, identify keyframes, and/or produce chain-of-causation reasoning labels.
- Extensible APIs; use your favorite server or local VLM

github.com/NVlabs/alpamayo-coc-autolabeler

Post-Training Recipes

Finetune Alpamayo for *your* tasks and data

- Post-train Alpamayo 1 and 1.5 with supervised fine-tuning and open-loop reinforcement learning.
- Extend existing recipes, develop your own, contribute back to the ecosystem.

github.com/NVlabs/alpamayo-recipes

New at CVPR 2026!

https://huggingface.co/blog/drmapavone/nvidia-alpamayo-2

NVIDIA

NVIDIA DRIVE Hyperion Robotaxi Platform

Announcing Alpamayo 2 Super Robotaxi Reasoning Model

Alpa Simulators

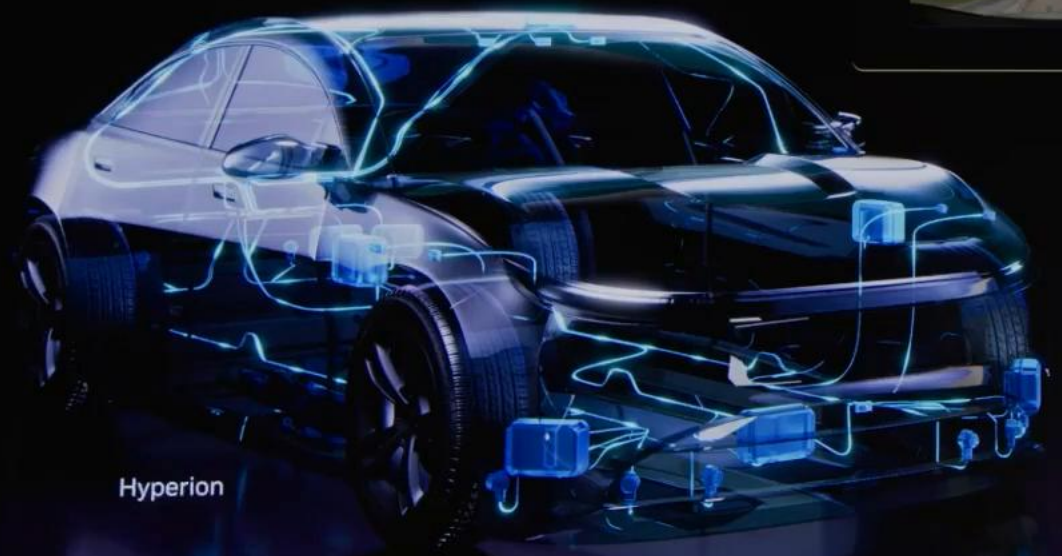


Alpamayo and OmniDreams Open Models

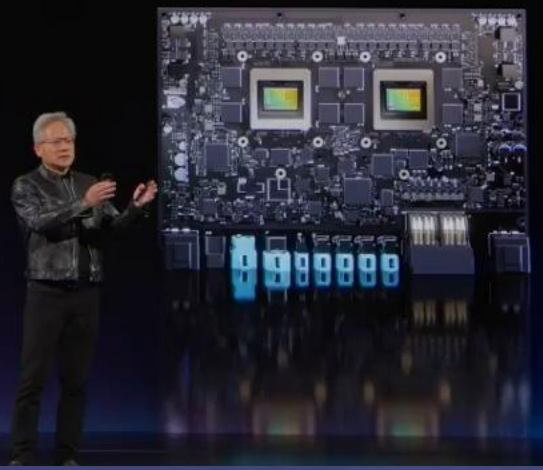


Training Data Generation

NVIDIA Halos OS



Hyperion



DRIVE AGX Thor

更多免费行业报告 并购家 www.ipoipo.net

Agenda

4:20 – 4:40pm: Training at Scale: The Physical AI Dataset

4:50 – 5:15pm: Autonomous Driving With Reasoning Models

5:25 – 6:10pm: Closing the Development Loop: Simulation and Training With AlpaSim + AlpaGym

Raffle #1

6:20 – 6:40pm: Deploying With Confidence: Physical AI Safety With Halos

6:50 – 7:20pm: Inside the Alpamayo Ecosystem: Partners, Open Research, and What to Build Next

Raffle #2

7:20pm: Closing Remarks

Mixer



Talk to us this week at CVPR!



Ed Schmerling



Yurong You



Max Igl



Yuxiao Chen



Apoorva Sharma



Riccardo Mariani



Matt Cragun



Boris Ivanovic



Yan Wang



Wenjie Luo

更多免费行业报告 并购家 www.ipoipo.net