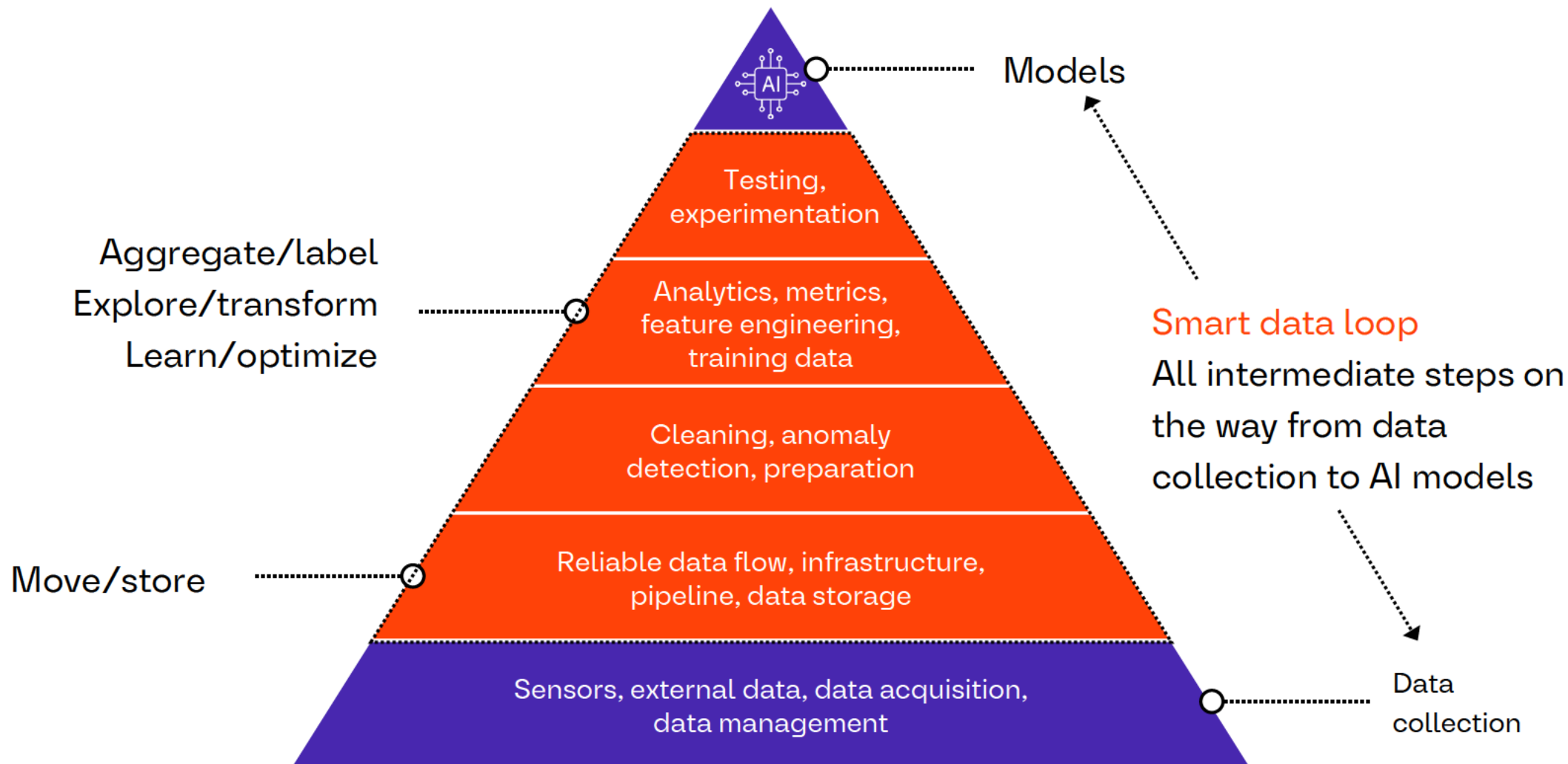


Beyond justbetterDATA

From data strategy to autonomous reliability at scale

The data science hierarchy

AI performance is constrained by the lowest data layer



The industry is moving to E2E driving Dominant AI paradigm

	pushes large-scale end-to-end neural driving
	integrates foundation models into perception & planning
	evolves toward road intelligence models
	is integrating AI-driven driving stacks
	is investing in data-centric autonomy systems
	is enabling foundation-model-based autonomy platforms

What connects all of them?



Massive data scaling



Intelligent selection



Continuous learning loop

E2E driving & scaling

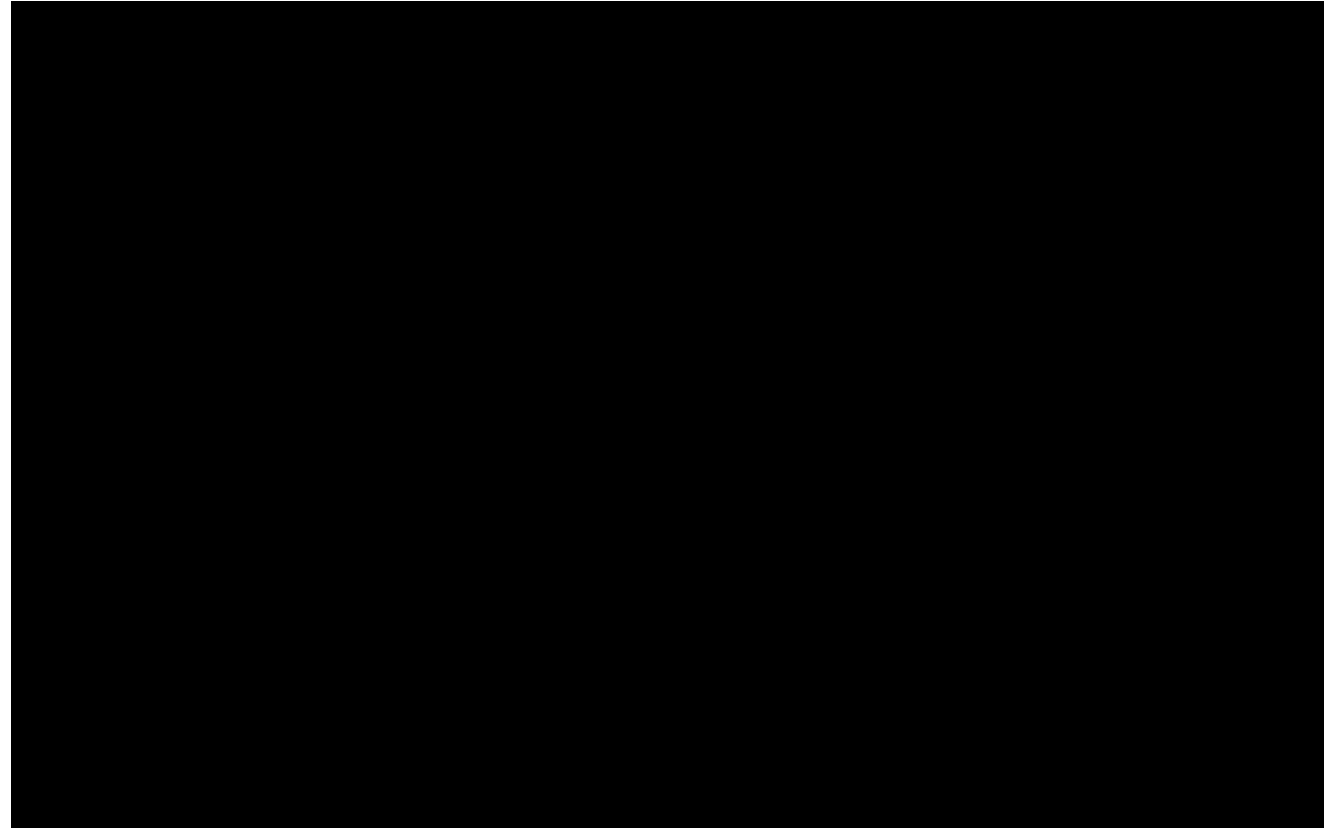
Dominant AI paradigm in autonomous mobility



Performance scales with data:

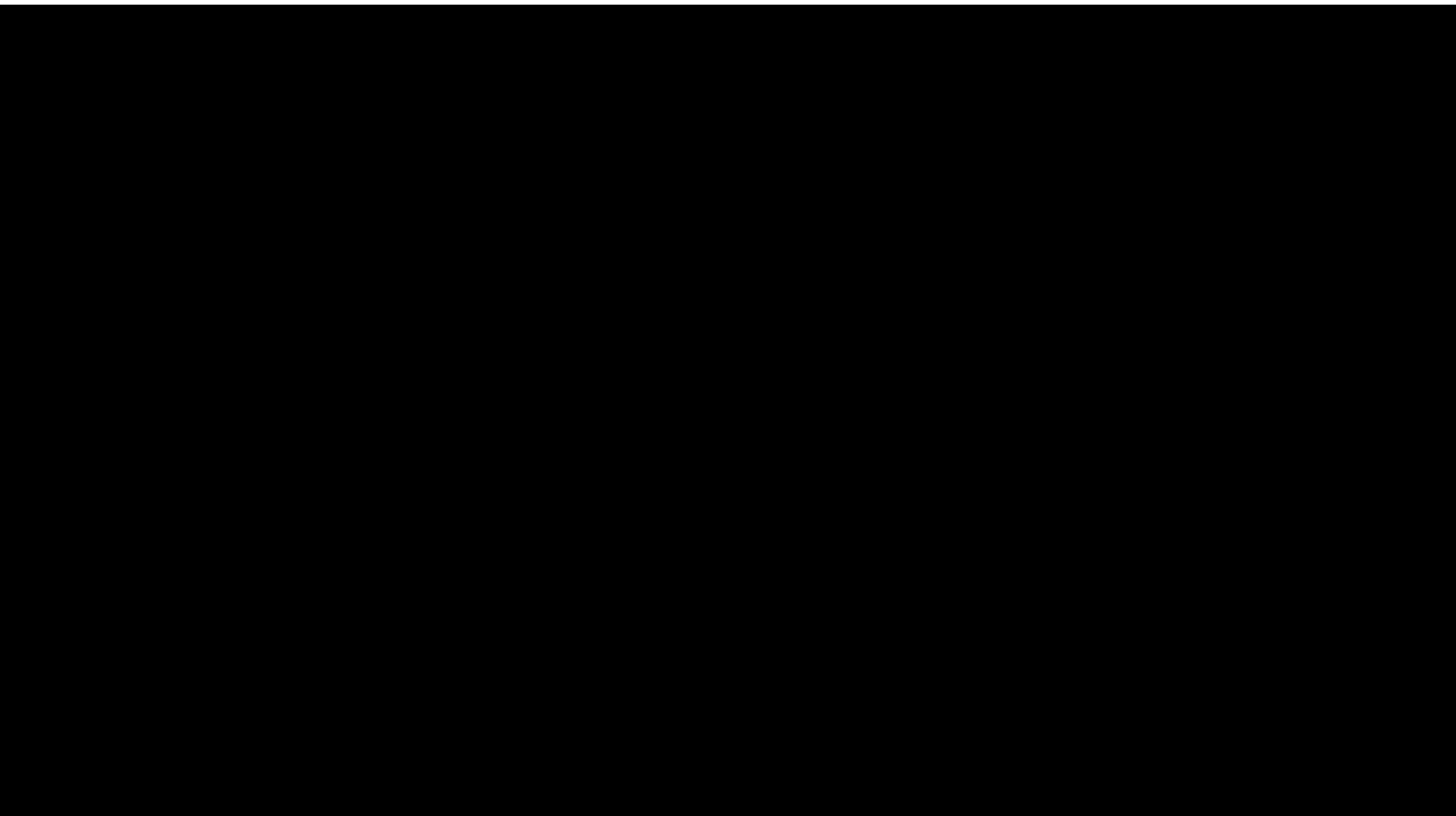
NVIDIA now studies scaling laws for autonomous driving.

Tesla estimates that roughly 10 billion miles are required to achieve safe unsupervised driving.



The Real bottleneck

Not compute, not models, but data



Data matters

0.1% still breaks the system

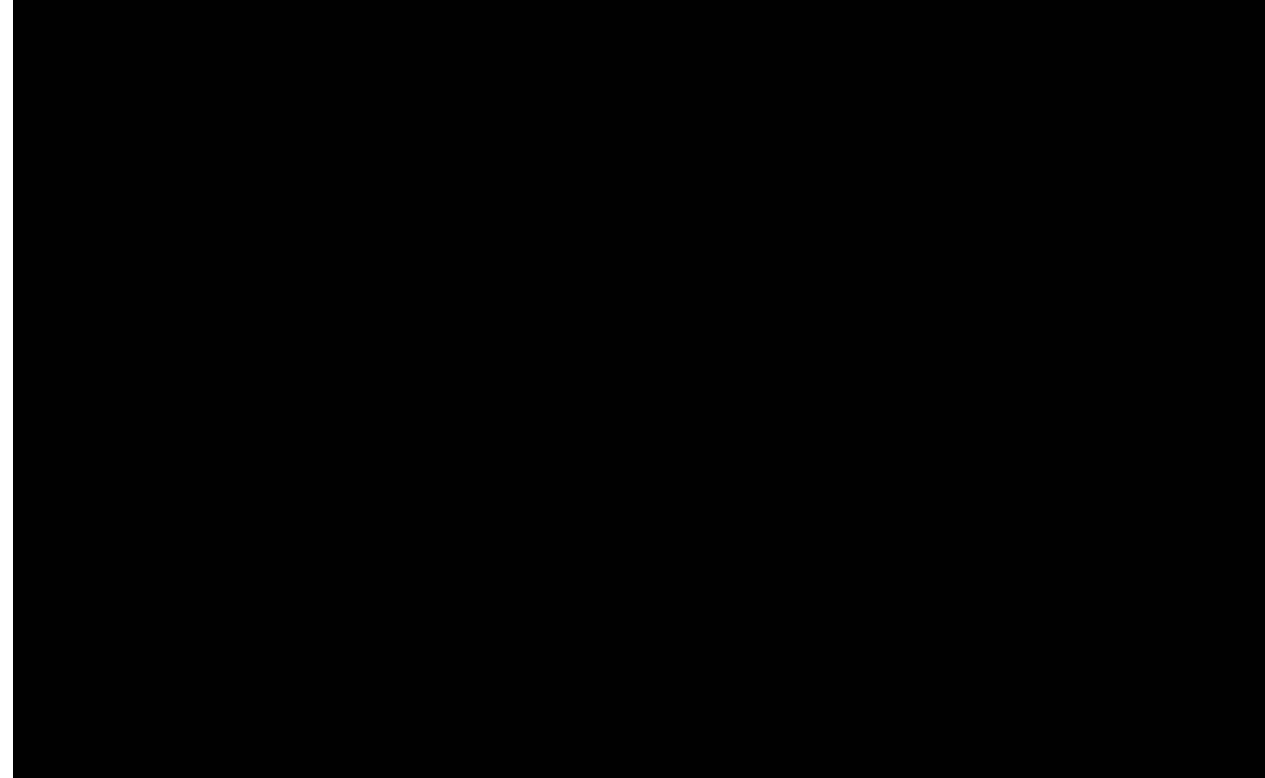
In 2026, a Level-4 autonomous vehicle struck a child in a school zone due to unexpected pedestrian behavior and occlusion.

Technical Breakdown

Rare alignment of common factors caused sensor occlusion and unpredictable child movement, exposing gaps in AI training data.

Data and Validation Gap

Incident highlights need for prioritizing rare scenario data to improve autonomous vehicle validation processes.



[Federal probe launched into Waymo crash in Santa Monica school zone | Fox News](#)

Waymo's E2E driving benchmark

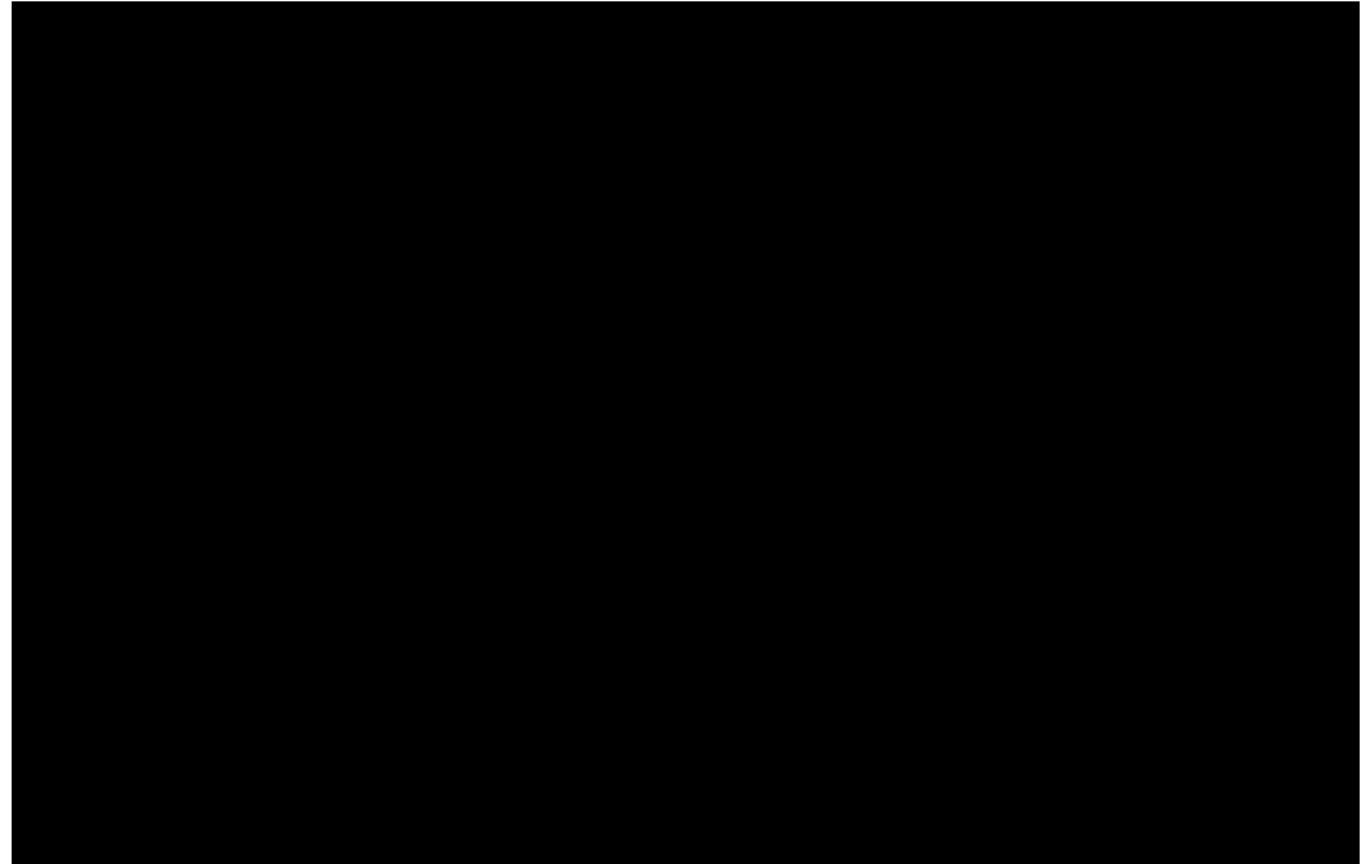
Targeting the 0.03% long-tail scenarios

In early 2025, Waymo released a new open dataset designed for E2E driving models.

Goal: evaluate performance in long-tail driving scenarios.



Rare, safety-critical moments that account for less than **0.03%** of real-world driving time yet often determine the overall safety of an autonomous system.



[Waymo's End-to-End Driving Benchmark: Targeting the 0.03% Long-Tail Scenarios – Making World Models Safe for Autonomous Systems](#)

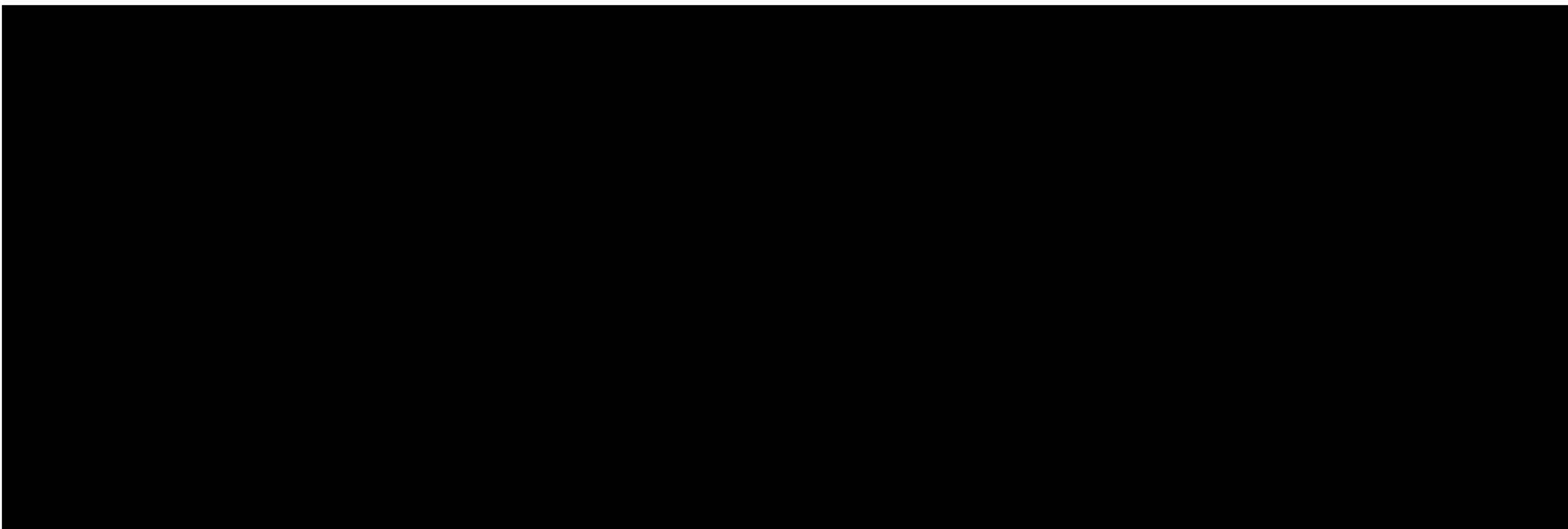
Foundation models need foundation data

Shifting the bottleneck from models to data



Several major players now explicitly frame autonomous driving as

Foundation model + **data scaling** problem not merely as an algorithm or stack integration

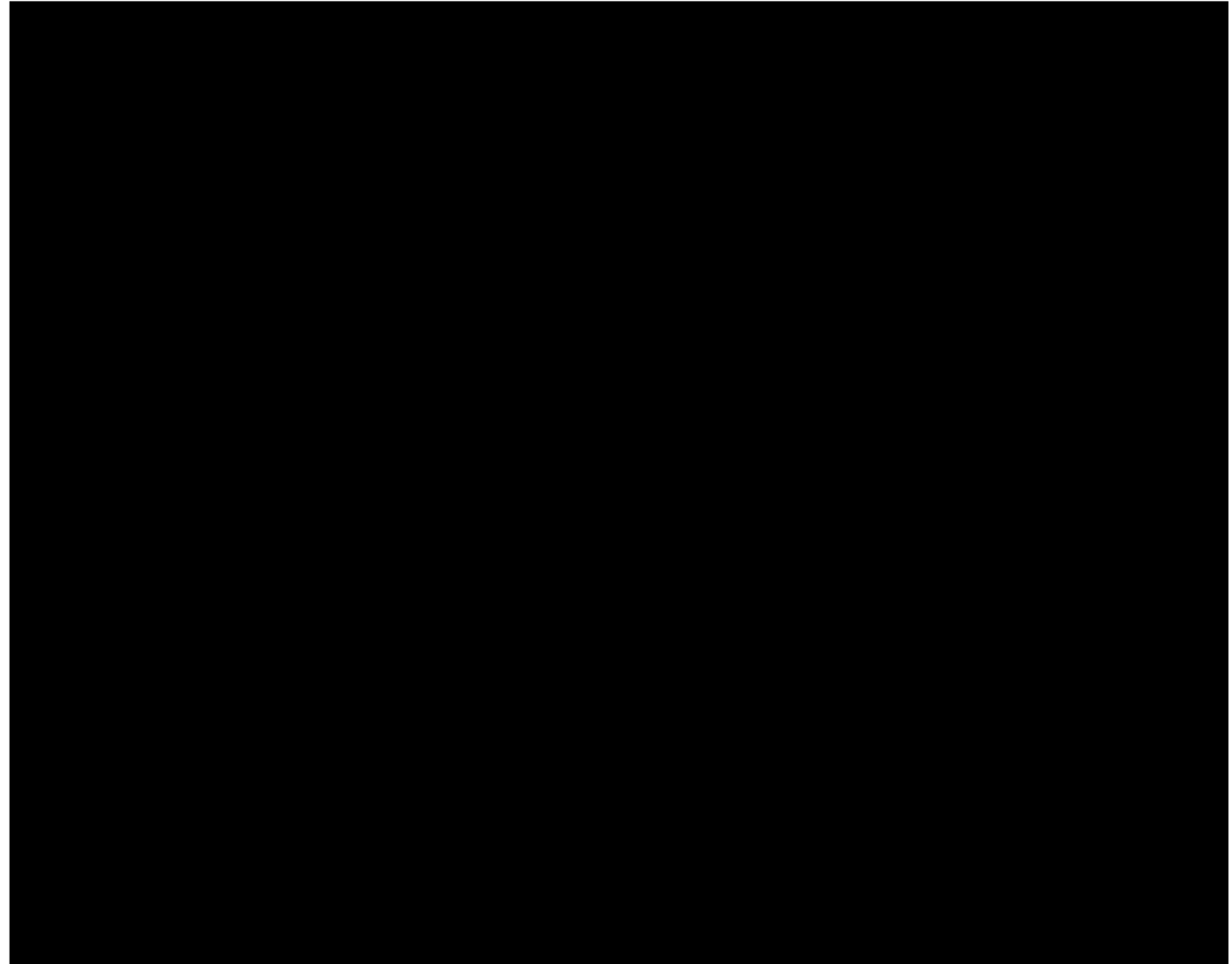


From E2E models to E2E intelligence systems

Data strategy



**E2E driving models
require
E2E data selection**



[WOD-E2E: Waymo Open Dataset for End-to-End Driving in Challenging Long-tail Scenarios](#)

Europe's autonomous advantage will be data, not KM

Data strategy for safe, scalable AI



Federated data ecosystem

Europe's autonomous advantage relies on a federated data ecosystem combining data from OEMs, suppliers, fleets, and researchers.

Trusted data sharing

Trusted data spaces and marketplaces ensure privacy, regulatory compliance, and data sovereignty across Europe.

Innovative data approaches

Scenario-driven synthetic data and federated validation loops enhance autonomous driving model safety and accuracy.

Thank you for your attention

Dr. Claudio Longo

claudio.longo@aumovio.com

