Foundation models

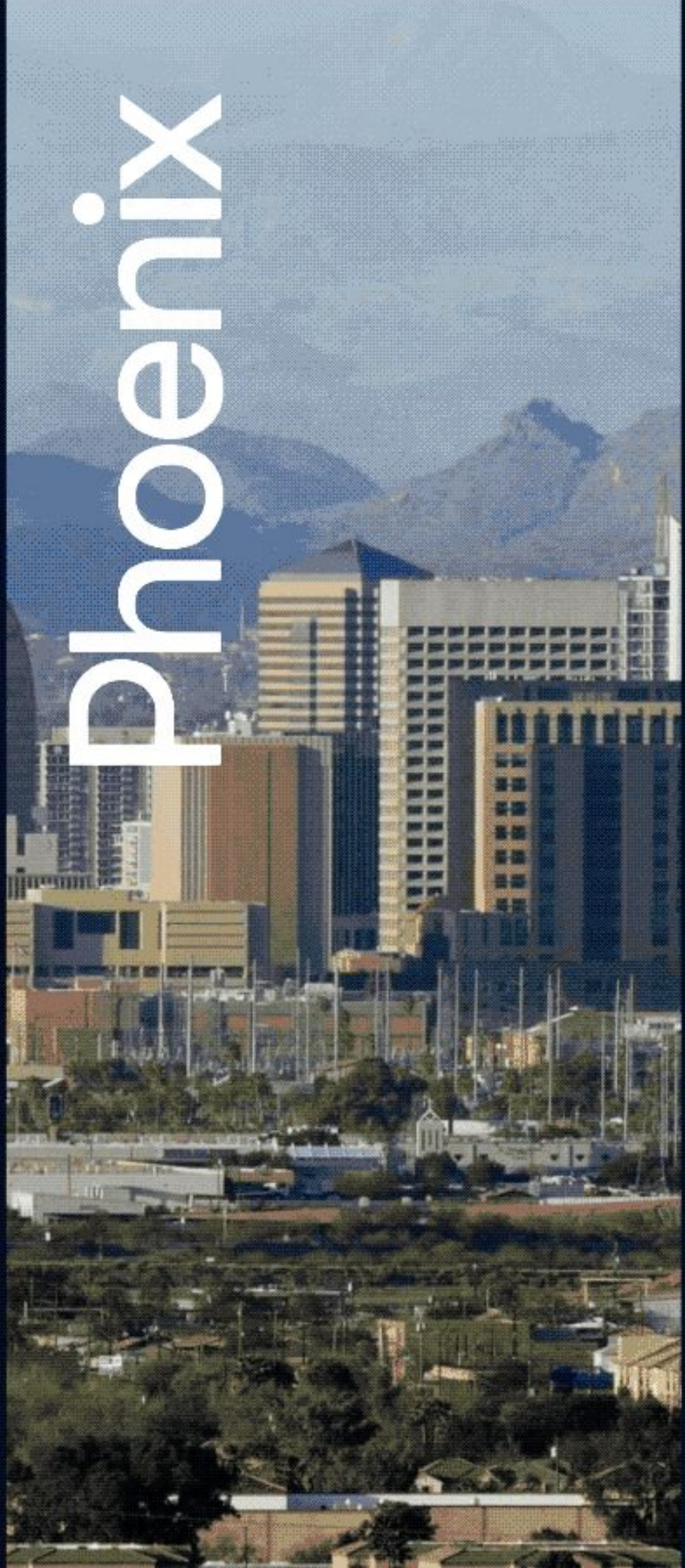
For autonomous driving

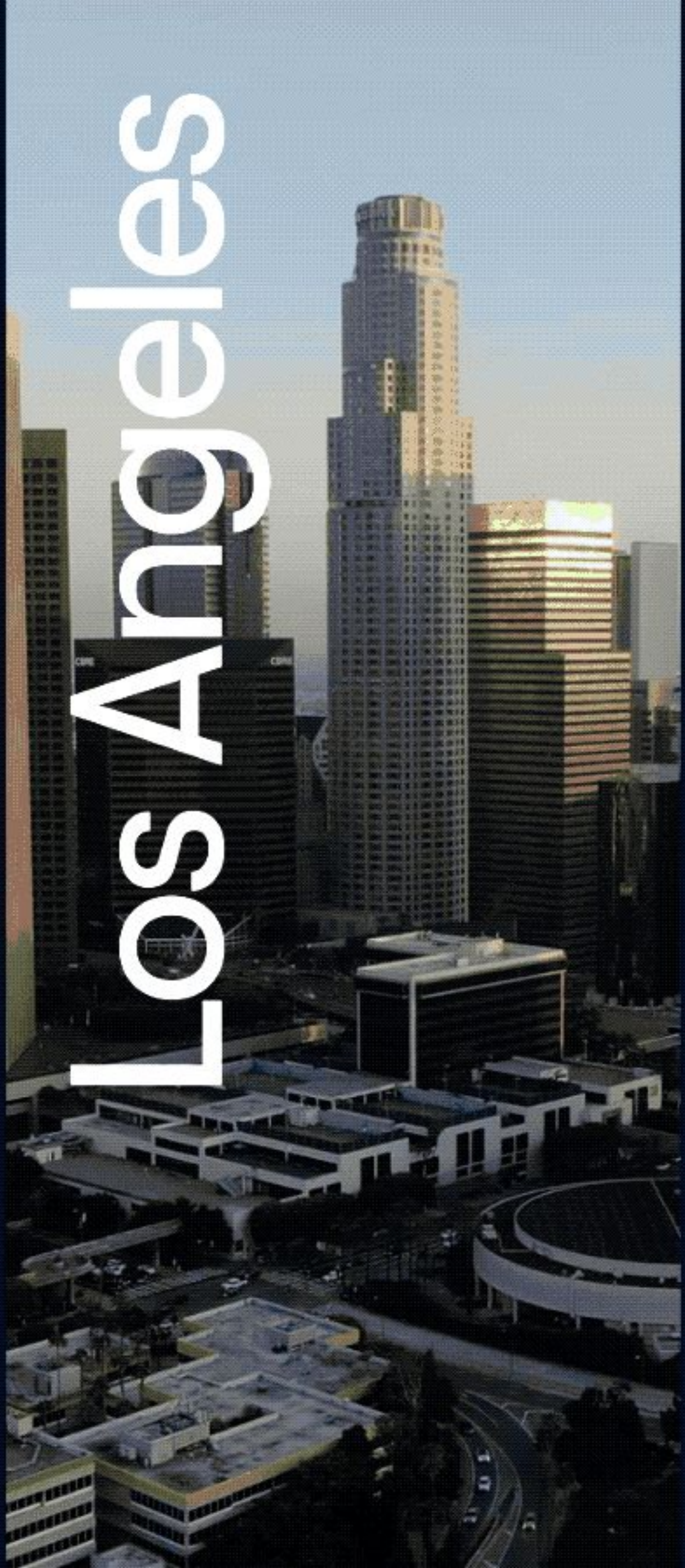
Vincent Vanhoucke
Distinguished Engineer
Waymo

OUR MISSION –

Be the world's most trusted driver.

San Francisco


Phoenix


Los Angeles

Austin

Up next

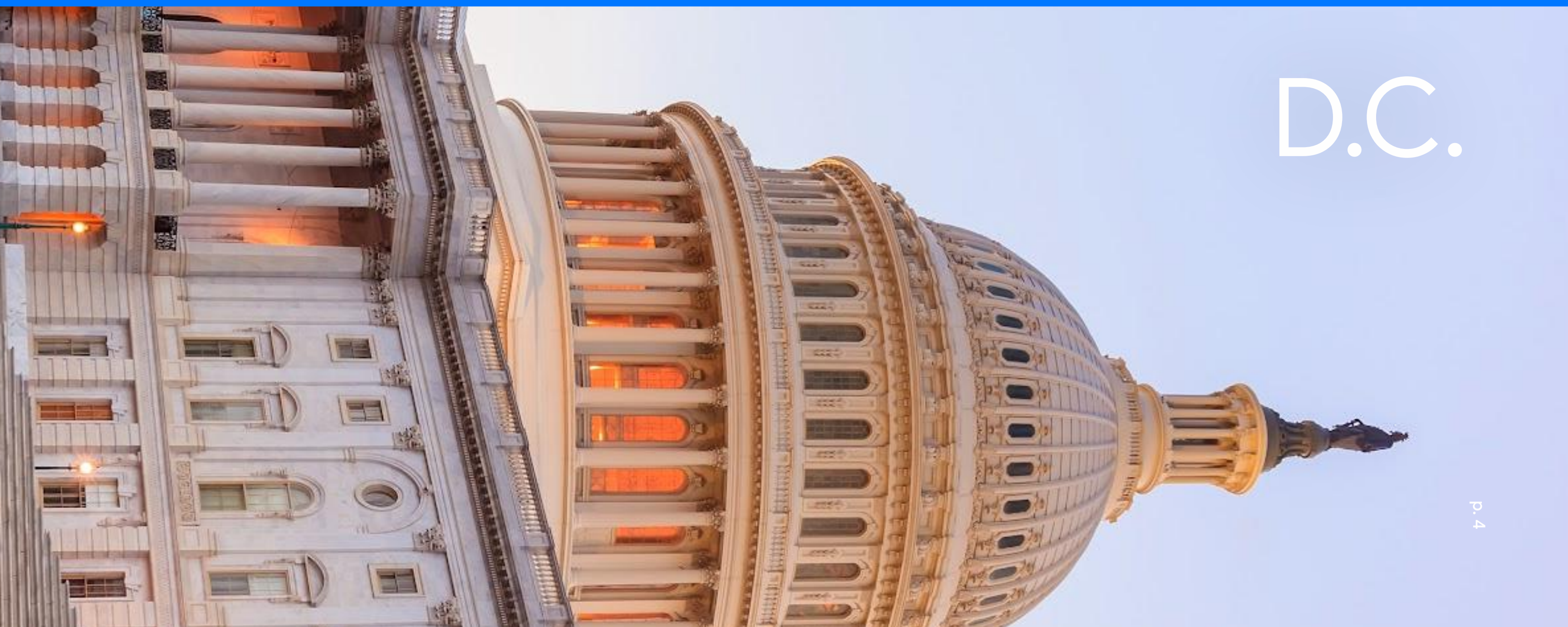
WAYMO ONE × Uber

Atlanta

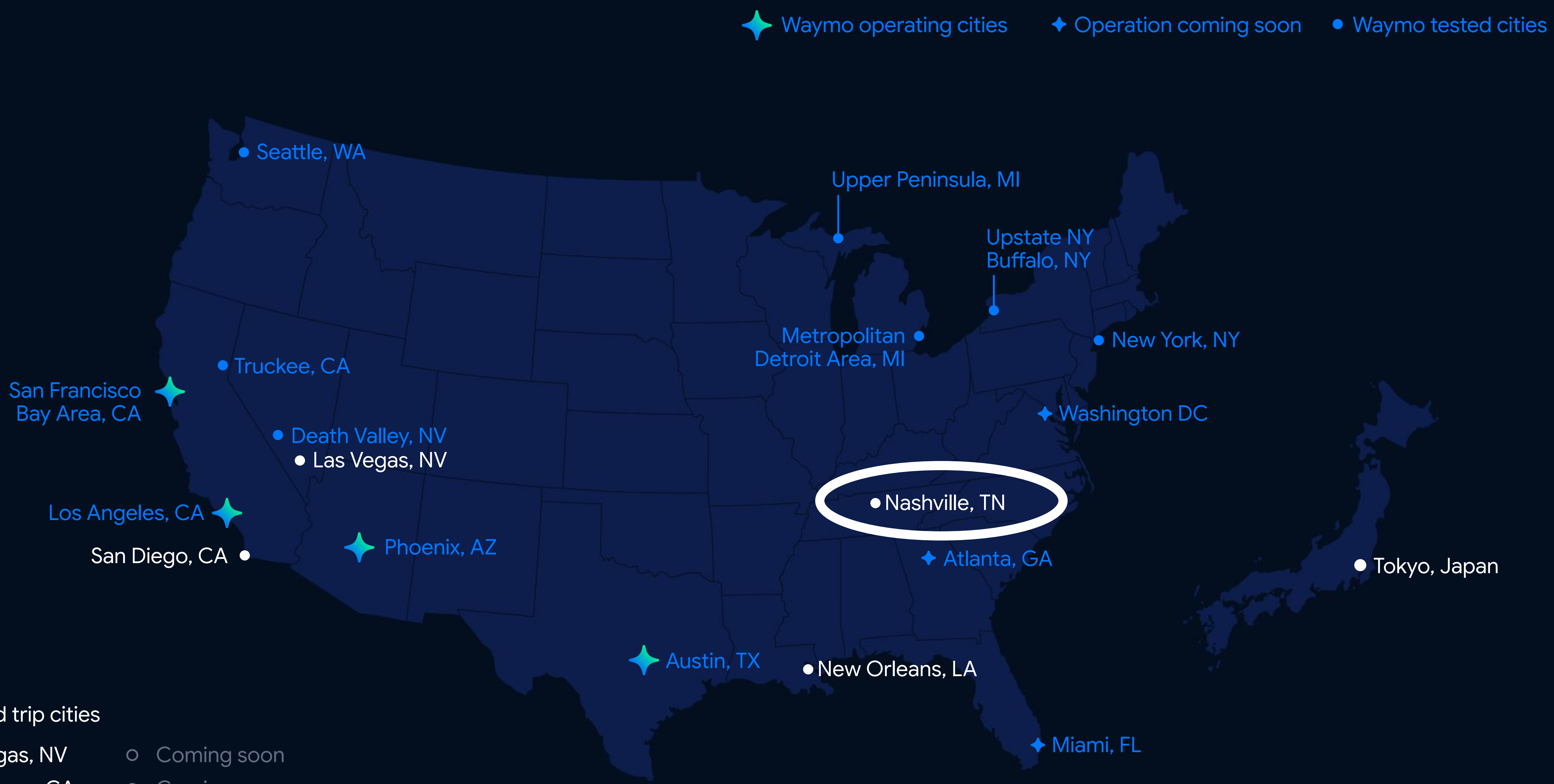
WAYMO ONE

Miami

WAYMO ONE



D.C.



2025 Road trip cities

- ● Las Vegas, NV
- ● San, Diego, CA
- ● New Orleans, LA
- ● Nashville, TN
- ○ Coming soon
- ○ Coming soon
- ○ Coming soon
- ○ Coming soon
- ○ Coming soon
- ○ Coming soon

Road trip

WAYMO ONE

Now serving over

250,000

paid trips per week

Protecting Vulnerable Road Users

Compared to human drivers over 56.7 million miles in our operational cities, Waymo Driver had:

92%

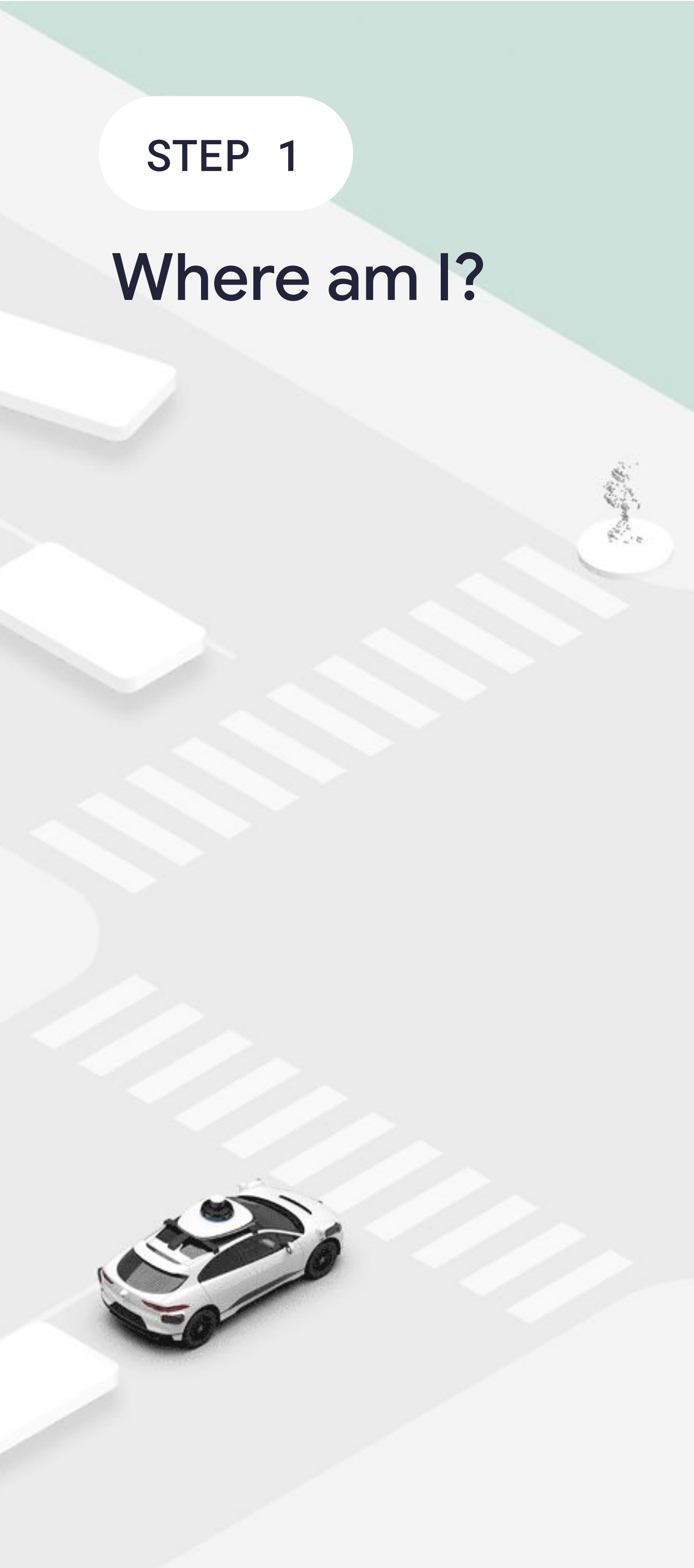
Fewer crashes with injuries to pedestrians

82%

Fewer crashes with injuries to cyclists

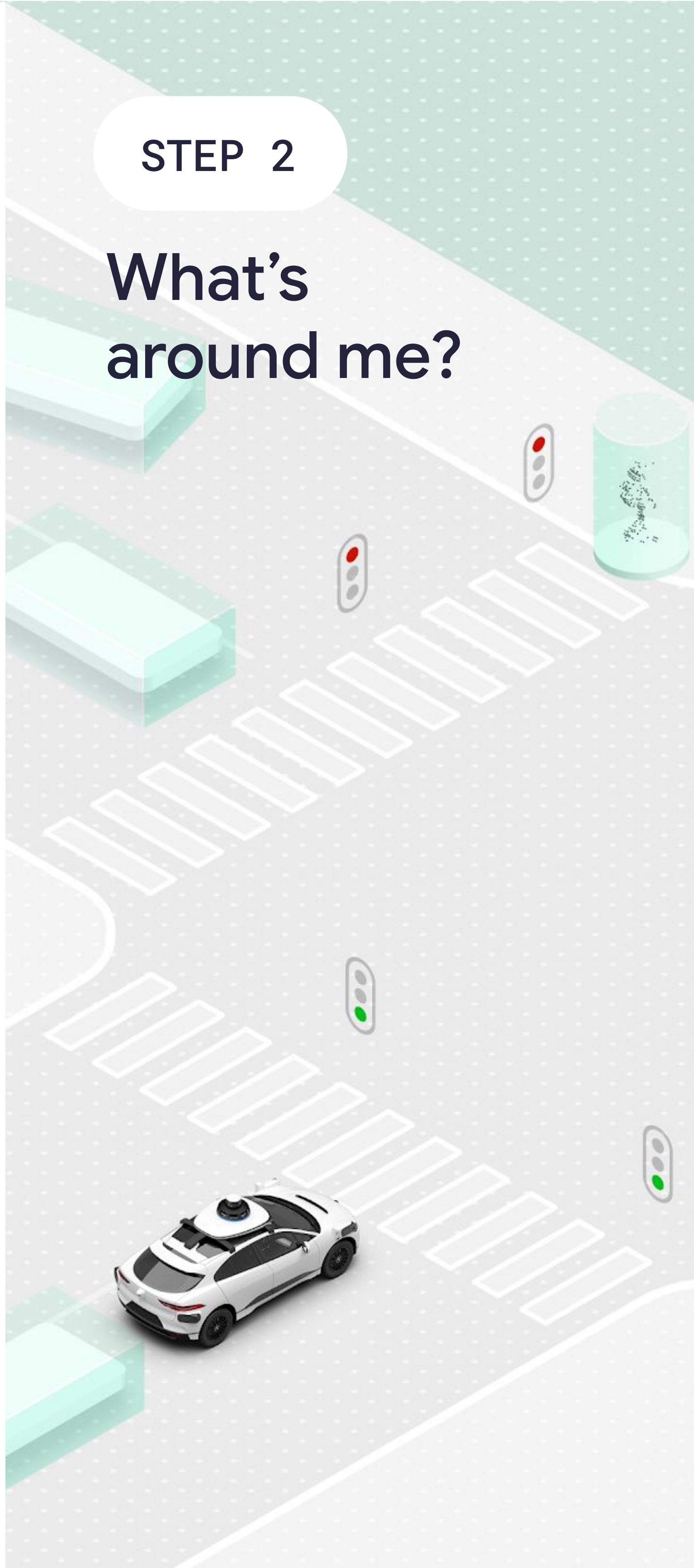
82%

Fewer crashes with injuries to motorcyclists



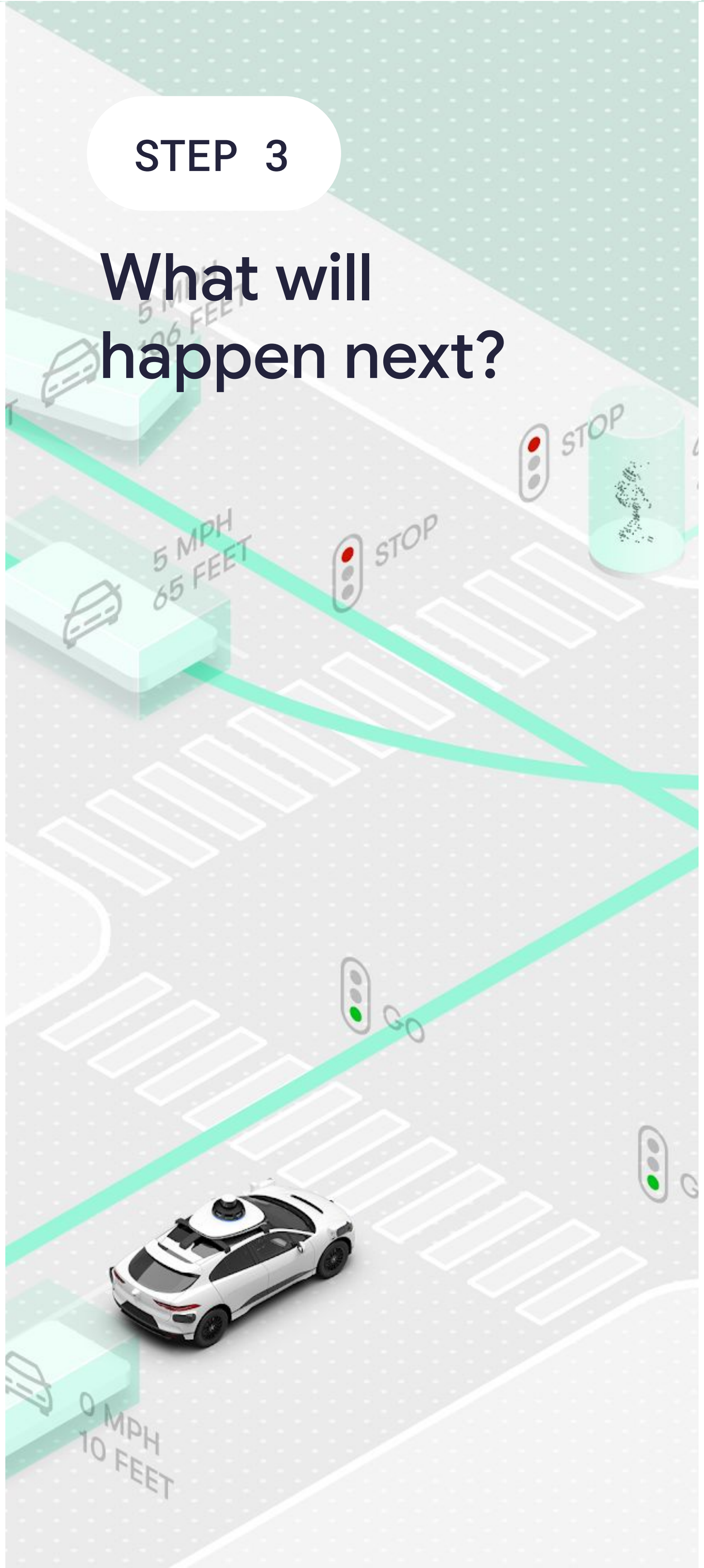
STEP 1

Where am I?



STEP 2

What's around me?

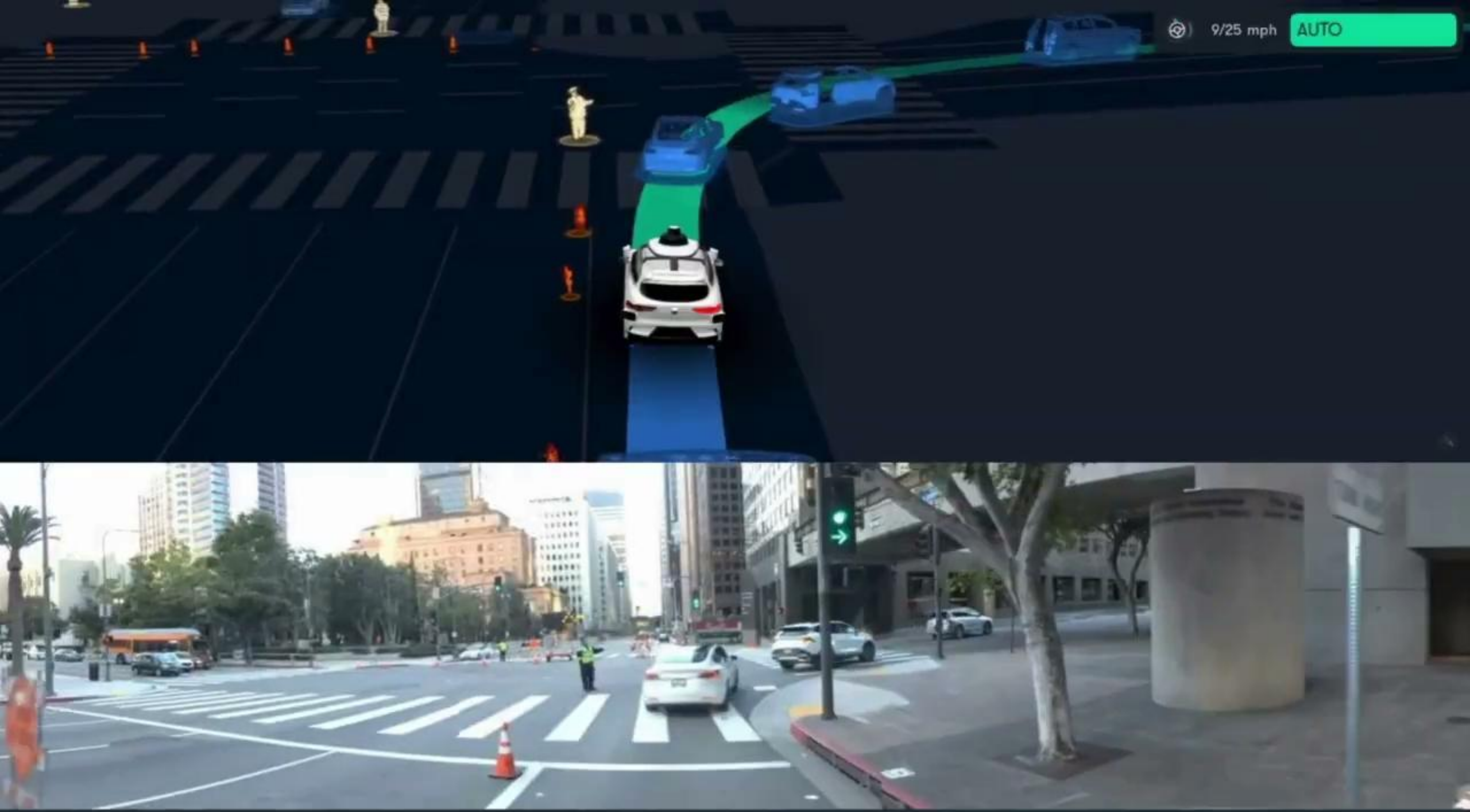


STEP 3

What will happen next?

STEP 4

What should I do?

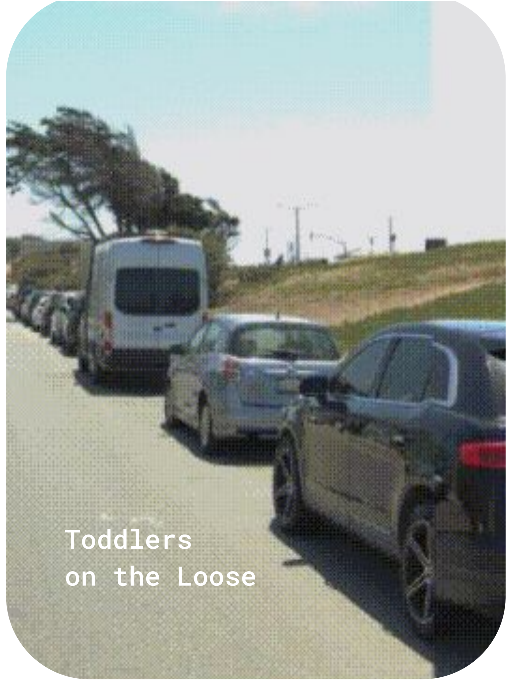


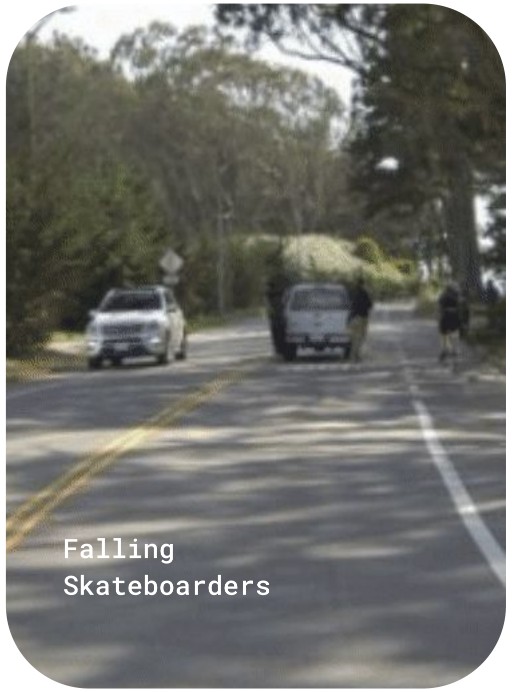
9/25 mph
AUTO

Long tail of driving scenarios presents significant challenges

Unusual Behaviors

Stray Cyclists


Toddlers on the Loose

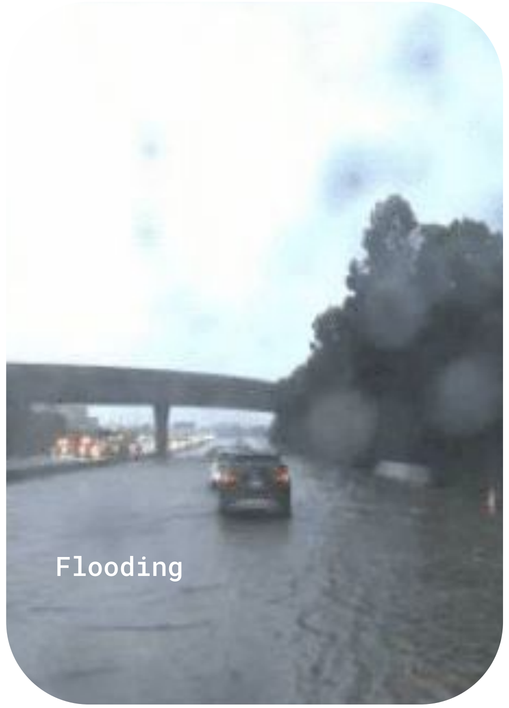

Falling Skateboarders

Foreign Objects on the Road

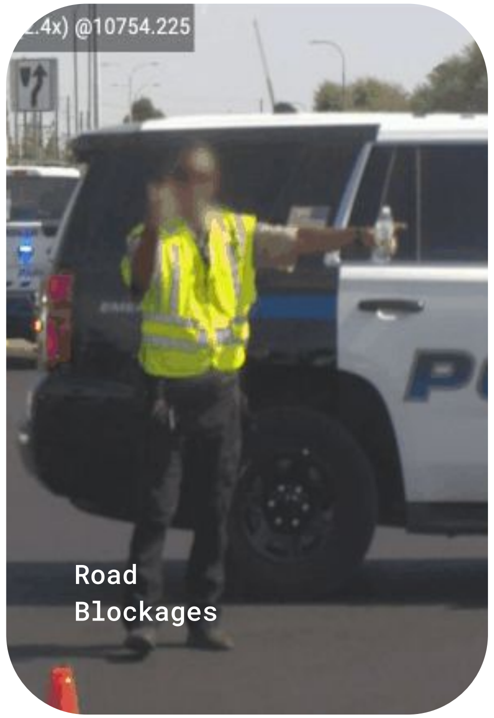
A BBQ Grill Falling Off Truck

Extreme Weather

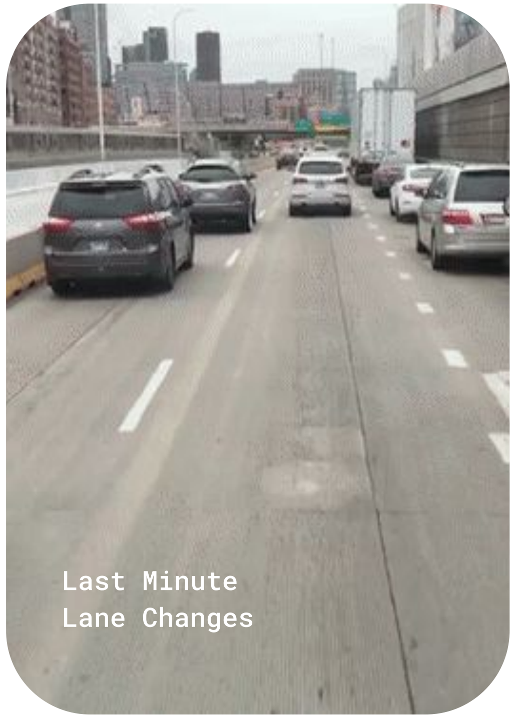
Falling Trees


Flooding

Unique Interactions



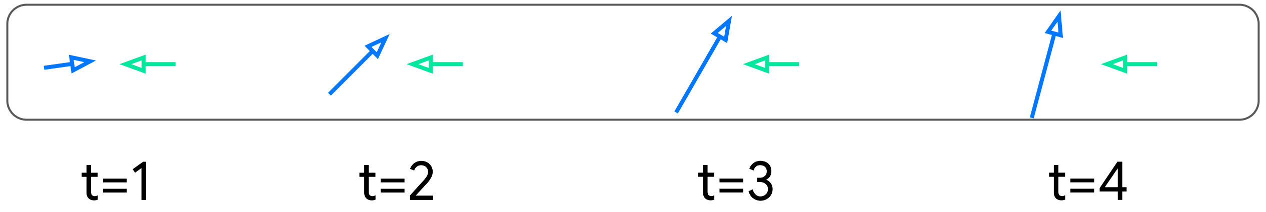
Road Blockages

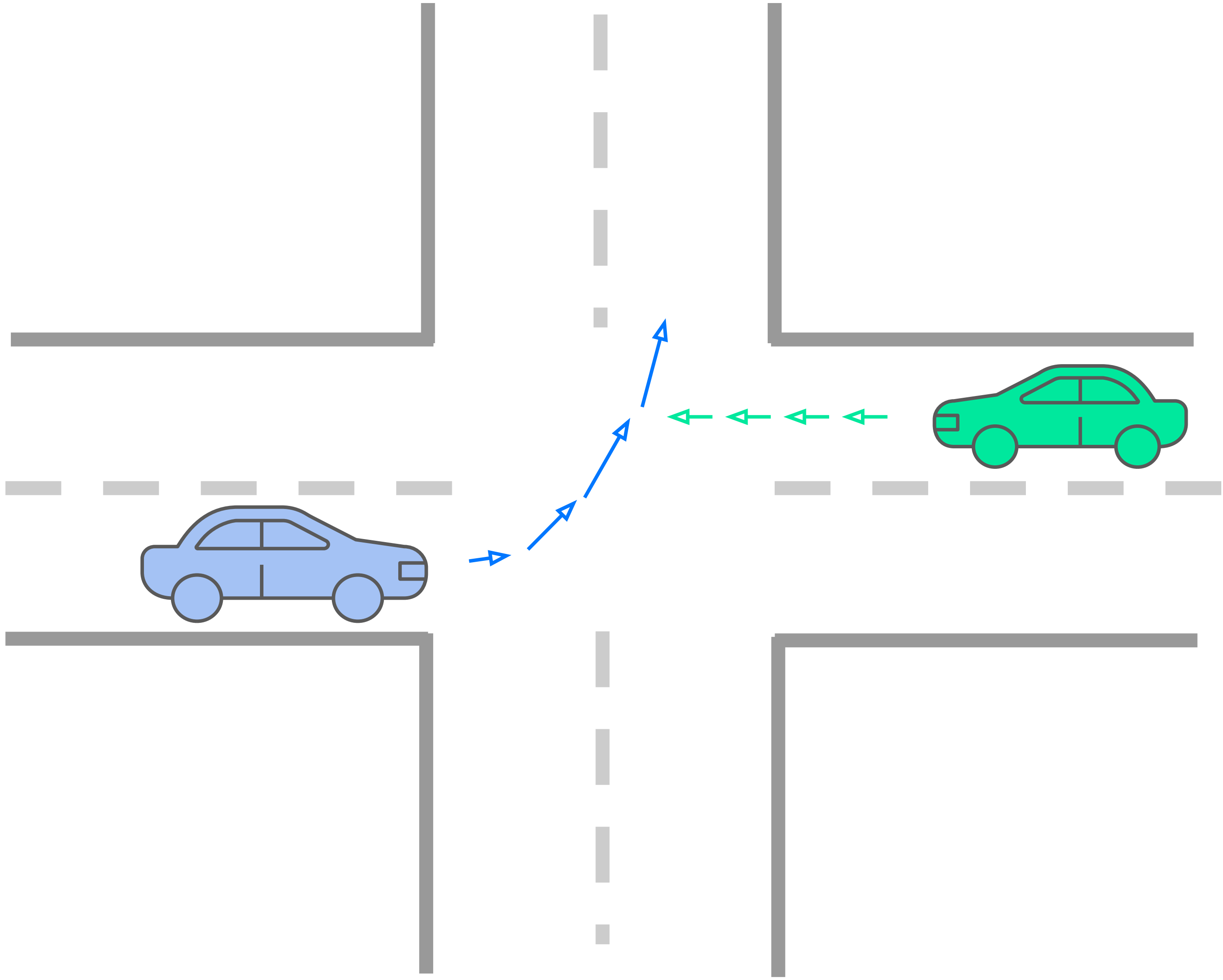

Last Minute Lane Changes

Driving as a Conversation

- Trajectories as sentences in a **new language**
- Vocabulary consists of **state / motion words** (vectors)
- Like language, trajectories have **local continuity and global context**
- Model architecture is **similar to** that of a **LLM**

Motion token sequence:





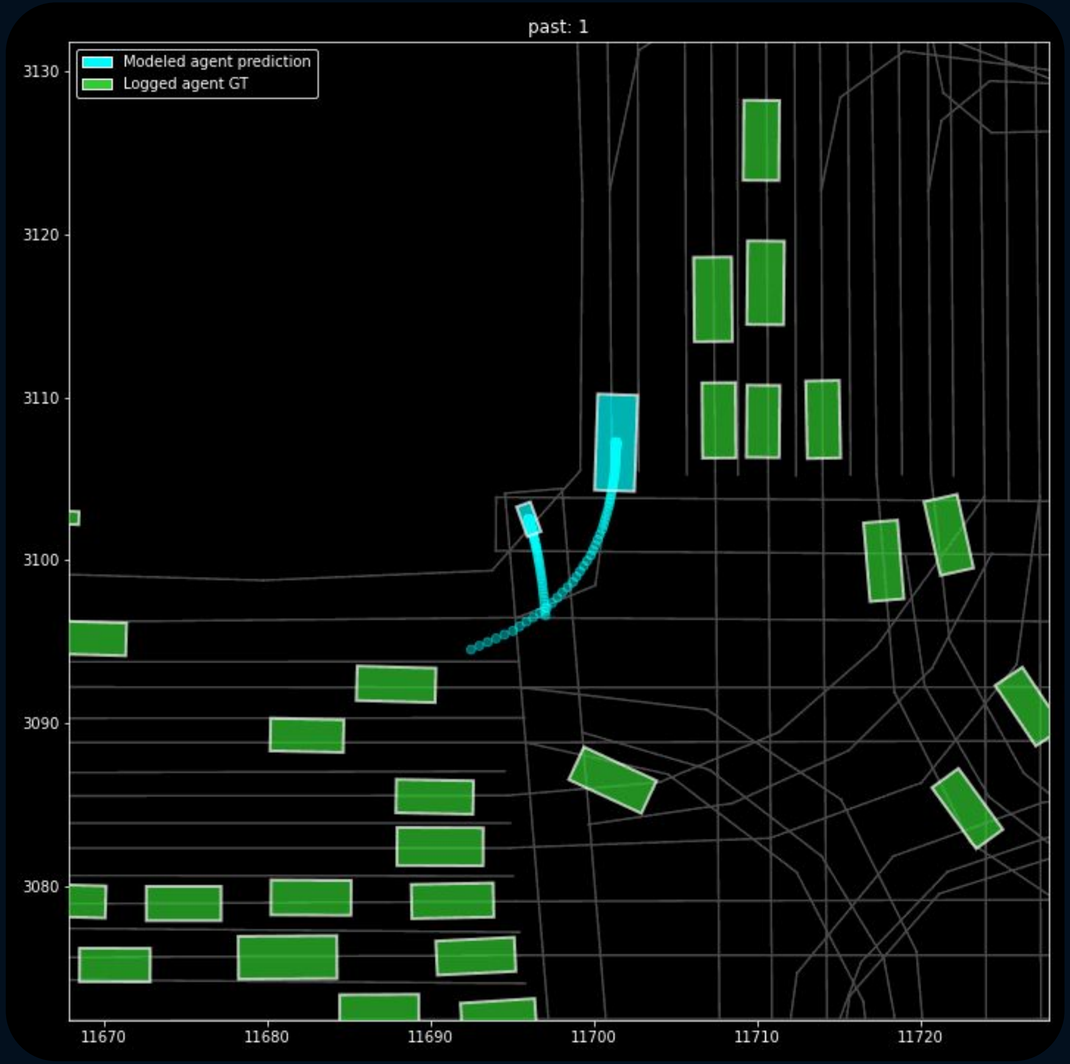
MotionLM: Multi-Agent Motion Forecasting as Language Modeling

Ari Seff, Brian Cera, Dian Chen, Mason Ng, Aurick Zhou, Nigamaa Nayakanti, Khaled S. Refaat, Rami Al-Rfou, Benjamin Sapp

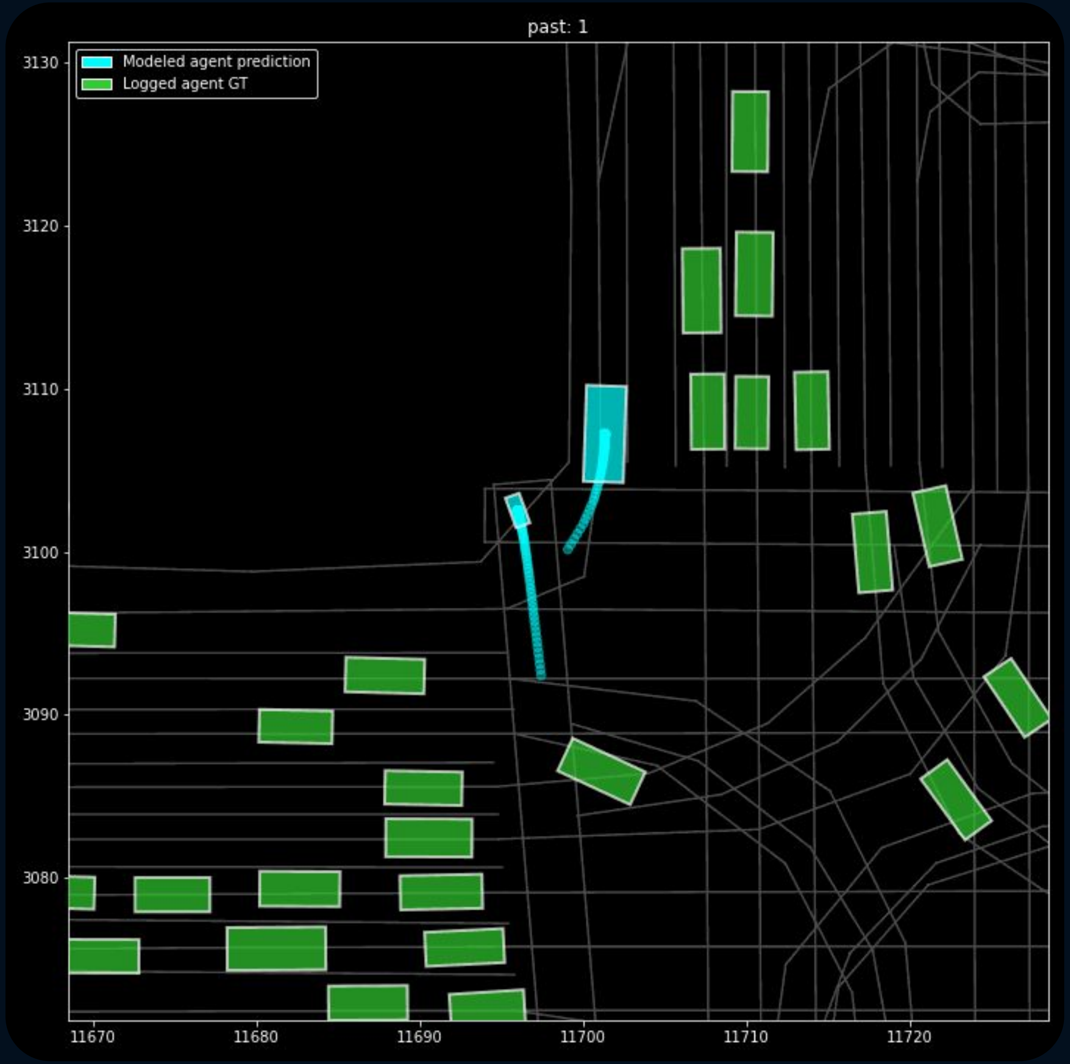
ICCV 2023

Examples

MARGINAL PREDICTION

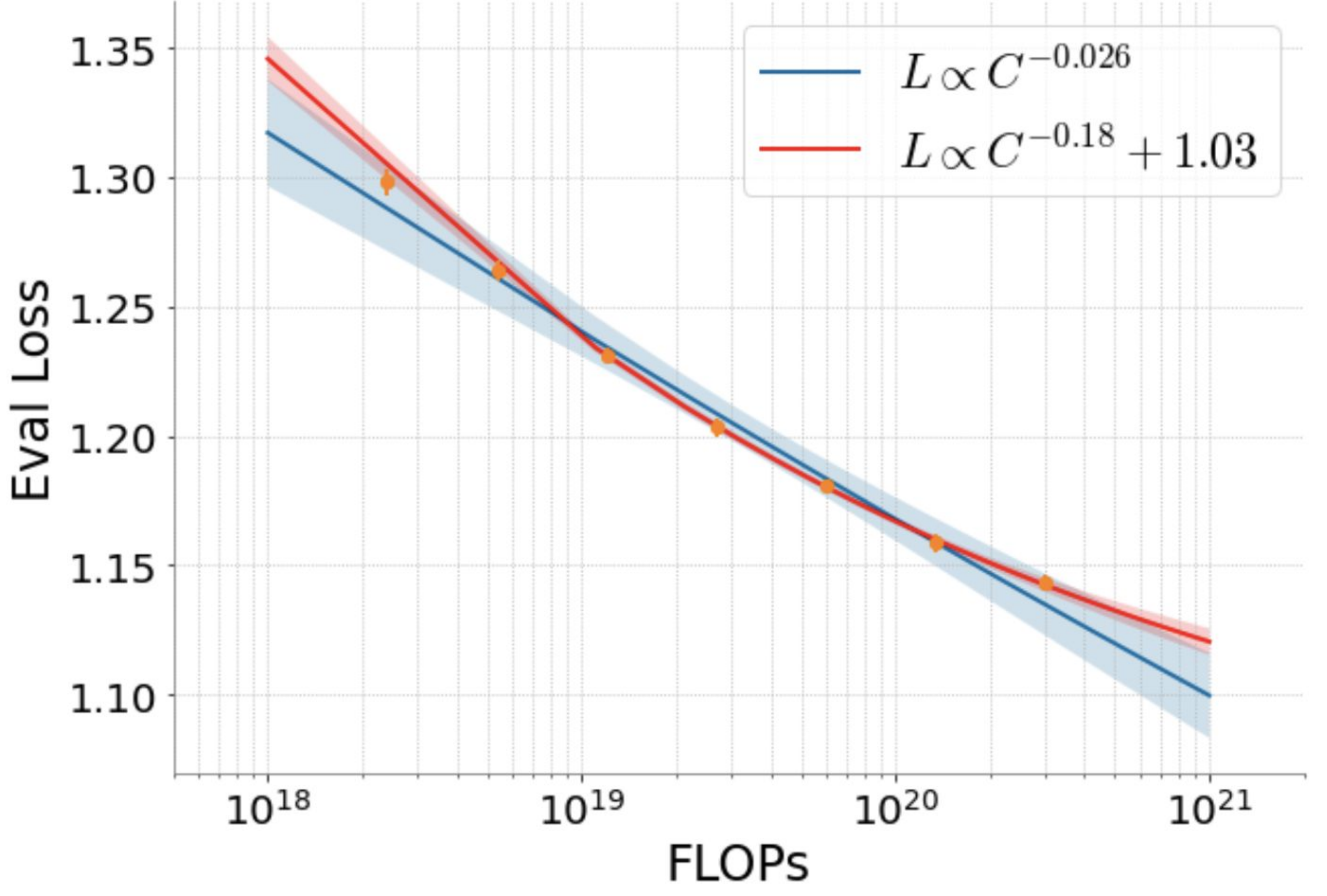


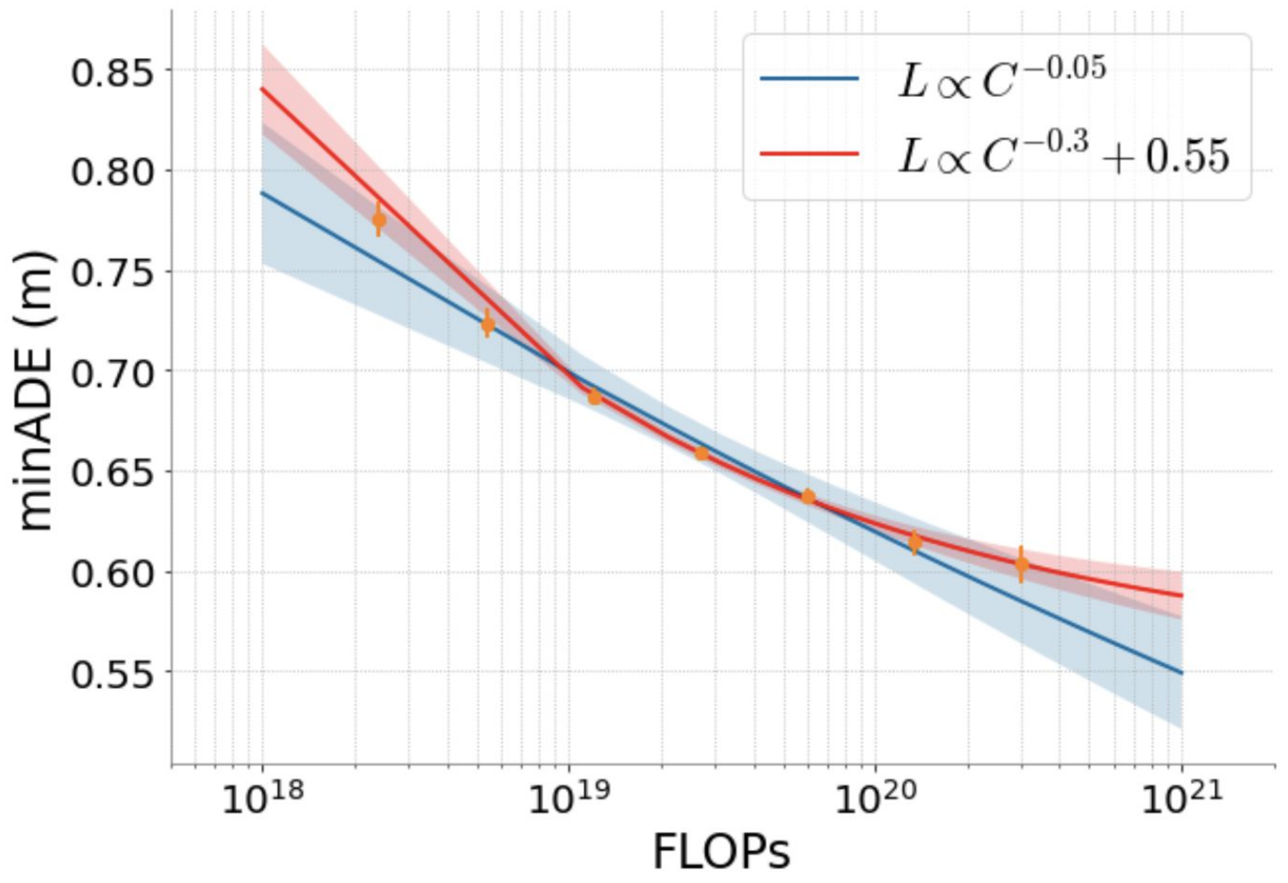
CAUSAL JOINT “DIALOGUE”



Scaling Laws

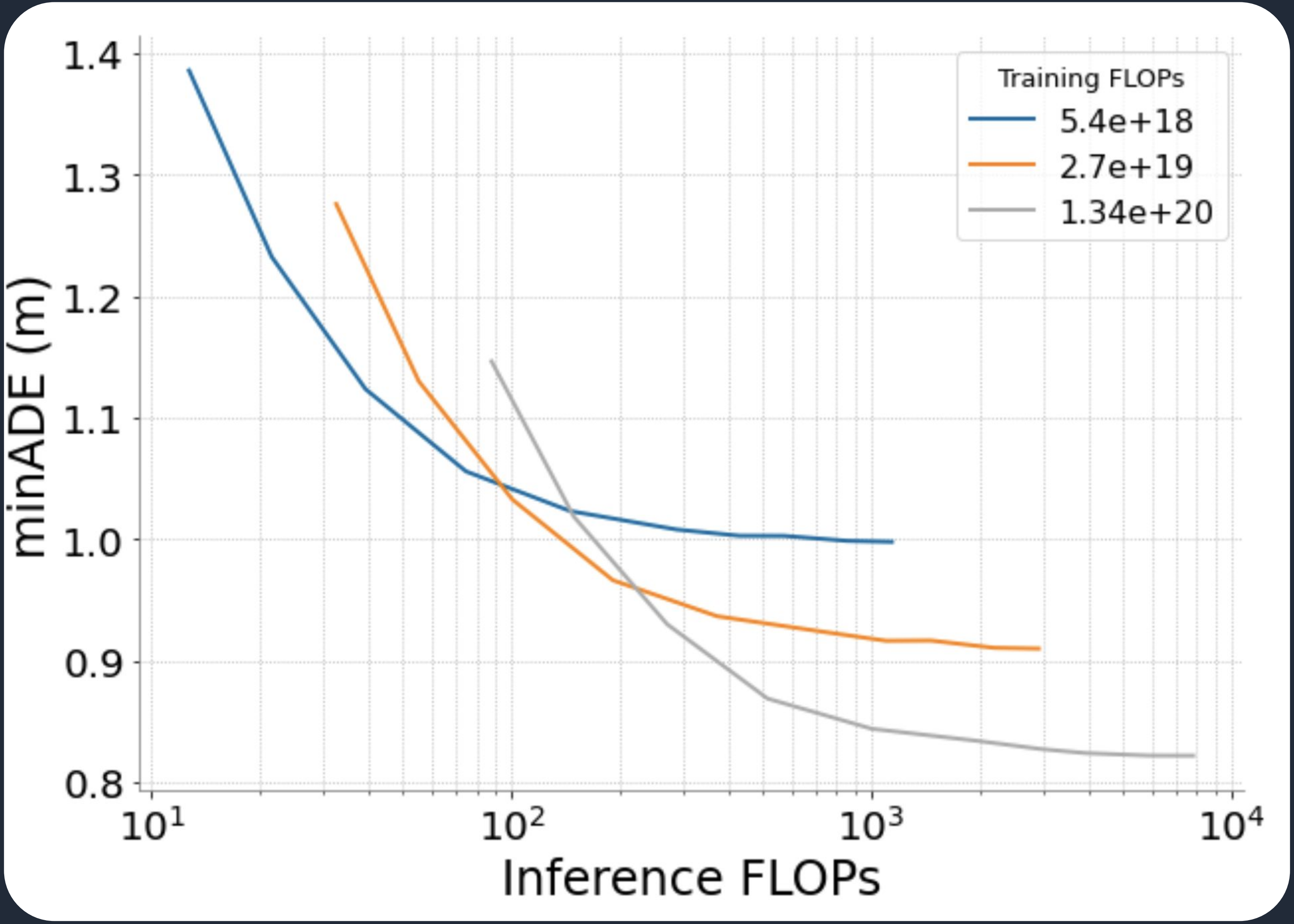
Once you have a good architecture, performance scales with model size and data





Inference Scaling Laws

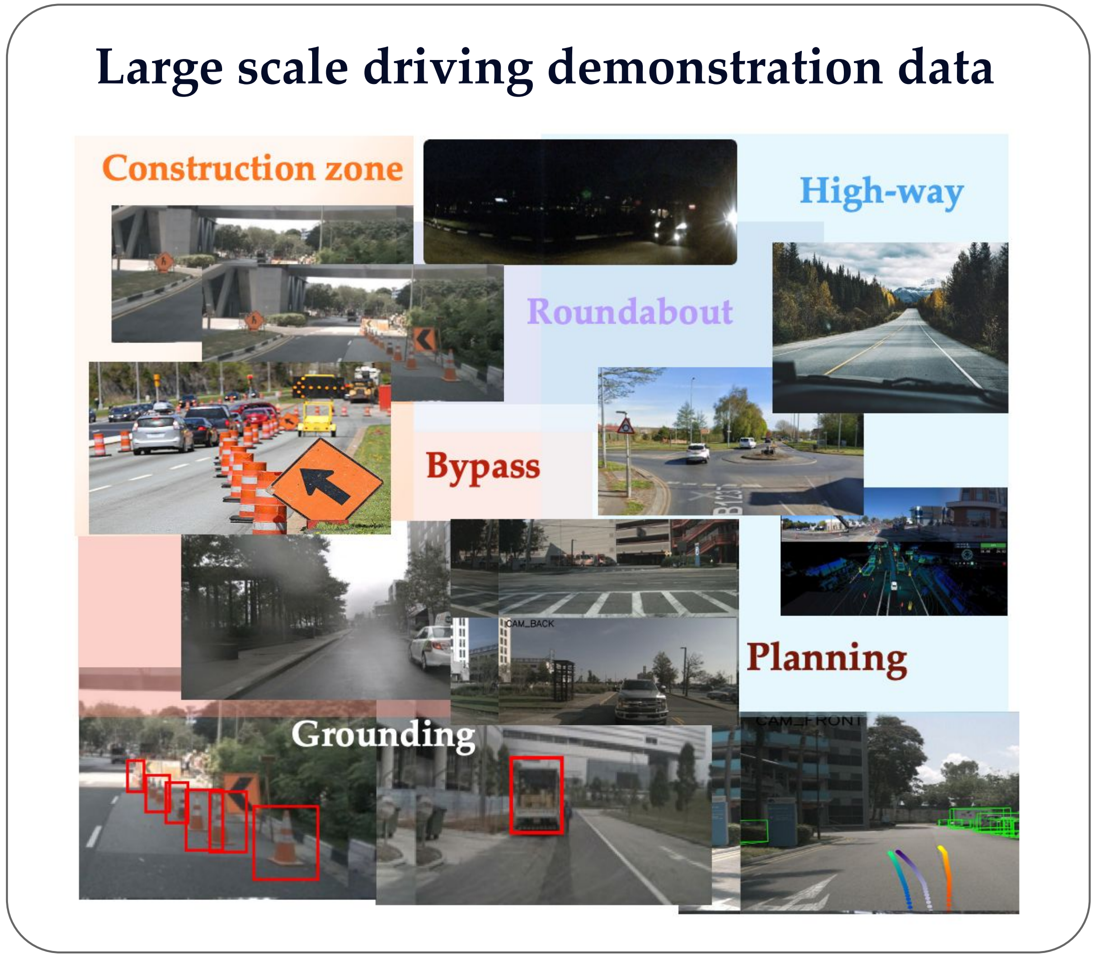
Performance scales with the amount of compute used at runtime



For more on motion scaling laws, don't miss **Ben Sapp**'s talk tomorrow at the Workshop on Distillation of Foundation Models for Autonomous Driving!

(NEW!) Now on ArXiv

Post-training Preference Alignment


Large scale driving demonstration data

Let's remember what the expert did and copy them!

$$\max_{\theta} \mathbb{E}_{(\xi,c)\sim\mathcal{D}} \left[\Pi_{t=0}^{T} \mathbb{P}_{\theta}(\boldsymbol{a}_t^* | \boldsymbol{a}_{<t}^*; c)\right]$$

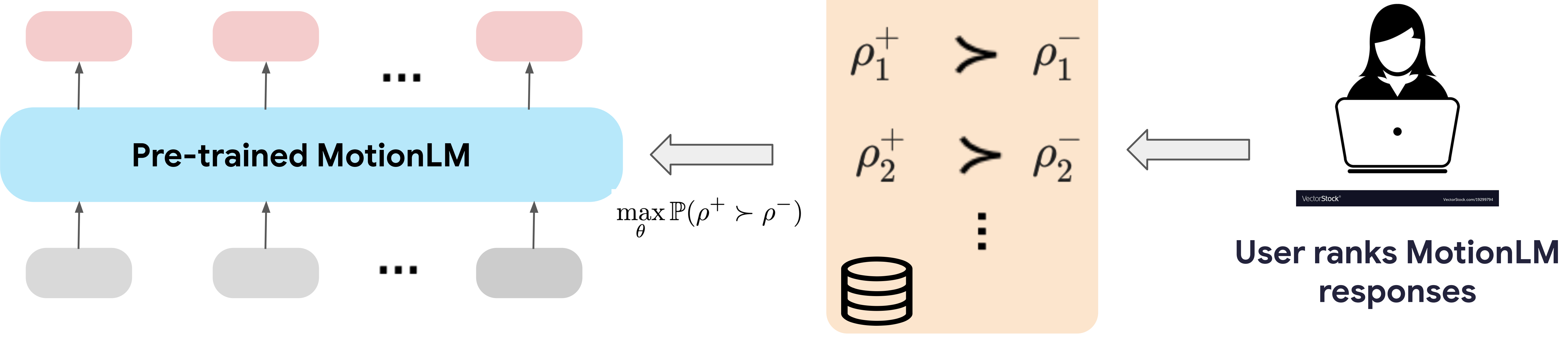
Misalignment: by optimizing an incomplete or mis-specified objective, these models lead to **undesirable** behaviors at best and **safety hazards** at worst!

Direct Post-Training Preference Alignment for Multi-Agent Motion Generation Model Using Implicit Feedback from Pre-training Demonstrations

Thomas (Ran) Tian, Kratarth Goel

ICLR 2025, Spotlight

Post-training Preference Alignment



Pre-trained MotionLM

$$\max_{\theta} \mathbb{P}(\rho^+ \succ \rho^-)$$

$$\rho_1^+ \succ \rho_1^-$$

$$\rho_2^+ \succ \rho_2^-$$

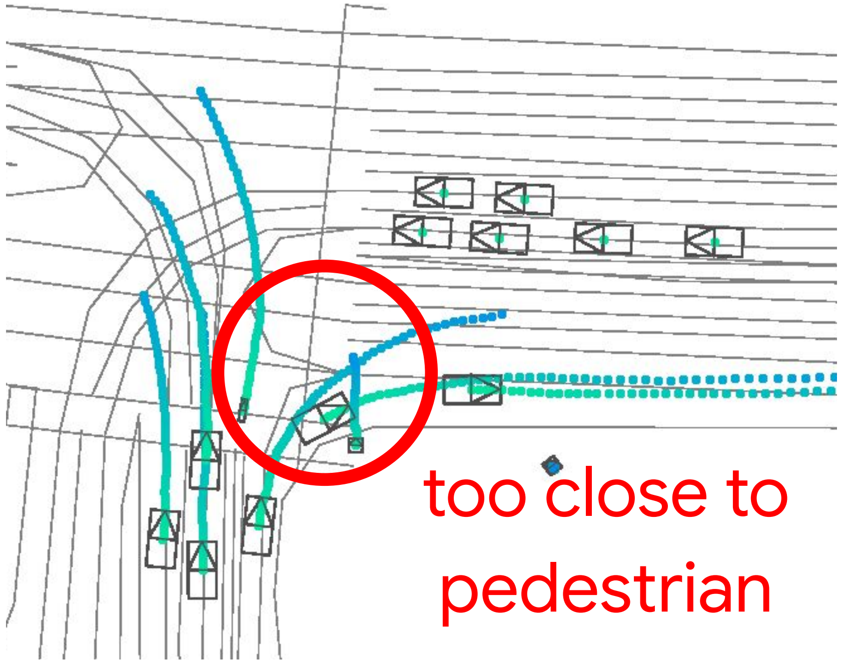
$$\vdots$$

VectorStock® VectorStock.com/19299794

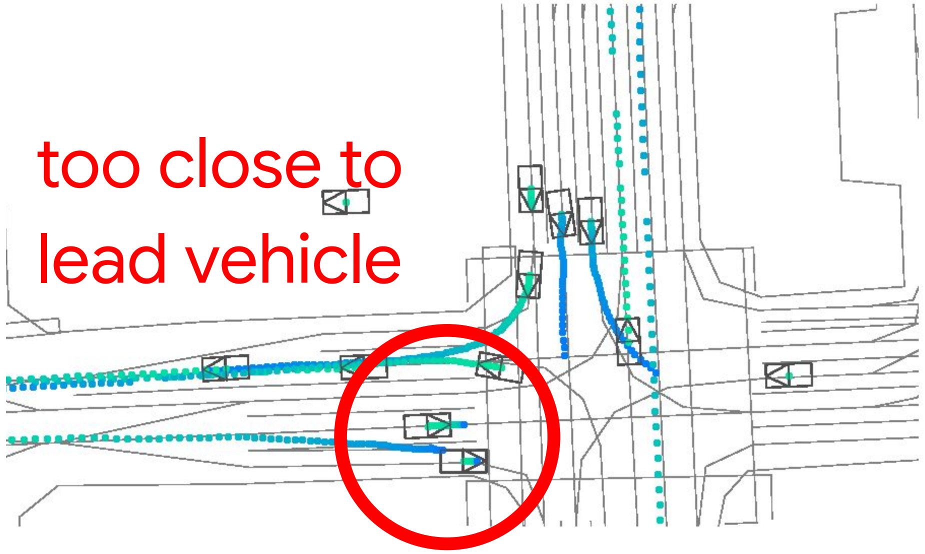
User ranks MotionLM responses

Reconcile the disparity between the next-token prediction objective and human preferences.

Pre-trained traffic simulation model

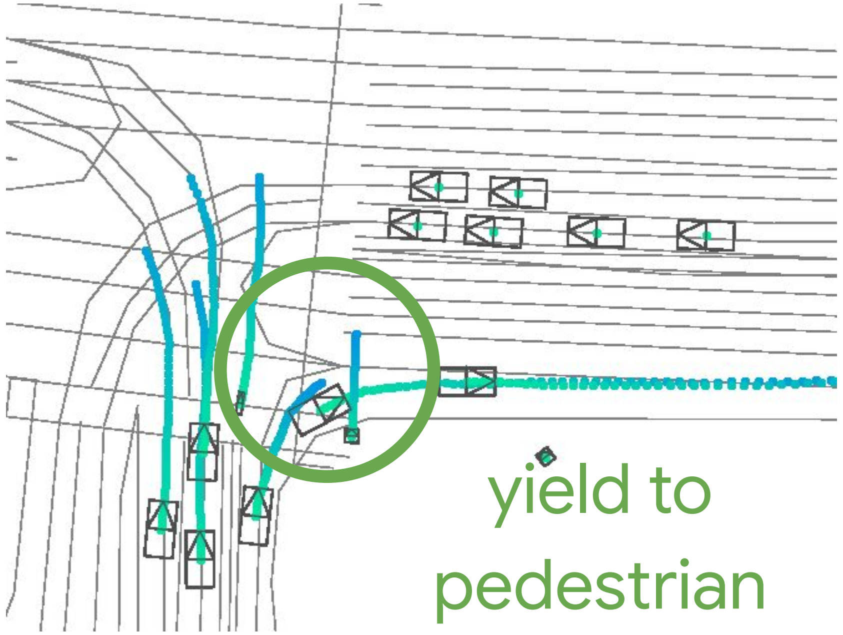


too close to pedestrian

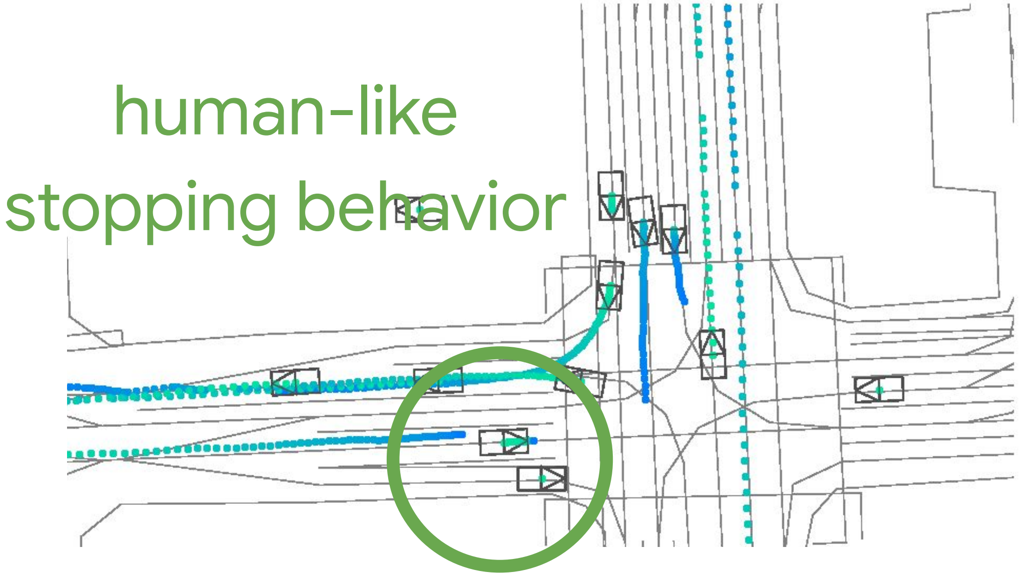


too close to lead vehicle

After post-training alignment



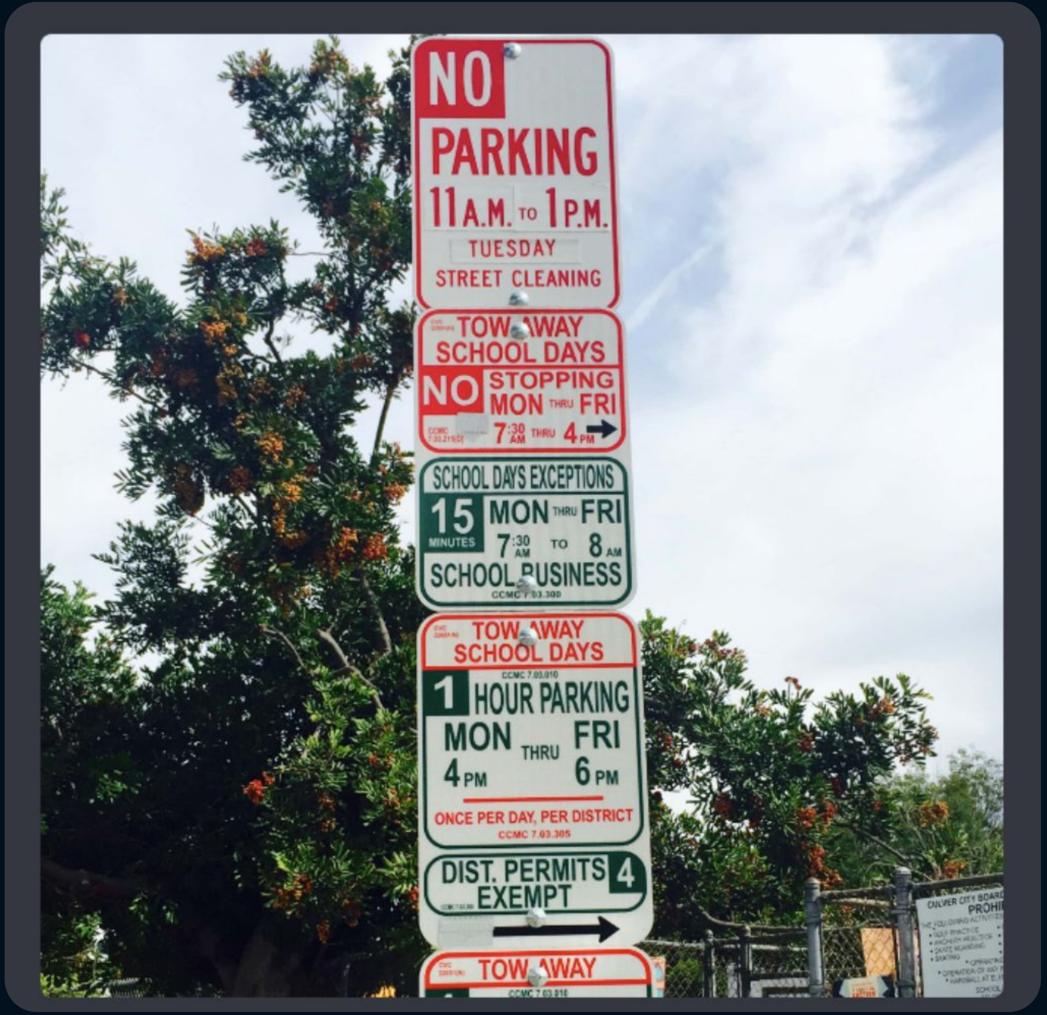
yield to pedestrian



human-like stopping behavior

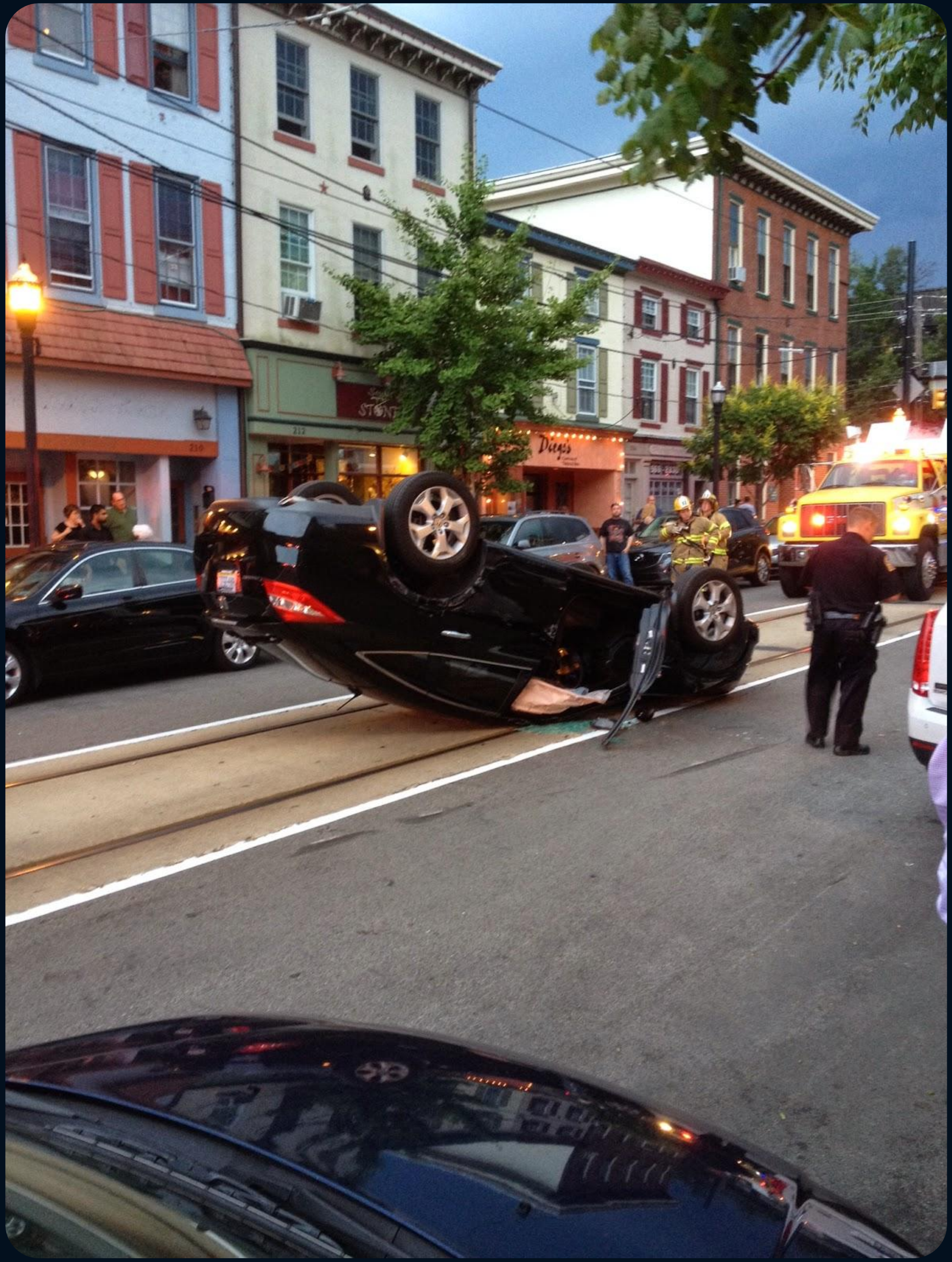
Waymo | Confidential & Proprietary

Leveraging language understanding and general reasoning capabilities of LLM/VLMs



It's Wednesday at 4pm. Can I park at this spot right now? Tell me in 1 line.

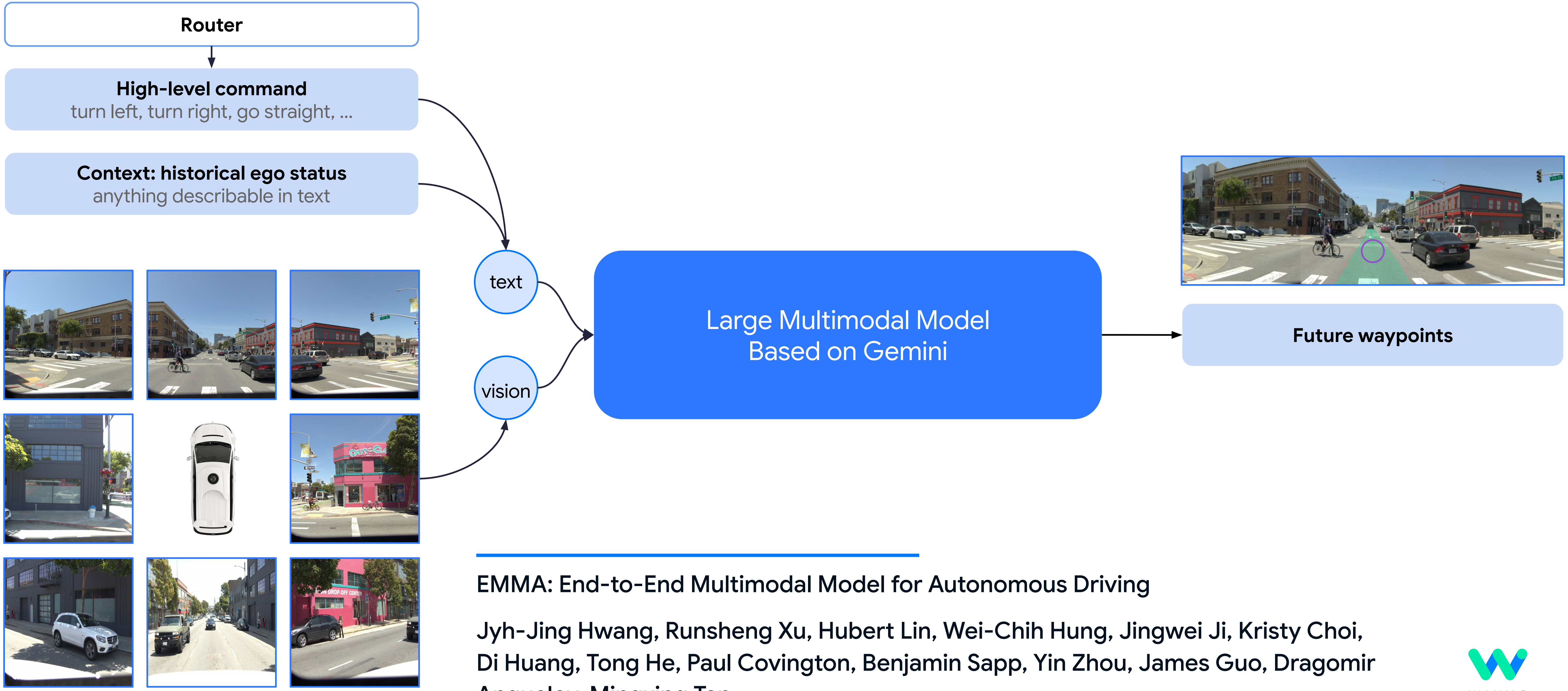
VLM: Yes, you can park for up to 1 hour starting at 4pm.



I am driving on the street, this is what I see from the front window. Tell me what's happening and what I should pay special attention to. And tell me what I should do and specify the reasons.

VLM: There's been a serious accident! A car has flipped onto its roof and is blocking the streetcar tracks. Emergency responders are on the scene. Pay close attention to the emergency personnel and any instructions they may give. You should slow down significantly, proceed with extreme caution, and be prepared to stop completely if necessary. The most important thing is to avoid hindering the emergency response and to ensure your own safety and the safety of others around you. If possible, consider an alternate route to avoid the area entirely.

Multimodal Models for AV motion prediction



Router

High-level command
turn left, turn right, go straight, ...

Context: historical ego status
anything describable in text

text

vision

Large Multimodal Model
Based on Gemini

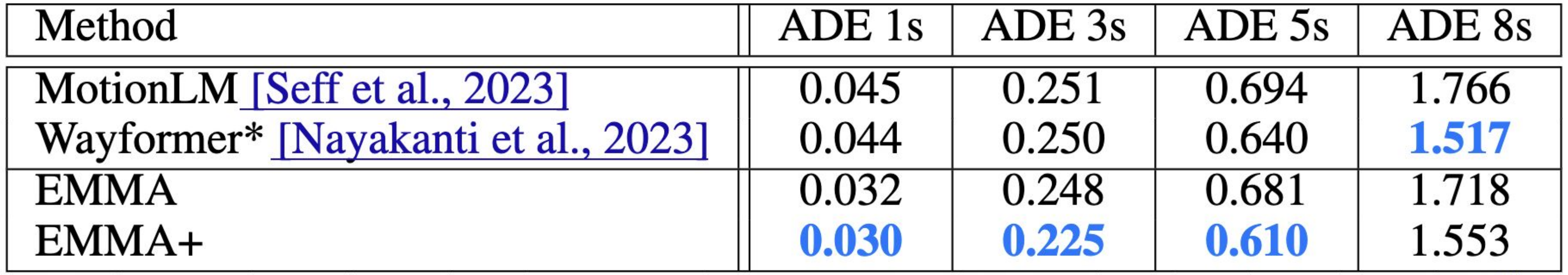
Future waypoints

EMMA: End-to-End Multimodal Model for Autonomous Driving

Jyh-Jing Hwang, Runsheng Xu, Hubert Lin, Wei-Chih Hung, Jingwei Ji, Kristy Choi, Di Huang, Tong He, Paul Covington, Benjamin Sapp, Yin Zhou, James Guo, Dragomir Anguelov, Mingxing Tan

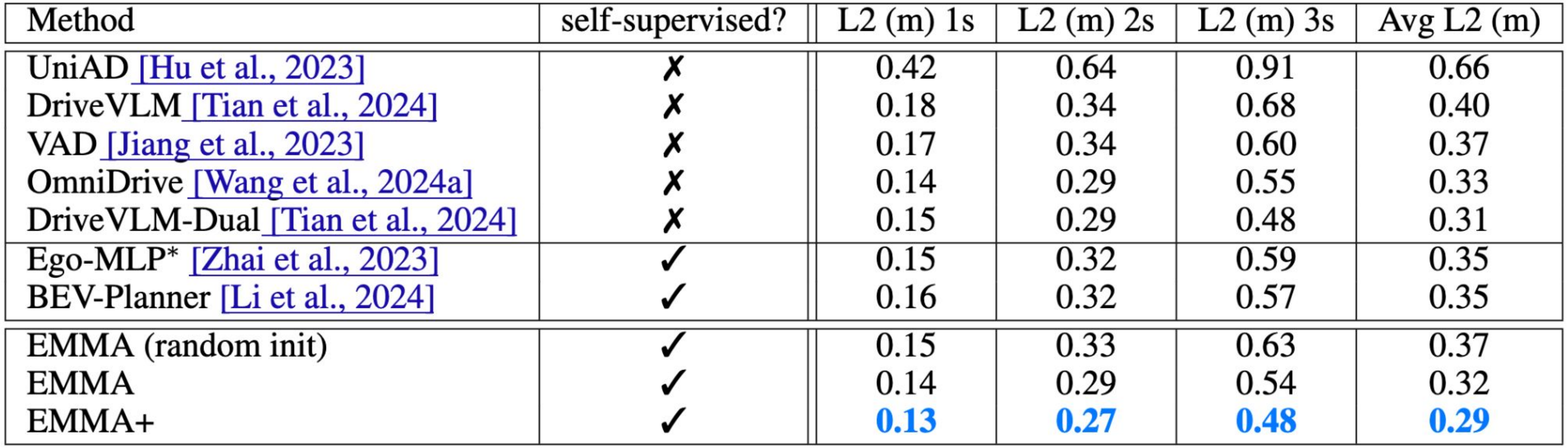
Competitive Academic Benchmark Results

WAYMO OPEN MOTION DATASET



Method	ADE 1s	ADE 3s	ADE 5s	ADE 8s
MotionLM [Seff et al., 2023]	0.045	0.251	0.694	1.766
Wayformer* [Nayakanti et al., 2023]	0.044	0.250	0.640	**1.517**
EMMA	0.032	0.248	0.681	1.718
EMMA+	**0.030**	**0.225**	**0.610**	1.553

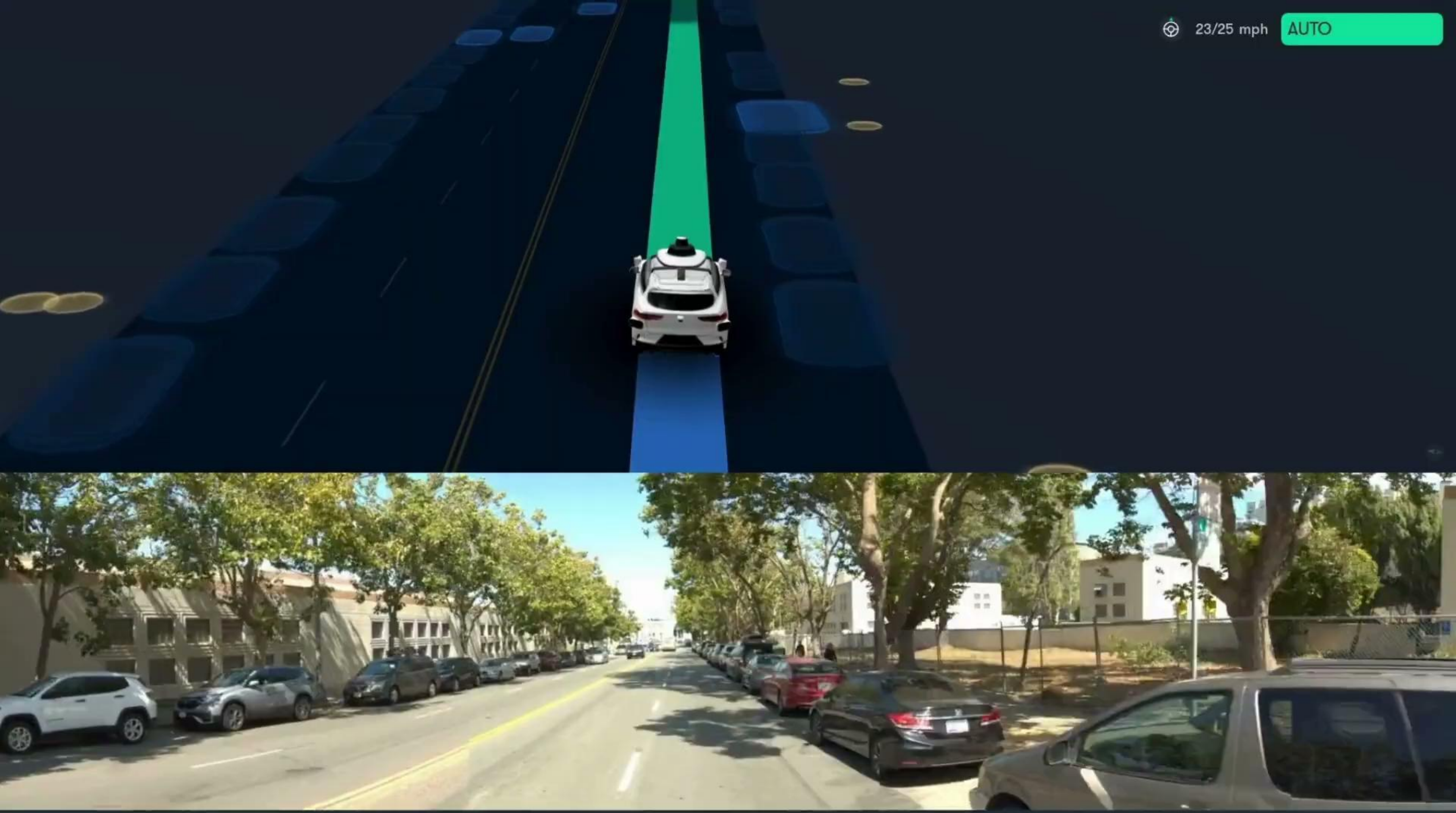
NUSCENES DATASET



Method	self-supervised?	L2 (m) 1s	L2 (m) 2s	L2 (m) 3s	Avg L2 (m)
UniAD [Hu et al., 2023]	✗	0.42	0.64	0.91	0.66
DriveVLM [Tian et al., 2024]	✗	0.18	0.34	0.68	0.40
VAD [Jiang et al., 2023]	✗	0.17	0.34	0.60	0.37
OmniDrive [Wang et al., 2024a]	✗	0.14	0.29	0.55	0.33
DriveVLM-Dual [Tian et al., 2024]	✗	0.15	0.29	0.48	0.31
Ego-MLP* [Zhai et al., 2023]	✓	0.15	0.32	0.59	0.35
BEV-Planner [Li et al., 2024]	✓	0.16	0.32	0.57	0.35
EMMA (random init)	✓	0.15	0.33	0.63	0.37
EMMA	✓	0.14	0.29	0.54	0.32
EMMA+	✓	**0.13**	**0.27**	**0.48**	**0.29**

Foundation models open up new, scalable avenues for deep semantic understanding of the world and of human behavior.

Ultimately, these advances are in service of building a product that can delight riders and earn their trust every day.


23/25 mph
AUTO

If you're interested in building **foundation models** for robots operating in the real world **today**.

Join us on this ride!

vanhoucke@waymo.com

CVPR 2025

Come chat with me at Booth 1223!

June 11-15, 2025

CVPR Nashville JUNE 11-15, 2025