

# Foundation models

---

# For autonomous driving

---

Vincent Vanhoucke  
Distinguished Engineer  
Waymo



OUR MISSION –

# Be the world's most trusted driver.





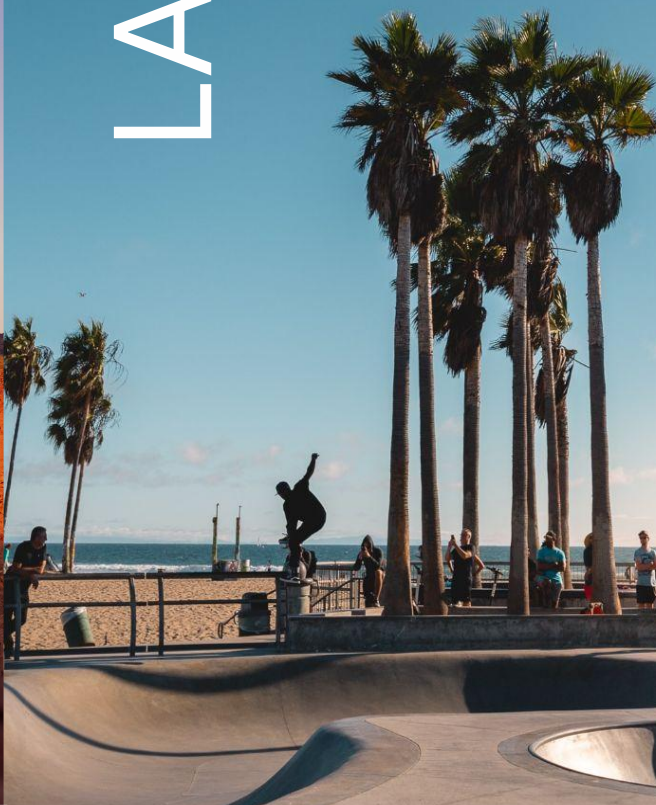
SF



PHX



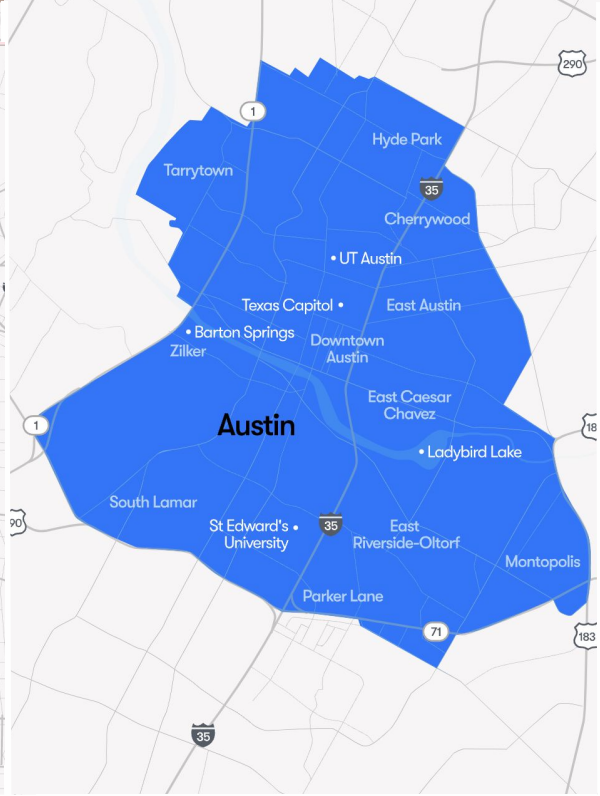
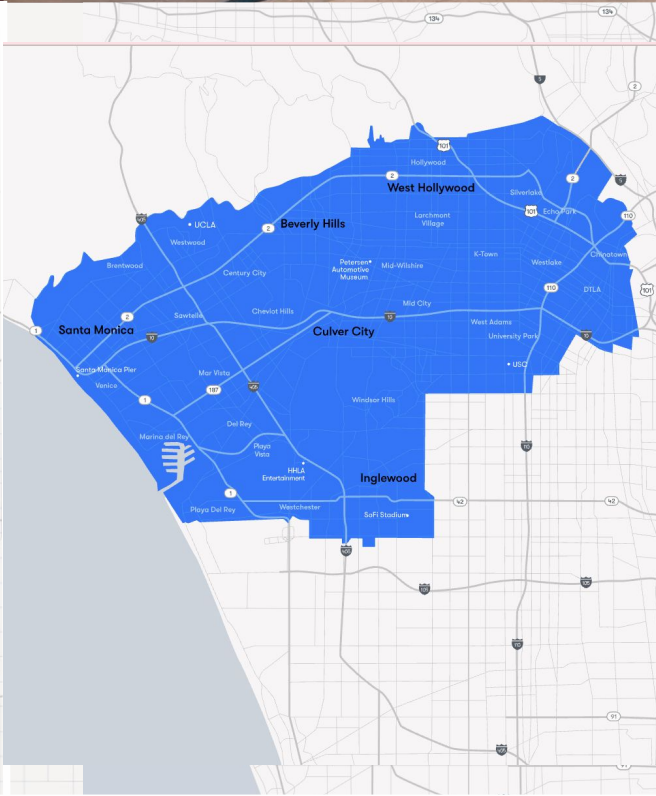
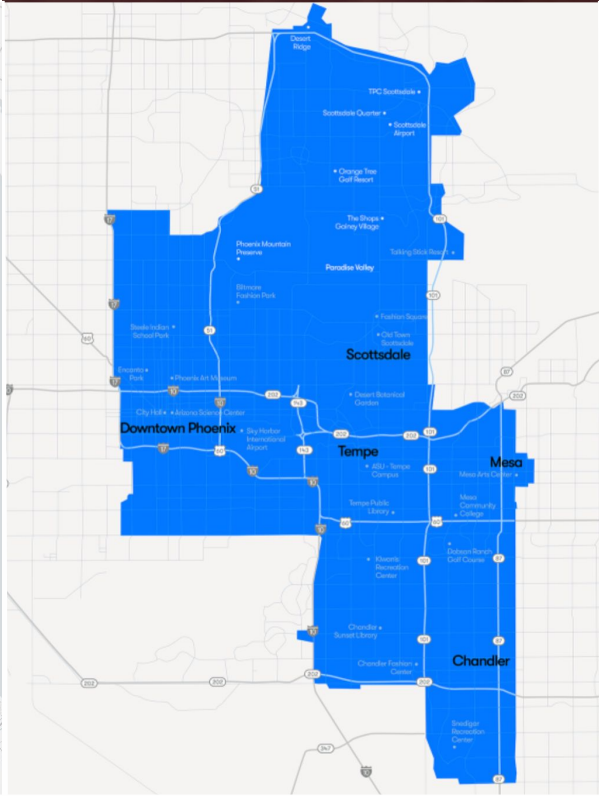
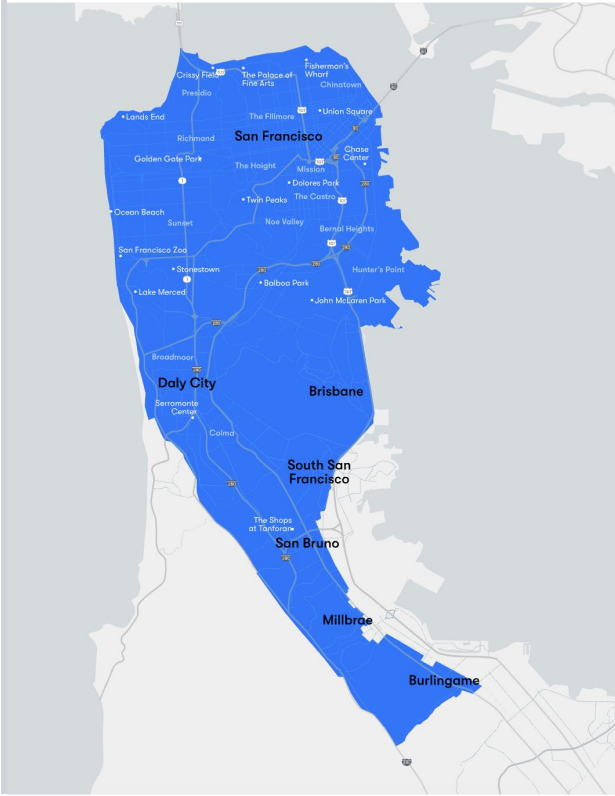
LA



ATX



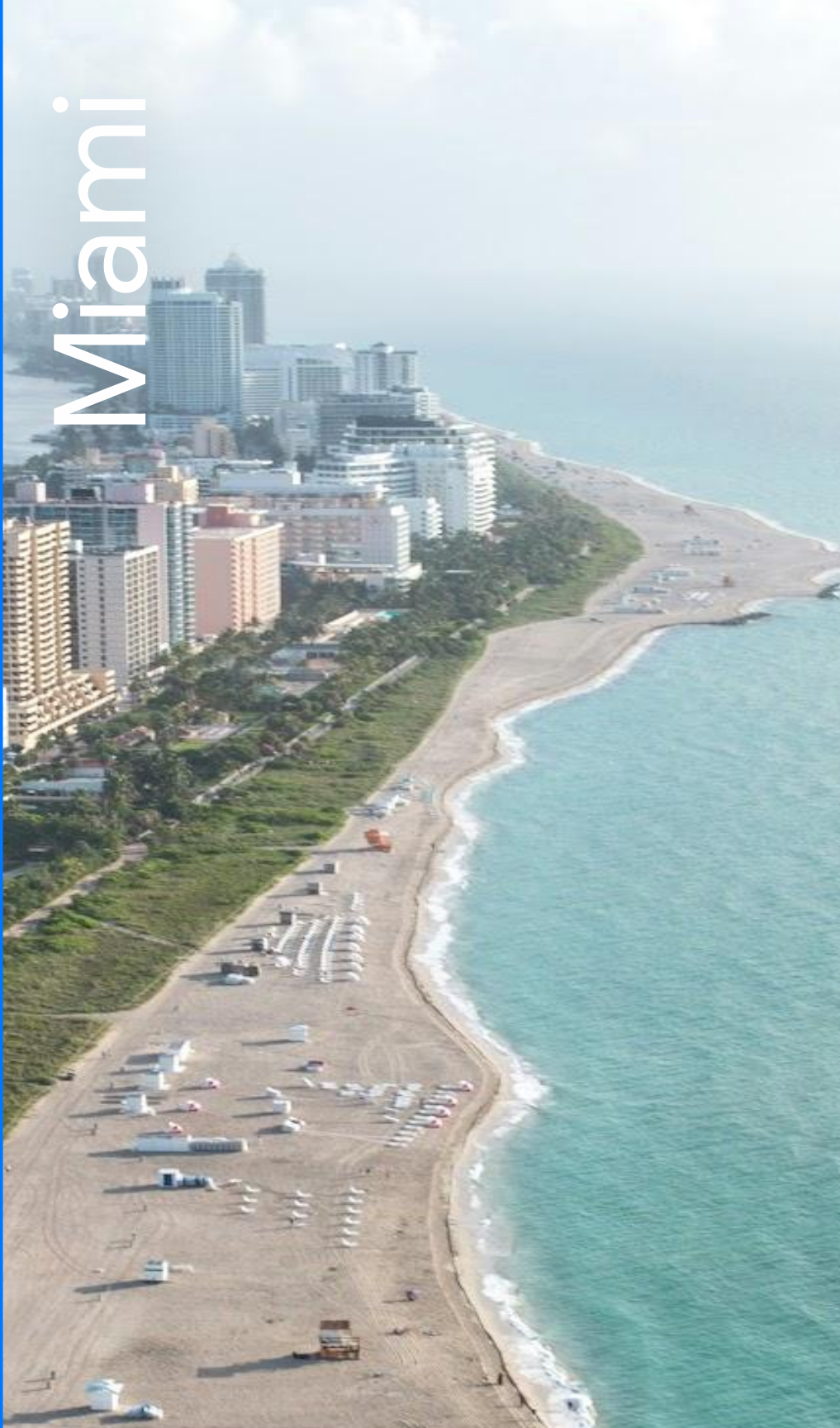
ATL





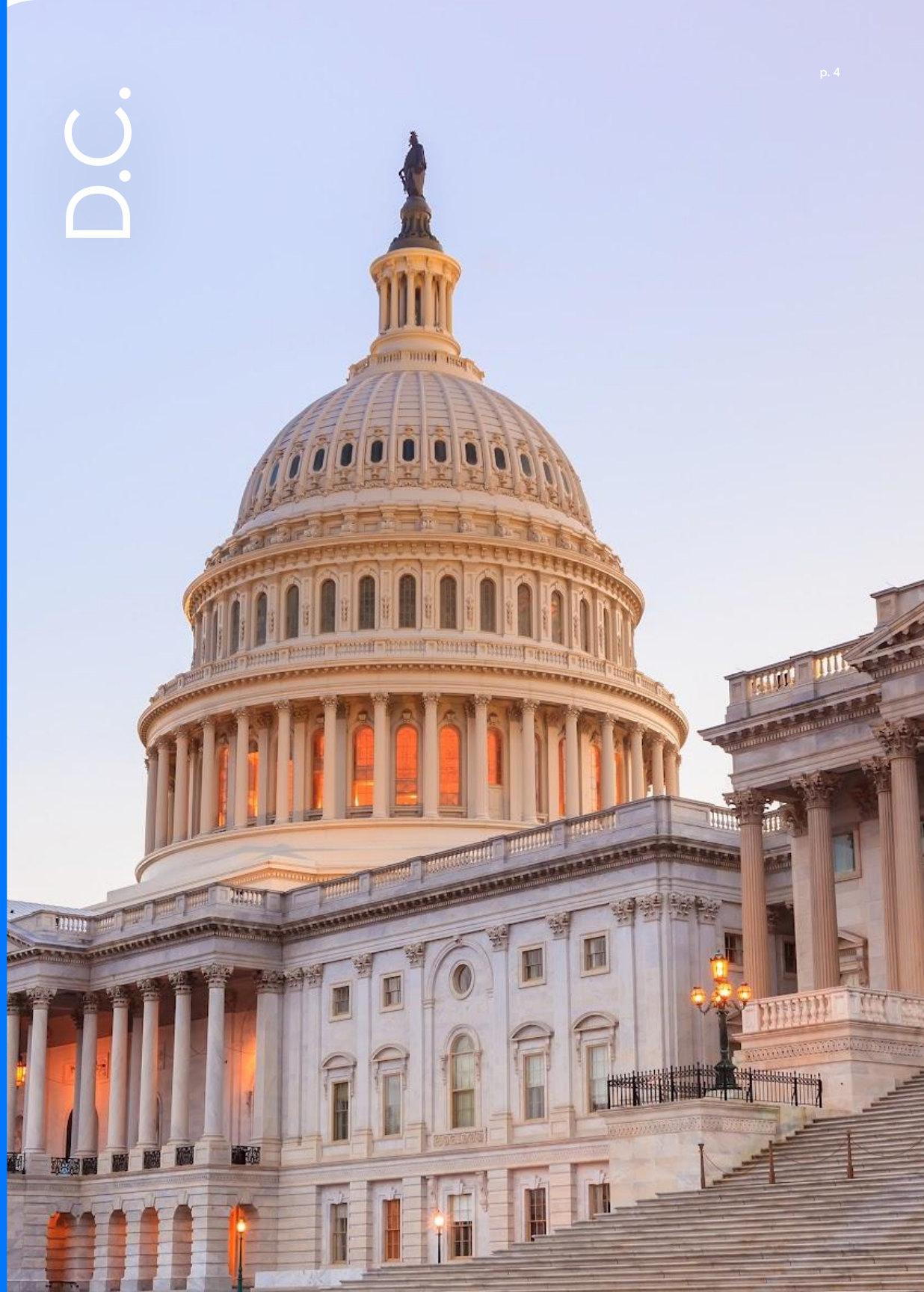
Up next

WAYMO ONE



Miami

WAYMO ONE



D.C.



✦ Waymo operating cities   ✦ Operation coming soon   ● Waymo tested cities



#### 2025 Road trip cities

- Las Vegas, NV
- San Diego, CA
- New Orleans, LA
- Nashville, TN
- Dallas, TX
- Boston, MA
- Houston, TX
- San Antonio, TX
- Orlando, FL
- Philadelphia, PA

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



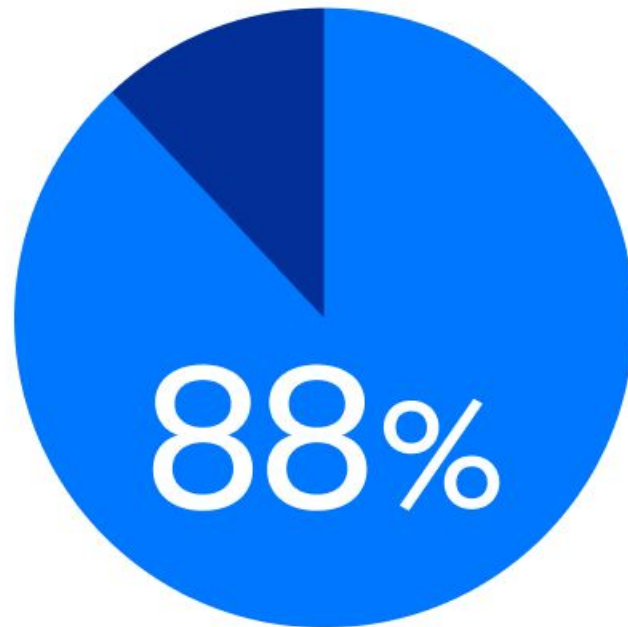
# Insurance Data: Swiss Re Analysis

Swiss Re examined liability claims and found that the Waymo Driver is significantly safer than human drivers when it comes to crash causation.

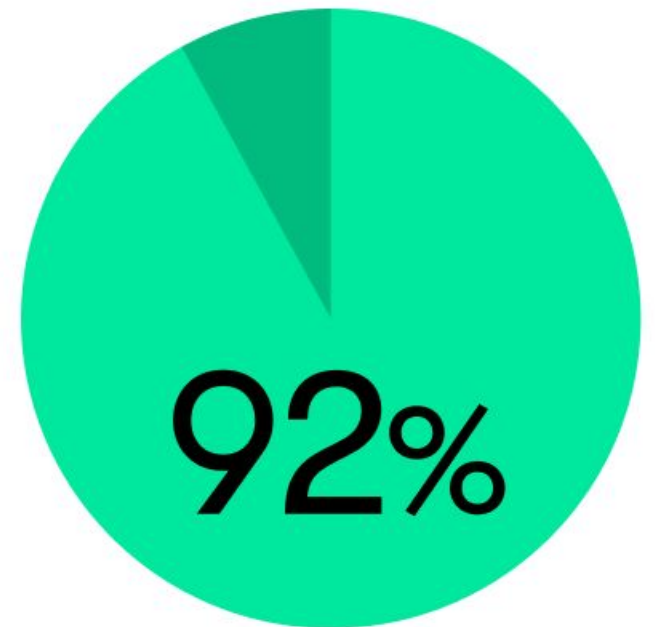
Compared to human drivers over **25.3M miles**, the Waymo Driver:

---

Reduced the frequency of  
property damage claims by

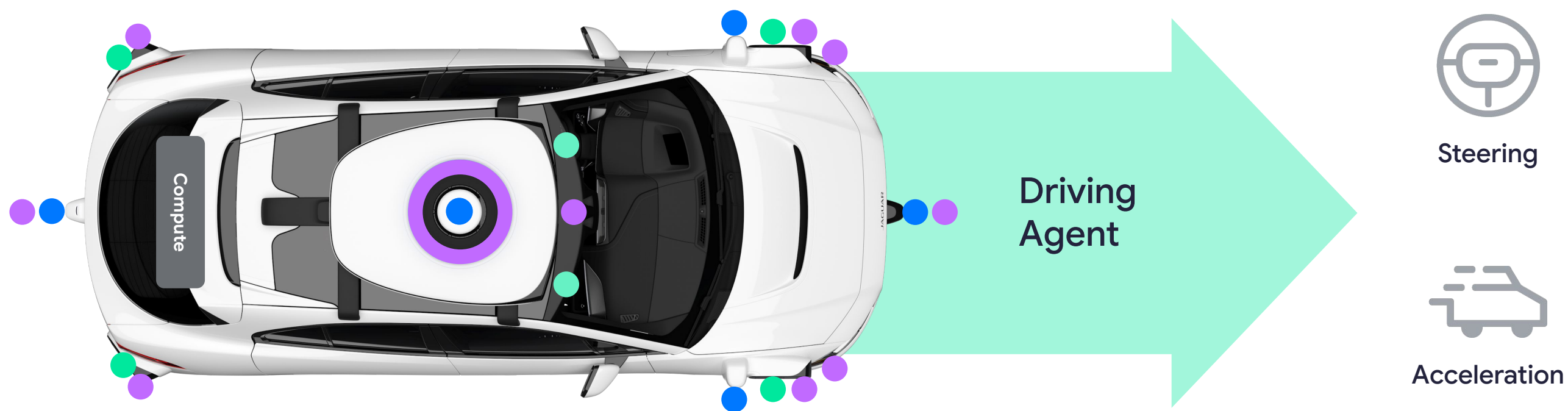


Reduced bodily injury  
claims by





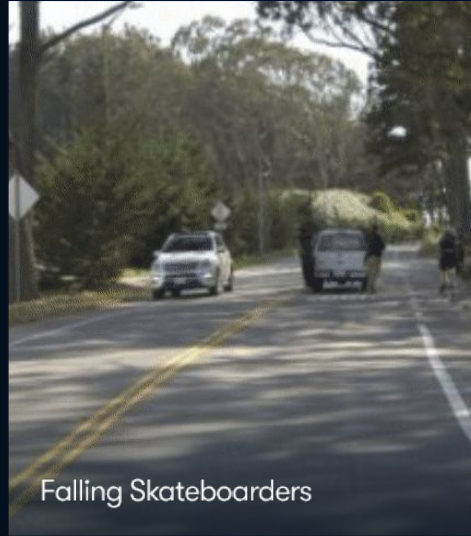
# Autonomous driving, explained



- Lidar system
- Vision system
- Radar system

# There is a long tail of complex scenarios we need to handle

## Unusual behavior



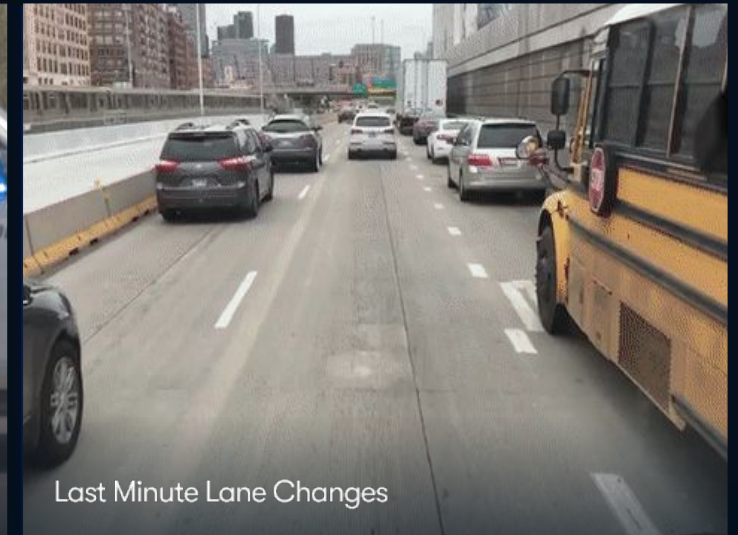
## Foreign objects on the road



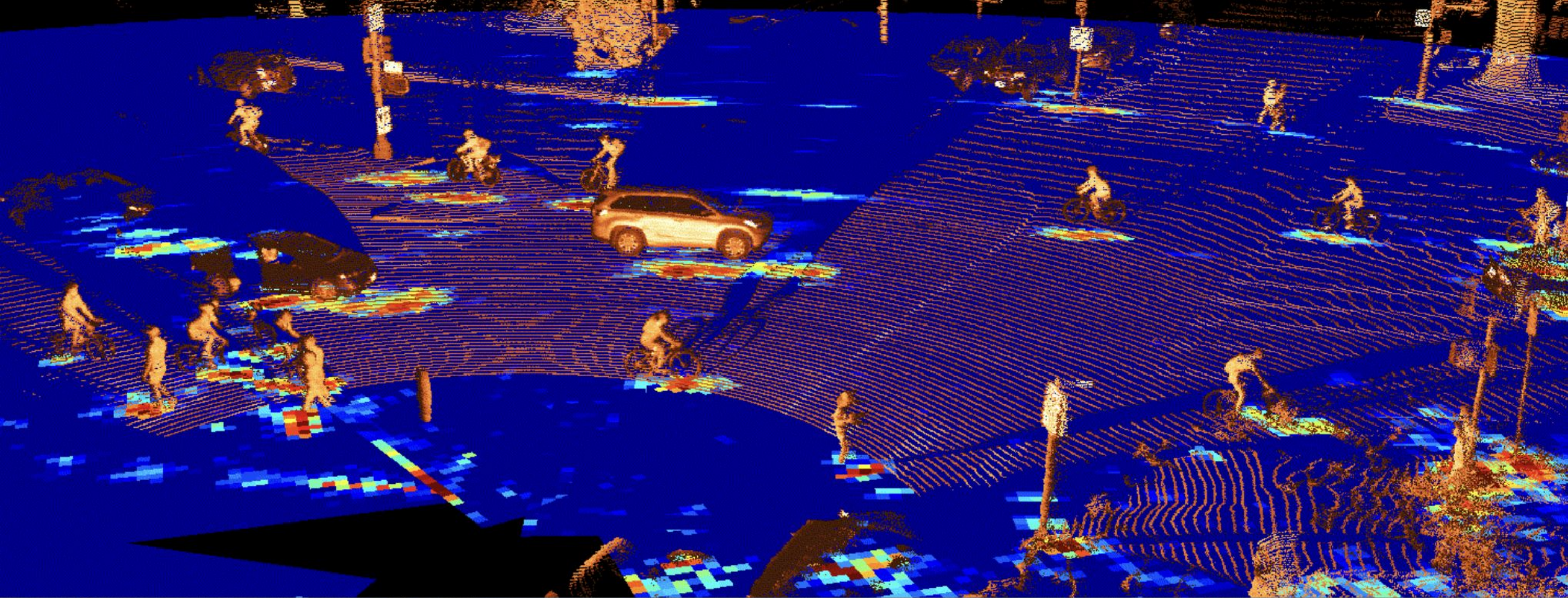
## Extreme weather



## Unique interactions









# Waymo Driver's superhuman sensing ability

HD CAMERA



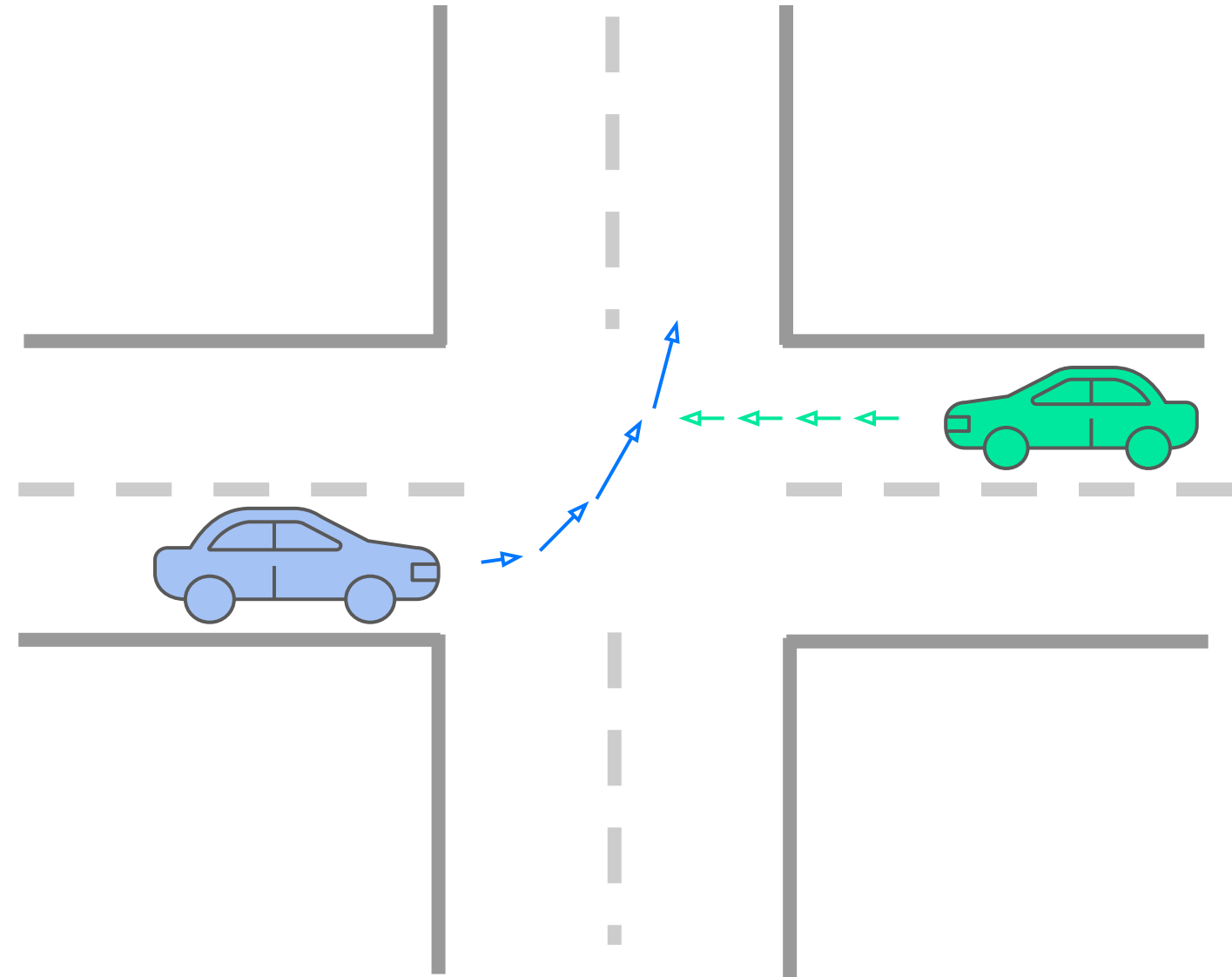
LIDAR



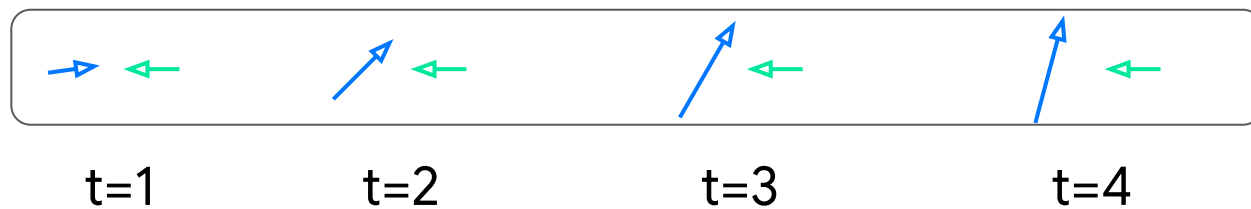


# Driving as a Conversation

- 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 **Large Language Model**



Motion token sequence:



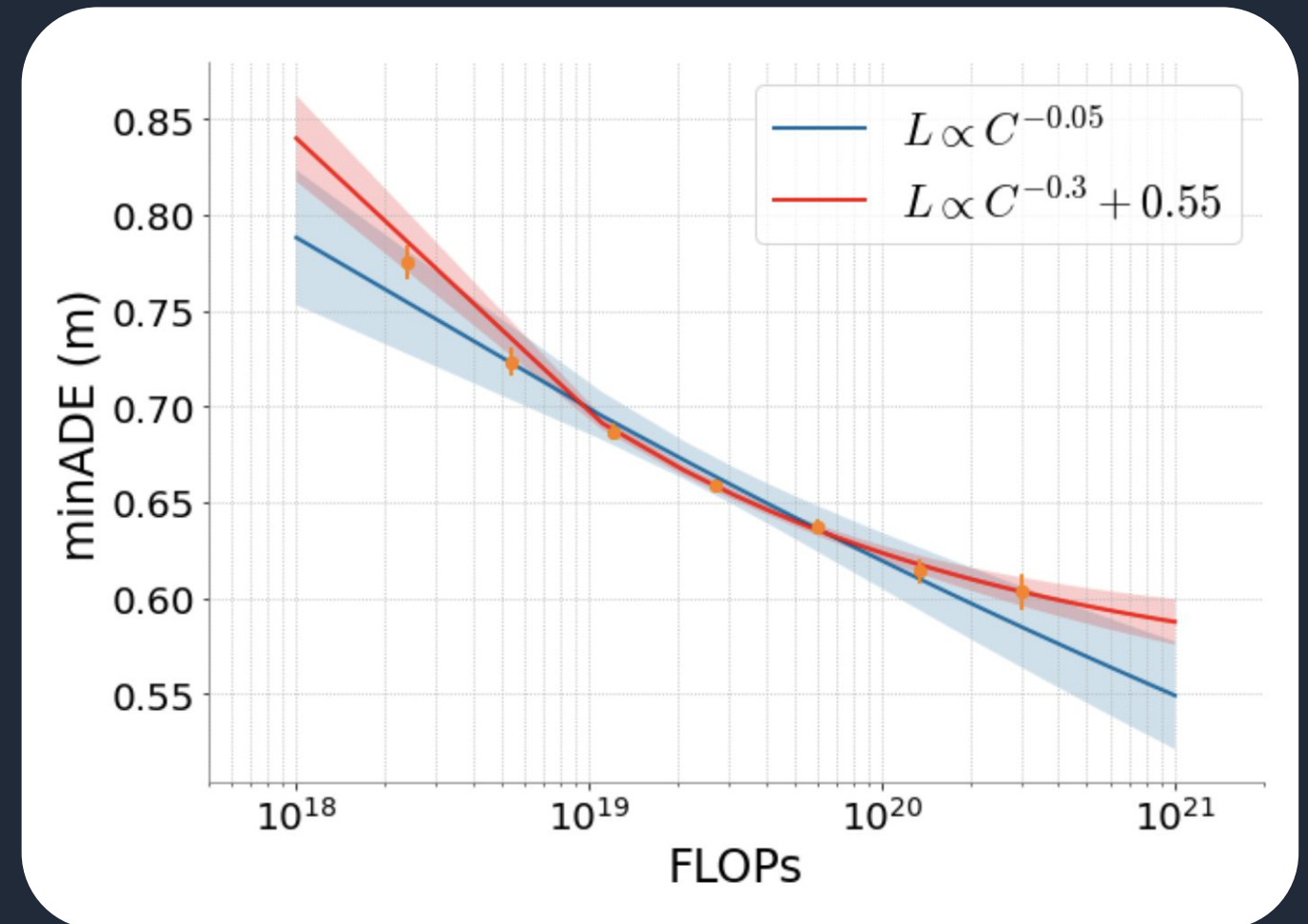
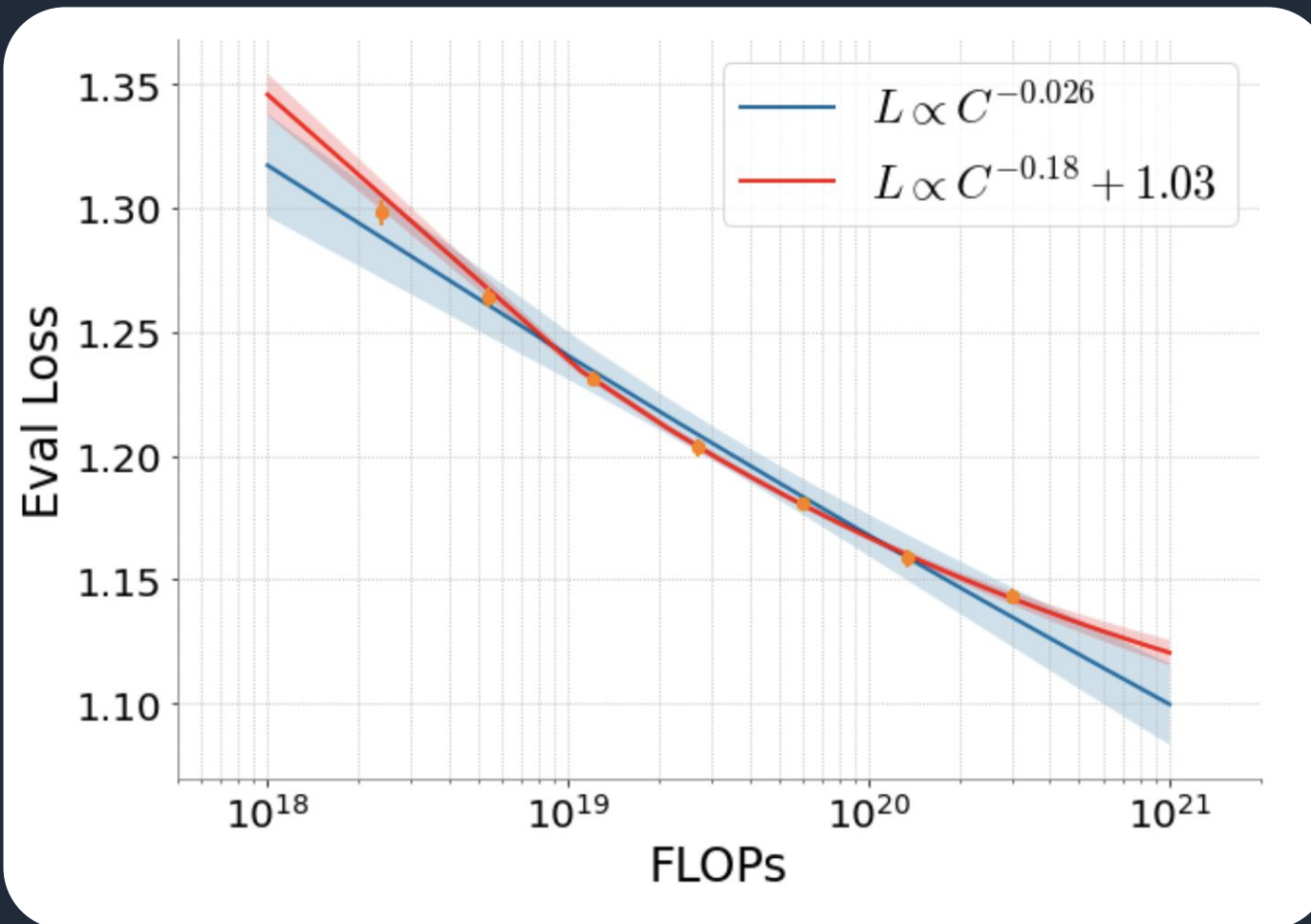
## MotionLM: Multi-Agent Motion Forecasting as Language Modeling

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

ICCV 2023

## Scaling Laws

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



## Scaling Laws of Motion Forecasting and Planning

Mustafa Baniodeh, Kratharth Goel, Scott Ettinger, Carlos Fuertes, Ari Seff, Tim Shen, Cole Gulino, Chenjie Yang, Ghassen Jerfel, Dokook Choe, Rui Wang, Vinutha Kallem, Sergio Casas, Rami Al-Rfou, Benjamin Sapp, Dragomir Anguelov, 2025



# Language understanding and reasoning



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.

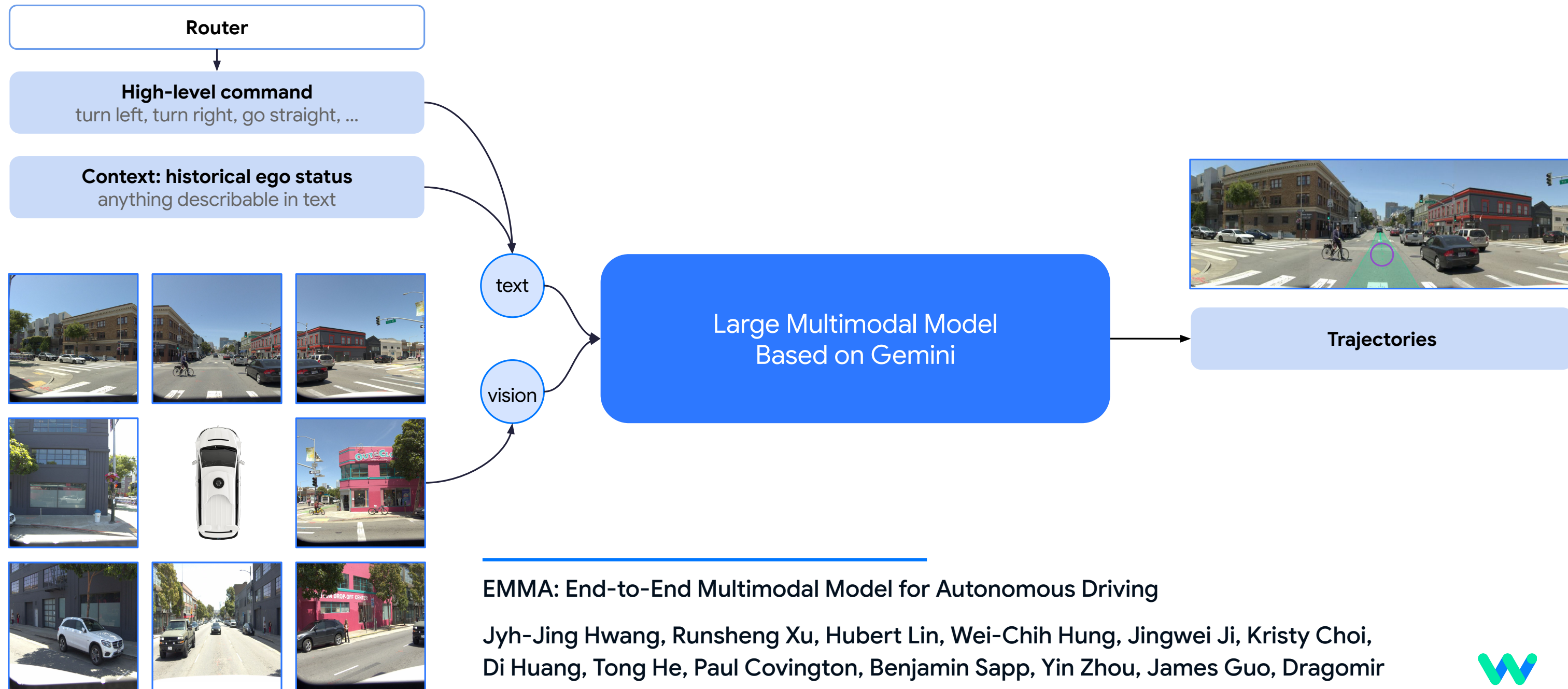


I am driving on the street, this is what I see from the front window. Tell me what's happening and what I should pay special attention to. And tell me what I should do and specify the reasons.

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 Motion Prediction



## 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



# What makes the autonomous-vehicle AI problem so challenging



Complex physical  
environment

High-performance  
requirements

Real-time  
computation

# Thank you<sup>✦✦</sup>

