Gatik
AI RESEARCH GROUP

September 24, 2024

Middle Mile & Foundation Models

Aleksandr Petiushko, Head of AI Research

01

GATIK

02

MIDDLE MILE

03

FOUNDATION MODELS

04

CONCLUSION

Gatik

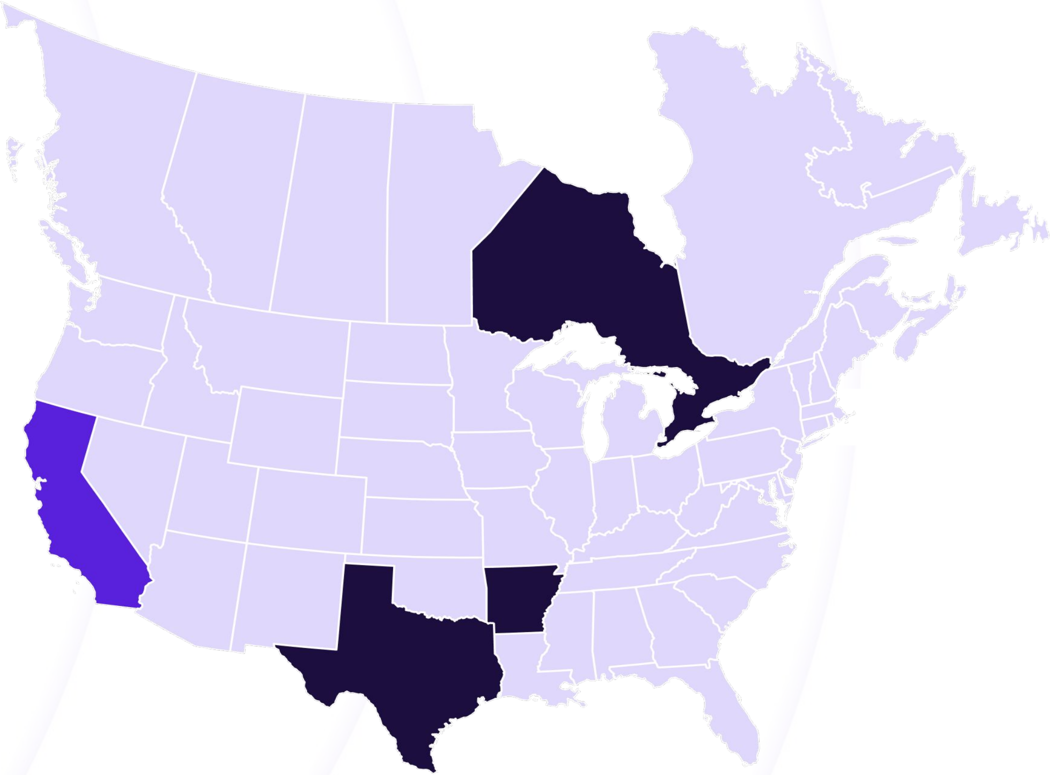
Company Overview

Our Mission

To deliver goods safely and efficiently using our fleet of autonomous trucks.


Gatik
Gatik

Our Locations



California, US – HQ
Mountain View

Texas, US
Fort Worth

Arkansas, US
Bentonville

Ontario, Canada
Toronto

Our Strong Strategic Partnerships Enable Commercialization At Scale

OEM and Manufacturing Partners

Fleet Partners

Key Safety Partners

EDGE CASE RESEARCH

Commercial Partners

Gatik

Middle Mile

Pros and Cons

Middle Mile Specifics

There are a lot of differences between the common long haul trucking as well as robotaxis/last mile delivery.

ODD

Routing

Autonomy

Data

Validation

MIDDLE MILE

Operational Design Domain

A lot of **differences** in comparison to long haul (highway) trucking:

- Vulnerable Road Users (**VRU**s)
- Includes **semi-urban** use case
  - E.g., intersections
- **Lower speed**

NIST ADS SAFETY MEASUREMENT AND OPERATIONAL DESIGN DOMAIN

Image source

Routing

We can concentrate on the **predefined routes**:

- Less and fixed routes means a **smaller ODD**
- **Route selection** can target different **objectives**
  - E.g., safe roads, optimizing time / costs
- **Regularity** of trips: positive **feedback** loop

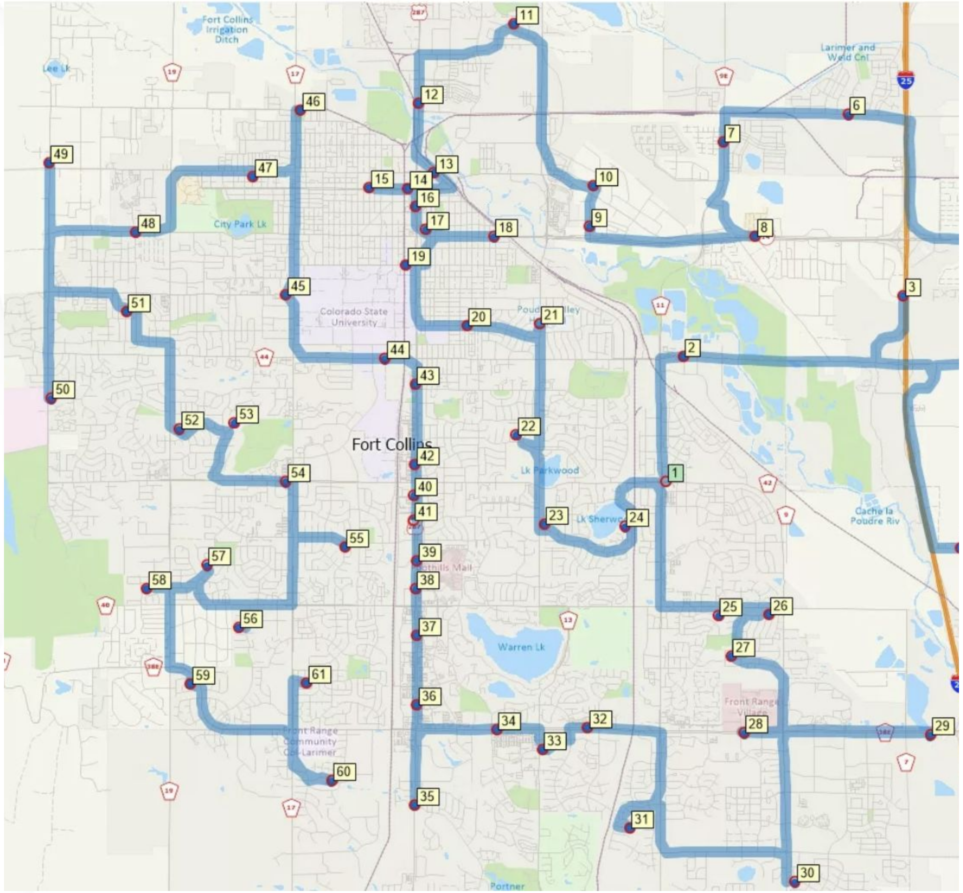


Image source

Autonomy I

- **Mapping**: **less** burden on *online* mapping
  - Still need to be able to **detour** and **re-route**
- **Perception**: **less** large open space *artifacts*, but more focus on VRU detection and tracking
  - Overall trucking problem: **sensor positions** are quite **different** from those in the open source data

Image source

MIDDLE MILE

Autonomy II

- **Behavior**: closer to robotaxi/urban use case
  - E.g., occlusion detection and handling
- **Controls**: need in high precision / maneuverability in case of VRU

MIDDLE MILE

Data

- **Less coverage**, but
- **More** route-specific **behavior diversity**
- Still need to ***proactively*** collect data of reasonable area around the fixed routes

Image source

MIDDLE MILE

Validation

- **Smaller ODD** means **faster** validation cycle
  - Combinatorial explosion of all possible use cases is quite compact
- **Lower speed**: **better** safety expectation
- **VRUs**: still need in **generalizable** solution

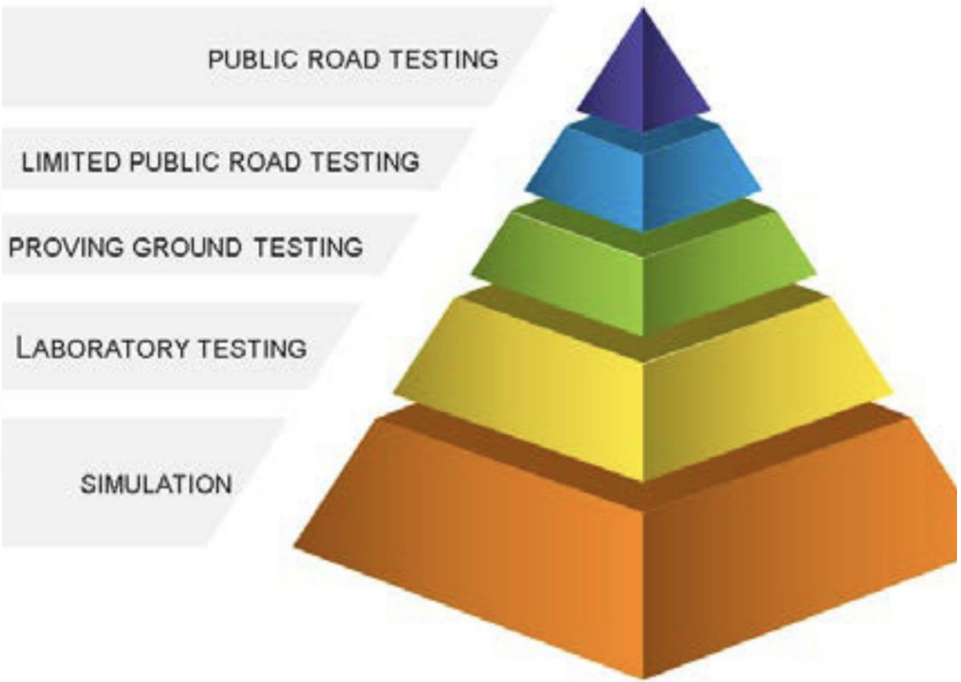


Image source

Foundation Models

Why?

Middle Mile: Challenges

Restriction != Simplicity

Actually, Middle Mile is a combination of everything (although on a smaller scale): domains, use cases, behaviors, etc.

Adjustability

Trucking use case is far from usual robotaxi domain: different sensors, different planning-control feedback loop, stricter requirements on perception.

Safety and OOD

No any collected data of a reasonable size would contain super rare safety-critical examples. AD deployment is dependent on its safety.

Adaptation

Need to quickly adapt to road works, closed roads, incorrect prior information. Moreover, even for one truck, its dynamics is different w.r.t. its load.

Why FMs for Middle Mile?

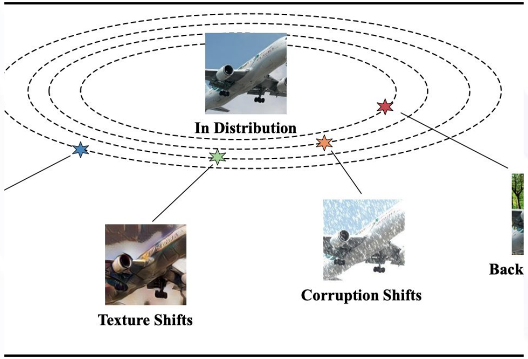


Image source

Generalization

Generalization to raw inputs, different locations, and multi-modal behaviors including out-of-distribution analysis.

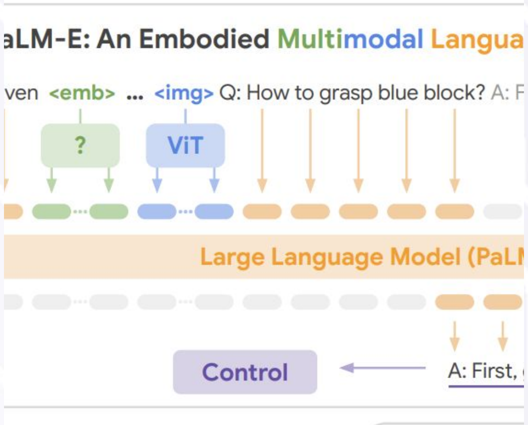
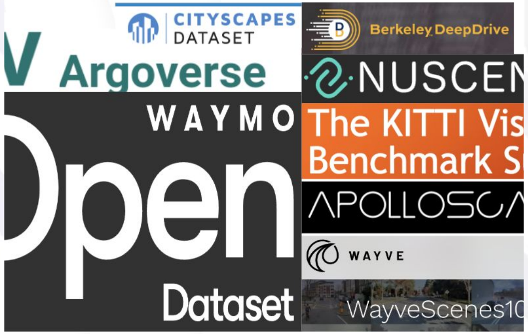


Image source

Flexibility

Through the Visual-Language Model unified interface, any data modality encoders can be used (and even in any combination and order!).



Data

No need in huge pre-training (it's already done!). Auto-labeling / open-vocabulary capabilities save a lot of resources.

Examples of Foundation Models in AD

DriveLM: Driving with Graph Visual Question Answering

Chonghao Sima[4,1*] Katrin Renz[2,3*] Kashyap Chitta[2,3] Li Chen[4,1]
Hanxue Zhang[1] Chengen Xie[1] Jens Beißwenger[2,3] Ping Luo[4]
Andreas Geiger[2,3†] Hongyang Li[1†]

LMDrive: Closed-Loop End-to-End Driving with Large Language Models

Hao Shao[1,2] Yuxuan Hu[3] Letian Wang[4]
Steven L. Waslander[4] Yu Liu[2,5] ✉ Hongsheng Li[1,3,5] ✉

DriveMLM: Aligning Multi-Modal Large Language Models with Behavioral Planning States for Autonomous Driving

Wenhai Wang[2,1*], Jiangwei Xie[3*], ChuanYang Hu[3*], Haoming Zou[4*], Jianan Fan[3*], Wenwen Tong[3*], Yang Wen[3*], Silei Wu[3*], Hanming Deng[3*], Zhiqi Li[5,1*], Hao Tian[3], Lewei Lu[3], Xizhou Zhu[6,3], Xiaogang Wang[2,3], Yu Qiao[1], Jifeng Dai[6,1✉]

DOLPHINS: MULTIMODAL LANGUAGE MODEL FOR DRIVING

Yingzi Ma[1] Yulong Cao[2] Jiachen Sun[3] Marco Pavone[2,4] Chaowei Xiao[1,2]

DriveGPT4: Interpretable End-to-end Autonomous Driving via Large Language Model

Zhenhua Xu[1] Yujia Zhang[2] Enze Xie[3*] Zhen Zhao[4] Yong Guo[3]
Kwan-Yee. K. Wong[1] Zhenguo Li[3] Hengshuang Zhao[1*]

DRIVEVLM: The Convergence of Autonomous Driving and Large Vision-Language Models

Xiaoyu Tian[1*] Junru Gu[1*] Bailin Li[2*] Yicheng Liu[1*] Yang Wang[2] Zhiyong Zhao[2]
Kun Zhan[2] Peng Jia[2] Xianpeng Lang[2] Hang Zhao[1†]

GPT-DRIVER: LEARNING TO DRIVE WITH GPT

Jiageng Mao[1] Yuxi Qian[1] Junjie Ye[1] Hang Zhao[2] Yue Wang[1]

VLP: Vision Language Planning for Autonomous Driving

Chenbin Pan[1*] Burhaneddin Yaman[2] Tommaso Nesti[2] Abhirup Mallik[2]
Alessandro G Allievi[2] Senem Velipasalar[1] Liu Ren[2]

The Problems with Foundation Models in AD

Efficiency

No one HW architecture allows at least 10 Hz inference now. With CoT, RAG, and other techniques the latency only increases.

Implicit bias

While there is a big hope of a FM to capture all aspects of driving, it would be quite hard to get the real understanding of its knowledge frontiers: hallucinations, uncertainty estimation, etc.

Pre-trained AD model

There is no pre-trained model that can be used as a plug-and-play solution. Open-source LLMs need careful fine-tuning.

AD Specifics

3D (and unstructured point clouds in particular) generalization performance is far from being at acceptable level. Much stricter requirements on hallucination avoidance (safety). Custom Chain-of-Thoughts (Localization+Mapping – Perception – Prediction – Planning – Control).

Conclusion

Conclusion

1. **Middle Mile** is a **unique intersection** of tasks, requirements, and challenges
2. **Foundation Models** are **eventually inevitable** for the Middle Mile – mostly because of their generalization capabilities
3. We still need to **pave the road** for the **safe** and **practical** applications of FM in AD

Thank you!

Gatik

Gatik