ResearchInChina
www.researchinchina.com

# End-to-end Autonomous Driving Industry Report, 2024-2025

Dec. 2024

# End-to-end intelligent driving research: How Li Auto becomes a leader from an intelligent driving follower

There are two types of end-to-end autonomous driving: global (one-stage) and segmented (two-stage) types. The former has a clear concept, and much lower R&D cost than the latter, because it does not require any manually annotated data sets but relies on multimodal foundation models developed by Google, META, Alibaba and OpenAI. Standing on the shoulders of these technology giants, the performance of global end-to-end autonomous driving is much better than segmented end-to-end autonomous driving, but at extremely high deployment cost.

Segmented end-to-end autonomous driving still uses the traditional CNN backbone network to extract features for perception, and adopts end-to-end path planning. Although its performance is not as good as global end-to-end autonomous driving, it has lower deployment cost. However, the deployment cost of segmented end-to-end autonomous driving is still very high compared with the current mainstream traditional "BEV+OCC+decision tree" solution.

## Representative of Global End-to-end Autonomous Driving: Waymo EMMA

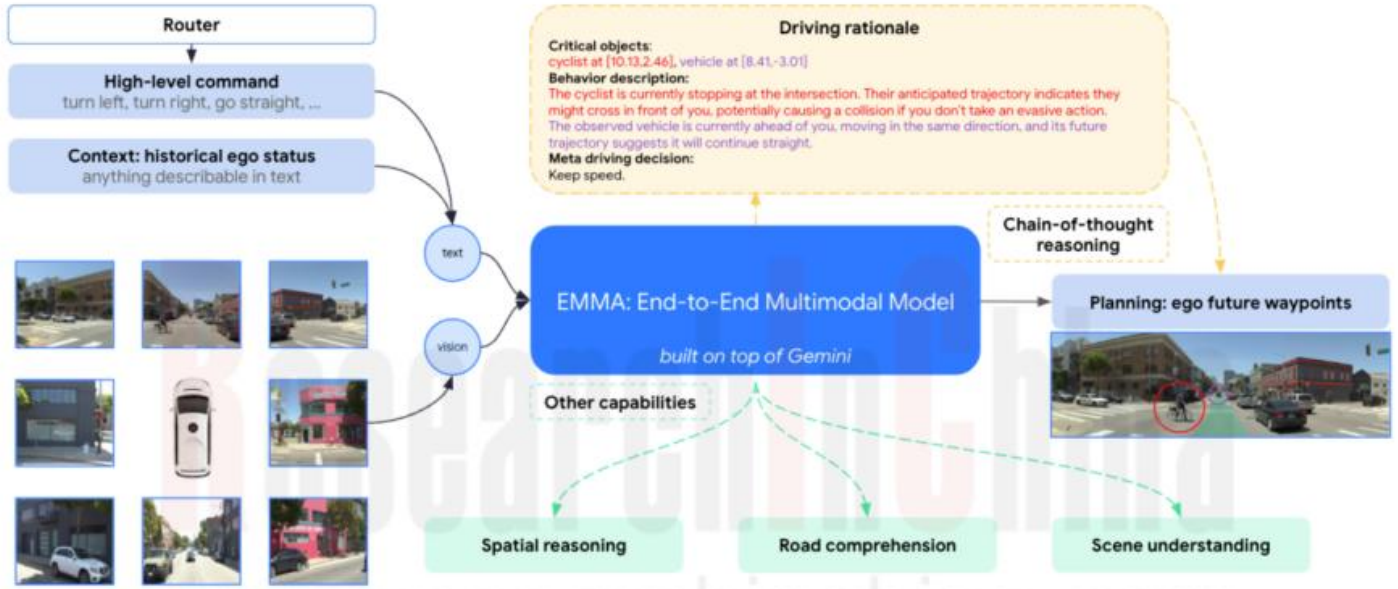



Figure 1: EMMA overview diagram. It takes 3 inputs (**left**): 1) a high-level command from the router, 2) historical status of the ego vehicle, and 3) surround-view camera videos. The model then predicts ego future trajectories (**right**) for motion planning that will be transformed into vehicle driving control signals. Further, we can ask the model to explain its rationale (**top right**) before predicting trajectories, which enhances both the performance and explainability of the model through chain-of-thought reasoning. Notably, we incorporate visual grounding into the rationale so that the model also predicts the accurate 3D/BEV location for critical objects. For completeness, we highlight three additional capabilities of our model (**bottom**).

*Source: Waymo*

ResearchInChina
www.researchinchina.com

As a representative of global end-to-end autonomous driving, Waymo EMMA directly inputs videos without a backbone network but with a multimodal foundation model as the core. UniAD is a representative of segmented end-to-end autonomous driving.

## Representative of Segmented End-to-end Autonomous Driving: UniAD

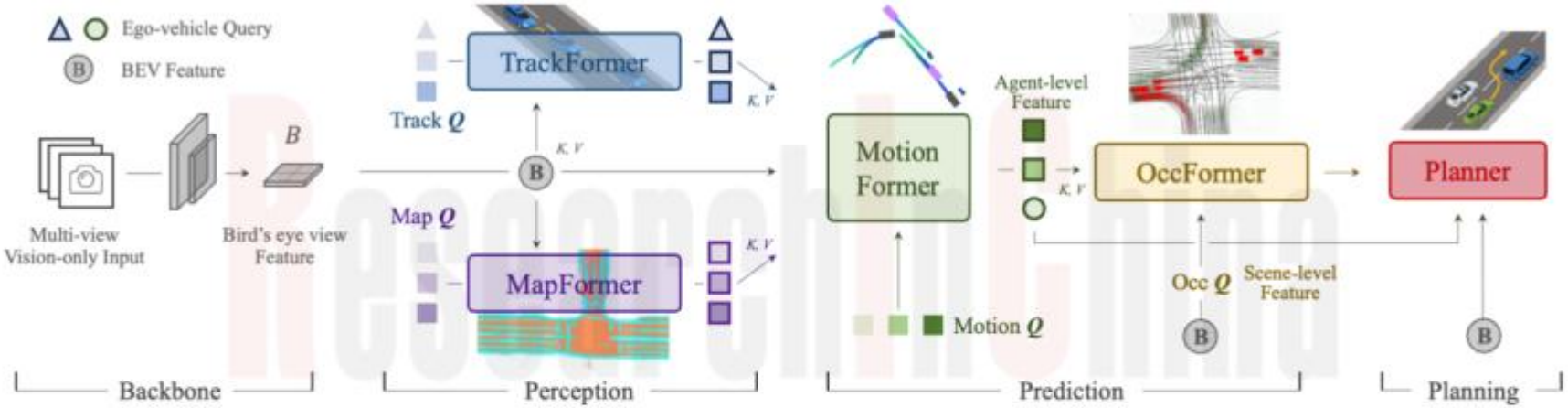



Figure 2. **Pipeline** of Unified Autonomous Driving (UniAD). It is exquisitely devised following planning-oriented philosophy. Instead of a simple stack of tasks, we investigate the effect of each module in perception and prediction, leveraging the benefits of joint optimization from preceding nodes to final planning in the driving scene. All perception and prediction modules are designed in a transformer decoder structure, with task queries as interfaces connecting each node. A simple attention-based planner is in the end to predict future waypoints of the ego-vehicle considering the knowledge extracted from preceding nodes. The map over occupancy is for visual purpose only.

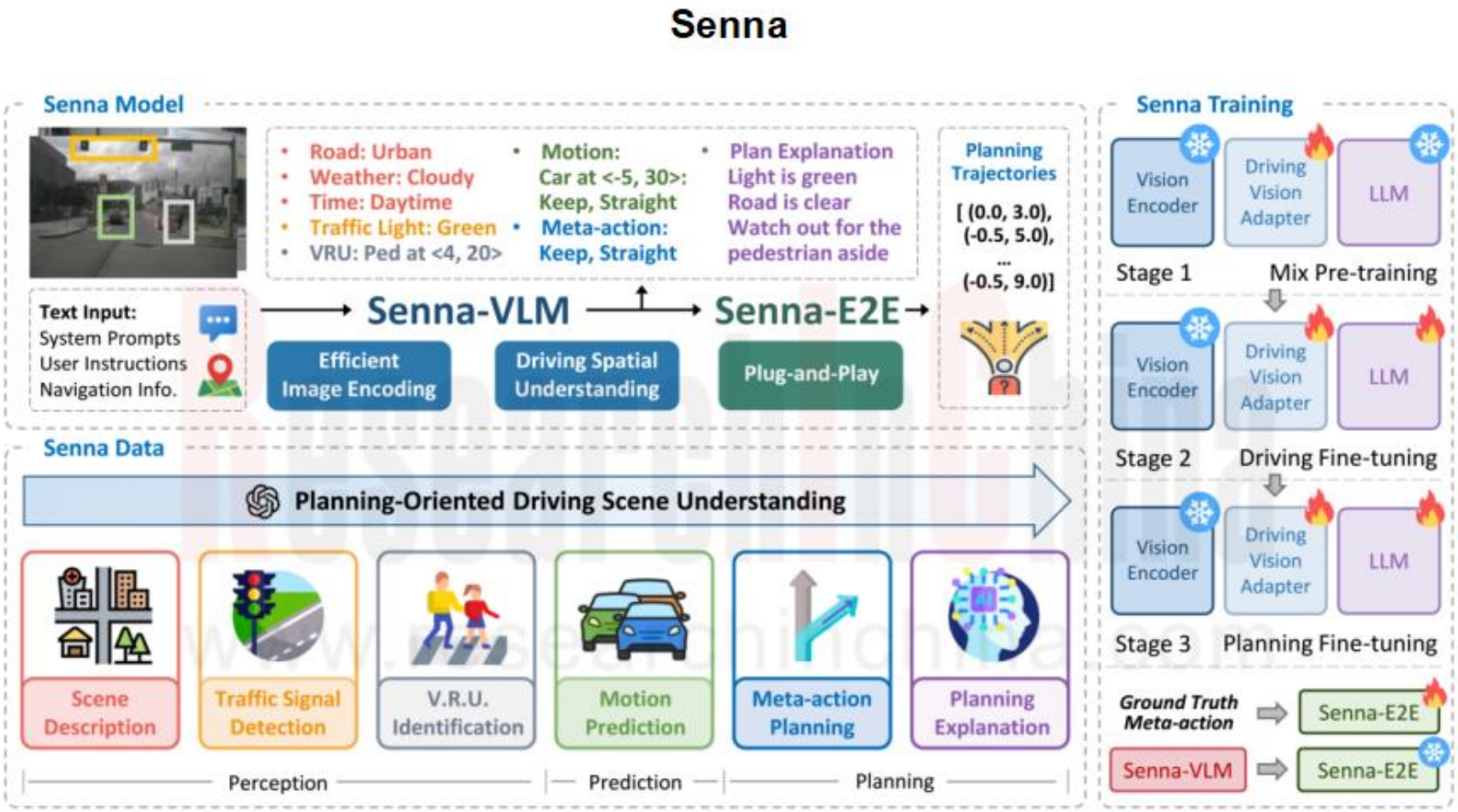
*Source: Planning-oriented Autonomous Driving*

# End-to-end autonomous driving researches are mainly divided into two categories

Based on whether feedback can be obtained, end-to-end autonomous driving researches are mainly divided into two categories: the research is conducted in simulators such as CARLA, and the next planned instructions can be actually performed; the research based on collected real data, mainly imitation learning, referring to UniAD. End-to-end autonomous driving currently features an open loop, so it is impossible to truly see the effects of the execution of one's own predicted instructions. Without feedback, the evaluation of open-loop autonomous driving is very limited. The two indicators commonly used in documents include L2 distance and collision rate.

L2 distance: The L2 distance between the predicted trajectory and the true trajectory is calculated to judge the quality of the predicted trajectory.
Collision rate: The probability of collision between the predicted trajectory and other objects is calculated to evaluate the safety of the predicted trajectory.

The most attractive thing about end-to-end autonomous driving is the potential for performance improvement. The earliest end-to-end solution is UniAD. A paper at the end of 2022 revealed that the L2 distance was as long as 1.03 meters. It was greatly reduced to 0.55 meters at the end of 2023 and further to 0.22 meters in late 2024. Horizon Robotics is one of the most active companies in the end-to-end field, and its technology development also shows the overall evolution of the end-to-end route. After UniAD came out, Horizon Robotics immediately proposed VAD whose concept is similar to that of UniAD with much better performance. Then, Horizon Robotics turned to global end-to-end autonomous driving. Its first result was HE-Driver, which had a relatively large number of parameters. The following Senna has a smaller number of parameters and is also one of the best-performing end-to-end solutions.

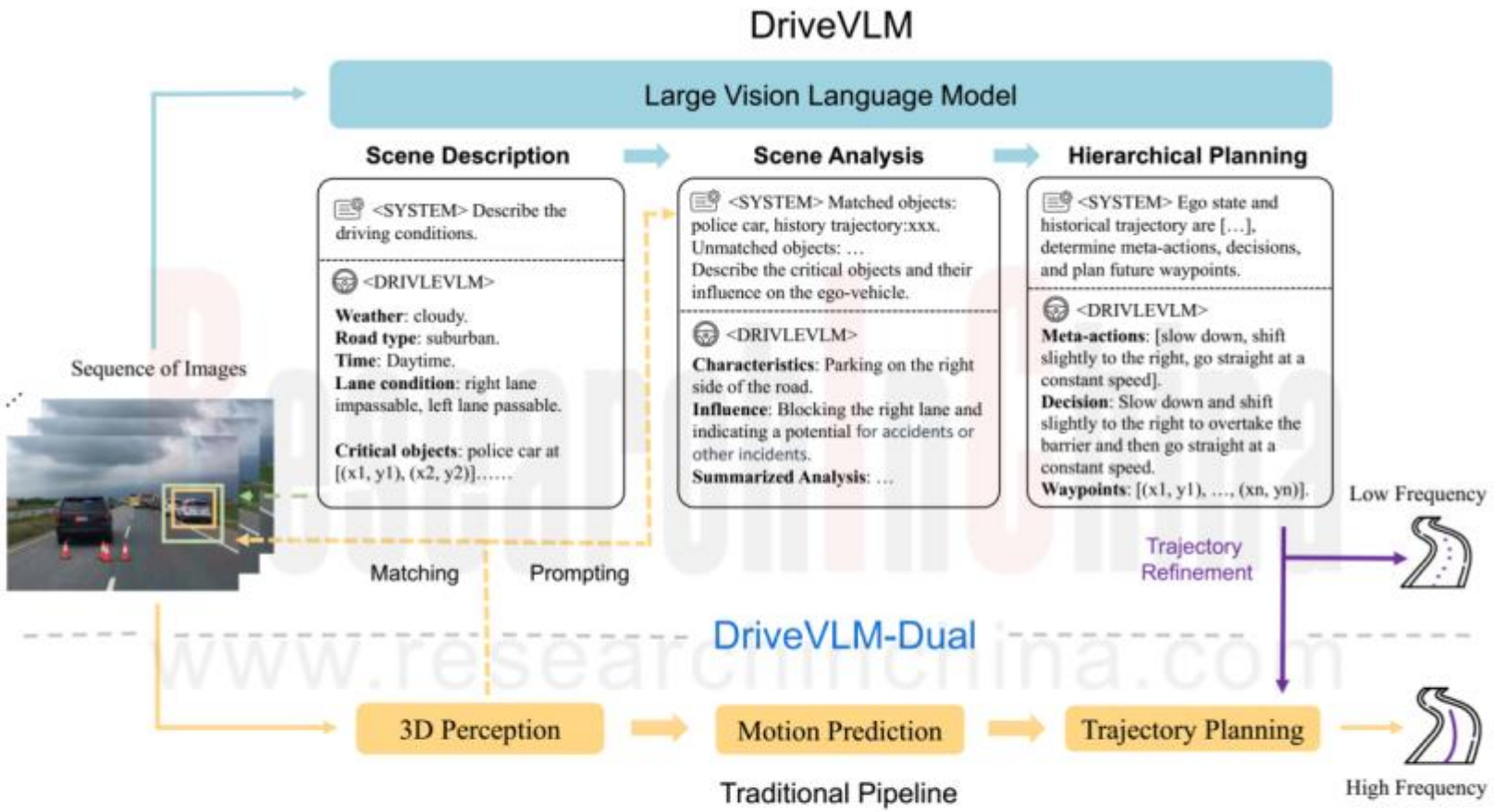



*Source: "Senna: Bridging Large Vision-Language Models and End-to-End Autonomous Driving*

# DriveVLM

The core of some end-to-end systems is still BEVFormer which uses vehicle CAN bus information by default, including explicit information related to the vehicle's speed, acceleration and steering angle, exerting a significant impact on path planning. These end-to-end systems still require supervised training, so massive manual annotations are indispensable, which makes the data cost very high. Furthermore, since the concept of GPT is borrowed, why not use LLM directly? In this case, Li Auto proposed DriveVLM.

As the figure below shows, the pipeline of DriveVLM from Li Auto mainly involves design of the three major modules: scenario description, scenario analysis, and hierarchical planning.




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

# The scenario description module of DriveVLM

The scenario description module of DriveVLM is composed of environment description and key object recognition. Environment description focuses on common driving environments such as weather and road conditions. Key object recognition is to find key objects that have a greater impact on current driving decision. Environment description includes the following four parts: weather, time, road type, and lane line.

Differing from the traditional autonomous driving perception module that detects all objects, DriveVLM focuses on recognizing key objects in the current driving scenario that are most likely to affect autonomous driving decision, because detecting all objects will consume enormous computing power. Thanks to the pre-training of the massive autonomous driving data accumulated by Li Auto and the open source foundation model, VLM can better detect key long-tail objects, such as road debris or unusual animals, than traditional 3D object detectors.

For each key object, DriveVLM will output its semantic category (c) and the corresponding 2D object box (b) respectively. Pre-training comes from the field of NLP foundation models, because NLP uses very little annotated data and is very expensive. Pre-training first uses massive unannotated data for training to find language structure features, and then takes prompts as labels to solve specific downstream tasks by fine-tuning.

DriveVLM completely abandons the traditional algorithm BEVFormer as the core but adopts large multimodal models. Li Auto's DriveVLM leverages Alibaba's foundation model Qwen-VL with up to 9.7 billion parameters, 448*448 input resolution, and NVIDIA Orin for inference operations.

# How does Li Auto transform from a high-level intelligent driving follower into a leader?

**How does Li Auto transform from a high-level intelligent driving follower into a leader?**

At the beginning of 2023, Li Auto was still a laggard in the NOA arena. It began to devote itself to R&D of high-level autonomous driving in 2023, accomplished multiple NOA version upgrades in 2024, and launched all-scenario autonomous driving from parking space to parking space in late November 2024, thus becoming a leader in mass production of high-level intelligent driving (NOA).

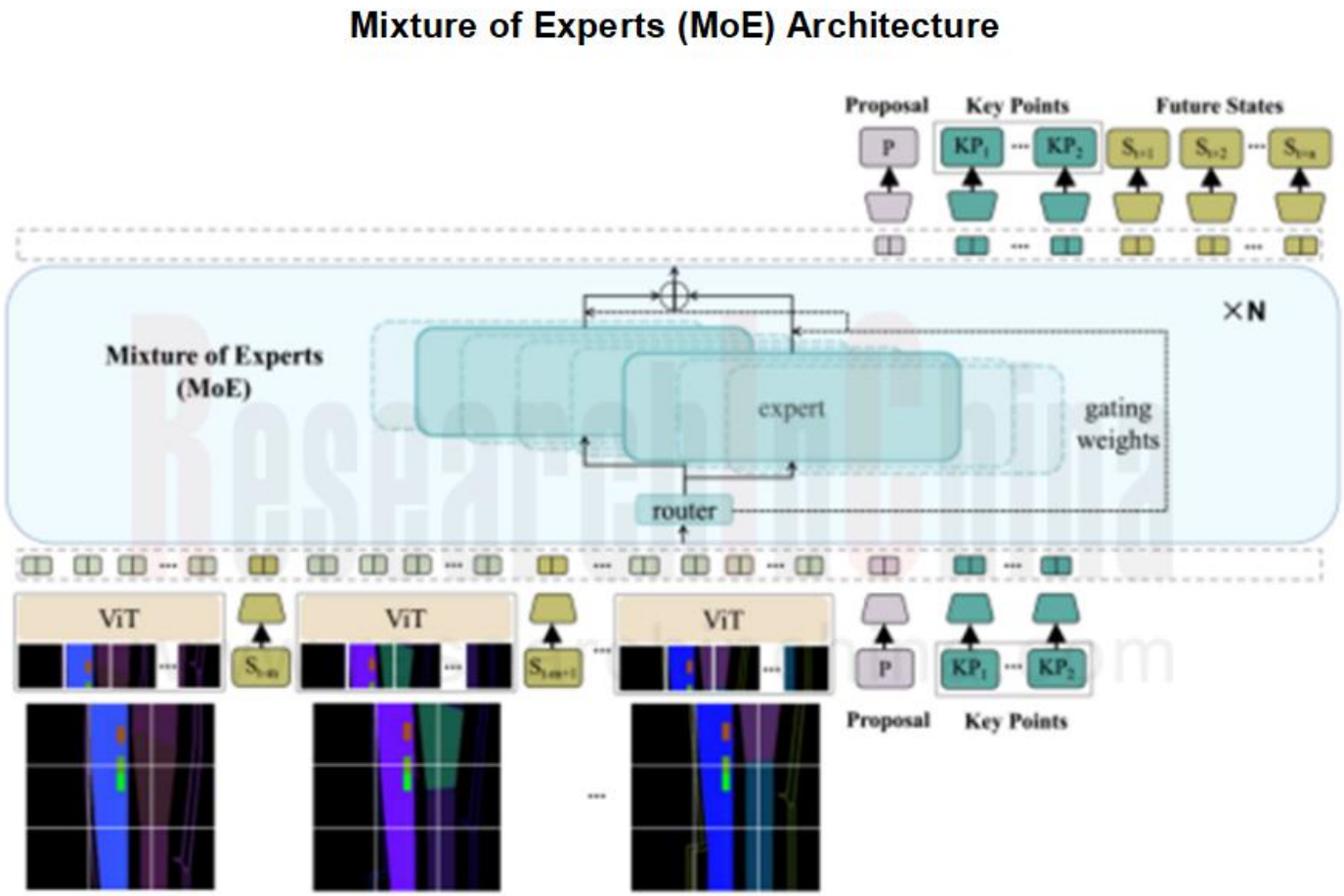
Reviewing the development history of Li Auto's end-to-end intelligent driving, in addition to the data from its own hundreds of thousands of users, it also partnered with a number of partners on R&D of end-to-end models. DriveVLM is the result of the cooperation between Li Auto and Tsinghua University.

In addition to DriveVLM, Li Auto also launched STR2 with Shanghai Qi Zhi Institute, Fudan University, etc., proposed DriveDreamer4D with GigaStudio, the Institute of Automation of Chinese Academy of Sciences, and unveiled MoE with Tsinghua University.

**Mixture of Experts (MoE) Architecture**

In order to solve the problem of too many parameters and too much calculation in foundation models, Li Auto has cooperated with Tsinghua University to adopt MoE Architecture. Mixture of Experts (MoE) is an integrated learning method that combines multiple specialized sub-models (i.e. "experts") to form a complete model. Each "expert" makes contributions in the field in which it is good at. The mechanism that determines which "expert" participates in answering a specific question is called a "gated network". Each expert model can focus on solving a specific sub-problem, and the overall model can achieve better performance in complex tasks. MoE is suitable for processing considerable datasets and can effectively cope with the challenges of massive data and complex features. That’s because it can handle different sub-tasks in parallel, make full use of computing resources, and improve the training and reasoning efficiency of models.

# Mixture of Experts (MoE) Architecture




*Source: Generalizing Motion Planners with Mixture of Experts for Autonomous Driving*

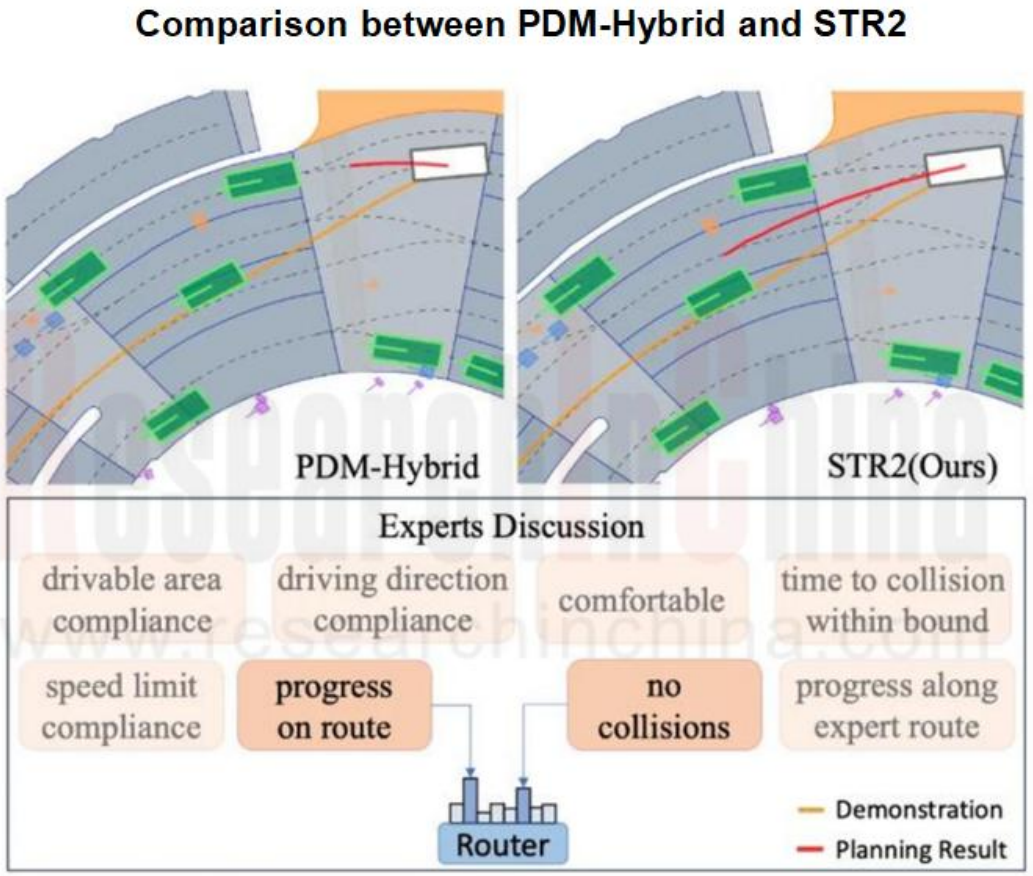
# STR2 Path Planner

STR2 is a motion planning solution based on Vision Transformer (ViT) and MoE. It was developed by Li Auto and researchers from Shanghai Qi Zhi Research Institute, Fudan University and other universities and institutions.

STR2 is designed specifically for the autonomous driving field to improve generalization capabilities in complex and rare traffic conditions.

STR2 is an advanced motion planner that enables deep learning and effective planning of complex traffic environments by combining a Vision Transformer (ViT) encoder and MoE causal transformer architecture.

The core idea of STR2 is to wield MoE to handle modality collapse and reward balance through expert routing during training, thereby improving the model's generalization capabilities in unknown or rare situations.




*Source: Generalizing Motion Planners with Mixture of Experts for Autonomous Driving*

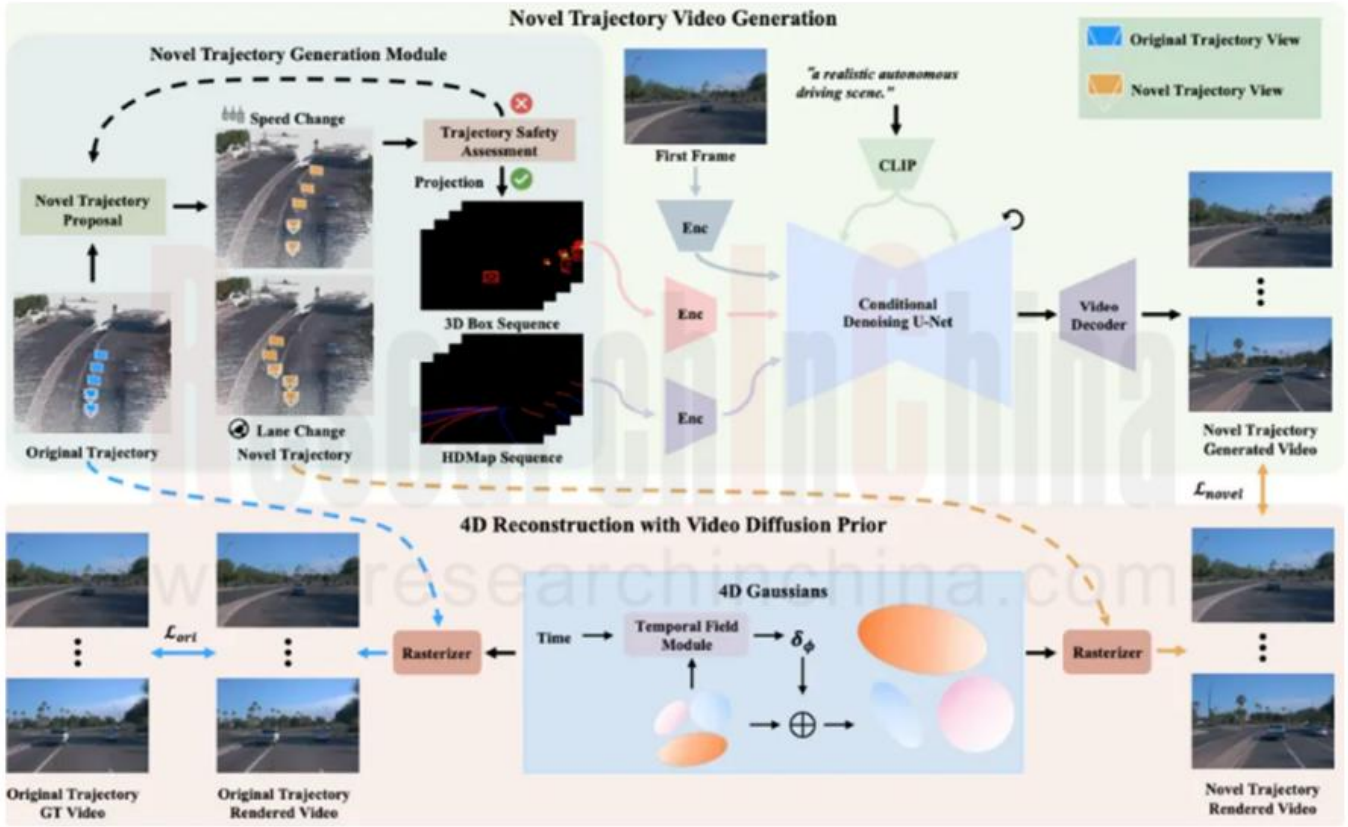
# DriveDreamer4D World Model

In late October 2024, GigaStudio teamed up with the Institute of Automation of Chinese Academy of Sciences, Li Auto, Peking University, Technical University of Munich and other units to propose DriveDreamer4D.

- DriveDreamer4D uses a world model as a data engine to synthesize new trajectory videos (e.g., lane change) based on real-world driving data.
- DriveDreamer4D can also provide rich and diverse perspective data (lane change, acceleration and deceleration, etc.) for driving scenarios to increase closed-loop simulation capabilities in dynamic driving scenarios.
- The overall structure diagram is shown in the figure. The novel trajectory generation module (NTGM) adjusts the original trajectory actions, such as steering angle and speed, to generate new trajectories. These new trajectories provide a new perspective for extracting structured information (e.g., vehicle 3D boxes and background lane line details).

* Subsequently, based on the video generation capabilities of the world model and the structured information obtained by updating the trajectories, videos of new trajectories can be synthesized. Finally, the original trajectory videos are combined with the new trajectory videos to optimize the 4DGS model.

**Video-based Trajectory Generation and 4D Reconstruction System**




*Source: DriveDreamer4D: World Models Are Effective Data Machines for 4D Driving Scene Representation*

# Table of Content (1)


**1. Foundation of End-to-end Autonomous Driving Technology**
1.1 Terminology and Concept of End-to-end Autonomous Driving
1.2 Introduction to and Status Quo of End-to-end Autonomous Driving
Background of End-to-end Autonomous Driving
Reason for End-to-end Autonomous Driving: Business Value
Difference between End-to-end Architecture and Traditional Architecture (1)
Difference between End-to-end Architecture and Traditional Architecture (2)
End-to-end Architecture Evolution
Progress in End-to-end Intelligent Driving (1)
Progress in End-to-end Intelligent Driving (2)
Comparison between One-stage and Two-stage End-to-end Autonomous Driving
Mainstream One-stage/Segmented End-to-end System Performance Parameters
Significance of Introducing Multi-modal models to End-to-end Autonomous Driving
Problems and Solutions for End-to-end Mass Production (1)
Problems and Solutions for End-to-end Mass Production (2)
Progress and Challenges in End-to-end Systems

1.3 Classic End-to-end Autonomous Driving Cases
SenseTime UniAD
Technical Principle and Architecture of SenseTime UniAD
Technical Principle and Architecture of Horizon Robotics VAD
Technical Principle and Architecture of Horizon Robotics VADv2
VADv2 Training
Technical Principle and Architecture of DriveVLM
Li Auto Adopts MoE
MoE and STR2
E2E-AD Model: SGADS
E2E Active Learning Case: ActiveAD
End-to-end Autonomous Driving System Based on Foundation Models

1.4 Foundation Models
1.4.1 Introduction
Core of End-to-end System - Foundation Models
Foundation Models (1) - Large Language Models: Examples of Applications in Autonomous Driving
Foundation Models (2) - Vision Foundation (1)
Foundation Models (2) - Vision Foundation (2)
Foundation Models (2) - Vision Foundation (3)
Foundation Models (2) - Vision Foundation (4)
Foundation Models (3) - Multimodal Foundation Models (1)
Foundation Models (3) - Multimodal Foundation Models (2)
1.4.2 Foundation Models: Multimodal Foundation Models
Development of and Introduction to Multimodal Foundation Models
Multimodal Foundation Models VS Single-modal Foundation Models (1)
Multimodal Foundation Models VS Single-modal Foundation Models (2)
Technology Panorama of Multimodal Foundation Models
Multimodal Information Representation
1.4.3 Foundation Models: Multimodal Large Language Models
Multimodal Large Language Models (MLLMs)
Architecture and Core Components of MLLMs
MLLMs - Mainstream Models
Application of MLLMs in Autonomous Driving

1.5 VLM & VLA
Application of Vision-Language Models (VLMs)
Development History of VLMs
Architecture of VLMs
Application Principle of VLMs in End-to-end Autonomous Driving
Application of VLMs in End-to-end Autonomous Driving

# Table of Content (2)


VLM→VLA
VLA Models
VLA Principle
Classification of VLA Models
Core Functions of End-to-end Multimodal Model for Autonomous Driving (EMMA)

1.6 World Models
Definition and Application
Basic Architecture
Generation of Virtual Training Data
Tesla's World Model
Nvidia
InfinityDrive: Breaking Time Limits in Driving World Models

1.7 Comparison between E2E-AD Motion Planning Models
Comparison between Several Classical Models in Industry and Academia
Tesla: Perception and Decision Full Stack Integrated Model
Momenta: End-to-end Planning Architecture Based on BEV Space
Horizon Robotics 2023: End-to-end Planning Architecture Based on BEV Space
DriveIRL: End-to-end Planning Architecture Based on BEV Space
GenAD: Generative End-to-end Model

1.8 Embodied Language Models (ELMs)
ELMs Accelerate the Implementation of End-to-end Solutions
Application Scenarios
Limitations and Positive Impacts

**2 Technology Roadmap and Development Trends of End-to-end Autonomous Driving**

2.1 Technology Trends of End-to-end Autonomous Driving
Trend 1
Trend 2
Trend 3
Trend 4
Trend 5
Trend 6
Trend 7

2.2 Market Trends of End-to-end Autonomous Driving
Layout of Mainstream End-to-end System Solutions
Comparison of End-to-end System Solution Layout between Tier 1 Suppliers (1)
Comparison of End-to-end System Solution Layout between Tier 1 Suppliers (2)
Comparison of End-to-end System Solution Layout between Other Autonomous Driving Companies
Comparison of End-to-end System Solution Layout between OEMs (1)
Comparison of End-to-end System Solution Layout between OEMs (2)
Comparison of NOA and End-to-end Implementation Schedules between Sub-brands of Domestic Mainstream OEMs (1)
Comparison of NOA and End-to-end Implementation Schedules between Sub-brands of Domestic Mainstream OEMs (2)
Comparison of NOA and End-to-end Implementation Schedules between Sub-brands of Domestic Mainstream OEMs (3)
Comparison of NOA and End-to-end Implementation Schedules between Sub-brands of Domestic Mainstream OEMs (4)

2.3 End-to-end Autonomous Driving Team Building
Impacts of End-to-end Foundation Models on Organizational Structure (1)
Impacts of End-to-end Foundation Models on Organizational Structure (2)

# Table of Content (3)

End-to-end Autonomous Driving Team Building of Domestic OEMs (1)
End-to-end Autonomous Driving Team Building of Domestic OEMs (2)
End-to-end Autonomous Driving Team Building of Domestic OEMs (3)
End-to-end Autonomous Driving Team Building of Domestic OEMs (4)
End-to-end Autonomous Driving Team Building of Domestic OEMs (5)
End-to-end Autonomous Driving Team Building of Domestic OEMs (6)
End-to-end Autonomous Driving Team Building of Domestic OEMs (7)
Team Building of End-to-end Autonomous Driving Suppliers (1)
Team Building of End-to-end Autonomous Driving Suppliers (2)
Team Building of End-to-end Autonomous Driving Suppliers (3)
Team Building of End-to-end Autonomous Driving Suppliers (4)

**3. End-to-end Autonomous Driving Suppliers**
3.1 MOMENTA
Profile
One-stage End-to-end Solutions (1)
One-stage End-to-end Solutions (2)
End-to-end Planning Architecture
One-stage End-to-end Mass Production Empowers the Large-scale Implementation of NOA in Mapless Cities
High-level Intelligent Driving and End-to-end Mass Production Customers

3.2 DeepRoute.ai
Product Layout and Strategic Deployment
End-to-end Layout
Difference between End-to-end Solutions and Traditional Solutions
Implementation Progress in End-to-end Solutions
End-to-end VLA Model Analysis
Designated End-to-end Mass Production Projects and VLA Model Features
Hierarchical Prompt Tokens
End-to-end Training Solutions
Application Value of DINOv2 in the Field of Computer Vision
Autonomous Driving VQA Task Evaluation Data Sets
Score Comparison between HoP and Huawei

3.3 Huawei
Development History of Huawei's Intelligent Automotive Solution Business Unit
End-to-end Concept and Perception Algorithm of ADS
ADS 3.0 (1)
ADS 3.0 (2): End-to-end
ADS 3.0 (3): ASD 3.0 VS. ASD 2.0
End-to-end Solution Application Cases of ADS 3.0 (1)
End-to-end Solution Application Cases of ADS 3.0 (2)
End-to-end Solution Application Cases of ADS 3.0 (3)
End-to-end Autonomous Driving Solutions of Multimodal LLMs
End-to-end Testing—VQA Tasks
Architecture of DriveGPT4
End-to-end Training Solution Examples
The Training of DriveGPT4 Is Divided Into Two Stages
Comparison between DriveGPT4 and GPT4V

3.4 Horizon Robotics
Profile
Main Partners
End-to-end Super Drive and Its Advantages
Architecture and Technical Principle of Super Drive
Journey 6 and Horizon SuperDrive? All-scenario Intelligent Driving Solution
Senna Intelligent Driving System (Foundation Model + End-to-end)

# Table of Content (4)


Core Technology and Training Method of Senna
Core Module of Senna

3.5 Zhuoyu Technology
Profile
R&D and Production
Two-stage End-to-end Parsing
One-stage Explainable End-to-end Parsing
End-to-end Mass Production Customers

3.6 NVIDIA
Profile
Autonomous driving solution
DRIVE Thor
Basic Platform for Autonomous Driving
Next-generation Automotive Computing Platform
Latest End-to-end Autonomous Driving Framework: Hydra-MDP
Self-developed Model Architecture

3.7 Bosch
Intelligent Driving China Strategic Layout (1)
Based on the End-to-end Development Trend, Bosch Intelligent Driving initiates the Organizational Structure Reform
Intelligent Driving Algorithm Evolution Planning

3.8 Baidu
Profile of Apollo
Strategic Layout in the Field of Intelligent Driving
Two-stage End-to-end
Production Models Based on Two-stage End-to-end Technology Architecture
Baidu Auto Cloud 3.0 Enables End-to-end Systems from Three Aspects

3.9 SenseAuto
Profile
UniAD End-to-end Solution
DriveAGI: The Next-generation Autonomous Driving Foundation Model and Its Advantages
DiFSD: SenseAuto's End-to-end Autonomous Driving System That Simulates Human Driving Behavior
DiFSD: Technical Interpretation

3.10 QCraft
Profile
"Driven-by-QCraft" High-level Intelligent Driving Solution
End-to-end Layout
Advantages of End-to-end Layout

3.11 Wayve
Profile
Advantages of AV 2.0
GAIA-1 World Model - Architecture
GAIA-1 World Model - Token
GAIA-1 World Model - Generation Effect
LINGO-2

3.12 Waymo
End-to-end Multimodal Model for Autonomous Driving (EMMA)
EMMA Analysis: Multimodal Input

# Table of Content (5)


EMMA Analysis: Defining Driving Tasks as Visual Q&A
EMMA Analysis: Introducing Thinking Chain Reasoning to Enhance Interpretability
Limitations of EMMA

3.13 GigaStudio
Introduction
DriveDreamer
DriveDreamer 2
DriveDreamer4D
3.14 LightWheel AI
Profile
Core Technology
Core Technology Stack
Data Annotation and Synthetic Data

**4. End-to-end Autonomous Driving Layout of OEMs**
4.1 Xpeng's End-to-end Intelligent Driving Layout
End-to-end System (1): Architecture
End-to-end System (2): Intelligent Driving Model
End-to-end System (3): AI+XNGP
End-to-End System (4): Organizational Transformation
Data Collection, Annotation and Training

4.2 Li Auto’s End-to-end Intelligent Driving Layout
End-to-end Solutions (1)
End-to-end Solutions (2)
End-to-end Solutions (3)
End-to-end Solutions (4)
End-to-end Solutions (5)
End-to-end Solutions (6)
End-to-end Solutions: L3 Autonomous Driving
End-to-end Solutions: Building of a Complete Foundation Model
Technical Layout: Data Closed Loop

4.3 Tesla’s End-to-end Intelligent Driving Layout
Interpretation of the 2024 AI Conference
Development History of AD Algorithms
End-to-end Process 2023-2024
Development History of AD Algorithms (1)
Development History of AD Algorithms (2)
Development History of AD Algorithms (3)
Development History of AD Algorithms (4)
Development History of AD Algorithms (5)
Tesla: Core Elements of the Full-stack Perception and Decision Integrated Model
"End-to-end" Algorithms
World Models
Data Engines
Dojo Supercomputing Center

4.4 Zeron’s End-to-end Intelligent Driving Layout
Profile
End-to-end Autonomous Driving System Based on Foundation Models (1)
End-to-end Autonomous Driving System Based on Foundation Models (2) - Data Training
Advantages of End-to-end Driving System

4.5 Geely & ZEEKR’s End-to-end Intelligent Driving Layout
Geely’s ADAS Technology Layout: Geely Xingrui Intelligent Computing Center (1)

# Table of Content (6)

Geely’s ADAS Technology Layout: Geely Xingrui Intelligent Computing Center (2)
Geely’s ADAS Technology Layout: Geely Xingrui Intelligent Computing Center (3)
Xingrui AI foundation model
Application of Geely’s Intelligent Driving Foundation Model Technology
ZEEKR’s End-to-end System: Two-stage Solution
ZEEKR Officially Released End-to-end Plus
ZEEKR’s End-to-end Plus
Examples of Models with ZEEKR’s End-to-end System

4.6 Xiaomi Auto’s End-to-end Intelligent Driving Layout
Profile
End-to-end Technology Enables All-scenario Intelligent Driving from Parking Spaces to Parking Spaces
Road Foundation Models Build HD Maps through Road Topology
New-generation HAD Accesses End-to-end System
End-to-end Technology Route

4.7 NIO’s End-to-end Intelligent Driving Layout
Intelligent Driving R&D Team Reorganization with an Organizational Structure Oriented Towards End-to-end System
From Modeling to End-to-end, World Models Are the Next
World Model End-to-end System
Intelligent Driving Architecture: NADArch 2.0
End-to-end R&D Tool Chain
Imagination, Reconstruction and Group Intelligence of World Models
NSim
Software and Hardware Synergy Capabilities Continue to Strengthen, Moving towards the End-to-end System Era

4.8 Changan Automobile’s End-to-end Intelligent Driving Layout
Brand Layout
End-to-end System (1)
End-to-end System (2)
Production Models with End-to-end System

4.9 Mercedes-Benz's End-to-end Intelligent Driving Layout
Brand New "Vision-only Solutions without Maps, L2++ All-scenario High-level Intelligent Driving Functions"
Brand New Self-developed MB.OS
Cooperation with Momenta

4.10 Chery’s End-to-end Intelligent Driving Layout
Profile of ZDRIVE.AI
Chery’s End-to-end System Development Planning

# Contact

**Beijing Headquarters**

TEL: 13718845418
Email: report@researchinchina.com


Website: ResearchInChina

WeChat: Zuosiqiche

**Chengdu Branch**

TEL: 028-68738514
FAX: 028-86930659


zuosiqiche