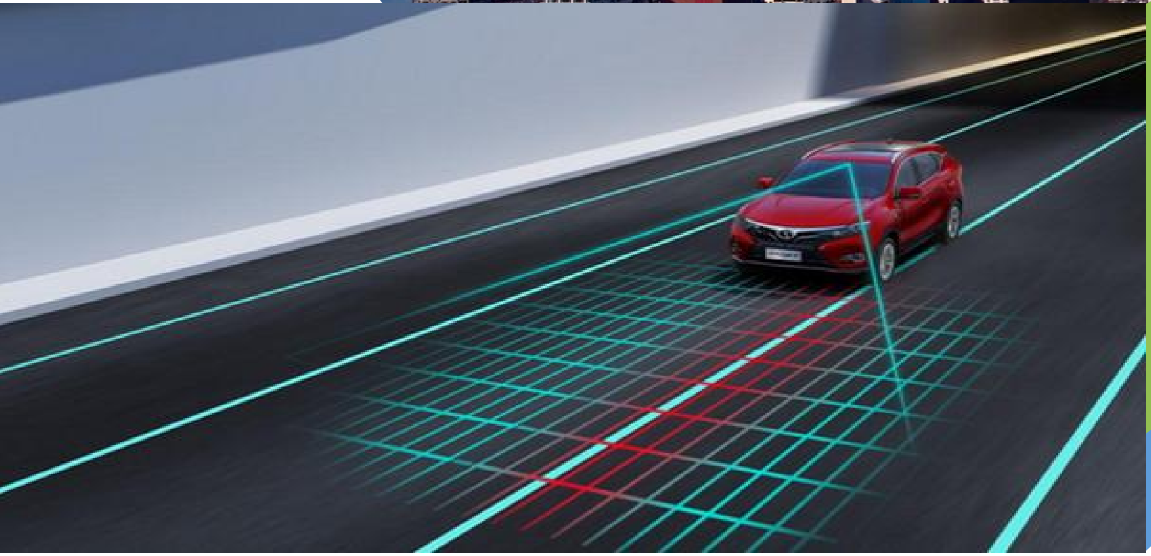


ResearchInChina
www.researchinchina.com

End-to-end Autonomous Driving (E2E AD) Research Report, 2024

May 2024

End-to-end Autonomous Driving Research: status quo of End-to-end (E2E) autonomous driving

1. Status quo of end-to-end solutions in China

An end-to-end autonomous driving system refers to direct mapping from sensor data inputs (camera images, LiDAR, etc.) to control command outputs (steering, acceleration/deceleration, etc.). It first appeared in the ALVINN project in 1988. It uses cameras and laser rangefinders as input and a simple neural network to generate steering as output.

In early 2024, Tesla rolled out FSD V12.3, featuring an amazing intelligent driving level. The end-to-end autonomous driving solution garners widespread attention from OEMs and autonomous driving solution companies in China.

Compared with conventional multi-module solutions, the end-to-end autonomous driving solution integrates perception, prediction and planning into a single model, simplifying the solution structure. It can simulate human drivers making driving decisions directly according to visual inputs, effectively cope with long tail scenarios of modular solutions and improve the training efficiency and performance of models.

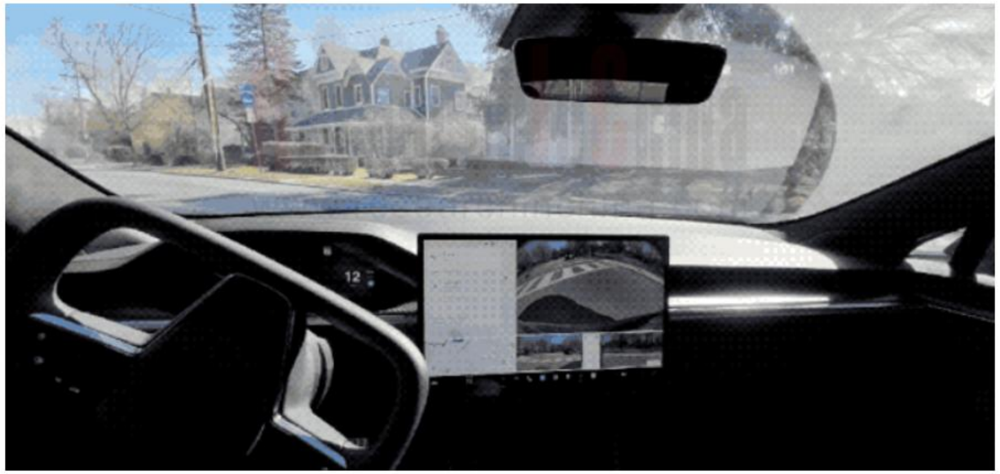
Comparison between Conventional Multi-module Solutions and End-to-end Solution (Part)

Solution type	Modular autonomous driving	End-to-end autonomous driving
Drive type	Rule-based, coding for making rules.	Data-driven, using massive data to train the system.
Task mode	Multi-task learning	Single-task learning
Model	There are several modules, each adopting an independent model.	Use an end-to-end integrated foundation model
Generalization	Poor	Good

Source: ResearchInChina

Some OEMs' Planning for End-to-end Solution Implementation and Mass Production

Actual Test of FSD V12.3



Some OEMs' Planning for End-to-end Solution Implementation and Mass Production

OEM	Application Time	Status Quo of Implementation	Solution Features
NIO	H1 2024	R&D started in H2 2023, and mass production is expected in 2024.	-
Xpeng	2024	The plan to introduce the solution on vehicles was announced in January 2024.	Build a 600 PFLOPS GPU cluster for training.
Li Auto	H1 2024	The foundation model was launched in the first half of the year, and the end-to-end solution is planned to reach L3.	Full process modeling
Xiaomi	2024	In late 2023, an end-to-end perception and decision model was announced. In March 2024, Xiaomi SU7 was equipped with the model.	Generate road topology in real time; recognize static agents in real time
Geely	2024	Cooperation with PhiGent Robotics; expected SOP in 2024	Use the dynamic scene graph to predict possible collisions of agents
Jiyue	2024	Iterate the VTA foundation model, and develop and train the BEV end-to-end perception model	Realize coverage of all road elements; generate road topology in real time

Source: ResearchInChina

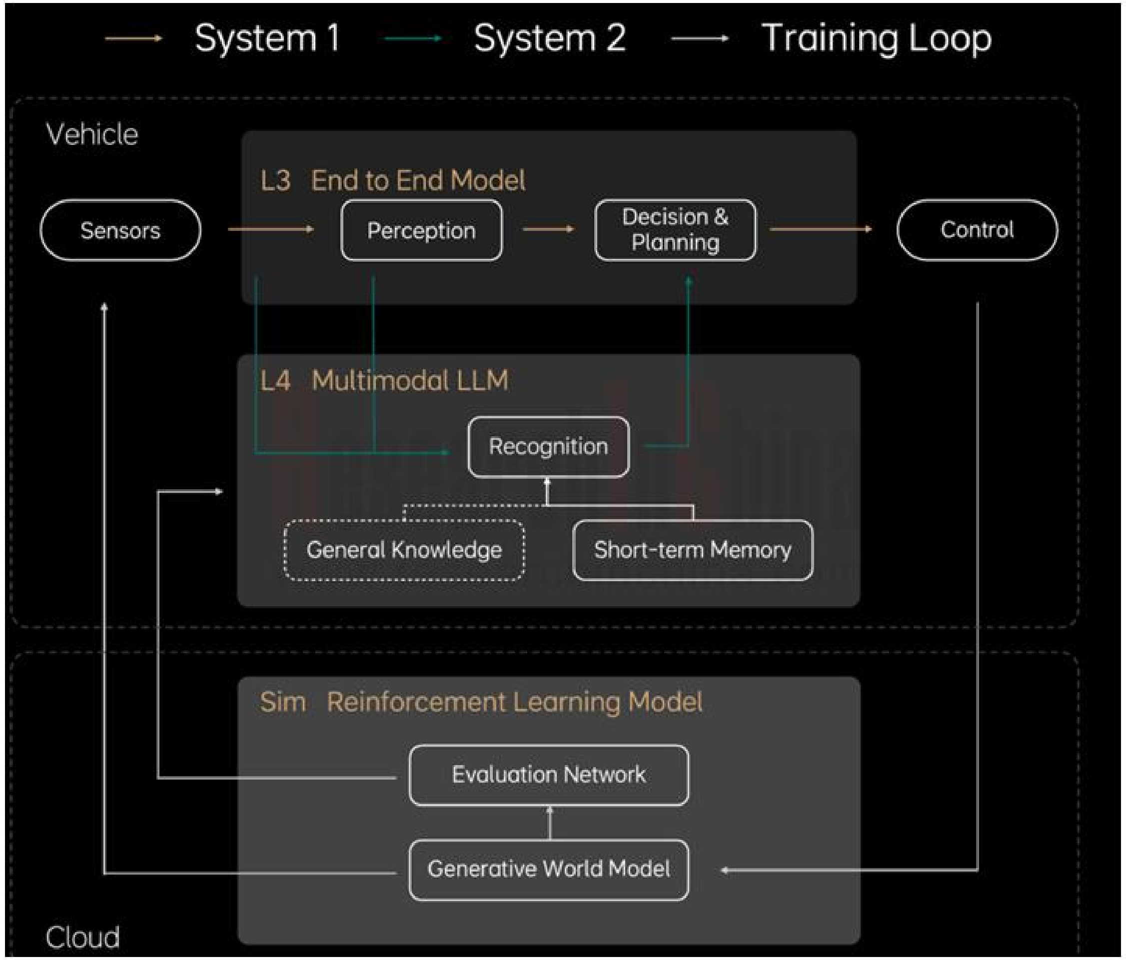
Li Auto's end-to-end solution

Li Auto believes that a complete end-to-end model should cover the whole process of perception, tracking, prediction, decision and planning, and it is the optimal solution to achieve L3 autonomous driving. In 2023, Li Auto pushed AD Max3.0, with overall framework reflecting the end-to-end concept but still a gap with a complete end-to-end solution. In 2024, Li Auto is expected to promote the system to become a complete end-to-end solution.

Li Auto's autonomous driving framework is shown below, consisting of two systems:
Fast system: System 1, Li Auto's existing end-to-end solution which is directly executed after perceiving the surroundings.
Slow system: System 2, a multimodal large language model that logically thinks and explores unknown environments to solve problems in unknown L4 scenarios.

In the process of promoting the end-to-end solution, Li Auto plans to unify the planning/forecast model and the perception model, and accomplish the end-to-end Temporal Planner on the original basis to integrate parking with driving.

Li Auto's Autonomous driving Framework



Source: Li Auto

Data becomes the key to the implementation of end-to-end solutions

2. Data becomes the key to the implementation of end-to-end solutions.

The implementation of an end-to-end solution requires processes covering R&D team building, hardware facilities, data collection and processing, algorithm training and strategy customization, verification and evaluation, promotion and mass production. Some of the sore points in scenarios are as shown in the table:c

The integrated training in end-to-end autonomous driving solutions requires massive data, so one of the difficulties it faces lies in data collection and processing.

First of all, it needs a long time and may channels to collect data, including driving data and scenario data such as roads, weather and traffic conditions. In actual driving, the data within the driver's front view is relatively easy to collect, but the surrounding information is hard to say.

During data processing, it is necessary to design data extraction dimensions, extract effective features from massive video clips, make statistics of data distribution, etc. to support large-scale data training.

Some Sore Points in Implementation of End-to-end Solutions in Scenarios

Difficulty	Specific impact
Computing power	End-to-end solutions adopt a complex network structure, which needs massive data for training and high-compute chips. Insufficient compute may lead to the simplification of network structure design and algorithms, which then affects the overall performance of end-to-end solutions.
	The operation of end-to-end solutions requires high frame rate and low latency, but the low computing power at the vehicle end makes it difficult to apply the solutions to vehicles, and large-scale pruning and other processes are thus a must.
Data	Data is difficult to acquire: • The collected data is not true enough and is in skewed distribution, making it difficult to carry out large-scale end-to-end training. • Simulation and model training require special scenario data (e.g., corner cases and tunnel scenarios), which occur infrequently and are difficult to gather into data sets.
	Low data quality and difficult processing: • It is difficult to control the quality of data collected in shadow mode of production vehicles, and more invalid data may be collected instead. • Data cleaning and feature extraction involve a lot of manpower and material costs.
Training strategy	• How to arrange the integrated annotation strategy • How to obtain the true value of multi-label probability distribution of end-to-end output
Verification & assessment	It is difficult to conduct open-loop evaluation.
Interpretability	Fail to meet expectations
Team building	• End-to-end technical staff are inadequate • The current organizational division of labor is incompatible with that under the end-to-end architecture, and it is necessary to redistribute the organizational structure and responsibilities.

Source: ResearchInChina

As of March 2024, DeepRoute.ai's end-to-end autonomous driving solution has been designated by Great Wall Motor and involved in the cooperation with NVIDIA. It is expected to adapt to NVIDIA Thor in 2025. In the planning of DeepRoute.ai, the transition from the conventional solution to the "end-to-end" autonomous driving solution will go through sensor pre-fusion, HD map removal, and integration of perception, decision and control.

Data Layout of DeepRoute.ai

Link	Layout
Data acquisition	• Source: Cooperative automakers • Target data: Screen out the driving data of drivers with more than 6 years of driving experience and no violation of traffic rules within 3 years on different complex road sections; collect their steering wheel angle and speed, and pedal opening and speed; train the model according to the driving environment at that time.
Data processing	Use the data processing experience in iteration of foundation models.

Source: ResearchInChina

DriveDreamer, an autonomous driving model of GigaStudio, is capable of scenario generation, data generation, driving action prediction and so forth. In the scenario/data generation, it has two steps:

When involving single-frame structural conditions, guide DriveDreamer to generate driving scenario images, so that it can understand structural traffic constraints easily. Extend its understanding to video generation. Using continuous traffic structure conditions, DriveDreamer outputs driving scene videos to further enhance its understanding of motion transformation.

DriveDreamer Is Capable of Continuous Driving Video Generation and Seamless Alignment with Text Prompts and Structured Traffic Restrictions

Source: GigaStudio

3. End-to-end solutions accelerate the application of embodied robots.

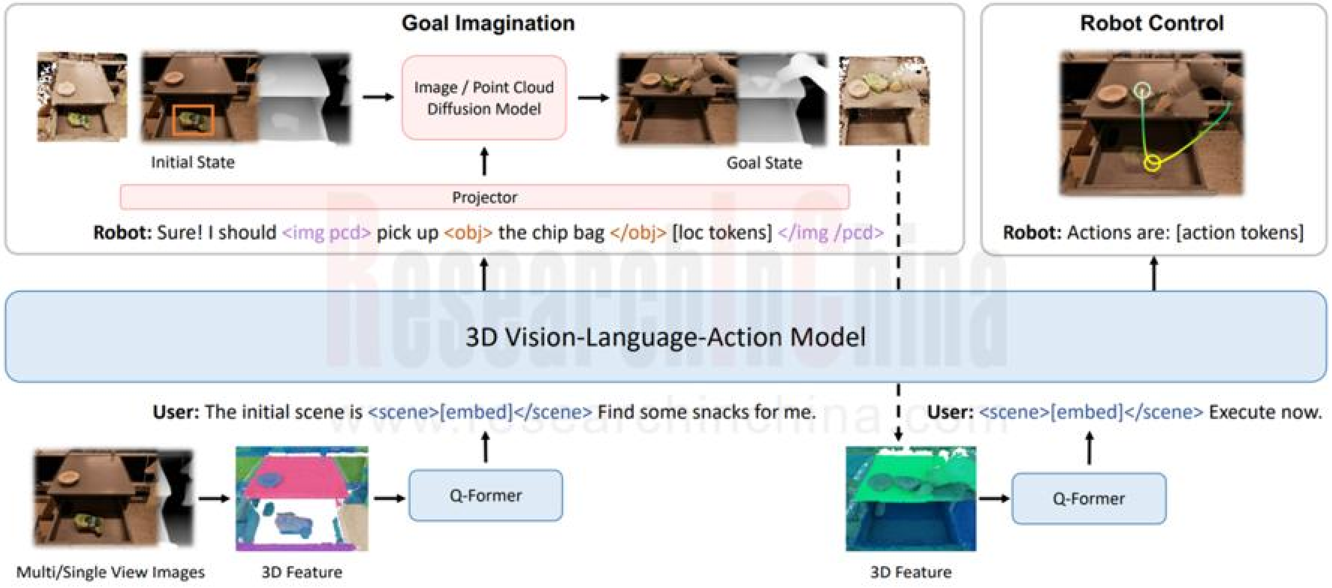
In addition to autonomous vehicles, embodied robots are another mainstream scenario of end-to-end solutions. From end-to-end autonomous driving to robots, it is necessary to build a more universal world model to adapt to more complex and diverse real application scenarios. The development framework of mainstream AGI (General Artificial Intelligence) is divided into two stages:

Stage 1: the understanding and generation of basic foundation models are unified, and further combined with embodied artificial intelligence (embodied AI) to form a unified world model;

Stage 2: capabilities of world model + complex task planning and control, and abstract concept induction gradually evolve into the era of the interactive AGI 1.0.

In the landing process of the world model, the construction of an end-to-end VLA (Vision-Language-Action) autonomous system has become a crucial link. VLA, as the basic foundation model of embodied AI, can seamlessly link 3D perception, reasoning and action to form a generative world model, which is built on the 3D-based large language model (LLM) and introduces a set of interactive markers to interact with the environment.

3D-VLA solution



Source: University of Massachusetts Amherst, MIT-IBM Watson AI Lab and Other Institutions.

Some manufacturers of humanoid robots adopt end-to-end solutions

As of April 2024, some manufacturers of humanoid robots adopting end-to-end solutions are as follows:

How to apply end-to-end solutions to some embodied robots

Vendor	Robot	End-to-end technology application
Figure	Figure 01	For E2E-VLM configured with OpenAI, all behaviors are driven by the neural network visual motion transformer strategy, and pixels are directly mapped to actions.
Deepmind	RT-2	The VLA model uses an end-to-end solution to directly output robot actions.
Tesla	Optimus Gen 2	It features end-to-end learning, and the algorithm as a whole is very similar to the E2E-AD solution. Next, it may carry the General World Model built by Tesla.
Udeer·AI	Udeer·AI Intelligent Cleaning Robot	The end-to-end Large Physical Language Model (LPLM) is used

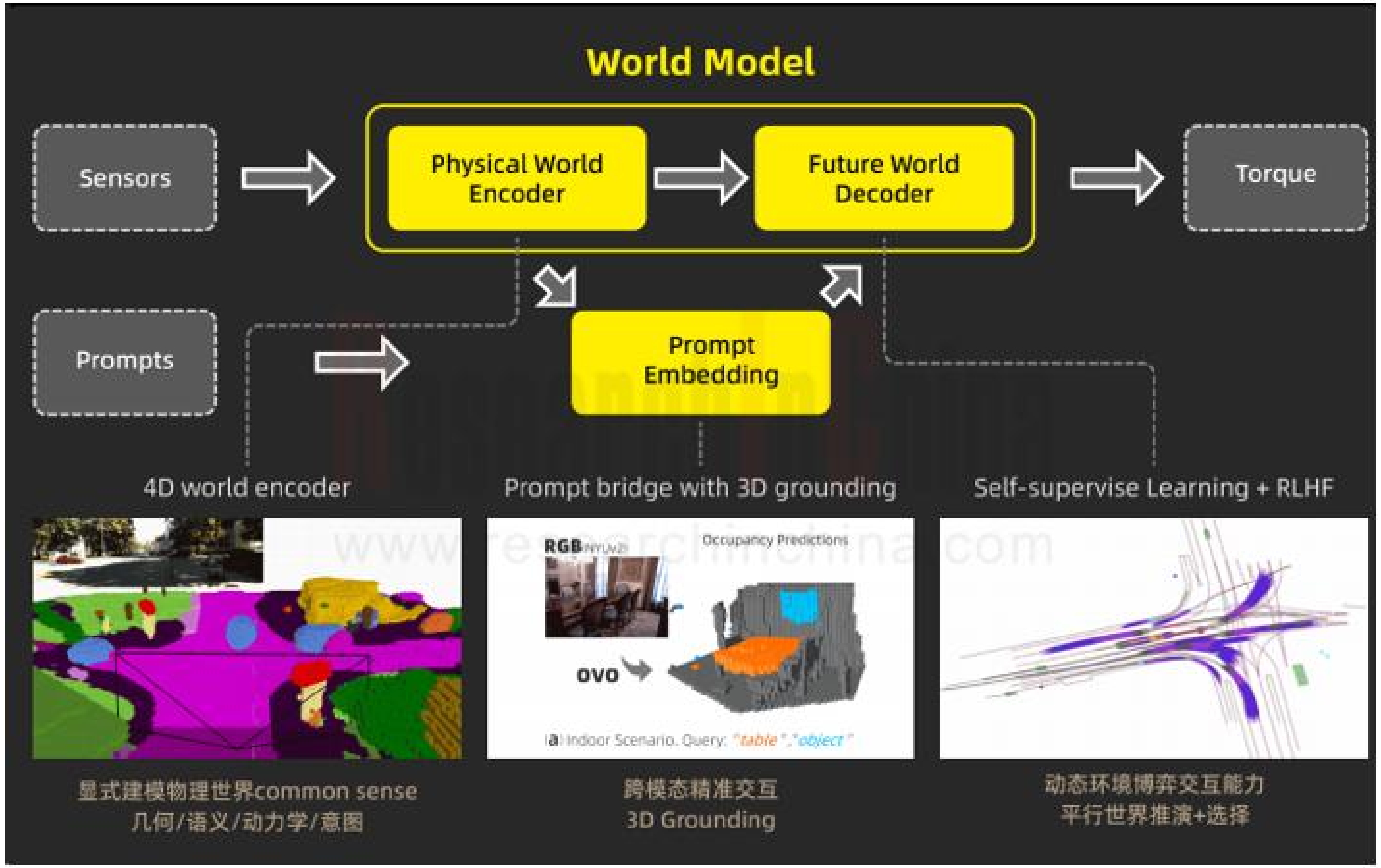
Source: ResearchInChina

For example, Udeer·AI's Large Physical Language Model (LPLM) is an end-to-end embodied AI solution that uses a self-labeling mechanism to improve the learning efficiency and quality of the model from unlabeled data, thereby deepening the understanding of the world and enhancing the robot's generalization capabilities and environmental adaptability in cross-modal, cross-scene, and cross-industry scenarios.

LPLM abstracts the physical world and ensures that this kind of information is aligned with the abstract level of features in LLM. It explicitly models each entity in the physical world as a token, and encodes geometric, semantic, kinematic and intentional information.

In addition, LPLM adds 3D grounding to the encoding of natural language instructions, improving the accuracy of natural language to some extent. Its decoder can learn by constantly predicting the future, thus strengthening the ability of the model to learn from massive unlabeled data.

Architecture of LPLM



Source: Udeer·AI

Table of Content (1)

Table of Content (2)

Table of Content (3)

Contact

Beijing Headquarters

TEL: 13718845418
Email: report@researchinchina.com

Website: ResearchInChina

WeChat: Zuosiqiche

Chengdu Branch

TEL: 028-68738514
FAX: 028-86930659

zuosiqiche