

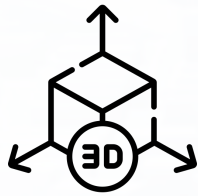
# 技术趋势分析

# 《端到端具身智能体》2025 八大技术趋势



## 世界模拟引擎

生成 / 重建 / 闭环反馈



## 空间智能

空间感知 / 推理想象



## 群体智能

多智能体 / 车路云协同



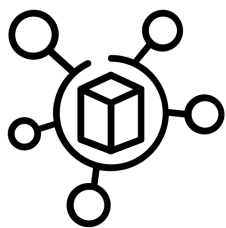
## 安全与风险

可解释性 / 价值对齐

当下趋势

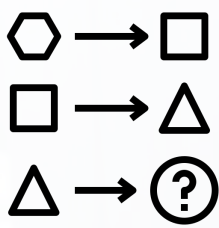
端到端、多模态、具身智能体大模型

**AGI**



## 涌现能力

海量数据 / 统一表征



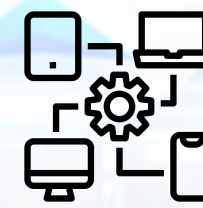
## 因果推理

长时记忆 / 层级规划



## 闭环反馈

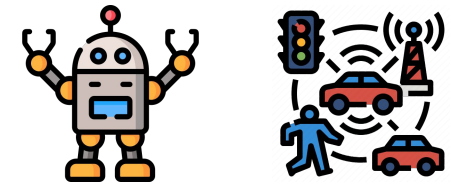
增量学习 / 终身学习



## 双系统

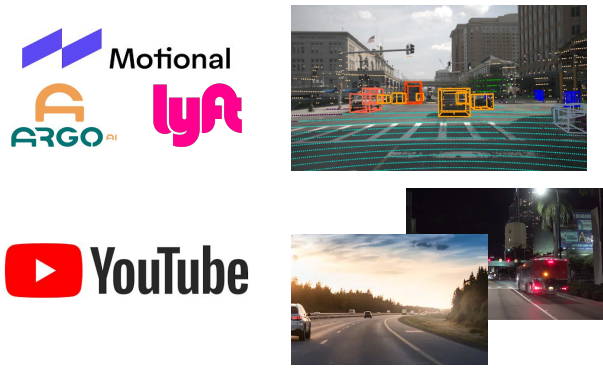
低功耗 / 低时延

# Towards Intelligent, Reliable and Generalizable Autonomy

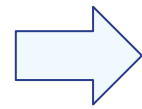
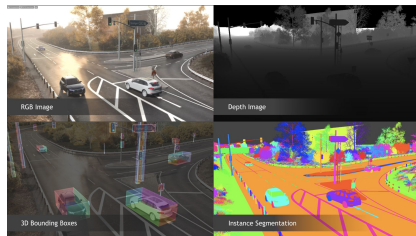


## Data-centric Pipeline

### Data Collection

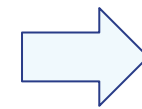
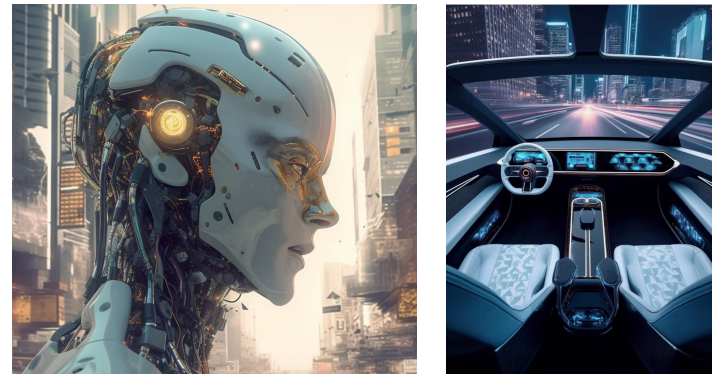


### Data Generation



## Pre-training DriveCore

### Foundation Model

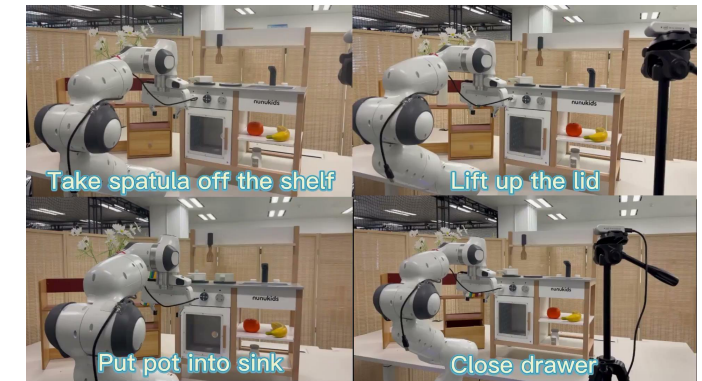


## Applications

### Autonomous Driving



### Embodied AI



## Integrated and General AGI for autonomous driving

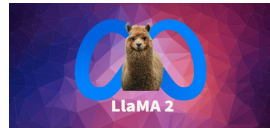
How to formulate?  
What's the objective goal?  
**GenAD (our on-going project)**

# Foundation Models

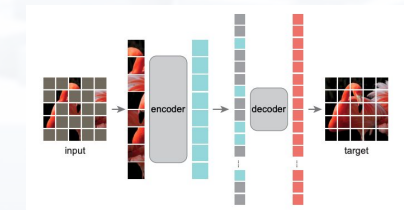
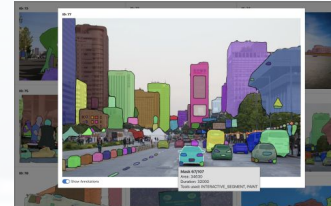
NLP (LLM)



ChatGPT



General CV



AD System

- Language Interpreter
- Driving Knowledge
- Any more?

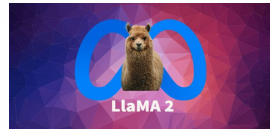
- Vision Abstractor
- Auto-labeling
- Any more?

# Foundation Models (cont'd)

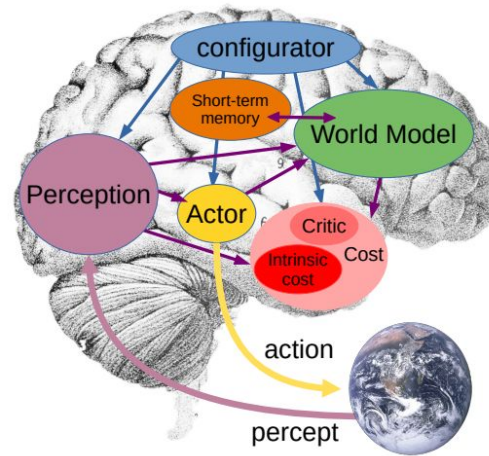
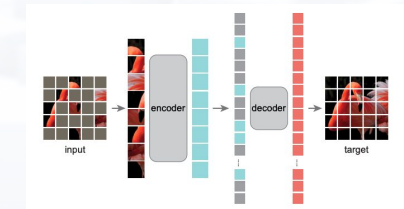
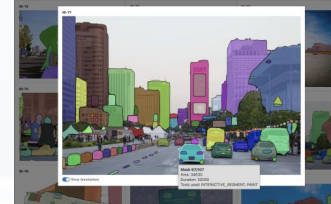
NLP (LLM)



ChatGPT



General CV

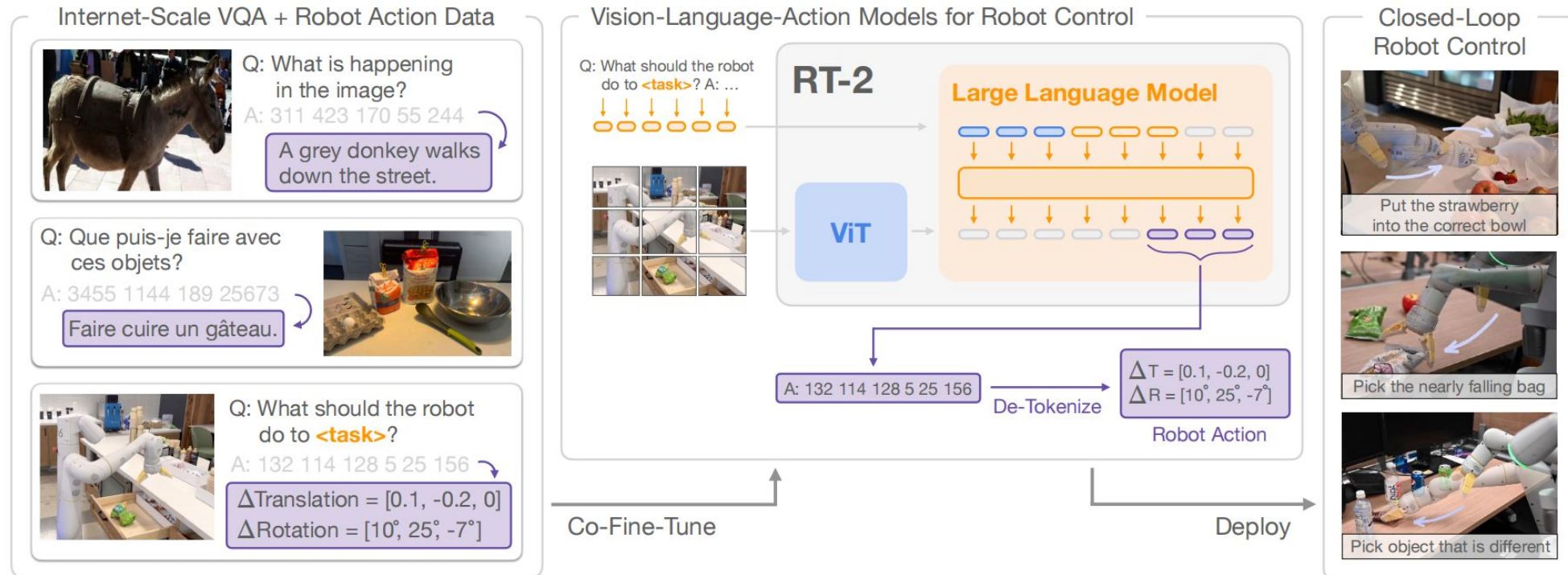


AD System



- Multimodality
- Intelligence
- Generalization

# Insight from Robotics / Embodied AI






- How vision-language models trained on Internet-scale data can be incorporated directly into **end-to-end robotic control**
- Goal: to **boost generalization** and enable emergent semantic reasoning

- Robotic tasks naturally fits into language at dissecting tasks step by step using language (prompt).
- Is it the **right way** to open the language tool box as does in **Robotics for Autonomous Driving**?






**Key ingredient(s): huge amount of data (not public) + language prompt to dissect tasks**

# Analogy to General Domains in CV/NLP/Robotics

## General Large Models

Domain	Method Abbreviation	Institute / Time	Data Scale	Public?
NLP (LLM)	GPT-4	 OpenAI / 2023.3	13T tokens	✗
	LLaMA 2	 Meta / 2023.7	2T tokens	✓
Vision	ViT-22B	 Google / 2023.2	4B images	✗
Vision Language (LLM backend)	BLIP-2	 Salesforce / 2023.1	129M images-text pairs	✓

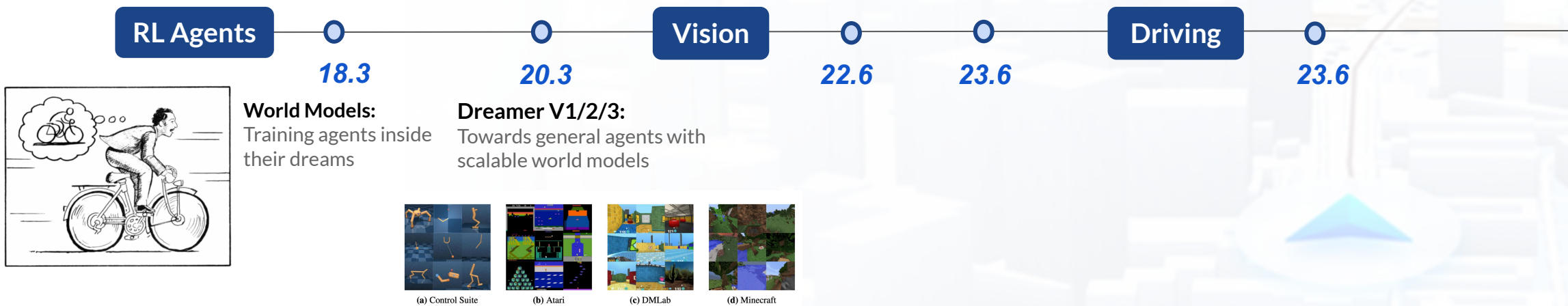
## Industrial Large Models (Application)

Autonomous Driving nuScenes: 4.5h	DriveAGI (GenAD)	 OpenDriveLab / 2023.11	2000 h videos (public)	✓
	GAIA-1	 Wayve / 2023.6	4700 h videos	✗
	World Model Demo	 Tesla / 2023.6	Unknown (Large-scale)	✗
Robotics (LLM backend)	PaLM-E	 Google / 2023.3	Unknown (Large-scale)	✗
	RT-2	 DeepMind / 2023.7	1B img-text pairs / 13 robots / 17 months	✗

If taken seriously for AD: lots of compute (at least 200 A100s) + massive amount of data (at least 10k hours of diverse, high-quality data)

# Trending: Recent Work on World Model

From simulated agents to real-world driving systems

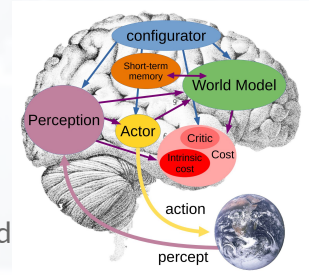




# Trending: Recent Work on World Model

From simulated agents to real-world driving systems

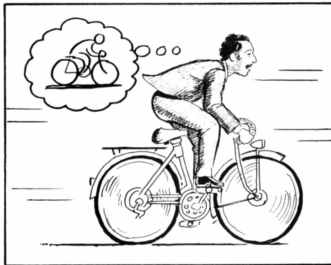
**Position Paper**  
(by LeCun)  
Positioning the developments of world models



RL Agents

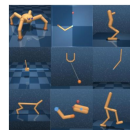
18.3

**World Models:**  
Training agents inside their dreams

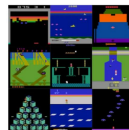


20.3

**Dreamer V1/2/3:**  
Towards general agents with scalable world models



(a) Control Suite



(b) Atari



(c) DMLab



(d) Minecraft

Vision

22.6

**I-JEPA:**  
Capturing visual knowledge in self-supervised manner



Driving

23.6

**Scaling up world models on large corpus of realistic driving videos**

**General World Model:** inhouse data collected around the globe

**GAIA-1:** 4700 hours of driving videos collected in London



World model to generate videos of the driving scenario. Then what?  
Is it useful for downstream tasks? (To be validated)

# Personal Take on Foundation Models into Autonomous Driving

---

End-to-end  
Auto Driving

## Pros:

1. Scalability
2. Global optimization
3. Easy-to-embed Infra

## For:

- Generalization/Robustness
- Performance
- Feasibility for deployment

# Personal Take on Foundation Models into Autonomous Driving

Research  OpenAI

## Video generation models as world simulators

### Mind-blowing Part



End-to-end  
Auto Driving

### Pros:

1. Scalability
2. Global optimization
3. Easy-to-embed Infra

### For:

- Generalization/Robustness
- Performance
- Feasibility for deployment

### Weakness Samples



Some rumors:

- 0.8M GPUs
- 50B video clips from Microsoft (ref: Youtube has 13B videos)
- This a side project from OpenAI

# Personal Take on Foundation Models into Autonomous Driving

Research  OpenAI

## Video generation models as world simulators

### Mind-blowing Part

End-to-end  
Auto Driving

### Pros:

1. Scalability
2. Global optimization
3. Easy-to-embed Infra

### For:

- Generalization/Robustness
- Performance
- Feasibility for deployment



### Weakness Samples



Some rumors:

- 0.8M GPUs
- 50B video clips from Microsoft (ref: Youtube has 13B videos)
- This a side project from OpenAI



**Towards Intelligent, Reliable and Generalizable System**

Data-driven

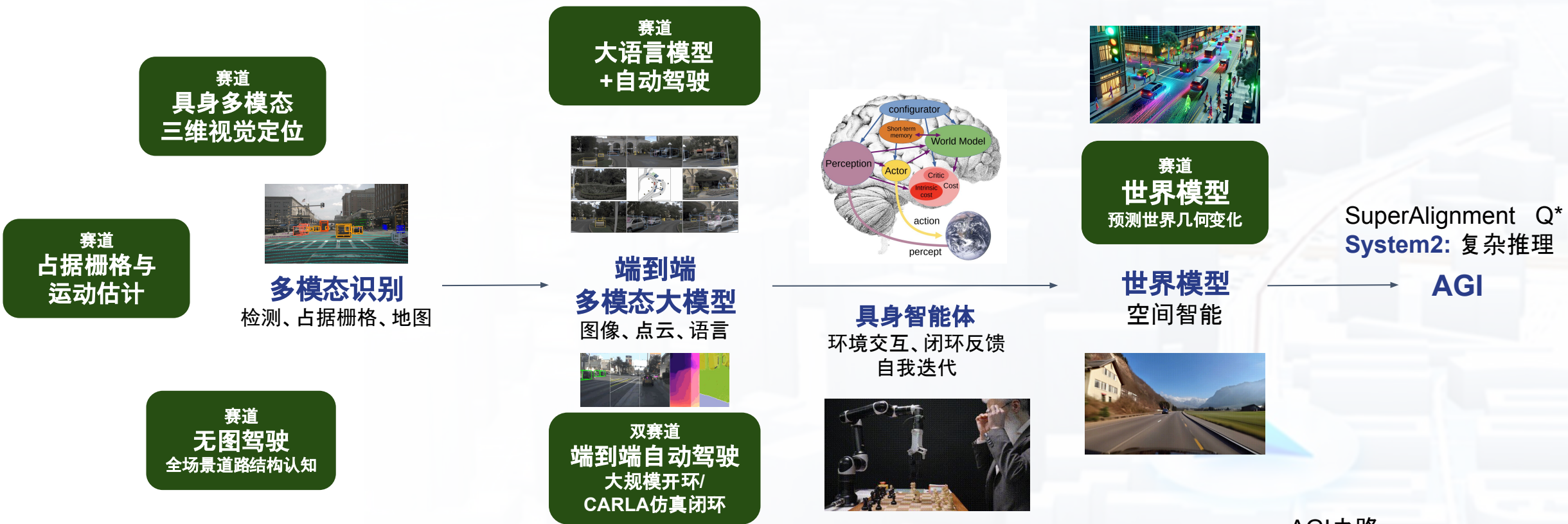
Alg-driven

Metric-driven

- Scaling data in all levels with self-supervised learning
- Simulating the physical world → **Interaction** between agents and env/physical world
- Rule of thumbs from foundation models
- Authentic evaluation metric. → Pixel-level **not** suffice. Actions require latent abstractions. Depends on task.
- Guarantee reliability and safety.

# 赛事总结

# 国际自动驾驶挑战赛 | 赛事背景



2020 - 2022  
早期阶段:感知、识别  
各模态简单融合

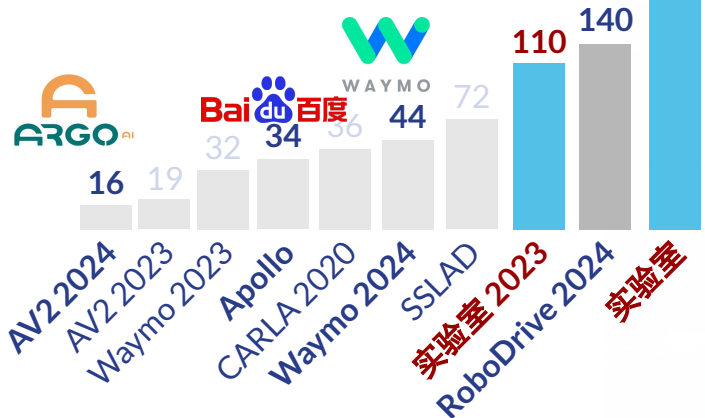
2023 - 2025  
端到端、多模态、大一统  
具身、空间智能

2025 -  
主动探索物理世界  
可交互、可反馈、可持续

AGI之路

# 国际自动驾驶挑战赛 | 赛事总结

## 参赛队伍数目



上海人工智能实验室  
Shanghai Artificial Intelligence Laboratory

> 26k  
关注人数

28  
国家/地区

> 92k  
网页浏览量

3000+  
提交次数

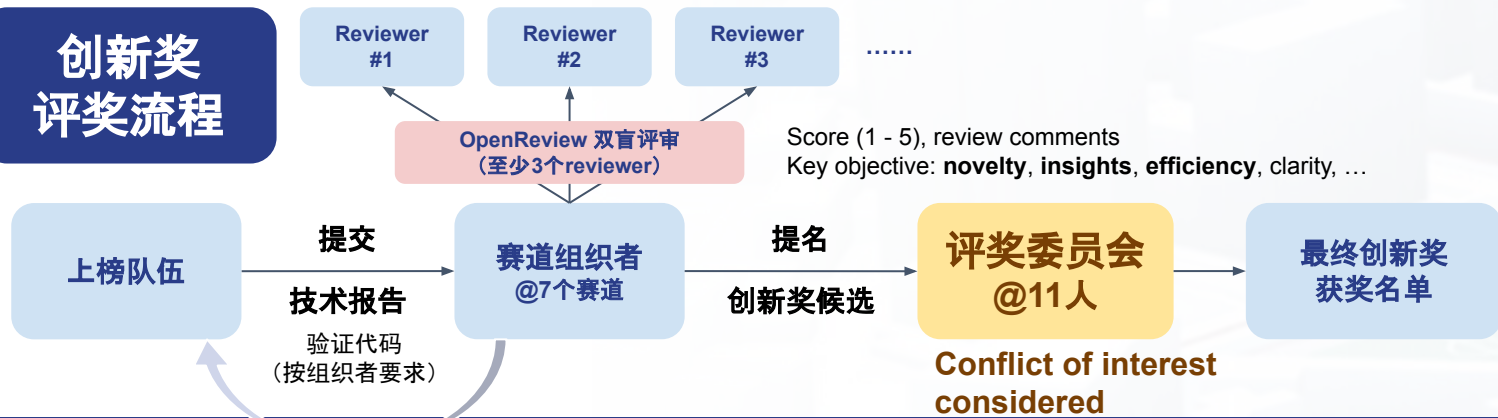
> 117k  
社交媒体热度

480+  
注册参赛队伍

## 全球参与



## 创新奖评奖流程



Reviewer #1, Reviewer #2, Reviewer #3, ...  
OpenReview 双盲评审 (至少3个reviewer)  
Score (1 - 5), review comments  
Key objective: **novelty, insights, efficiency, clarity, ...**

