



# 极端场景下的自动驾驶大模型 可信感知与安全决策

**Trustworthy** Perception and Safe Decision-Making of Large Models for **Autonomous Driving** in **Corner Cases**

CVCI 2025 | Multimodal Large Foundation Models and Autonomous Driving

讲者: 冷佳旭 (Jiaxu Leng)

会前 Workshop 报告

2025年10月28日星期二



# 报告提纲

可信场景分析

极端场景挑战

类脑智能计算

团队工作简介

未来发展趋势

# /01 自动驾驶 | 蓬勃发展，生机盎然

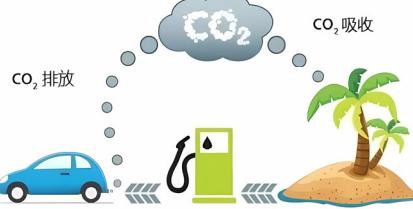
可信感知

极端场景

类脑计算

团队工作

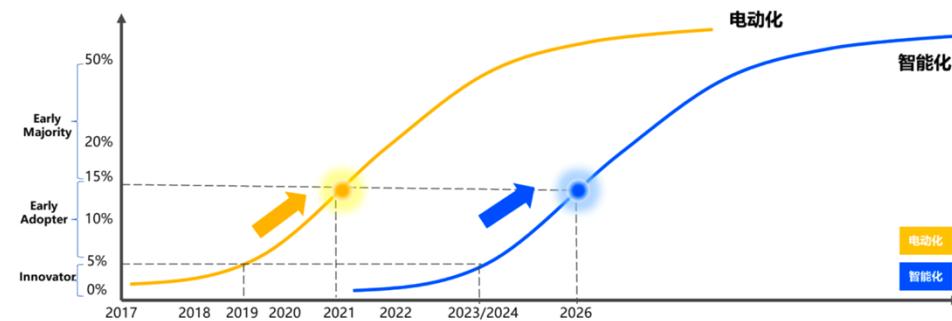
未来展望



“碳中和、碳达峰”  
双碳国家战略目标



汽车行业“新四化”  
(**电动化**、**网联化**、**智能化**、**共享化**)



个人安全    节省时间    方便出行    降低开支    绿色环保



聚焦“33618”现代制造业集群体系“提能级”  
重庆深入推进新型工业化  
发展新质生产力

重庆市建设世界级智  
能网联新能源汽车产  
业集群发展规划

重庆市自动驾驶和  
车联网创新应用行  
动计划



全国首个**自动驾驶**公交车项目落地重庆

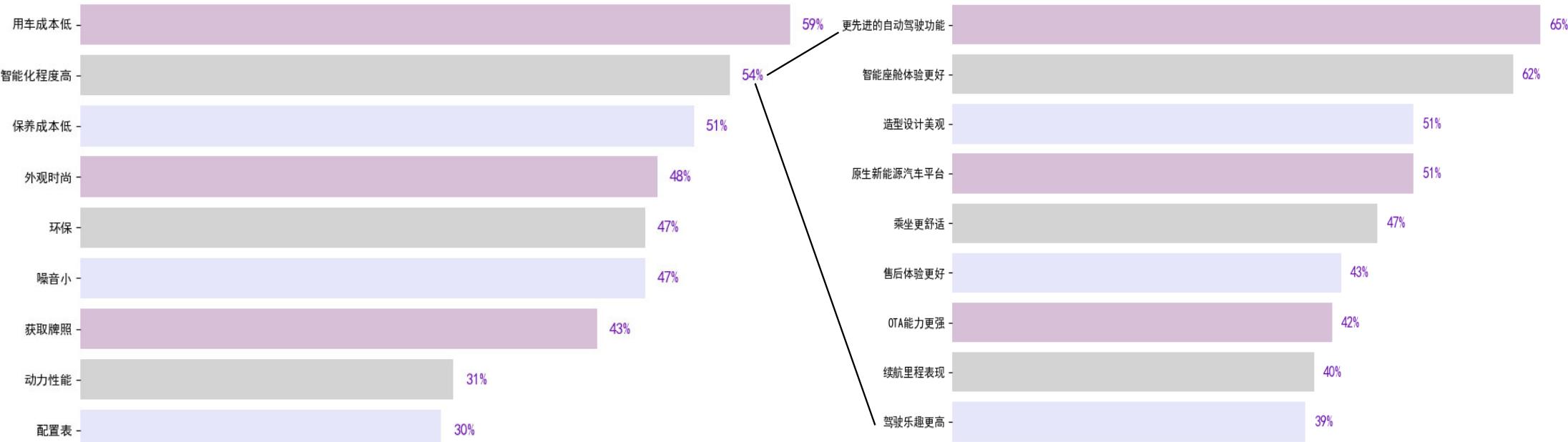
随着人工智能技术发展及国家地方的高度重视，  
汽车产业**智能化**成为电动化后又一个历史性机遇

# /01 汽车智能化成购车关键因素



- 国内汽车工业在电动化和智能化领域的快速发展，正改变消费者购车偏好，越来越多**消费者将智能化水平作为购车的关键因素**。
- 智能化的重要性日益凸显，推动汽车制造商加速智能技术的研发和投入。**智能化已成为新能源汽车竞争的主要领域**，未跟上步伐的品牌可能会逐渐失去市场竞争力。

购车时“智能化”考量因素高居第二



来源：麦肯锡《2024麦肯锡中国消费者洞察》

# /01 智能等级 → 人的参与程度

可信感知

极端场景

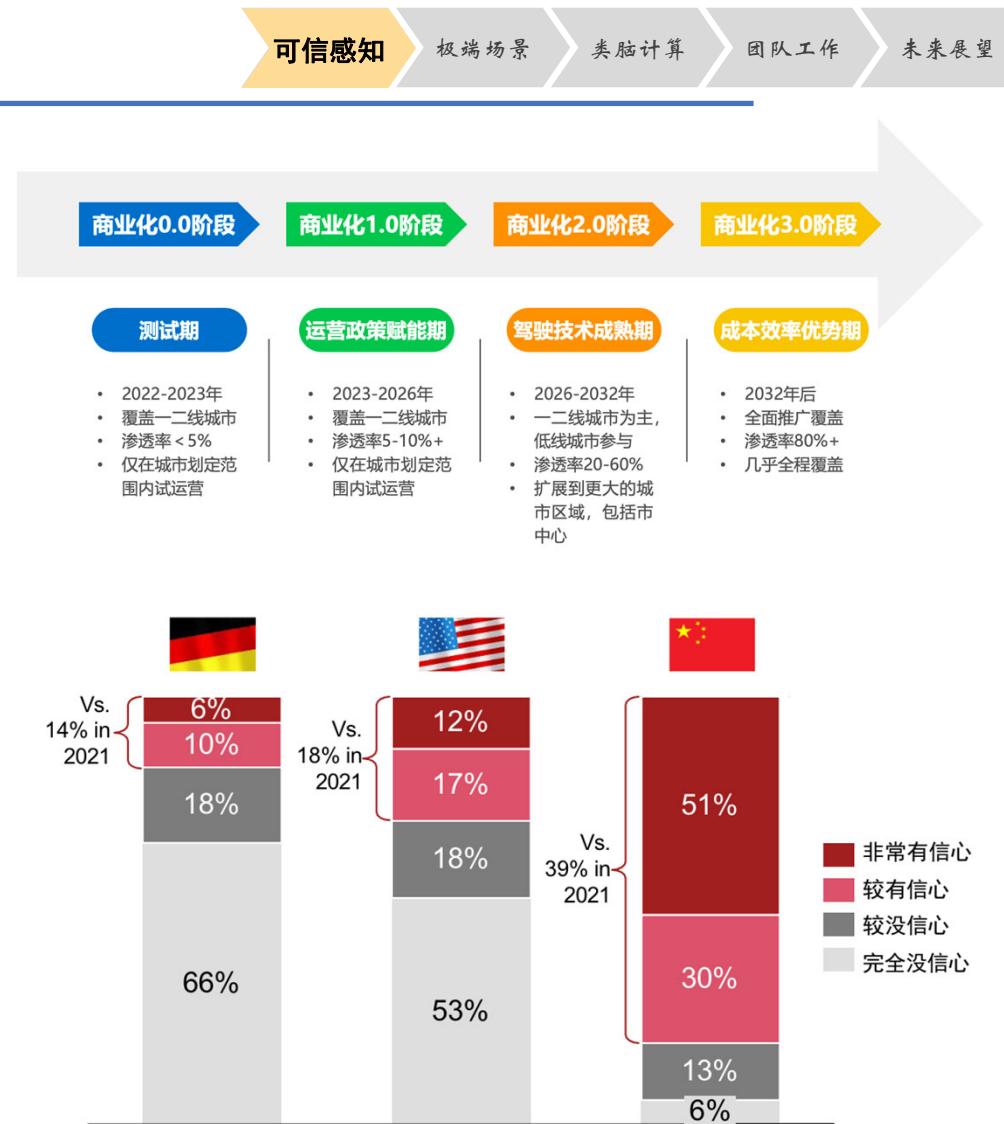
类脑计算

团队工作

未来展望



2022-2023年中国在售新车自动驾驶搭载率预测 @共研产业研究院&36氪研究院



2023 数字化汽车报告 @罗兰贝格 @思略特

# /01 极端场景下的感知失效与安全隐患

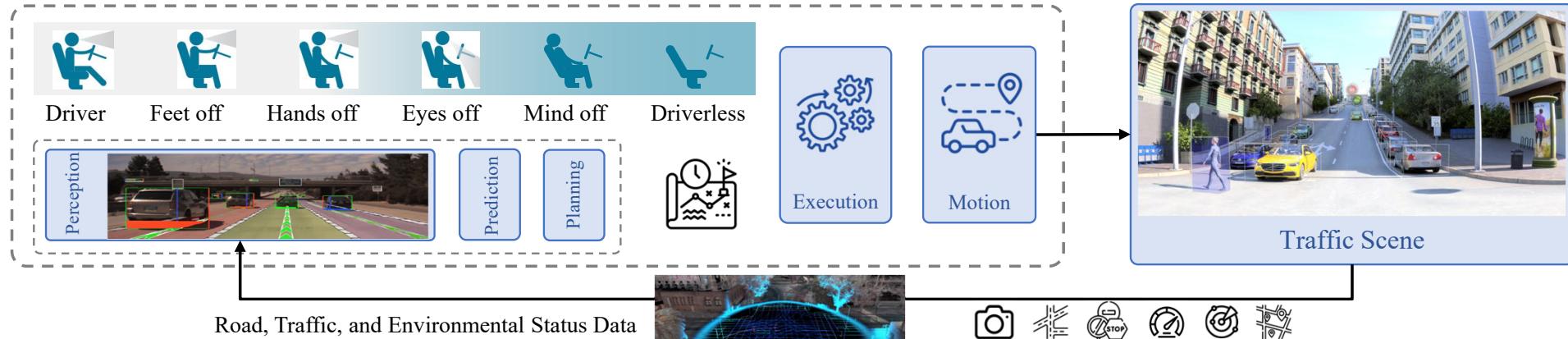
可信感知

极端场景

类脑计算

团队工作

未来展望



不可预测的极端情况可能导致无法接受的后果





# 报告提纲



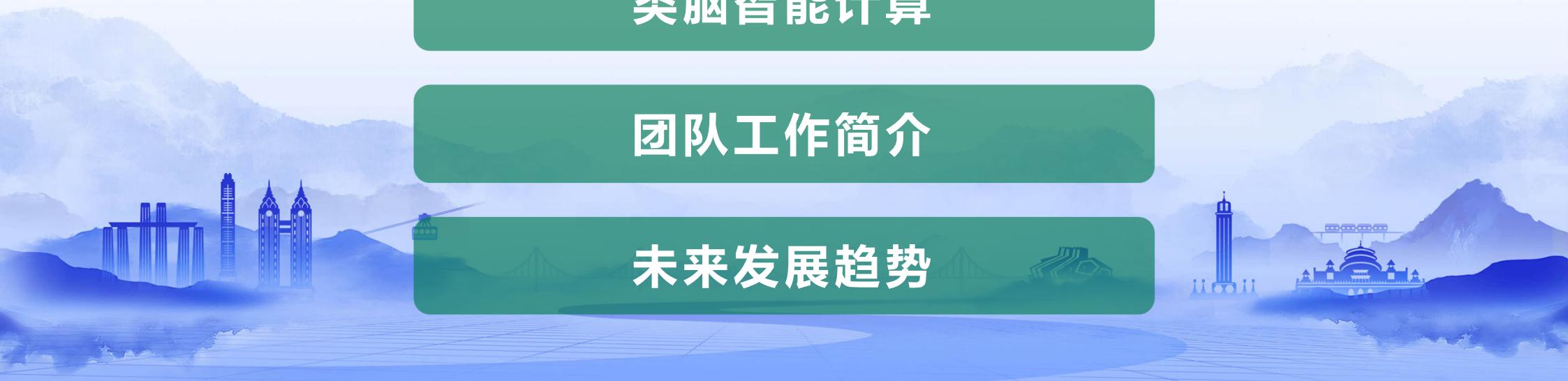
可信场景分析

极端场景挑战

类脑智能计算

团队工作简介

未来发展趋势



## /02 关键挑战 | 极端场景感知与理解

可信感知

极端场景

类脑计算

团队工作

未来展望



夜晚的城市街道

拥挤车道的下雨天



极端场景数据匮乏

极端场景难理解



决策可解释性差

## /02 面临挑战 ① | 极端场景数据匮乏

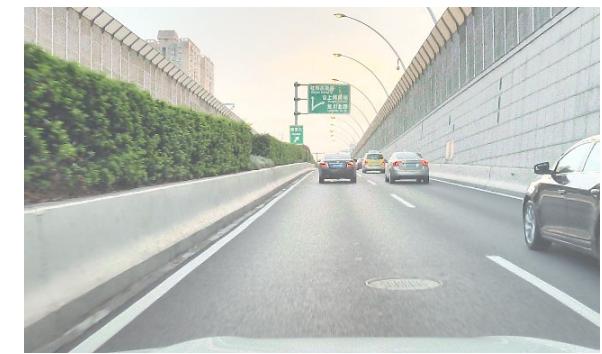
可信感知

极端场景

类脑计算

团队工作

未来展望



自动驾驶的数据采集与标注成本高，现有数据集无法穷尽真实世界的极端驾驶场景  
(如恶劣天气、低光照场景、光照突变、复杂交互等)。

## /02 面临挑战 ② | 极端场景难理解

可信感知

极端场景

类脑计算

团队工作

未来展望



在极端驾驶场景中，**语义信息复杂**，如道路边界与障碍物难以区分，或非规范交通行为（如电动车逆行、行人突然横穿）打破常规模式，导致模型在目标识别、语义理解上易发生**混淆**。

## /02 面临挑战 ③ | 决策可解释性差

可信感知

极端场景

类脑计算

团队工作

未来展望

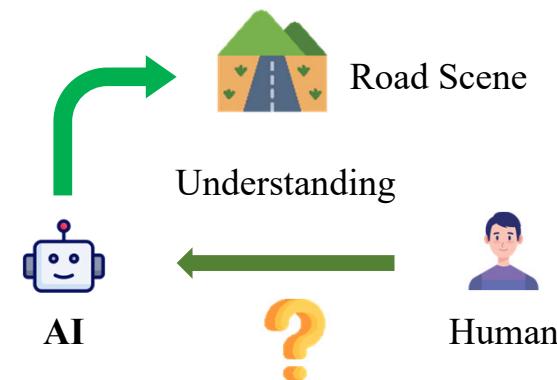


Why?  
E2EAD  
(端到端)



控制

可解释性差导致在极端场景下，端到端模型的**下限不可控**。例如，系统无法明确预知大风吹来的塑料袋以及纸箱子会诱导模型做出怎样的决策。



当前自动驾驶决策多为“**黑盒**”模型，其内部逻辑与机理**难以理解**。这种可解释性的缺乏不仅阻碍了系统调试、优化，更严重削弱了用户、监管及社会公众对其可靠性的信任。



# 报告提纲



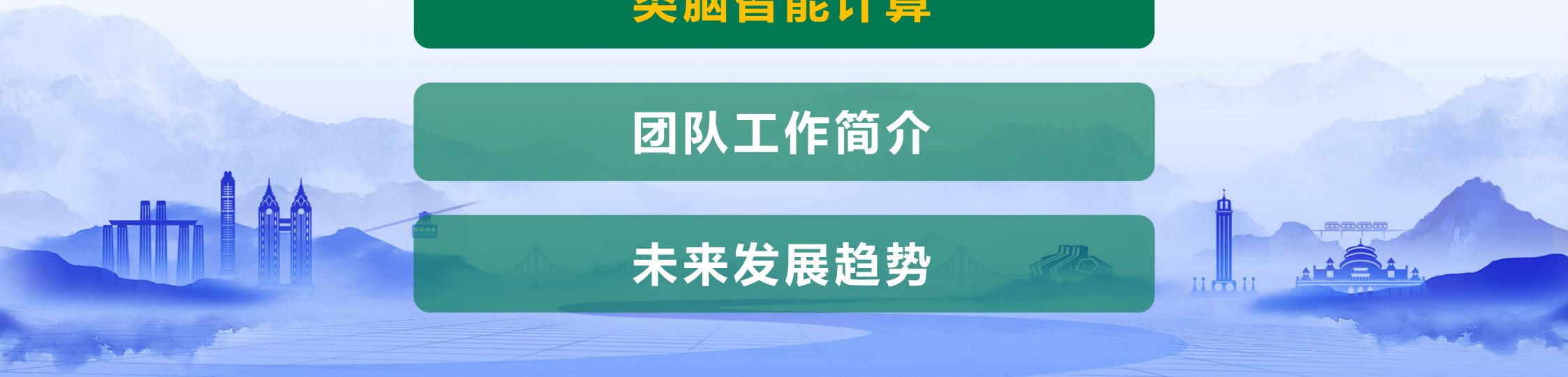
可信场景分析

极端场景挑战

类脑智能计算

团队工作简介

未来发展趋势



# /03 人工智能：模拟人的智能

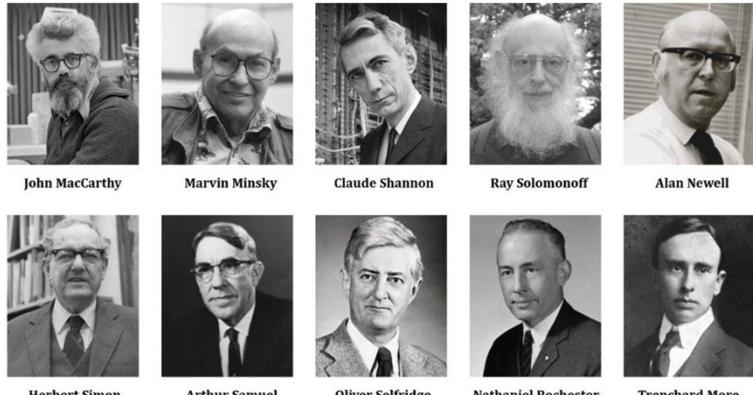
可信感知

极端场景

类脑计算

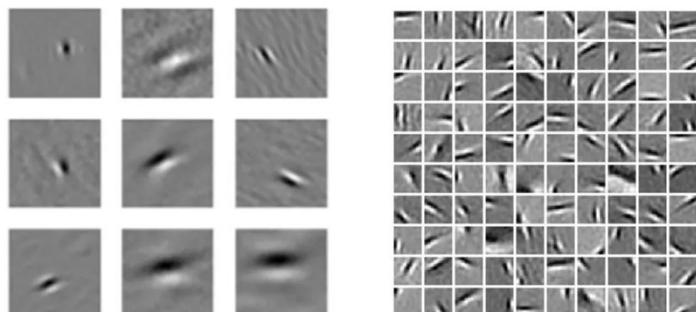
团队工作

未来展望



达特茅斯会议，1956

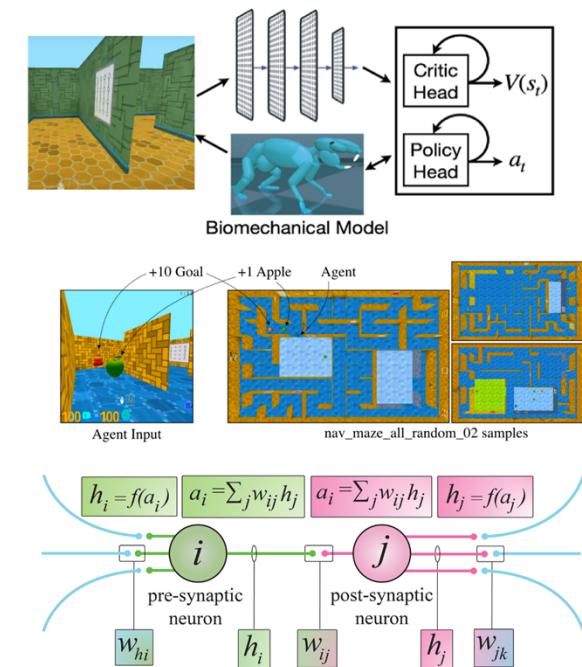
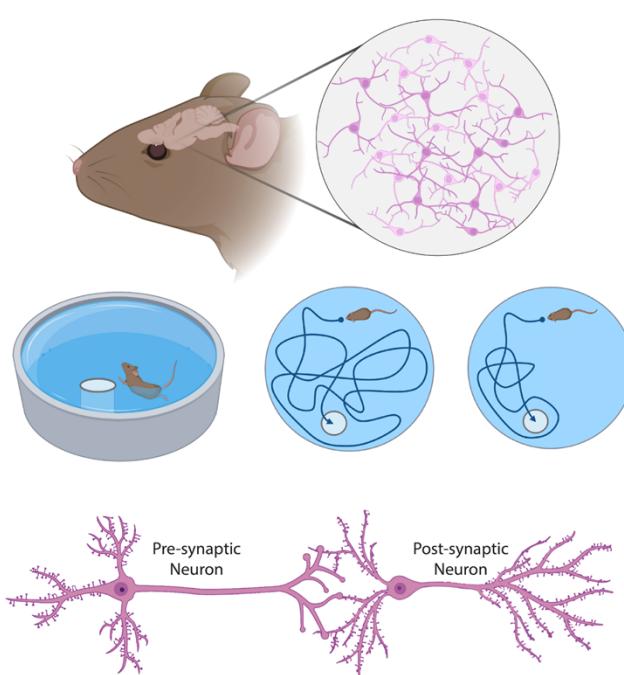
人工智能：让机器像人那样认知、思考和学习，  
即用计算机来模拟人的智能。



哺乳类初级视觉皮层简  
单细胞感受野

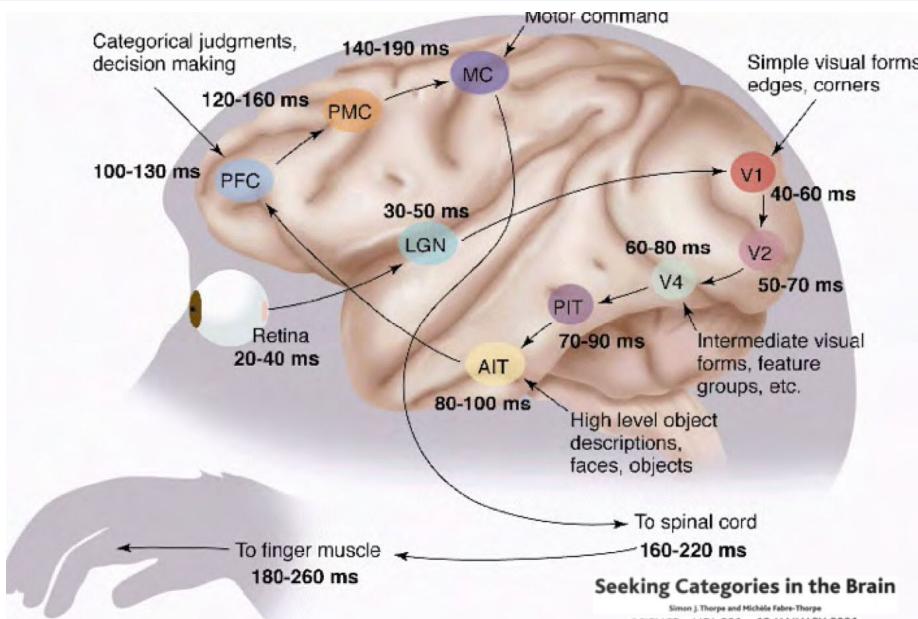
利用稀疏性训练得  
到的滤波器

Schmidgall, S., et al. "Brain-inspired learning in artificial neural networks: A review." *arXiv preprint arXiv:2305.11252*.

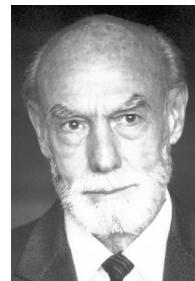


结构形式模仿 → 功能行为模仿 → 学习机制模仿

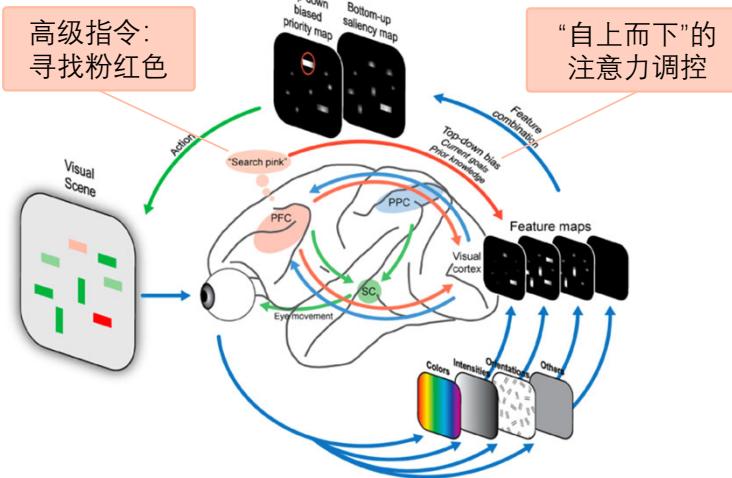
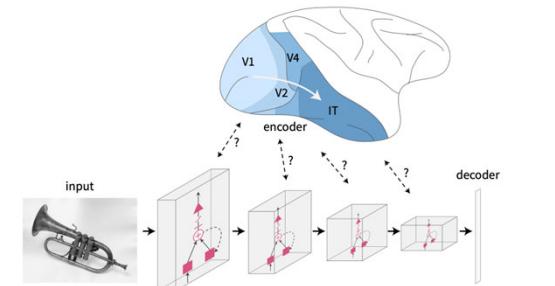
从信息处理与智能本质角度审视生物脑信息处理，借鉴其原理产生新型  
**类脑智能计算**技术，是实现人工智能创新的重要源泉



诺贝尔生理学或医学奖（1981年）  
@ David Hubel & Torsten Wiesel & Roger Sperry



从视觉处理机制可以看出高层的特征是低层特征的组合，从低层到高层的特征表达越来越抽象和概念化。



针对自动驾驶**极端场景数据匮乏**问题，借鉴**层级感知**机制，实现从低级特征到高级语义的多层融合与抽象表达。

Katsuki F, Constantinidis C. Bottom-up and top-down attention: different processes and overlapping neural systems. Neuroscientist 20:509-21, 2014.

Chen, Yao, et al. "Task difficulty modulates the activity of specific neuronal populations in primary visual cortex." Nature neuroscience 11.8 (2008): 974-982.

## /03 研究思路 (2) : 联想记忆

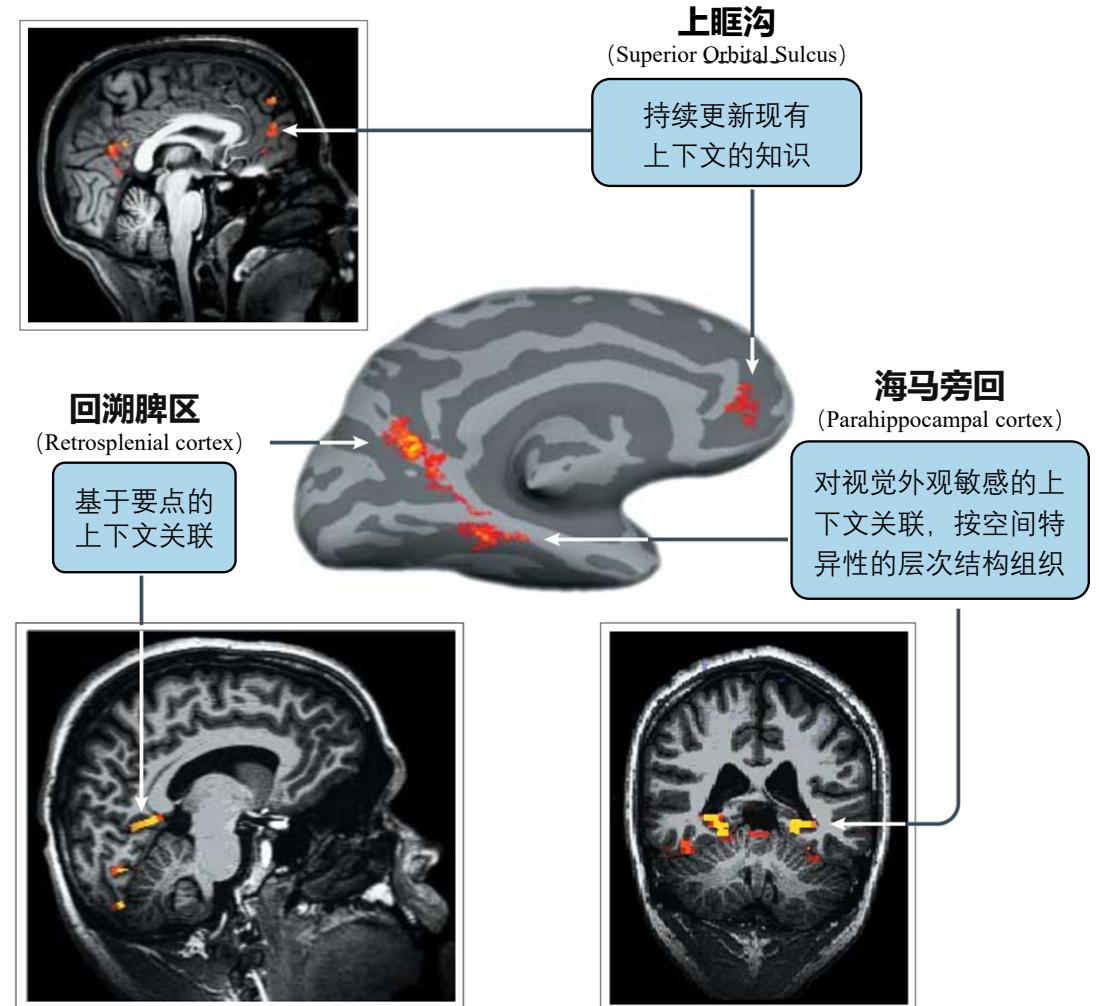
可信感知

极端场景

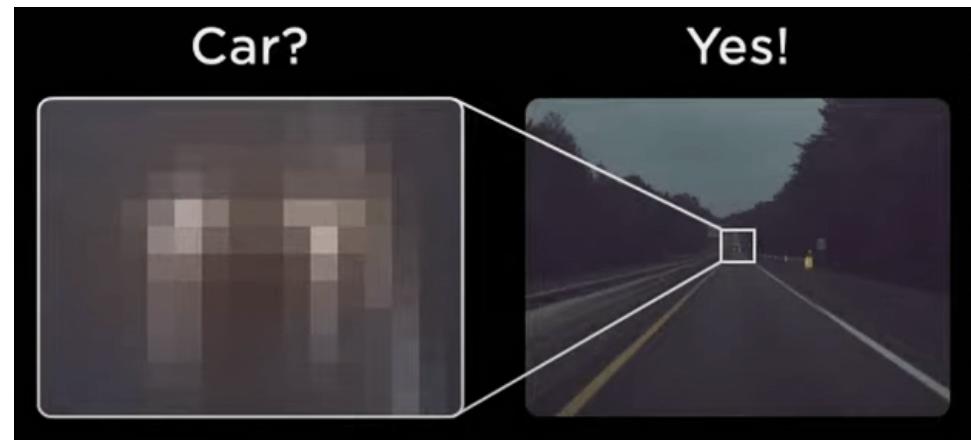
类脑计算

团队工作

未来展望



在人类感知中，基于视觉（object-VGG）、视觉语义（CLIP）和语义（SGPT）的深度特征表示与对象及其部件之间的相关性存在差异，**上下文语义信息**在识别物体时发挥着重要作用。



针对自动驾驶中**极端场景难理解**问题，借鉴**联想记忆**机制，从知识库中检索相近的知识，辅助极端场景理解。

Bar, Moshe. "Visual objects in context." *Nature Reviews Neuroscience* 5.8 (2004): 617-629.

Shoham, Adva, et al. "Using deep neural networks to disentangle visual and semantic information in human perception and memory." *Nature Human Behaviour* (2024): 1-16.

# /03 研究思路 (3) : 反馈学习

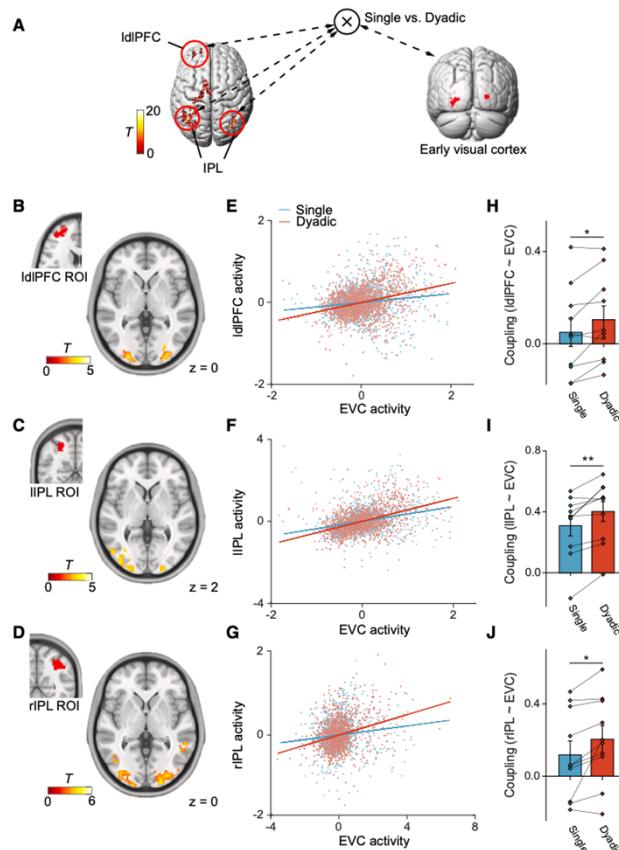
可信感知

极端场景

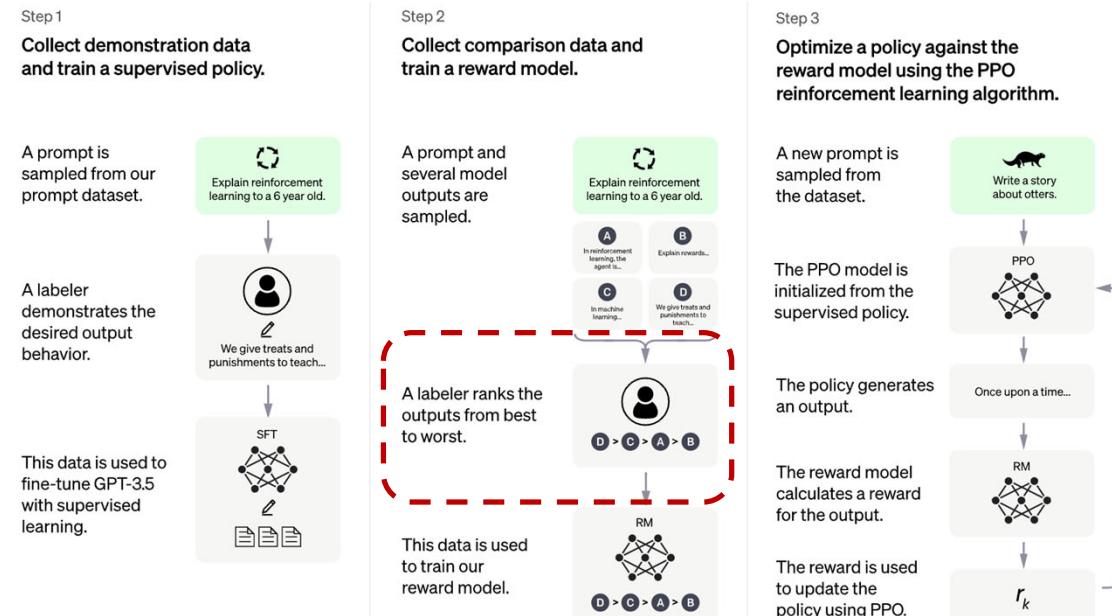
类脑计算

团队工作

未来展望



与单一训练相比，PPI分析发现在双人训练期间，作为种子区的IdIPFC (B)、IIPPL (C) 和rIPL (D) 与作为目标区的早期视觉皮层之间的功能联系增加。



在chatGPT中，人类反馈强化学习使语言模型能够将在文本数据的一般语料库上训练的模型与复杂的人类价值观的模型对齐。

针对自动驾驶**决策可解释性差**的问题，借鉴**反馈学习**机制，引入基于人类反馈的强化学习算法（RLHF），鼓励模型进行深入思考推理。

Zhang, Y., Bi, K., Li, J., Wang, Y., & Fang, F. (2023). Dyadic visual perceptual learning on orientation discrimination. *Current Biology*, 33(12), 2407-2416.



# 报告提纲



可信场景分析

极端场景挑战

类脑智能计算

团队工作简介

未来发展趋势



## /04 团队工作简介：类脑计算

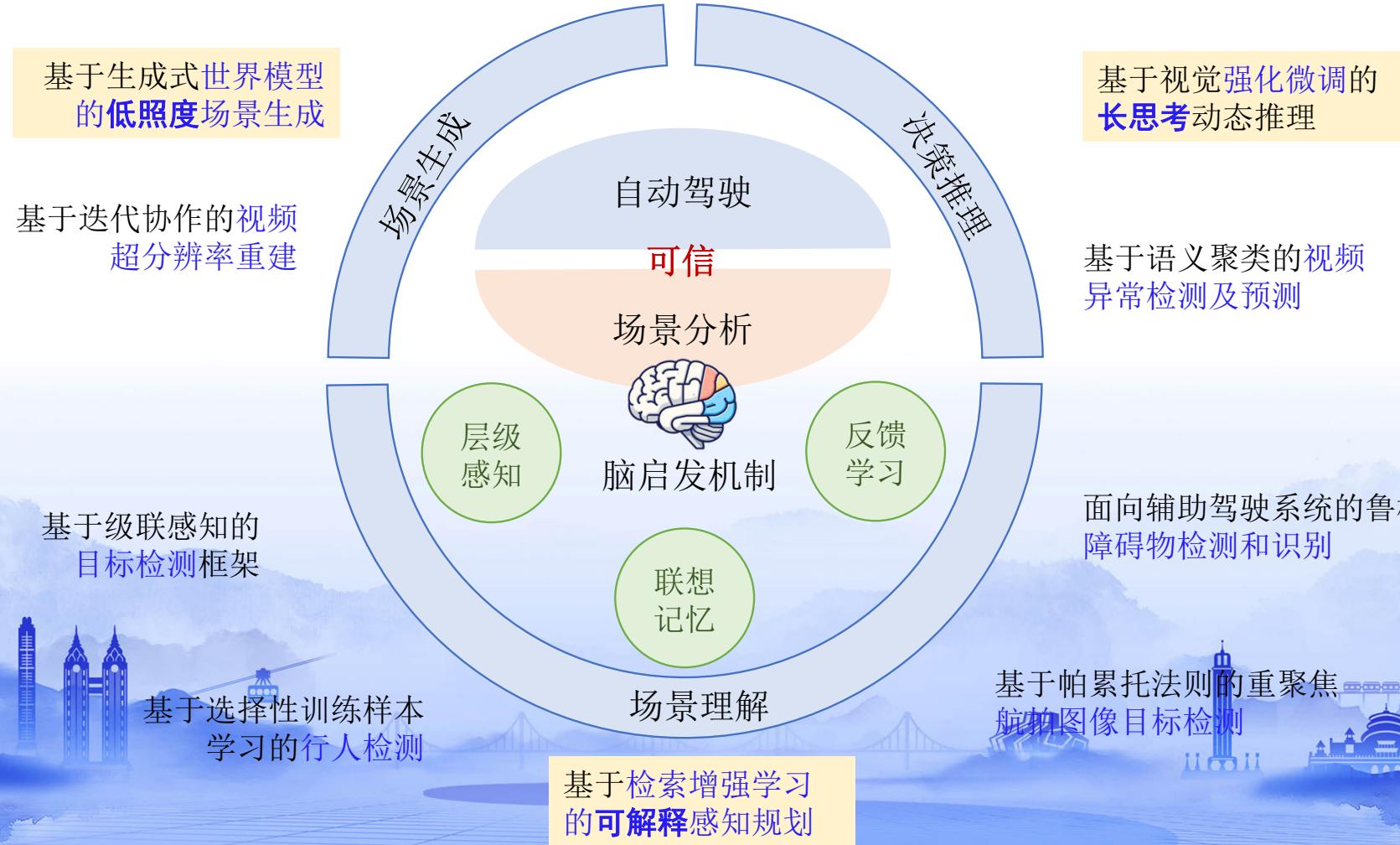
可信感知

极端场景

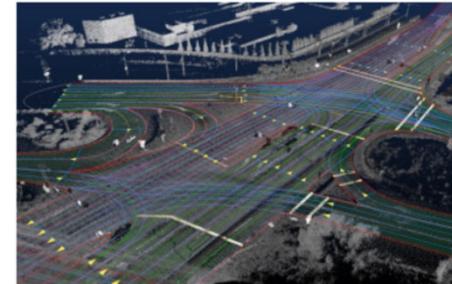
类脑计算

团队工作

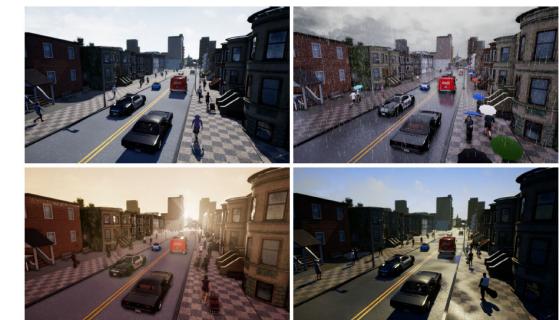
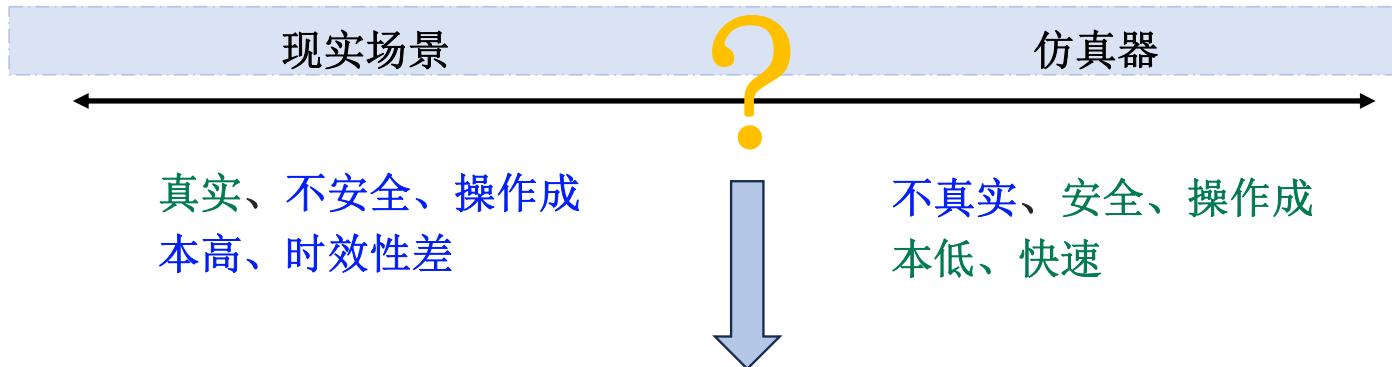
未来展望



- ✓ 场景生成：从源头出发解决高质量驾驶场景数据匮乏的问题



现实场景下的车辆数据采集与标注挑战：高成本、安全风险、数据时效滞后



依赖物理引擎的仿真平台  
(如CARLA) 的场景保真度有限

- ✓ 生成式世界模型：智驾数据可控生成，获取高价值训练数据，使能感知仿真

- 多数交通事故发生在**极端场景**（如低光照、极端天气条件等）下，自动驾驶系统的安全性验证需要大量且多样的极端驾驶场景数据，然而真实世界中此类极端场景数据**获取困难且成本高昂**。



雾天



雨天

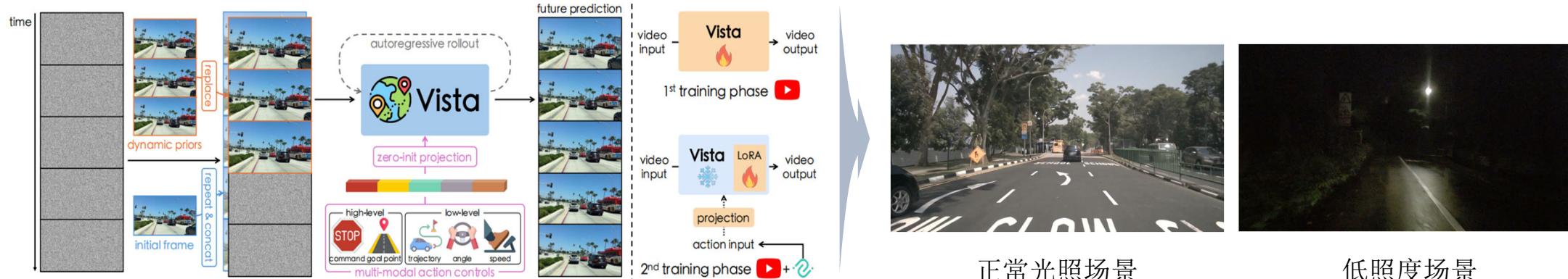


雪天



夜晚

- 现有方法对极端场景生成效果不佳，例如：**低照度场景**结构纹理模糊。



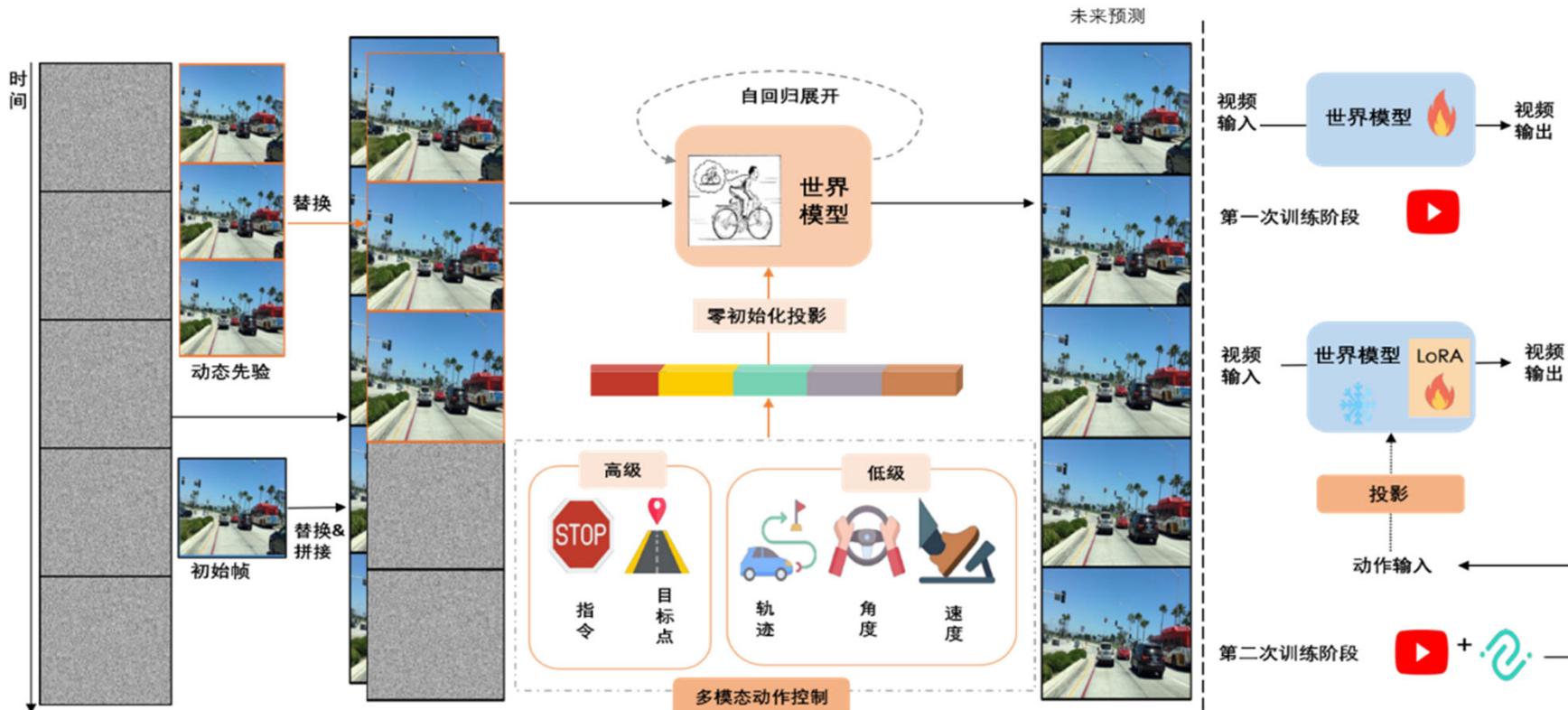
世界模型Vista 架构图[1]

[1] Vista: : A Generalizable Driving World Model with High Fidelity and Versatile Controllability. 2024, CVPR.

- 现有数据中低光照场景的数据较少，导致模型对低光照场景的学习不足。此外，与白天场景相比，**低照度场景的特征不够清晰**，难以提取丰富的驾驶场景信息。



- 研究动机：针对现有世界模型在生成低光照驾驶场景时出现**细节缺失、物体边缘模糊**以及**动态一致性不足**等问题，引入多粒度层级感知视角，探索全局结构理解—局部细节还原—时序动态保持三个层次上的协同建模，使模型能够在复杂低照度环境下感知多尺度语义与光照层次，从而提升生成场景的真实性、层次感与时空一致性。



- 基于**运动一致性约束**的损失函数，促进模型对动态实例和结构信息的学习理解，从而提升高分辨率下动态预测的**视频质量**
- 基于**先验信息**的视频预测方法，通过历史帧引导与高效学习策略，提升长时间预测的一致性，实现从**高层意图**到**低层操作**的精确且可控的**多层次**动作调控

□ 多重损失联合优化：

$$L_{diffusion} = \mathbb{E}_{z, \sigma, \hat{n}} [\sum_{i=1}^K \|D_\theta(\hat{n}_i; \sigma) - z_i\|^2],$$

$$L_{structure} = \mathbb{E}_{z, \sigma, \hat{n}} [\sum_{i=1}^K \|\mathcal{F}(D_\theta(\hat{n}_i; \sigma)) - \mathcal{F}(z_i)\|^2],$$

$$\begin{aligned} L_{dynamics} = & \mathbb{E}_{z, \sigma, \hat{n}} [\sum_{i=2}^K sg[\|D_\theta(\hat{n}_i; \sigma) - \\ & D_\theta(\hat{n}_{i-1}; \sigma) - (z_i - z_{i-1})\|] \odot \|D_\theta(\hat{n}_i; \sigma) - z_i\|^2], \end{aligned}$$

$$L_{Distribution} = \mathbb{E}_{z, \hat{n}} [W_2^2 (q_\theta(z|\hat{n}), p(z))],$$

其中， $W_2^2 = \|\mu_\theta - 0\|^2 + \text{tr}(\Sigma_\theta + I - 2(\Sigma_\theta^{1/2}))$

$$\begin{aligned} L_{final} = & L_{diffusion} + \alpha L_{dynamics} + \beta L_{structure} + \\ & \gamma L_{Distribution} \end{aligned}$$

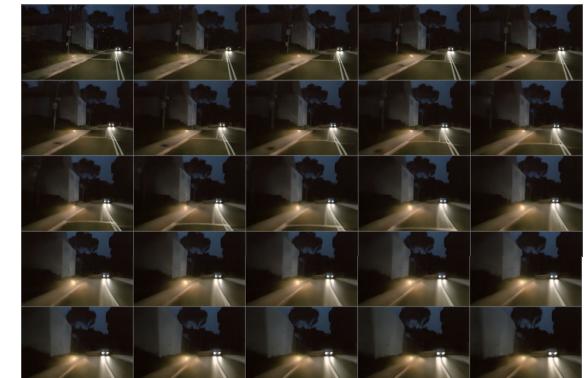
### 层次监督提升模型物理理解能力

不依赖于标准的扩散损失，引入多个损失函数来增强动力学并保留结构细节等物理知识。同时使用约束增强域间特征一致性。

阶段一视觉效果



阶段二视觉效果



低光照场景生成结果



## /04 基于检索增强学习的可解释感知规划（联想记忆）

可信感知

极端场景

类脑计算

团队工作

未来展望

- ✓ 理解场景：智能驾驶系统需要具备可信的场景分析能力，模型不仅要能够“看见”物体，更要“**理解**”每个物体的意义及其相互关系。

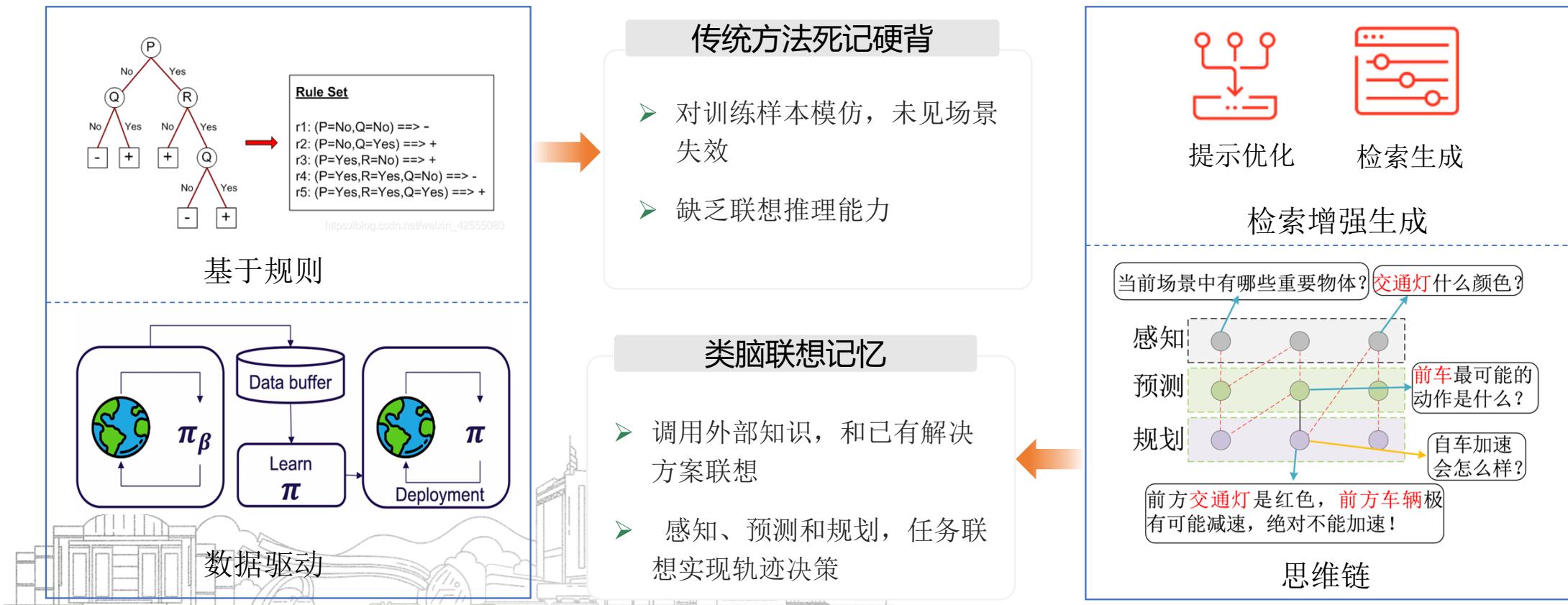


现实场景下的复杂场景理解挑战：**复杂道路、复杂车辆交互、极端天气、特殊目标**



无法理解复杂场景可能导致严重后果，制约自动驾驶技术实际落地部署

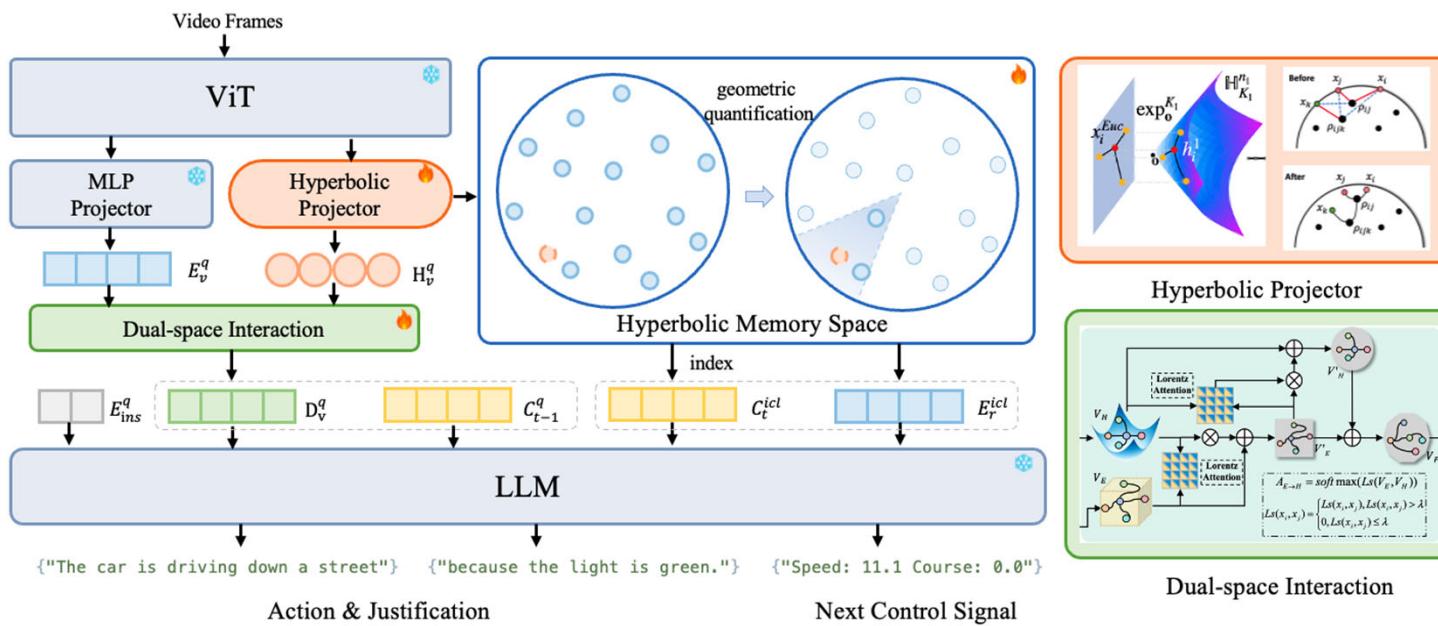
- 现有方法本质是对已有样本的模仿，无法穷举复杂场景，研究**思维链**和**检索生成**技术，利用自动驾驶任务间和场景间的关联性，模拟人脑的**联想记忆机制**。



## 挑战问题：

- ◆ 自动驾驶场景之间存在复杂的非均匀层次关系难以在欧氏空间中有效表征。
- ◆ 不同场景的不确定性差异使得其参考价值及所需参考信息难以量化。

## 网络结构：



## 网络结构：

- ◆ 利用场景相关语义相似性来检索历史样本，使模型能够挖掘有价值的历史场景。
- ◆ 引入双曲空间建模样本之间的关联关系，从而更好地保持场景语义的层级结构与全局一致性。
- ◆ 在仅使用极少训练数据的情况下，取得了与最先进方法相当的结果。

# /04 基于检索增强学习的可解释感知规划（联想记忆）

可信感知

极端场景

类脑计算

团队工作

未来展望

## 实验结果：

- 本次竞赛由香港科技大学、华为等研究机构发起，旨在推动更加可靠且可解释的自动驾驶代理的发展，吸引了香港中文大学、北京大学、清华大学等众多顶尖高校和研究机构的参与。

Team Name	Date	Final Score	General Perception	Region Perception	Driving Suggestion
llmforad	08-15	72.12	58.70	83.41	74.26
OpenDriver	08-09	69.72	54.41	83.00	71.76
NexusAD	08-15	68.97	57.58	84.31	65.02
123	08-02	68.79	52.98	83.07	70.32
LLMAnything	08-15	68.43	56.24	82.66	66.38
NTHUCVLab	08-15	67.85	55.16	82.88	65.50
Swift Unity Expedition	08-15	67.68	57.38	84.37	61.30
GY	07-26	66.79	46.32	84.50	69.54

ECCV 2024 Workshop 比赛获得第3名，  
并受邀于会议现场做口头报告。

Method	GP ↑	RP ↑								DS ↑	FS ↑
		ALL	Vehicle	VRU	Sign	Light	Cone	Barrier	Other		
MiniGPT-v2-7B* [7]	11.58	15.93	18.74	13.58	15.71	17.78	15.34	13.02	14.41	10.00	12.50
Shikra-7B* [8]	12.24	22.94	28.29	17.88	20.00	15.56	21.23	20.00	19.67	10.20	15.12
LLaVA-1.5-7B* [40]	19.30	42.06	46.67	38.47	39.14	48.89	50.83	30.93	33.82	23.16	28.17
Qwen-VL-7B* [3]	18.22	26.62	35.48	24.16	20.86	23.33	19.61	17.56	25.86	22.06	22.30
LLaVA-Instruct-7B [39]	17.76	51.70	52.68	44.96	53.71	47.78	57.21	48.49	51.12	37.38	35.61
DeepSeek-VL-7B [38]	28.84	58.13	59.42	46.79	63.14	51.11	70.93	54.07	51.51	52.72	46.56
MiniCPM-V-2.6-8B [25]	43.98	67.78	72.37	62.63	62.00	66.67	76.23	54.53	65.00	62.84	58.20
CODA-VLM-8B* [34]	55.04	77.68	78.79	73.80	64.86	73.33	86.18	78.72	68.75	58.14	63.62
GLM4V-9B [20]	36.99	73.50	76.61	62.91	66.57	67.78	83.28	67.79	69.80	52.08	54.19
LLaVA-1.5-13B* [40]	24.54	42.41	53.62	36.79	33.71	46.67	41.27	30.41	33.82	27.90	31.62
LLaVA-NeXT-13B* [41]	29.86	53.63	55.51	47.08	54.00	60.00	70.34	40.47	46.45	31.92	38.47
CogVLM2-19B [23]	33.72	69.55	68.99	62.99	61.43	64.44	79.02	67.21	69.08	50.44	51.24
InternVL-V1-5-20B* [12]	38.38	61.53	63.77	53.14	50.57	57.78	80.34	46.86	57.11	41.18	47.03
InternVL-2.0-26B [11]	40.50	73.33	76.69	68.47	64.86	68.89	83.68	61.45	70.39	61.36	58.40
Gemini-Pro*† [48]	25.24	51.38	49.03	42.77	37.43	42.22	69.56	45.70	51.32	27.40	34.67
GPT-4V*† [1]	57.50	56.26	60.89	40.58	49.43	54.44	66.08	50.17	53.16	63.30	59.02
Qwen-VL-Max*† [57]	47.51	74.21	77.20	69.71	71.43	76.67	78.09	64.48	76.45	62.12	61.28
GPT-4o*† [1]	54.80	79.10	80.27	71.68	79.14	62.22	85.25	78.14	76.51	67.40	67.10
LLaVA-NeXT-7B [41]	24.16	64.29	66.91	59.20	66.29	64.44	74.02	54.83	58.95	45.06	44.50
+ NexusAD w/o RCL	49.98	80.17	82.05	77.66	68.57	73.33	86.37	78.84	73.55	63.98	64.71
+ NexusAD (ours)	58.32	82.44	82.75	81.82	72.86	75.56	87.25	84.19	76.32	69.52	70.09
Qwen-2VL-7B [57]	40.82	69.42	74.23	64.16	65.71	66.67	71.81	58.72	70.99	47.32	52.52
+ NexusAD w/o RCL	50.98	80.63	82.46	76.86	70.57	77.78	86.37	79.59	75.00	64.20	65.27
+ NexusAD (ours)	58.64	82.15	83.45	77.59	70.57	80.00	87.30	81.92	77.43	70.56	70.45
InternVL-2.0-8B [11]	38.98	70.29	73.38	66.42	60.86	70.00	80.29	58.43	67.57	51.64	53.64
+ NexusAD w/o RCL	51.32	80.79	82.29	79.56	72.57	65.56	85.74	80.81	73.95	62.72	64.94
+ NexusAD (ours)	57.74	82.56	84.11	80.51	75.42	75.56	87.35	83.20	75.07	71.22	70.51

## /04 基于检索增强学习的可解释感知规划（联想记忆）

可信感知

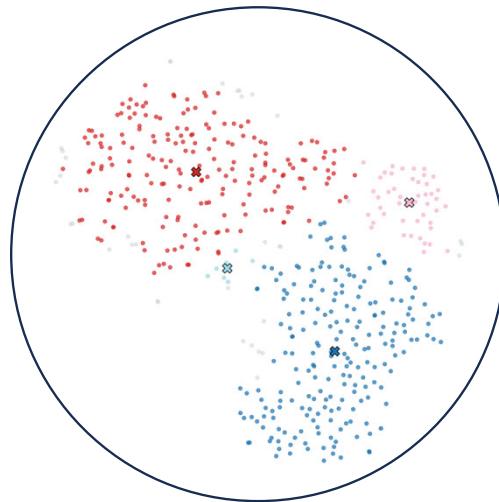
极端场景

类脑计算

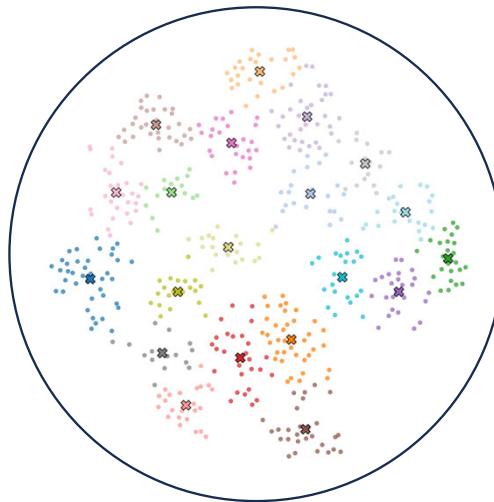
团队工作

未来展望

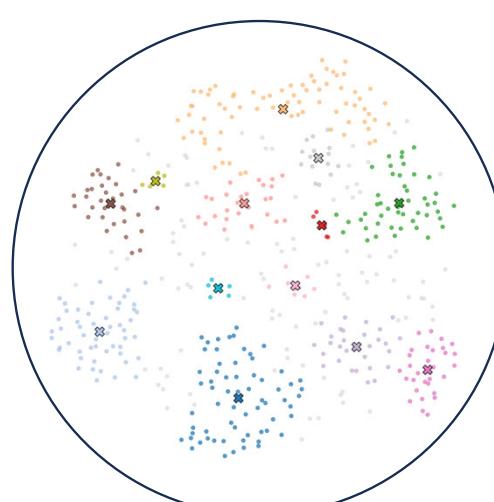
### 可视化分析&实验结果：



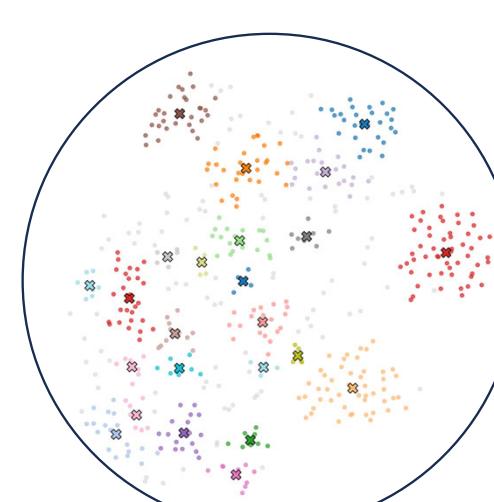
Vanilla Features



Euclidean Space



Hyperbolic Space



Optimized Features

Method	Course						Speed					
	RMSE↓	$A_{0.1} \uparrow$	$A_{0.5} \uparrow$	$A_{1.0} \uparrow$	$A_{5.0} \uparrow$	$A_{10.0} \uparrow$	RMSE↓	$A_{0.1} \uparrow$	$A_{0.5} \uparrow$	$A_{1.0} \uparrow$	$A_{5.0} \uparrow$	$A_{10.0} \uparrow$
ADAPT	5.87	54.49	86.39	91.06	97.36	98.20	2.68	11.77	31.79	47.48	92.75	95.87
DriveGPT4	4.57	69.22	79.14	84.47	95.72	96.74	1.09	56.93	77.77	87.97	99.00	99.57
RAG-Drive	4.48	74.32	88.69	93.12	98.30	99.10	0.69	51.12	85.54	94.49	99.81	99.91
NexusAD (Euc.)	4.25	75.82	89.53	93.99	98.36	99.15	0.62	60.09	90.42	96.29	99.81	99.91
<b>NexusAD</b>	<b>3.99</b>	<b>76.67</b>	<b>90.23</b>	<b>94.41</b>	<b>98.45</b>	<b>99.20</b>	<b>0.60</b>	<b>62.11</b>	<b>91.03</b>	<b>96.62</b>	<b>99.81</b>	<b>99.91</b>

## /04 基于视觉强化微调的长思考动态推理（反馈学习）

可信感知

极端场景

类脑计算

团队工作

未来展望



现有端到端自动驾驶模型**决策过程不透明**，缺乏**人类常识**与**复杂推理**能力，导致在面对需要深度理解、非常规决策的极端场景时，表现脆弱，规划失败率高，难以确保安全与可靠。



面对模糊、动态场景时（如人车混行、无信号交叉口），端到端模型易陷入“**快速但草率**”的决策模式，难以进行“**深思熟虑**”，严重影响安全与可信度。



决策难以解释



道路复杂多样



语义理解薄弱

亟需引入具备**深度思考与可解释性强**的新方法，以应对认知挑战，构建可靠**可信**系统

## 104 基于视觉强化微调的长思考动态推理（反馈学习）

可信感知

极端场景

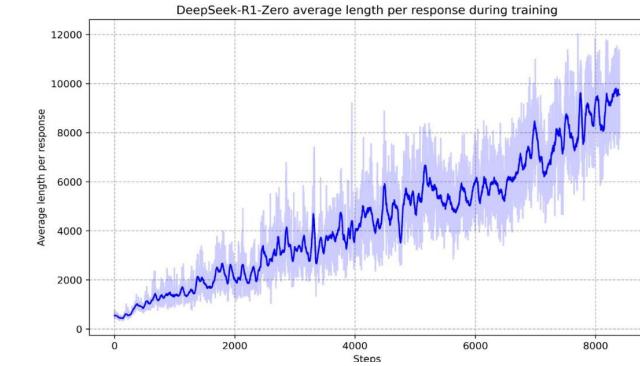
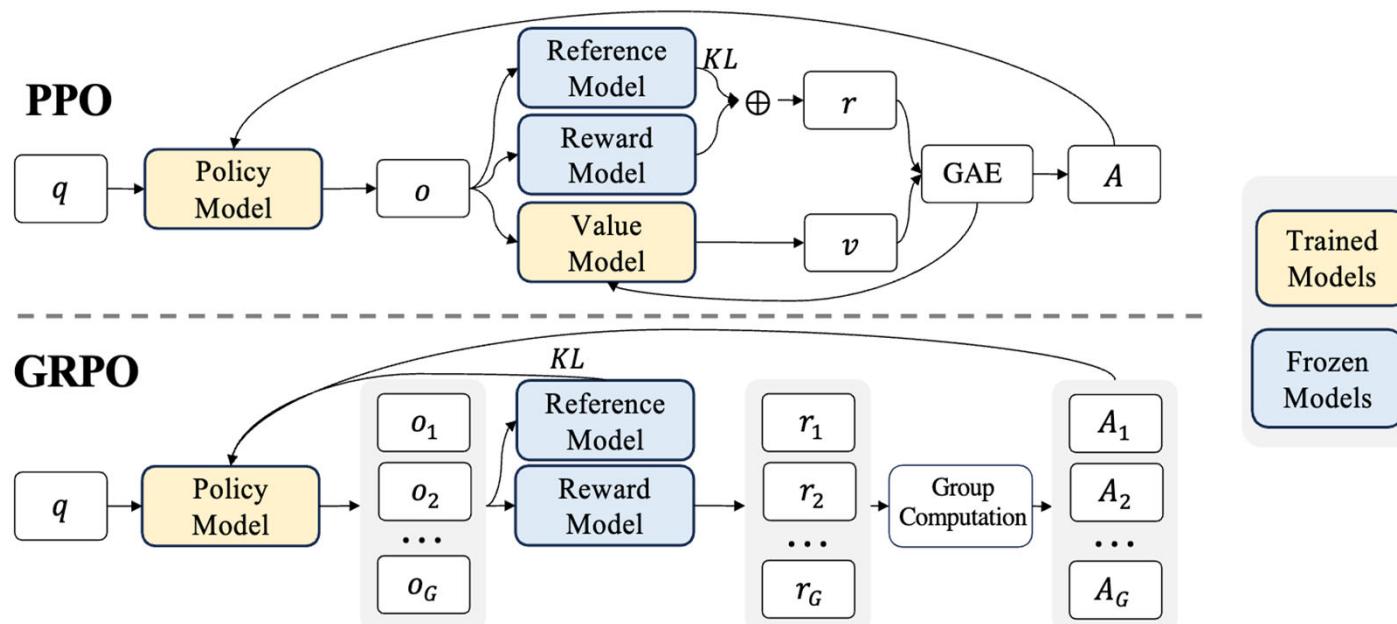
类脑计算

团队工作

未来展望

### 研究动机:

- 1) 利用**视觉语言模型** (VLM) 通用的世界知识、上下文理解和推理潜力，弥补传统模型不足。
- 2) 人类决策的启示：人类驾驶员在复杂情况下并非仅依赖直觉反应，而是**深度思考**的结果。这种深度思考的过程实际上是一种反馈学习机制。自动驾驶需要借鉴反馈学习，尤其在关键决策上，同样需要这种深度思考。



Question: If  $a > 1$ , then the sum of the real solutions of  $\sqrt{a - \sqrt{a + x}} = x$  is equal to  
Response: <think>  
To solve the equation  $\sqrt{a - \sqrt{a + x}} = x$ , let's start by squaring both ...  
 $(\sqrt{a - \sqrt{a + x}})^2 = x^2 \implies a - \sqrt{a + x} = x^2$ .  
Rearrange to isolate the inner square root term:  
 $(a - x^2)^2 = a + x \implies a^2 - 2ax^2 + (x^2)^2 = a + x \implies x^4 - 2ax^2 - x + (a^2 - a) = 0$   
...  
Wait, wait. Wait. That's an aha moment I can flag here.  
Let's reevaluate this step-by-step to identify if the correct sum can be ...  
We started with the equation:  
 $\sqrt{a - \sqrt{a + x}} = x$   
First, let's square both sides:  
 $a - \sqrt{a + x} = x^2 \implies \sqrt{a + x} = a - x^2$   
Next, I could square both sides again, treating the equation: ...

**DeepSeek** 在复杂推理任务上的成功表明，**强化学习**算法是实现模型**深度思考与规划**的可行路径。

# /04 基于视觉强化微调的长思考动态推理（反馈学习）

可信感知

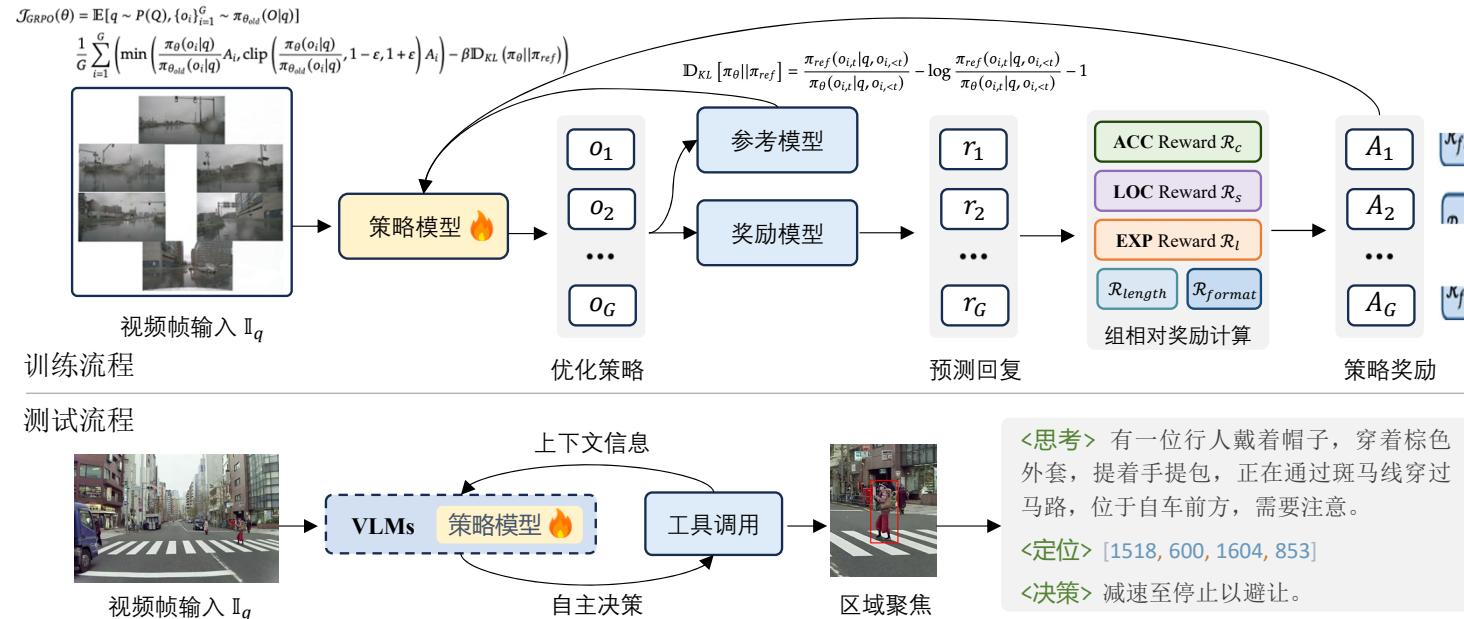
极端场景

类脑计算

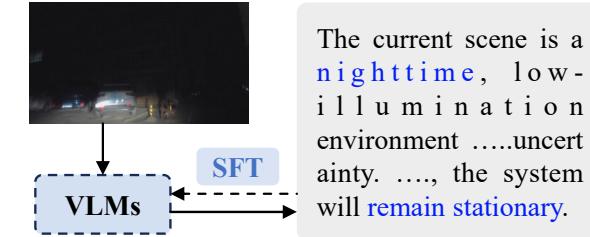
团队工作

未来展望

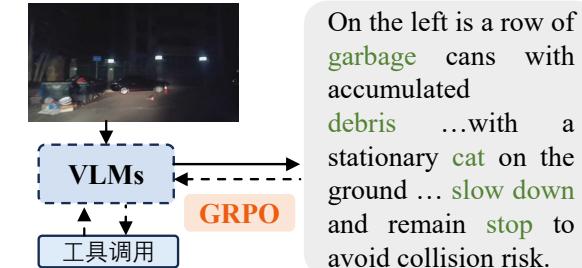
## ✓ 解决决策可解释性差问题



- 本研究提出一种基于策略引导的强化微调框架，构建面向风险目标的组合奖励函数，引导视觉语言模型聚焦重要目标并给出推理解释，从而实现更具可解释性的决策输出。
- 同时，引入工具调用机制与动态交互策略，引入上下文信息逐步优化模型在复杂驾驶情境下的响应质量与可视化表达能力，从而为智能体提供清晰、可信的人机交互支持，显著提升自动驾驶系统的可解释性与安全感知能力。



第一阶段：监督微调激活推理能力



第二阶段：强化微调扩展思维能力

GRPO  
强化微调



人类反馈  
学习

# /04 基于视觉强化微调的长思考动态推理（反馈学习）

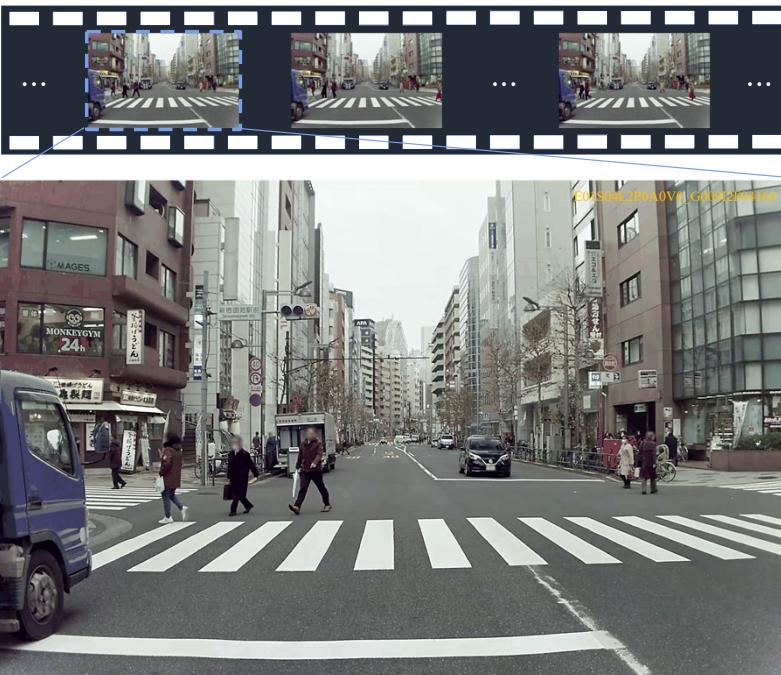
可信感知

极端场景

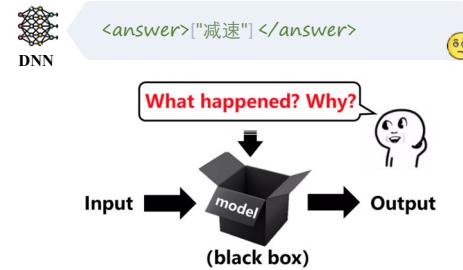
类脑计算

团队工作

未来展望



## □ 传统模型：



## □ 对话模型：

右侧斑马线上有行人正在过马路，其中一位穿着红色裙子的行人位于较右侧车道位置，其他行人也在陆续通过。左侧人行道上也有行人走动，但未处于过马路状态。

自车应立即减速并停车，礼让正在通过斑马线的行人，待行人全部安全通过后，再起步继续行驶。在行驶过程中，需注意观察周围交通状况，确保安全。



仅有通用解释，未能深入理解场景

## □ 推理模型：

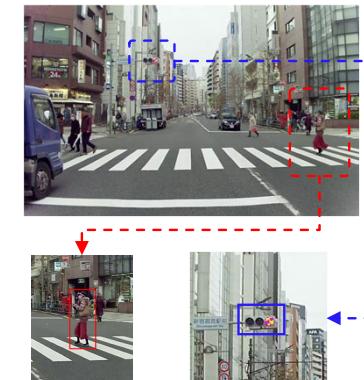
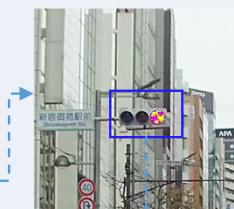


<think> 道路为双向车道，中间有明显的分隔线。左侧有一辆蓝色卡车，右侧对向有一辆黑色轿车。路口前方有行人正在通过斑马线，尤其有一名行人靠近自行车方向，需要特别注意。前方有交通信号灯，较为模糊，需要放大看一看：

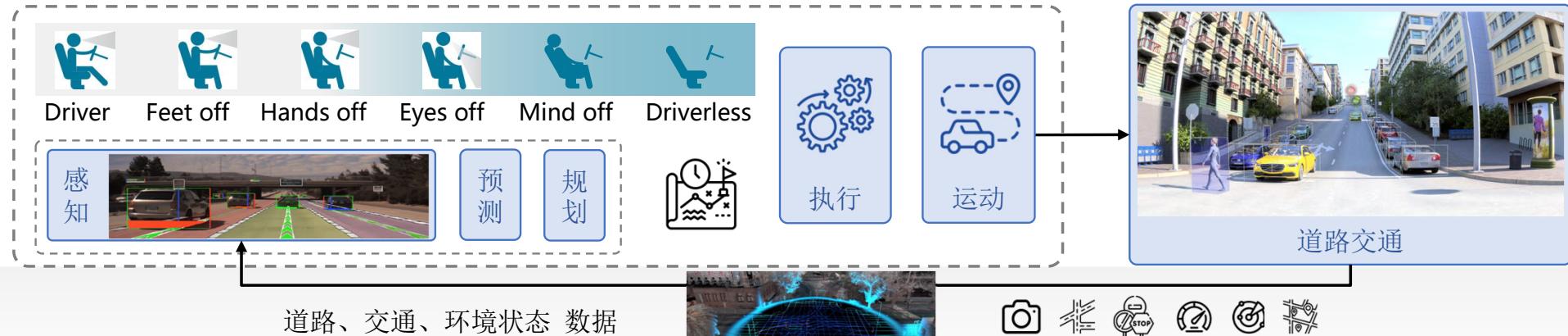
<seeking> [450, 166, 607, 306] </seeking>

可以看到当前仍然为红灯，虽然部分行人已通行过路口，但根据交通规则，车辆必须在红灯时完全停止，不得抢先通过。

<answer> 此时自行车应稳住刹车，停在停止线或安全距离外，耐心等待红灯转绿并确认所有行人完全通过后再准备启动。



不同于传统模型对于数据的记忆或简单解释，基于**视觉强化微调**的推理模型能够自主**调用相关工具**，实现对复杂场景的深入理解，并生成多种合理且可靠的驾驶策略。



目的

**驾驶场景生成****驾驶场景理解****驾驶决策推理**

机制

层级感知

联想记忆

反馈学习

方案

基于世界模型驱动的低  
照度场景生成基于检索增强学习的可  
解释感知规划基于视觉强化微调的长  
思考动态推理

应用

**自动驾驶大模型****智慧环卫大模型****智能机械臂平台****登月作诗大模型**

## 天瞳智行智驾平台展示



天瞳智行智驾平台，在国际挑战赛中获奖



★ CVPR

自动驾驶国际挑战赛创新奖



★ ECCV

Workshop Winner Solution



# 报告提纲



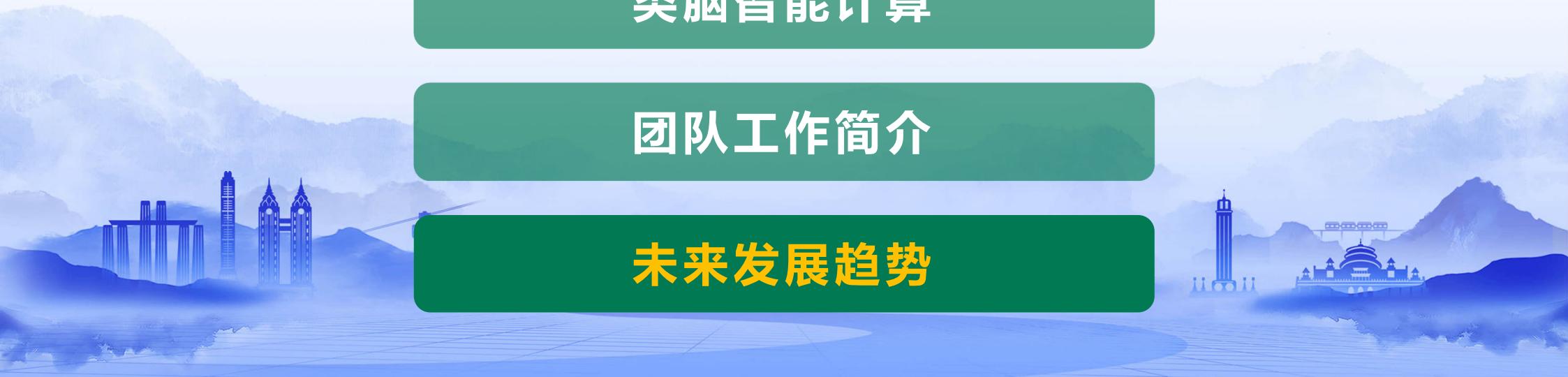
可信场景分析

极端场景挑战

类脑智能计算

团队工作简介

未来发展趋势



# /05 未来发展趋势 | 脑启发 → 脑参与 → 脑认知

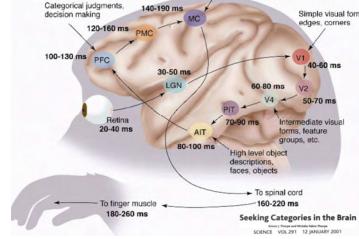
可信感知

极端场景

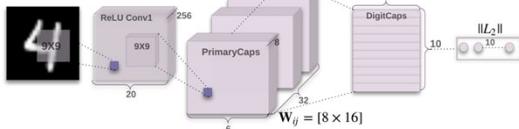
类脑计算

团队工作

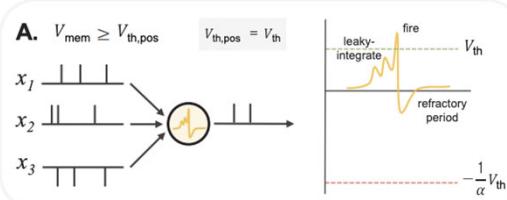
未来展望



模仿学习机制

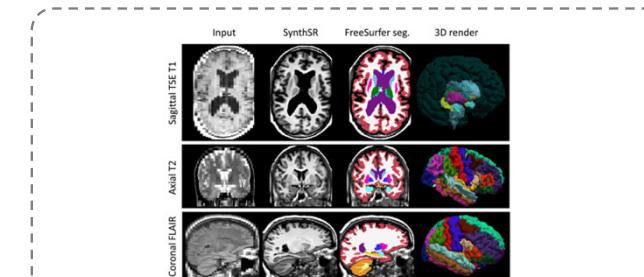
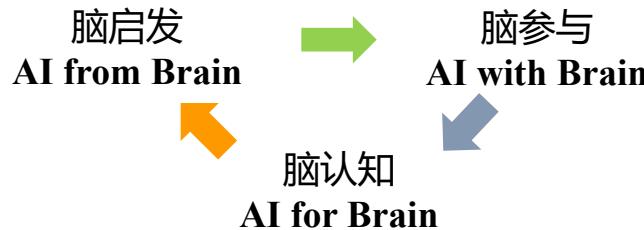


胶囊网络



脉冲神经网络

脑启发

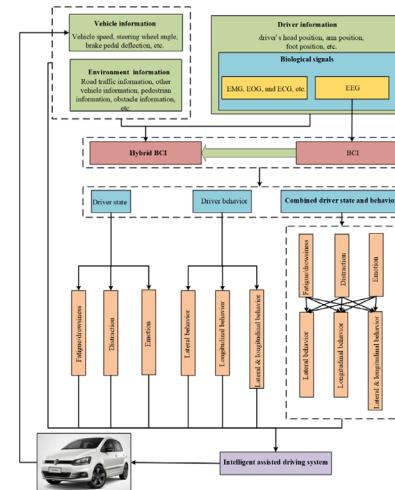


3D脑部渲染 (SynthSR)

脑认知



脑-控智能车辆  
利用大脑信号进行驾驶行为预判



基于脑电信号智能辅助驾驶系统

脑参与

## /05 未来发展趋势 | 脑机互鉴→三模协同

可信感知 极端场景 类脑计算 团队工作 未来展望

脑机互鉴的形式逐步拓展至多机器间的协同学习，推动形成大模型、中模型、小模型三模协同的智能学习体系。



边缘算力



多模态小模型



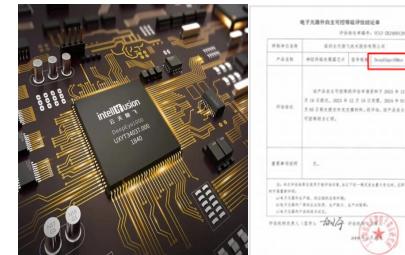
基座通用大模型



行业专业中模型 (具身智能)

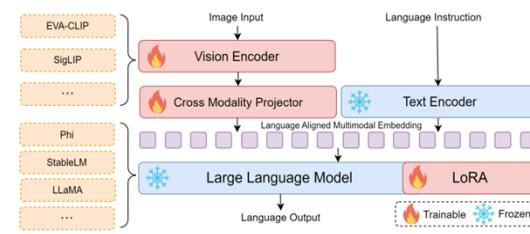


边端芯片小模型

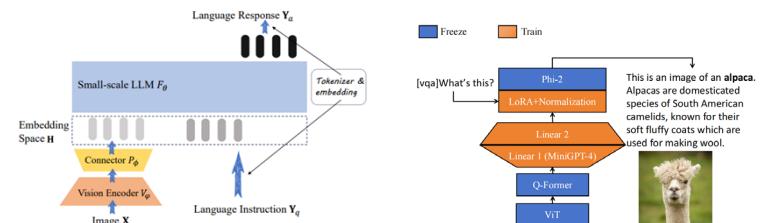


模型端侧部署: DeepEdge10max芯片

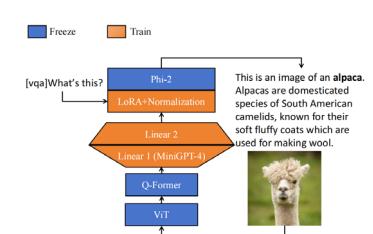
边缘算力通过就近提供计算、网络和智能等能力，加速经济转型升级，成为计算体系的新方向



Bunny



TinyLLaVA



TinyGPT-V

资源受限环境下的小型高效多模态模型

# Thanks for your listening and attention!

敬请批评指正！



团队微信公众号



个人主页

**CVCI 2025** | Multimodal Large Foundation Models and Autonomous Driving

讲者: 冷 佳 旭 (Jiaxu Leng)

**会前 Workshop 报告**

2025年10月28日星期二