

汽车行业专题报告： VLA和世界模型-通往高阶智能驾驶之路 ——辅助驾驶系列报告二

评级：推荐(维持)

戴畅(证券分析师)
S0350523120004
daic@ghzq.com.cn

陈飞宇(证券分析师)
S0350525060001
chenfy02@ghzq.com.cn

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

最近一年走势



相对沪深300表现

表现	1M	3M	12M
汽车	7.1%	11.5%	55.9%
沪深300	9.1%	16.2%	42.5%

相关报告

- 《汽车行业专题报告：辅助驾驶的AI进化论：站在能力代际跃升的历史转折点（推荐）*汽车*陈飞宇，戴畅》——2025-07-22
- 《汽车行业周报：全新问界M7 24小时小订突破15万台，吉利汽车合并极氪获股东高票通过（推荐）*汽车*戴畅》——2025-09-09
- 《汽车行业周报：2025成都车展开幕多家车企携新品参展，极氪9X开启预售（推荐）*汽车*戴畅》——2025-09-02
- 《汽车行业周报：蔚来全新ES8、吉利银河M9开启预售，中大型智能SUV市场迎来重磅新品（推荐）*汽车*戴畅》——2025-08-25
- 《汽车行业周报：新款问界M7将于9月上市，2025年7月乘用车批发销量同比增长14.7%（推荐）*汽车*戴畅》——2025-08-18

本篇报告解决了以下核心问题：1、VLA和世界模型是什么及特点？2、目前辅助驾驶产业玩家的路径差异性？3、辅助驾驶能力提升过程中的技术趋势和核心关注点是什么？

一、VLA模型（Vision-Language-Action）

- ◆ **定义：**VLA是一种端到端多模态人工智能架构，通过融合视觉输入（图像/视频）和自然语言指令，直接生成可执行的物理动作，实现从感知、理解到控制的闭环。
- ◆ **特点：**多模态融合-结合视觉语义与语言指令，增强场景理解与交互能力；语义推理与泛化-支持复杂场景的语义推理和长尾问题处理。
- ◆ **产业玩家：**理想汽车-自研Mind VLA模型，具备空间理解、思维、沟通与记忆、行为能力四大功能，基于Thor-U/Orin-X平台量产部署，支持语言控制驾驶，于2025年9月OTA全量推送至AD MAX车主；小鹏汽车：采用“VLA+OL”云端基模，通过蒸馏技术部署车端轻量化XVLA模型，于2025年9月OTA推送至G7 Ultra；元戎启行-DeepRoute IO 2.0于2025年8月发布，搭载自研的VLA模型，已达成5个定点，首批量产车将进入市场。

二、世界模型（World Model）

- ◆ **定义：**以视觉与运动数据为基础，通过生成式建模预测环境动态与行为后果，强化辅助驾驶系统在仿真和交互中的可靠性。
- ◆ **特点：**难例场景构建：通过生成式技术增强高密度 corner case 的覆盖；时空预测能力-对环境变化和车辆运动进行高精度预测。
- ◆ **产业玩家：**华为-乾崮ADS 4.0采用WEWA架构，构建难例场景库，提升泛化能力；蔚来-NWM模型支持多模态输入输出与自然语言交互，强化场景理解与控制；Momenta-R6飞轮大模型采用双模型架构，融合长短期记忆与真实数据闭环。

三、比较与结论：VLA与世界模型在技术上并非同级或对立关系。我们将发展路径分为两派，实质上是产业玩家在实现端到端能力之后，在能力优化侧重点上出现了分化。

- ◆ **通往高阶智驾的互补路径：**VLA 侧重语义融合与交互控制，通过语言增强可解释性与用户协同；世界模型侧重环境预测与仿真可靠性，通过生成技术提升安全冗余。
- ◆ **技术融合趋势明显：**双方均在向对方领域渗透：VLA 引入强化学习与仿真优化动作生成，世界模型扩展语言交互与多模态理解。
- ◆ **数据与安全为核心壁垒：**数据飞轮效应成为竞争关键-头部车企凭借高活跃用户积累海量里程数据，通过RLHF驱动模型高效迭代；安全性能量化用户体验：当前第一梯队MPA（平均事故里程）已达人类驾驶安全水平的6倍以上，MPI（平均接管里程）持续优化。

行业评级与投资建议：多模态VLA模型引入长序思维能力，世界模型提供仿真与强化训练环境，有效数据闭环+安全能力提升（MPA&MPI），共同构筑智能驾驶长期壁垒。考虑到L3辅助驾驶商用标准落地在即，多款L3级架构车型陆续上市，场景切割的阶段性功能实现，并向全场景覆盖的最终愿景发力，维持辅助驾驶行业“推荐”评级。智驾能力成为车企估值分化关键因素，短期关注**数据规模大、算法迭代快**的车企：**理想汽车、小鹏汽车、蔚来汽车、鸿蒙智行合作车企**，辅助驾驶解决方案供应商**Momenta、元戎启行**等。

风险提示：辅助驾驶政策法规推进不及预期；数据安全与隐私保护监管趋严；车企销量与技术搭载不及预期；行业竞争加剧导致技术研发投入承压；技术可靠性与长尾场景安全问题导致的信任与合规风险。

一、VLA模型：多模态融合驱动，重塑人车交互与决策控制

1.1 VLA：融合视觉、语言、动作的端到端人工智能模型

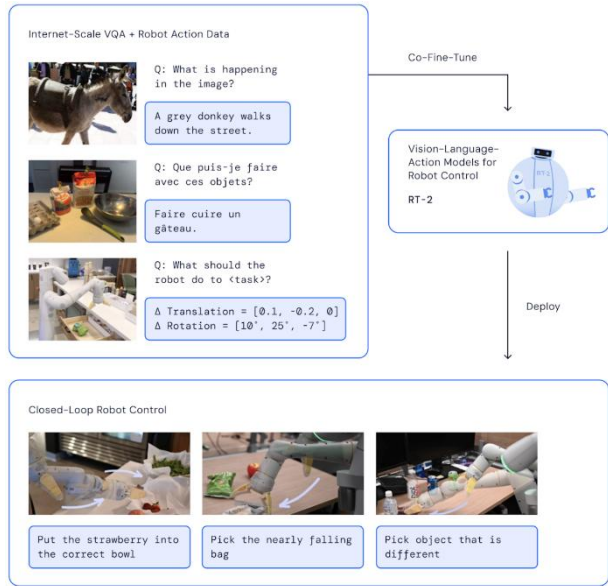
Vision-Language-Action (VLA) 模型是一种融合视觉 (Vision)、语言 (Language) 和动作 (Action) 三大模态的端到端人工智能模型。

- 它通过统一的多模态学习框架，将感知、推理与控制一体化，直接根据视觉输入（如图像、视频）和语言指令（如任务描述）生成可执行的物理世界动作（如机器人关节运动、车辆转向控制）。
- VLA概念形成于2021-2022年左右，由Google DeepMind的Robotic Transformer2（RT-2）等项目开创。

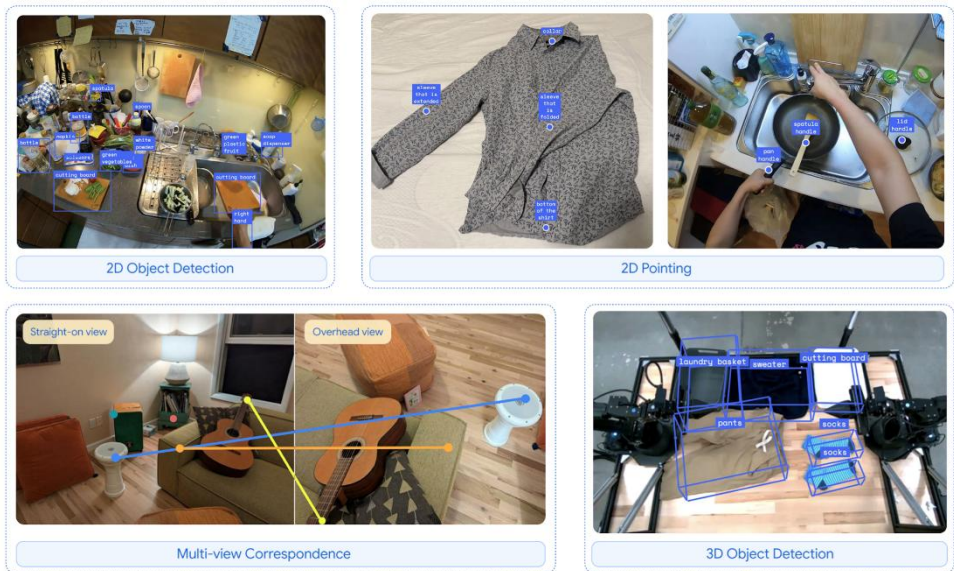
表、Google DeepMind 模型演变

公司 / 模型	时间	类型	泛领域应用	技术特征
Google DeepMind – RT-2	2023.07	VLA 模型、通过机器人控制模型	通用机器人控制 复杂任务泛化	对 VLM 模型的简单而有效的修改，显示出改进的泛化能力以及数据的语义和视觉理解能力，可以推理、解决问题和解释信息，支持物体操作和任务执行。
Google DeepMind – Gemini Robotics	2025.03	进化型 VLA	高阶机器人操作	基于 RT-2，增强 2D/3D 语义理解和机器人控制精度，支持复杂场景执行，面向下一代物理智能体。

图、Google DeepMind - RT-2：获取机器人相机图像并直接预测机器人要执行的动作



图、Google DeepMind - Gemini Robotics：擅长具身推理能力，包括检测物体和指向物体部件、查找相应点和检测 3D 物体



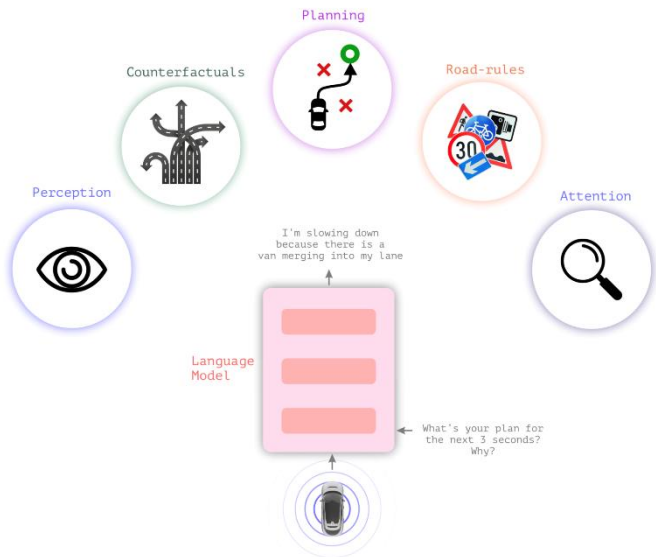
1.2 VLA应用于智能驾驶：WAYVE的早期尝试

Wayve成立于2017年，是一家在辅助驾驶领域进行具身智能研究的领先开发商，一直处于将大型语言模型应用于辅助驾驶的前沿，用自然语言来增强基础驱动模型的学习和可解释性。公司于2023年9月发布了其开环驾驶评论器LINGO-1研究模型，2024年4月推出的闭环驾驶模型LINGO-2，将语言与驾驶深度联系起来，以提供人工智能（AI）模型对驾驶场景理解的可见性，塑造人车交互的未来，以构建安全可靠的自动驾驶系统服务。

表、Wayve - LINGO 模型演变

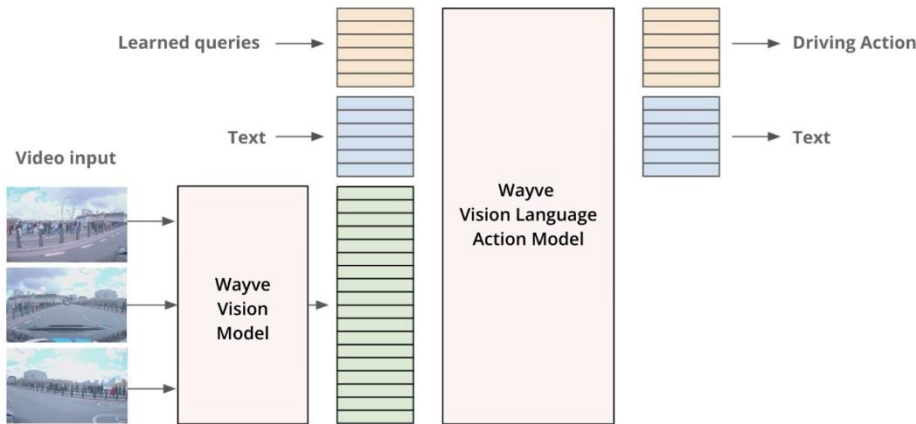
公司 / 模型	时间	类型	泛领域应用	技术特征
WAYVE - LINGO-1	2023.09	开环驾驶模型	1. 自动驾驶感知解释 2. 驾驶行为可解释性	利用视觉语言输入来执行视觉问答（VQA）并驱动评论描述场景理解、推理和注意力等任务—— 仅提供语言作为输出 。
WAYVE - LINGO-2	2024.04	闭环系统（VLAM）	1. 闭环自动驾驶控制 2. 人机交互指令驾驶	将 视觉和语言作为输入和输出 （包括驱动动作和语言）相结合，为其运动规划决策提供持续的驱动评论，可实现“语言驱动车辆行为”，为未来端到端自动驾驶提供基础。

图、WAYVE - LINGO-1：通过简单的提示更改来执行许多任务



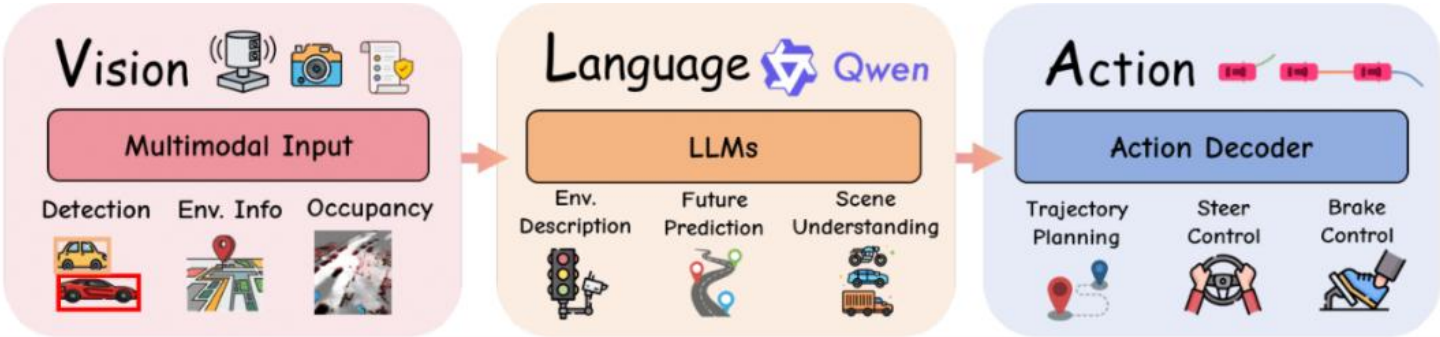
图、WAYVE - LINGO-2：用于驾驶的多模态 Transformer

LINGO-2 Architecture: Multi-modal Transformer for Driving



1.3 VLA核心架构模块：视觉、语言、动作三大编码器

图、VLA 架构概览：整合了视觉、语言和行动模块



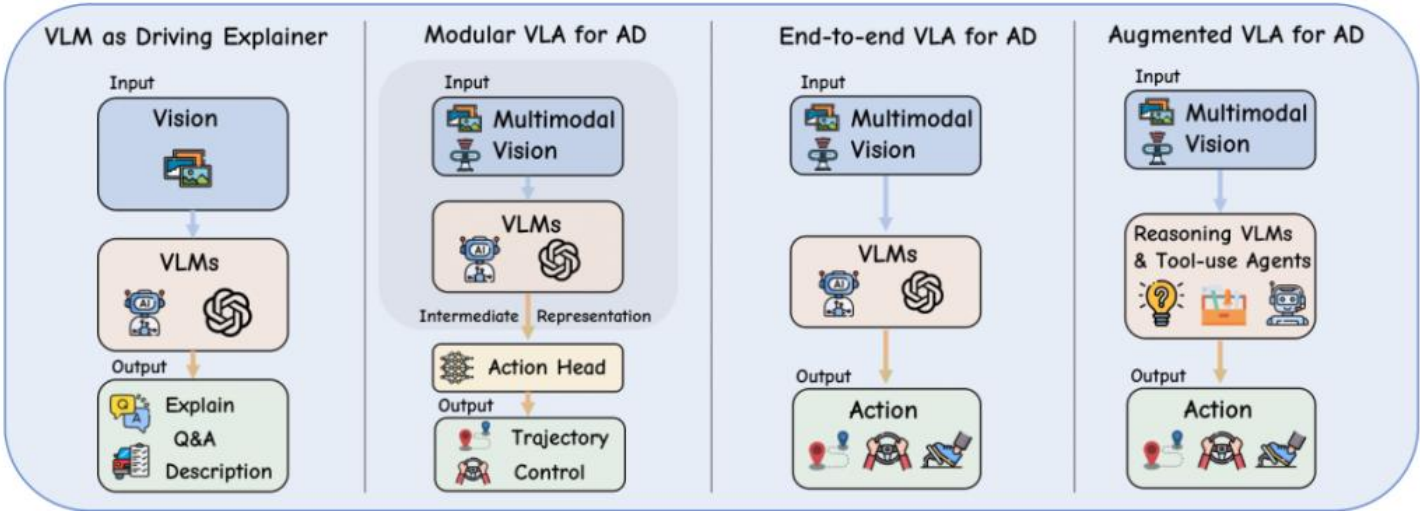
VLA的技术细节主要有三大核心模块：

- 1) 视觉编码器：采用自监督模型如DINOv2、CLIP，结合BEV投影和LiDAR融合技术。
- 2) 语言处理器：大语言模型（如LLaMA、Qwen、Vicuna、GPT等），通常通过LoRA等轻量化技术优化。
- 3) 动作解码器：包括自回归token、扩散规划器、分层控制器（高层策略→PID/MPC控制）。

图、VLA模型演进历程

从左至右依次为：Pre-VLA、模块化VLA、端到端VLA、以及推理增强的VLA

Progress of VLA Models for AD



学术界将VLA模型在智驾方面的发展归纳为四个主要阶段：

Pre-VLA：语言作为解释器LLM解码器用自然语言解释驾驶场景或推荐操作，但实际车辆控制还是由传统模块（PID控制器等）处理。无需直接输出控制。

模块化VLA：语言模型从被动的场景描述者，变成了模块化架构中主动的规划组件，类似：VLM + Action，Action是拆开的。例如采用混合专家架构，如“超车专家”或“走走停停专家”，利用语言提示根据上下文动态选择子规划器。

统一端到端VLA：单一网络直接映射多模态输入到控制或轨迹输出，国内叫法可以称为一段式VLA。但学术界也指出，端到端VLA的局限在于长程推理和复杂多步规划能力有限。

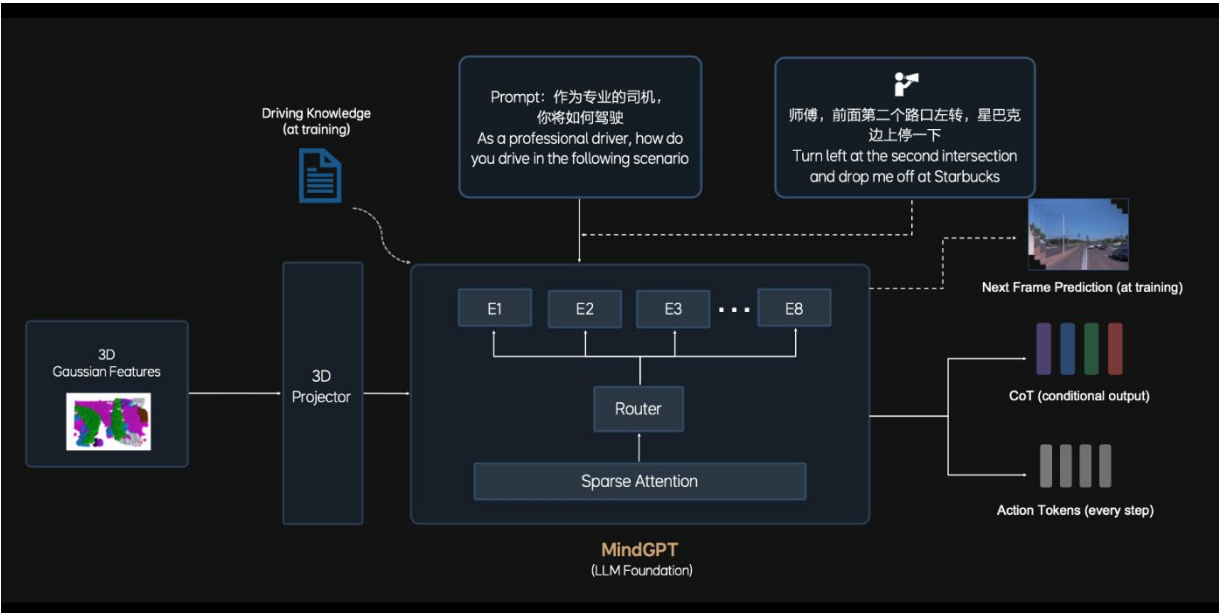
推理增强型VLA：语言模型处于控制闭环的核心，同时支持长时记忆和链式推理。这样它能在输出动作之前进行解释、预测并执行长远推理，真正实现代理司机的概念。

1.4 VLA的进阶能力：人机交互升级，实现推理透明化

短期：人机交互的能力，从座舱体验延伸到驾驶
实现智舱与智驾在功能层的协同创新

- 1) 输入-语言智能：用户可以直接与模型对话，模型可以自动拆解并执行任务
- 2) 理解-视觉智能：具备强大的文字理解及OCR识别能力
- 3) 输出-行为智能：自主切换快慢思考（慢思考通过思维链CoT实现透明化推理）；给出文字解释其驾驶决策偶成，解决端到端系统“黑盒”问题。

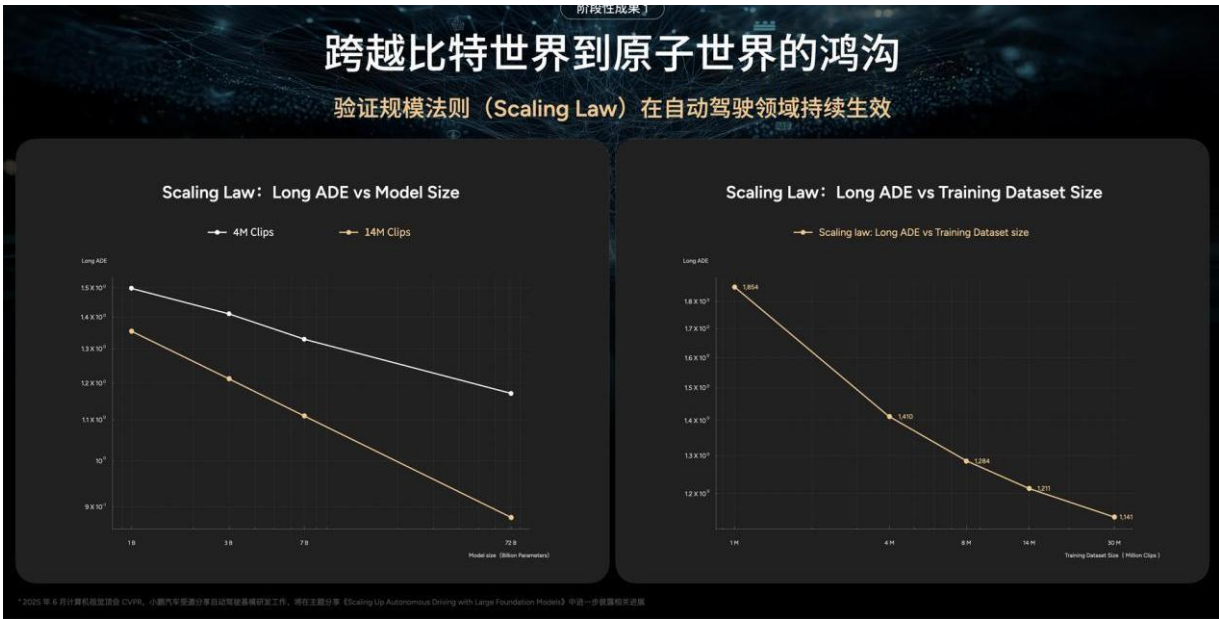
图、理想MindGPT：为实时边缘推理量身定制的设计



中长期：AI驱动的数据闭环是核心竞争力（数据仿真&强化学习）
从单车智能向车路云协同（中心化）

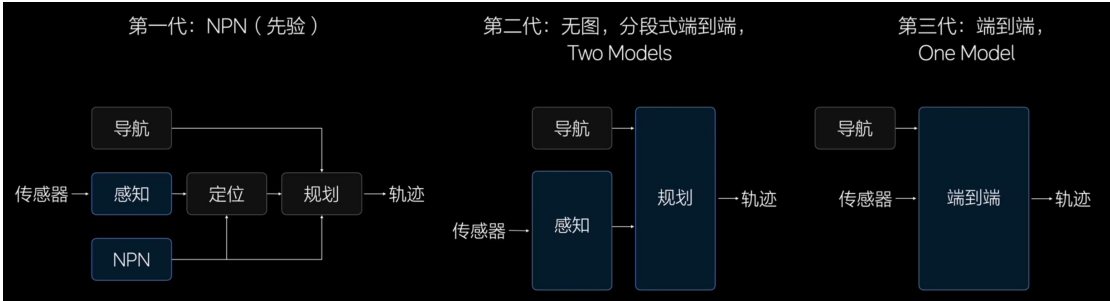
- 1) 世界模型就是仿真高阶形式，用以弥补数据缺失的问题，强化学习能够帮助模型进行自我进化
- 2) 规模法则在辅助驾驶领域持续生效：模型性能随参数规模、训练数据量提升而提升
- 3) 智驾算法的优化依赖于车端芯片与云端基座模型的协同升级与匹配

图、小鹏-验证规模法则在自动驾驶领域持续生效



1.5 理想智驾发展历程：从规则算法到多模态融合

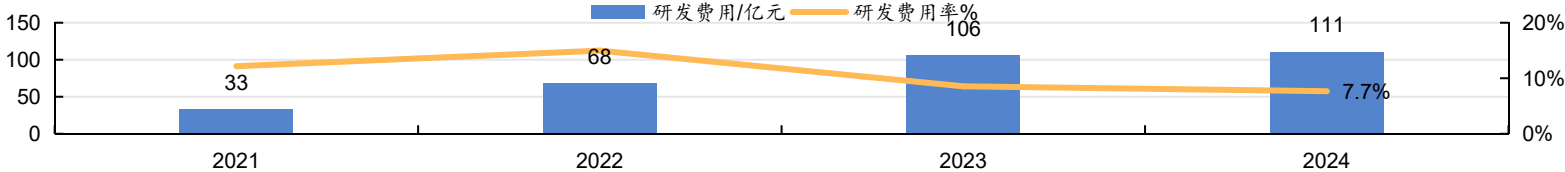
图、辅助驾驶进化过程



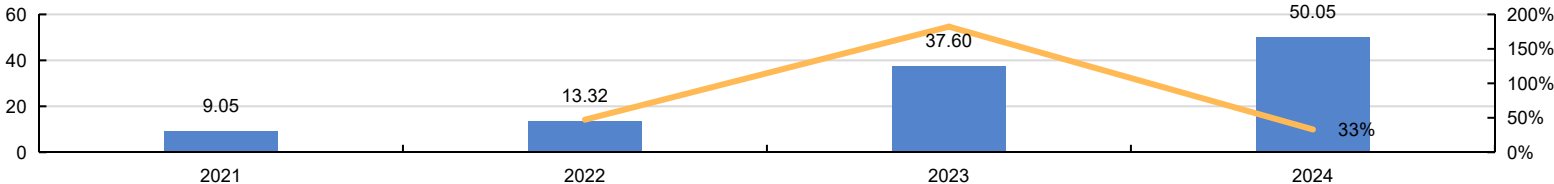
表、理想智能驾驶系统演进与算力平台迭代

	人工时代-规则算法期		AI时代-多模态融合		
	2021	2023	2024	2025	2026
特征	分段式规则算法（规划、控制、执行） 依赖高精地图	BEV感知 无图NOA 可用范围扩展	发布E2E+VLM 语言辅助识别 场景回答	发布Mind VLA, 包含MoE、 Sparse Attention、 Diffusion、RLHF等技术	强化学习 （提升类人决策）

图、理想2021年至2024年研发费用及研发费用率



图、理想汽车2021至2024年销量及同比增速



2024年7月，“端到端自动驾驶”团队成立，推进“端到端模型 + VLM + 世界模型”的三大体系。

2024年11月，智驾团队分为三组：“端到端”模型算法与落地、世界模型&技术预研以及量产研发。

2025年6月，原“研发与供应群组”和“销售与服务群组”整合并入新成立的“智能汽车群组”，全面负责智能汽车业务从战略规划到经营闭环的各项工作。全力转入AI，提出端到端闭环经营。

2025年7月，架构调整，新成立算力资源部（龙开文负责）和企业智能部（沈嵘负责），增强AI算力。

1.5 理想汽车的VLA：从E2E+VLM升级而来

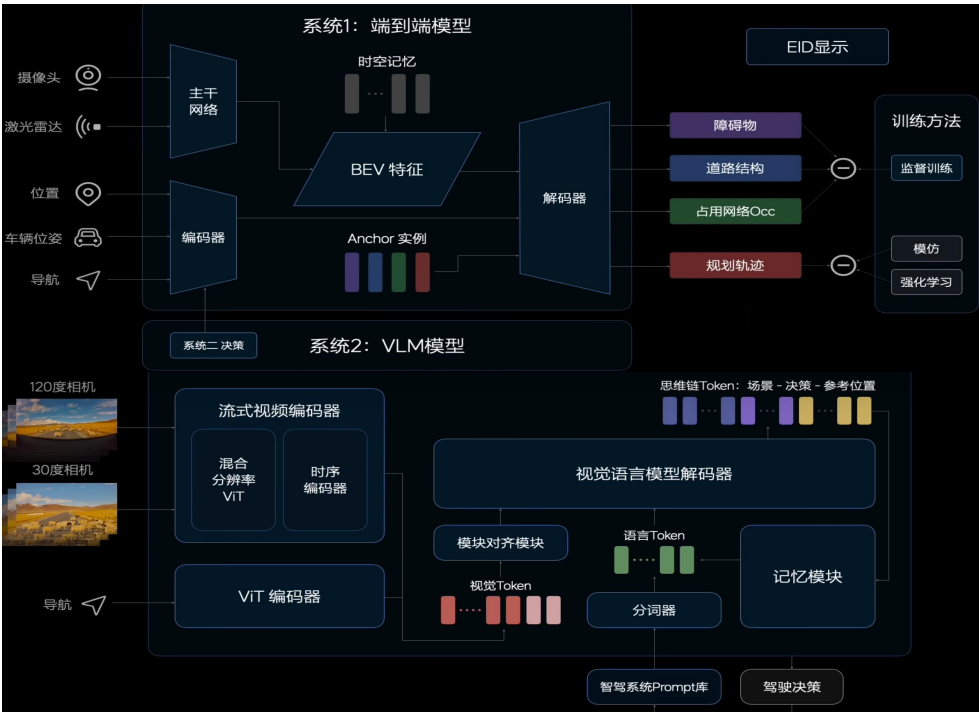
2024年的7月首次发布了一段式E2E+VLM的快慢双系统模型。在端到端基础上，VLM慢系统更像是副驾驶的角色，引入QWEN-LLM（通义千问大语言模型），能够给到端到端的快系统一些辅佐的决策，但是不直接生成规划轨迹，仅提供对于环境的理解和驾驶决策建议。创新性引入：流式视频编码器（能够缓存更长的视觉时序）、记忆模块（缓存多帧历史信息以执行超长时序的推理问题）、智驾Prompt库（协助系统一解决部分场景问题），以提供更丰富的上下文理解和人机交互体验。

2025 年 3 月份英伟达大会上首次披露了将MindVLA 架构。该架构将快慢系统融合于同一架构中，包含全新设计的视觉、语言、行为三大模块，通过新增高斯特征的 3D 感知能力强化场景理解；结合自研语言模型（MindGPT）及逻辑推理优化方案（MOE 架构 + Sparse Attention），输出合理的驾驶决策。此外，该架构具备思维链（COT）表达能力，可自主切换快慢思考模式，其输出的 Action Token 通过 Diffusion 模型优化，最终生成最佳驾驶轨迹。

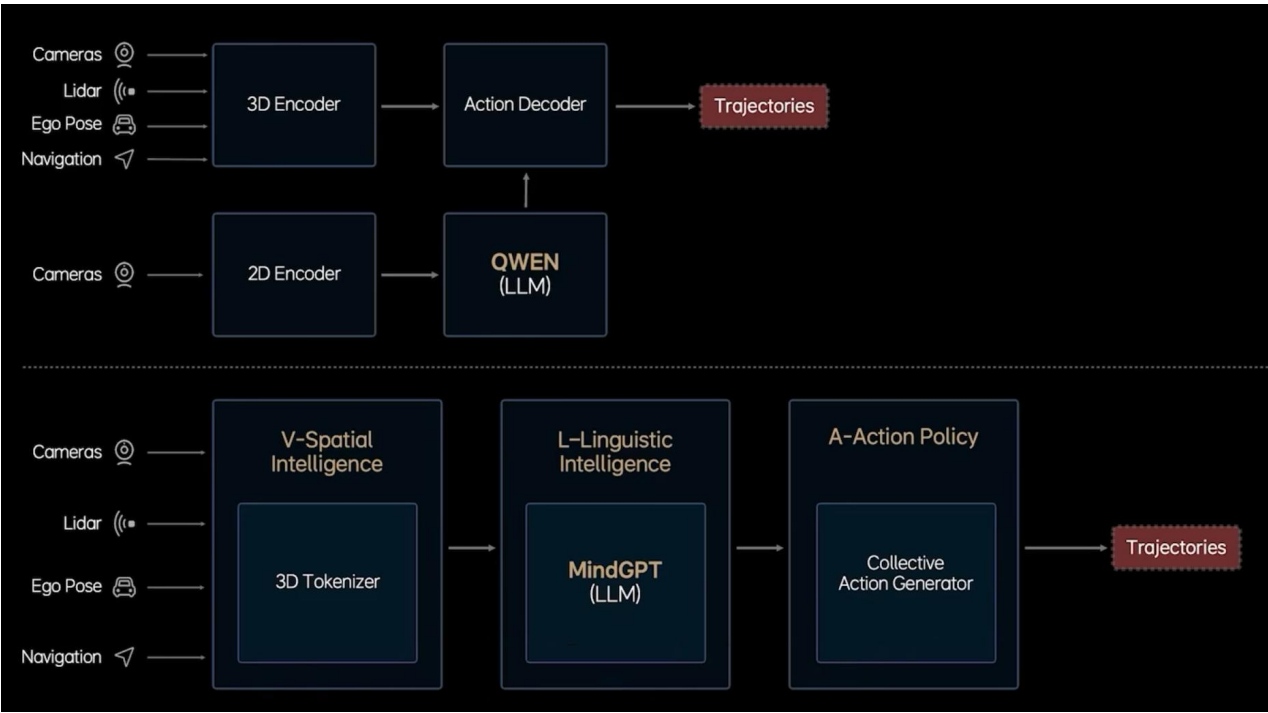
VLA具备空间理解能力、思维能力、沟通与记忆能力、行为能力，在输入、理解、输出都有全面增强，使得在输入端的语言信息能够控制车辆驾驶行为，结合语义理解和多模态感知，共同参与和影响规划轨迹的输出，提升复杂场景处理能力和长时序推理能力；同时能够以文字形式简要呈现推理过程，解决端到端的“黑盒”问题。

VLA司机大模型与 i8 同步量产交付，并于2025.9.10全量推送至ADmax车型（Thor-U 平台与 Orin-X 平台）。

图、理想辅助驾驶架构-系统1（E2E）+系统2（VLM）

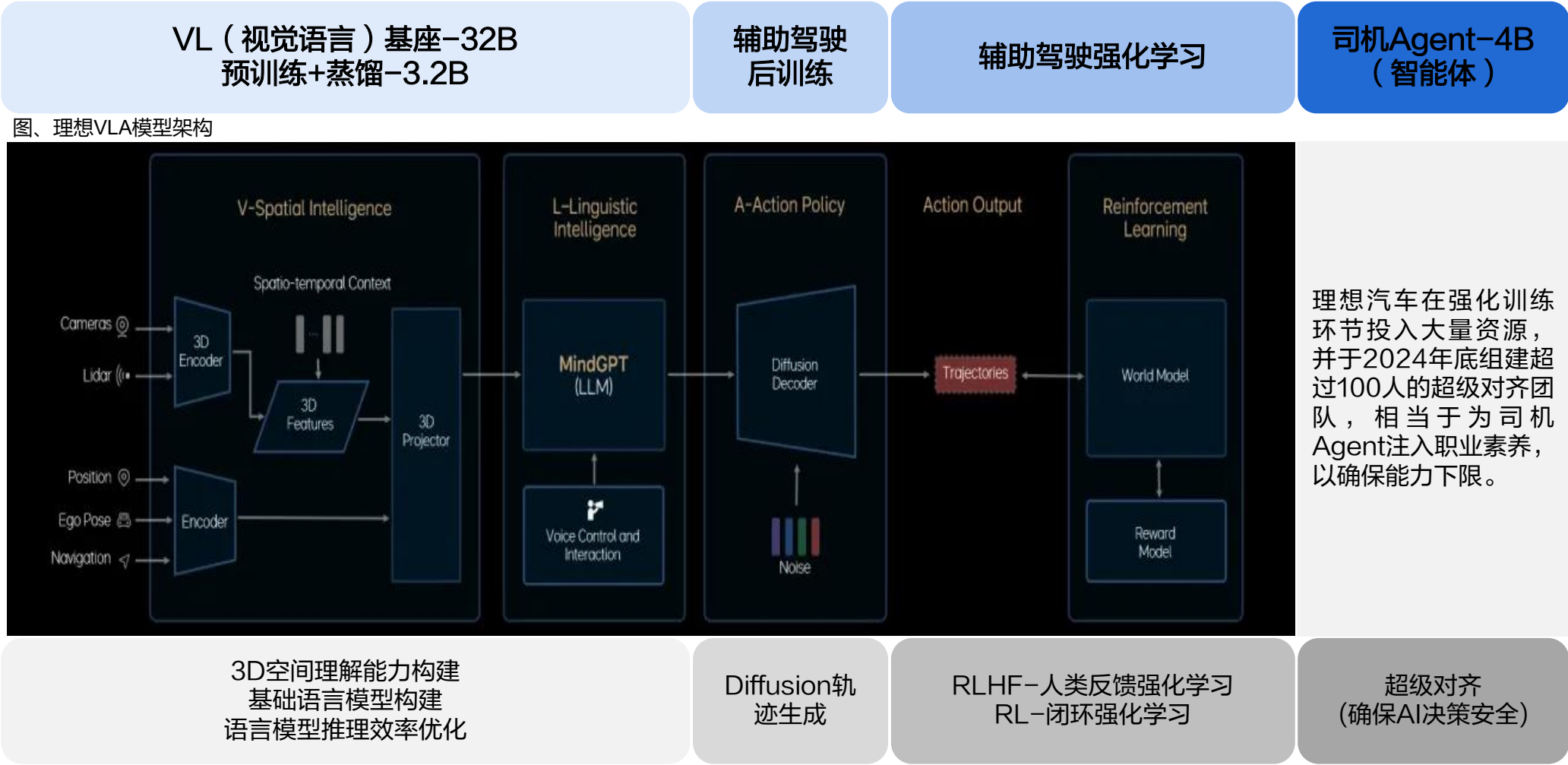


图、理想辅助驾驶架构-E2E+VLM到VLA



1.5 理想的Mind VLA：基于视觉语言基座与类人Agent架构

Mind VLA 是理想汽车自研的新一代自动驾驶大模型，将快思考与慢思考合并到一个大模型中，并赋予模型自主决定何时启动慢思维的能力，赋能车成为能够与用户沟通理解用户需求的司机。



1.6 小鹏VLA：2025.9 OTA推送G7 Ultra，新增人机共驾模式

小鹏汽车（2025.7.3）：首发本地端【VLA+OL + VLM】大模型，其中，VLA-OL模型构建了“运动型大脑+增强型小脑”的智驾架构

云端基模技术路线（云端训练，车端蒸馏）：采用强化学习、模型蒸馏的技术路线，先在云端训练超大模型，再通过蒸馏技术将其压缩至适配车端芯片的小模型（XVLA）。

小鹏的基座模型被定位为『物理AI终端的通用模型』，未来将同步赋能AI机器人和飞行汽车。这种多场景泛化能力，不仅摊薄研发成本，更构建了AI生态的协同效应。

图、小鹏云端模型工厂



具备持续强化学习的能力：强化学习能够帮助模型自我进化，学会处理训练数据中没有的长尾问题，实现安全的自动驾驶。只有足够强大的基座模型，才能被强化学习不断激发出能力上限。

自建AI基础设施

万卡智算集群算力高达10 EFLOPS
集群利用率常年保持90%以上

世界模型+规模法则

72B参数量基座大模型
海量驾驶数据
当前2000万Clips
2025年底将达2亿Clips

图、小鹏G7 Ultra VLA+VLM上车节奏



1.7 元戎启行VLA：已达成5个定点，首批量产车将进入市场

元戎启行（2025.8.26）：发布DeepRoute IO 2.0
该平台搭载自研的VLA模型，融合视觉感知、语义理解与动作决策三大核心能力，以“多模态+多芯片+多车型”适配为核心设计理念，支持激光雷达与纯视觉版本，可面向多种主流乘用车平台开展定制化部署。

基于车端有限算力本地部署模型

压缩词表
模型优化
硬件适配

数据是AI模型的基础

已达到千万级Clips数据规模

图、元戎启行VLA模型架构



元戎启行VLA模型的4大功能：空间语义理解、异形障碍物识别、文字类引导牌理解、记忆语音控车，这些功能将根据实际部署节奏逐步释放。

空间语义理解是本次发布的最大亮点：该功能可在视野受限的动态或静态盲区环境中（如公交车遮挡、复杂路口、桥洞等）感知潜在风险，主动对盲区进行“预防性预判”。系统可在风险出现前提前减速、稳妥通行，具备高度拟人化的防御性驾驶策略，为用户带来更安心的出行体验。

图、元戎启行VLA模型的4大功能



二、世界模型：数据闭环驱动，增强环境感知与仿真可靠性

2.1 世界模型：智慧的核心在于反事实推理

世界模型最早出现在机器学习领域：
2018年Jürgen Schmidhuber在NeurIPS发表了一篇以“world models”命名的文章，以认知科学中人脑mental model来类比世界模型，认为mental model参与了人类的认知、推理、决策过程。

该框架图有三个主要的模块组成：
即 Vision Model (V)，Memory RNN (M) 和 Controller (C)，
世界模型这种RNN架构，非常不适合GPU加速，所以前期发展缓慢。

世界模型最核心的能力在于反事实推理
即便对于数据中没有见过的决策，在世界模型中都能推理出决策的结果。

AI研究人员对world model的追求，是试图超越数据，进行反事实推理，回答what if问题能力的追求。这是一种人类天然具备，而当前的AI还做得很差的能力。一旦产生突破，AI决策能力会大幅提升，实现全自动驾驶等场景应用。

图、世界模型框架图（RNN）

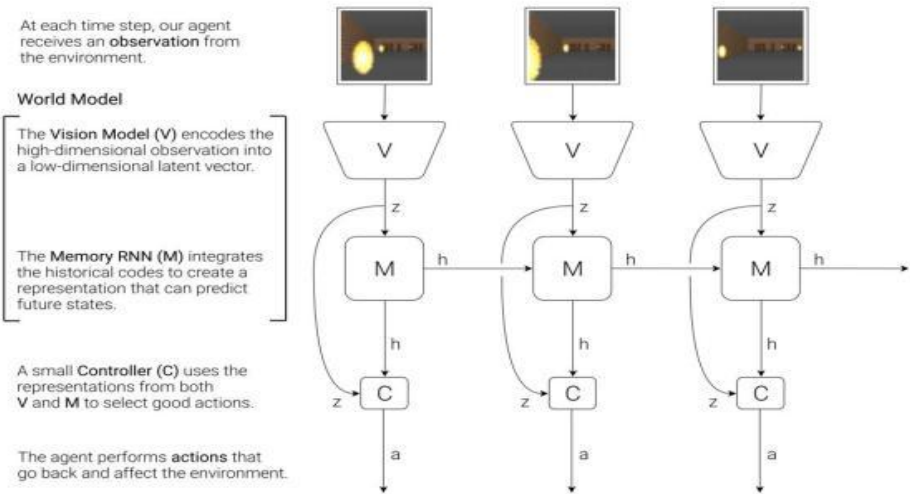
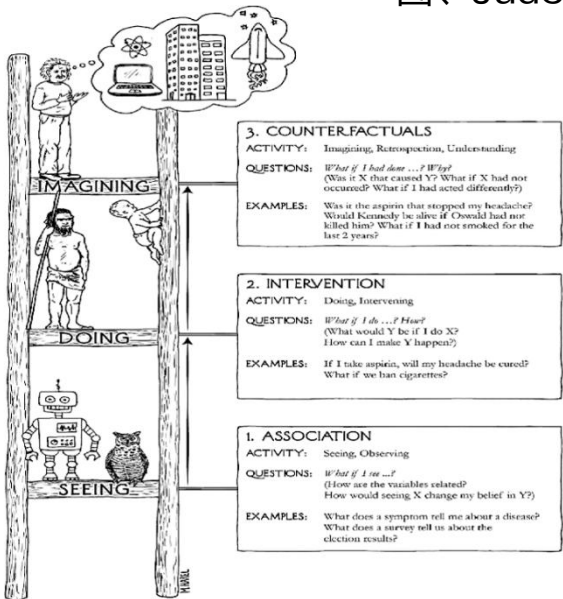


Figure 4. Our agent consists of three components that work closely together: **Vision (V)**, **Memory (M)**, and **Controller (C)**

图、Judea-因果阶梯



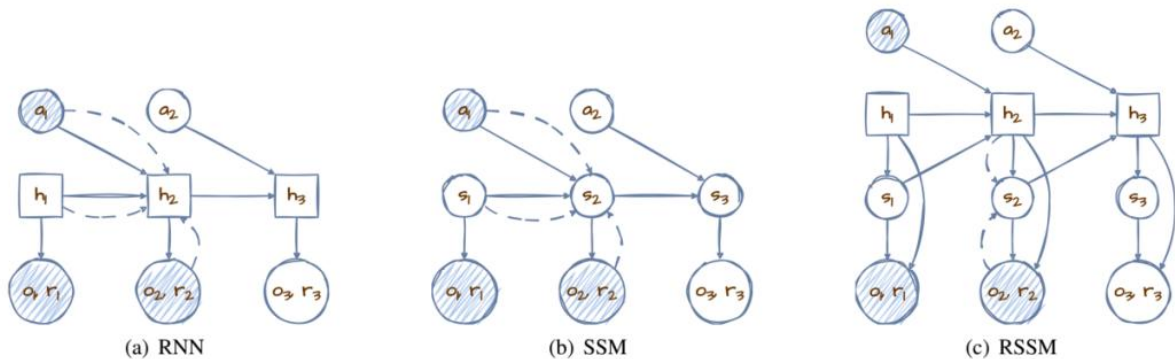
最上层是“反事实”，通过想象回答 what if 问题
中间层是“干预”，强化学习中的探索就是典型的干预
最下层是“关联”，也就是今天大部分预测模型主要在做的事

2.2 RSSM和JEPA是目前的主流世界模型架构

2019年，世界模型进一步演化出了RSSM架构-生成式预测

循环状态空间模型 (RSSM) 的创新之处在于它在图c中战略性地将状态分解为随机和确定性成分，有效地利用了确定性元素的预测稳定性以及随机元素的适应潜力。这种混合结构保证了强大的学习和预测能力，既适应了现实世界的不可预测性，又保持了信息的连续性。通过结合RNN的优势与状态空间模型 (SSM) 的灵活性，RSSM为世界模型建立了一个全面的框架，增强了它们在保持精确性和适应性的同时预测未来状态的能力。

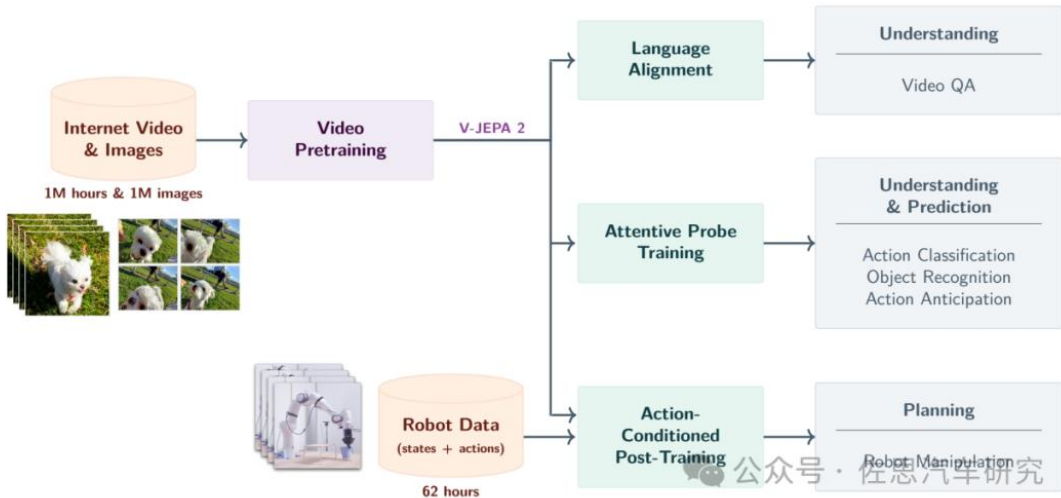
图、世界模型中RNN、RSM和RSSM体系结构的比较



JEPA 架构于2023 年提出
最新版本为 2025 年 6 月发布的 V-JEPA2
-表征式预测

JEPA的设计灵感源自人类认知方式。通过自我监督学习（self-supervised learning），从输入数据（如影片、传感器数据）中提取高维表征，并预测未来状态的表征，而非生成像素级细节。例如，在分析一段影片时，JEPA不会试图预测下一帧的每个像素，而是预测场景的抽象状态，这种方法大幅降低计算需求，并更接近人类的认知效率。

图、世界模型中JEPA-架构



2.3 世界模型：从“感知模仿”到“认知思考”的演进

世界模型通过自监督学习的方式从大量未标注的数据中提取出有价值的信息，从而增强模型的性能和效率

- 在驾驶场景生成方面，世界模型可以创造多样化、真实化的驾驶环境，丰富训练数据集，提高自动驾驶系统的鲁棒性和适应性。
- 在规划和控制策略方面，世界模型也可以帮助车辆进行未来预测和决策制定，提升行驶安全性和稳定性。

图、CVPR2023上Ashok分享Tesla的World Model研究进展



图、马斯克推崇的也是扩散模型diffusion



2.3 海外先验迭代-Wayve生成式世界模型

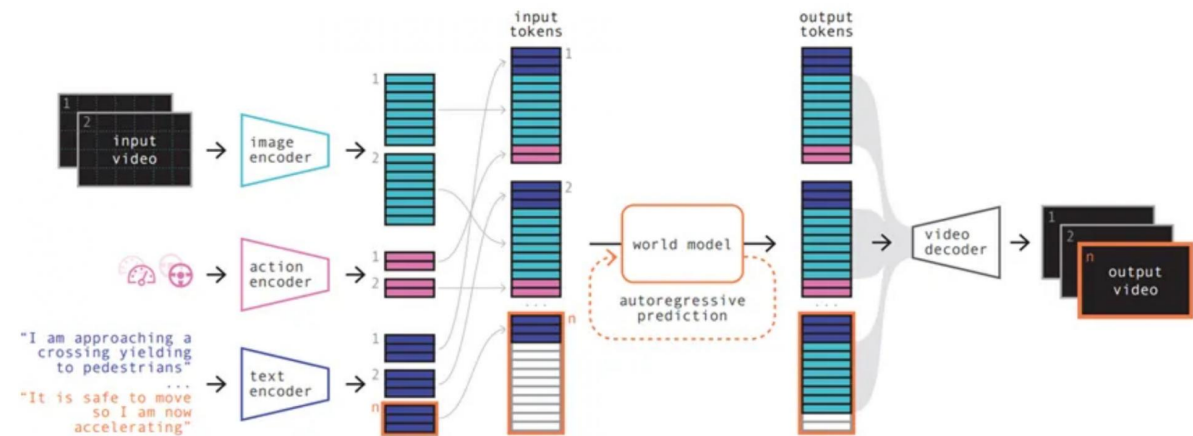
Wayve-GAIA-1（2023.6）生成式世界模型，开启自动驾驶环境建模新范式

GAIA-1 是用于端到端自动驾驶环境建模与决策的生成式世界模型，采用“视频+文本+动作”多模态输入方式，通过大规模英国城市驾驶数据训练，可学习预测视频序列后续帧，具备类大型语言模型（LLM）的自回归预测能力。

该模型突破标准生成视频模型局限，能理解并分解车辆、行人、道路布局、交通信号灯等核心驾驶概念，精准控制自车行为与场景特征；更可通过多样化驾驶数据学习世界潜在生成规则，综合自然结构与模式，生成高度逼真且多样的驾驶场景，为具身人工智能落地（系统与世界互动、理解并重现规则）迈出关键一步。

此外，GAIA-1 通过整合世界模型与驾驶模型，提升对现实场景的泛化能力，未来还将完善预测、模拟、鸟瞰图生成等方法；同时支持训练数据外的场景推断，能生成错误驾驶行为的模拟数据，为自动驾驶安全评估提供受控环境。

图、GAIA-1体系结构



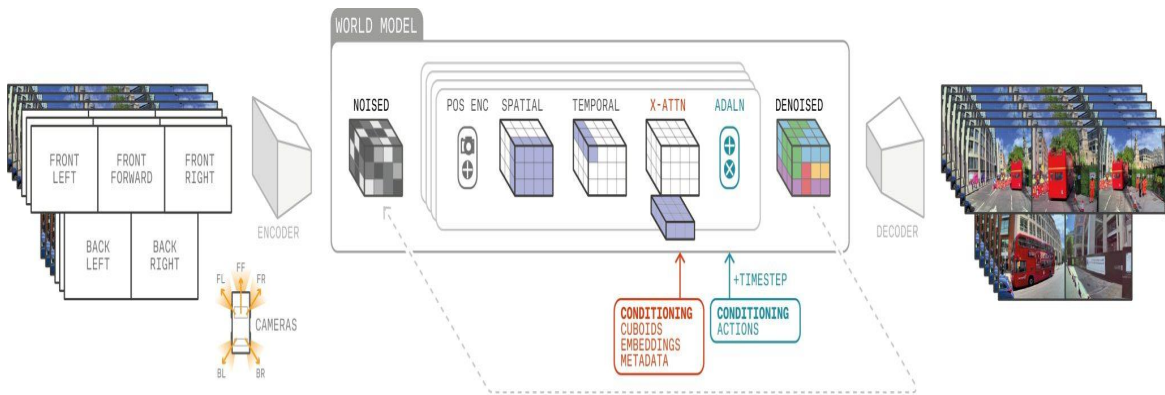
Wayve-GAIA-2（2025.3）场景能力大升级，推动自动驾驶泛化与验证进阶

GAIA-2 在 GAIA-1 基础上，核心升级为“更丰富真实的生成式场景能力”。由于真实世界复杂场景（尤其是低概率安全场景）采集成本高、难度大，GAIA-2 通过扩大地理覆盖范围，可生成英国、美国、德国等多国不同驾驶场景，让 AI 驾驶模型在合成数据中就能学习全球交通规则与道路标志。

同时，GAIA-2 支持时间（黎明到夜晚）、天气（晴天到雨雾）、道路类型（城市拥堵街、郊区乡间路、高速公路等）的精细控制，能高效生成全维度复杂场景，为 AI 驾驶系统提供更贴近真实世界的训练与验证环境。

此外，GAIA-2 具备“多视角同步生成”能力，可同时输出多达五个视角的视频，保证多摄像头视角在时间与空间上的一致性，助力依赖多传感器融合的自动驾驶系统更精准感知环境，做出更安全可靠的驾驶决策。

图、GAIA-2体系结构



2.4 鸿蒙智行 WEWA：构建难例场景库

乾崮智驾ADS 4-世界引擎+世界行为模型架构（WEWA架构）

于2025年4月发布，首版本9月开启规模推送，预计2025年底将新增车外语音控车、隔窗通话等功能

WE-难例扩散生成模型，安全强化学习

- 底层逻辑依托先进的扩散生成模型技术，通过AI生成难例扩散模型，以提供高质量高密度难例场景，并提高难例密度1000倍。智能驾驶的训练数据来源，实现从依赖有限、昂贵且难以获取的“人采实路数据”，向可无限生成、按需定制的“AI合成数据”的历史性跨越。
- 在云端长期实现AI与AI互相监督强化训练，构建了一套“安全优先”的强化学习（RL）机制，系统为AI智能体设定明确的奖惩规则。

WA-全模态感知，原生机模型

- 基于车辆多传感器融合（视觉、听觉、触觉等）收集的全模态感知信息进行直接训练
- 构建智驾原生基模型，显著降低了时延和资源需求，为高效精准决策打下了底层基础
- 运用MOE多专家架构可以根据不同场景调用不同的能力

图、华为WEWA架构



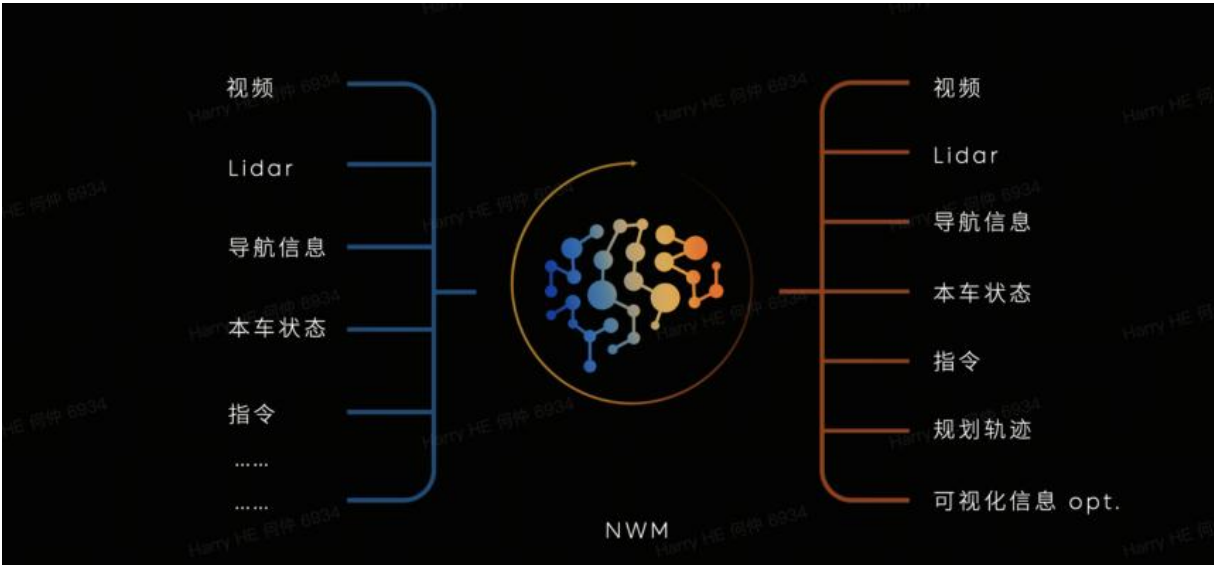
2.5 蔚来世界模型NWM：空间、时间理解及数据利用优化

蔚来在2024年的NIO IN上发布的蔚来世界模型（NWM），是可以全量理解信息、生成新的场景、预测未来可能发生的多元自回归生成模型。核心能力在于：认知世界后，在模型内进行想象重建；基于信息的理解，进行想象推演，从而规划轨迹。

- **空间理解**：生成式模型重构传感器输入，不再局限于“单一传感器数据”，而是像人类一样“脑补”环境全貌（比如雨天通过多模态数据推测前车距离）；
- **时间建模**：自回归模型动态推演长期时间序列（比如预判路口车辆3秒后的行驶轨迹）；
- **数据利用**：自监督学习摆脱“人工标注依赖”，用更少数据实现更优训练效果。

2025年5月30日，「蔚来世界模型 NWM」首个版本正式开启推送。在主动安全、高速领航、城区领航、智能泊车四大场景实现技术能力与用户体验的全面升级。蔚来世界模型首个版本实现的「停车场自主寻路」，可以在地下停车场通过自然语言交互，自主寻找停车场出口与指定楼栋。这个功能是自由交互、无图环境构建、环境理解能力的结合体，既有对复杂世界的认知与理解能力，还有高效实时推理的推理能力。即输入多模态数据，多模态包含图像和语言等，同时能输出多模态数据，包含图像和行动等等。所以，当蔚来世界模型打开语言输入时，它就**包含了VLA视觉语言行动模型的特征**。

图、蔚来世界模型信息输入与输出



图、蔚来世界模型对比

	空间理解 Spatial Cognition	时间理解 Temporal Cognition	使用海量数据 Extensive Data
常规端到端模型 Regular E2E Model	学习任务单一 抽取信息有损失 Single Learning Task Data Loss	无长时序建模能力 No Long Time-Series Modeling	轨迹监督信号信息密度低 依赖感知标注辅助训练 成本效率率低 Reliance on Auxiliary Training
蔚来世界模型 NWM 多元自回归生成模型 Multivariable Autoregressive Generative Model	生成模型重构传感器输入 抽取泛化信息 Extraction and Generalization	自回归模型 自动建模长时序环境 Automatic Long Time-Series Autoregressive Modeling	依赖自监督学习 无需人工标注 Self-Supervised Learning From Raw Data

2.6 Momenta R6 飞轮大模型：具备长短期记忆能力

Momenta: R6 飞轮大模型

2025年下半年将推出数据驱动的基于强化学习的一段式端到端，将感知与规划融合进单一模型，模拟人类驾驶的长期记忆能力。

同时，Momenta保留了DLP 模型（类似短期记忆机制），用于快速学习新数据。这一双模型架构协同作用，确保了训练数据的高质量和模型训练的低成本。

通过海量真实道路数据训练算法模型，并利用数据闭环持续迭代优化系统性能（截至2025.4累计合作量产车型超130款）。

图、Momenta R6 飞轮大模型记忆能力示意



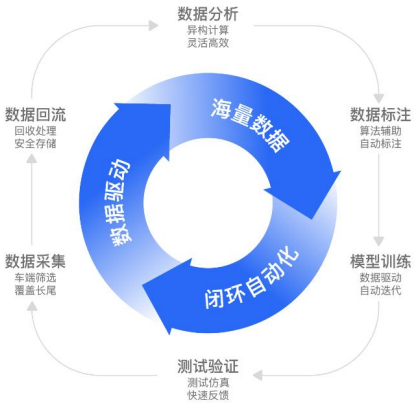
图、Flywheel 数据驱动闭环



图、闭环自动化（CLA）工具链

闭环自动化 (Closed Loop Automation)

CLA - 闭环自动化，是一整套让数据流推动数据驱动的算法自动迭代的工具链。CLA能自动筛选出海量黄金数据，驱动算法的自动迭代，让自动驾驶的飞轮越转越快。

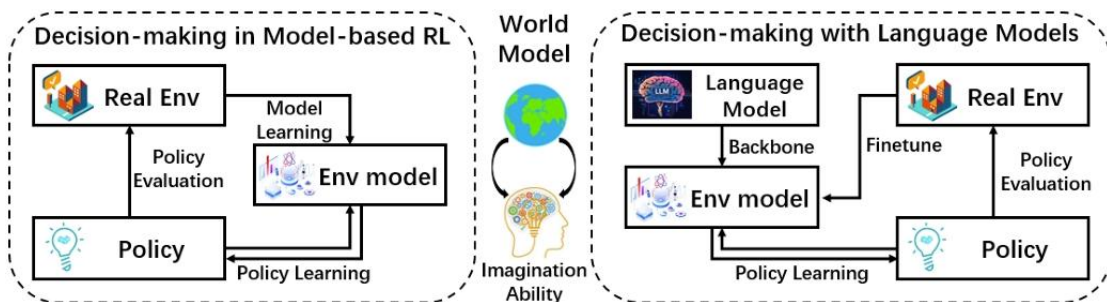


三、融合互补，共筑安全与体验新壁垒

当前世界模型主要聚焦在视觉-运动空间，而VLA将视觉与语言紧密耦合，多模态融合程度更深

- World Model（世界模型）本身并不一定牵涉语言模态，它更多是对视觉和运动数据的生成与理解。不过在具体实现上，世界模型也可以结合语言信息。核心是通过视频预测或其它方式，学习真实世界本质的时空规律。借助世界模型进行仿真（理解世界运行规律并生成有效数据难例数据）和 强化学习交互（提供虚拟环境和预测能力）。
- VLA天生就是多模态融合的产物，视觉和语言信息在模型内部被打通，必要时还可以接入语音、地图等模态，主张通过思维链理解事物，透过自然语言描述模型对场景的理解，提升泛化能力。

图、在决策任务中利用世界模型的两方案



VLA与世界模型在技术上并非同级或对立关系。我们将发展路径分为两派，实质上是产业玩家在实现端到端能力之后，在能力优化侧重点上出现了分化。

- 先重塑至VLA架构，融合语言模块，再提升强化学习的能力，如理想、小鹏、元戎启行
- 先搭建世界模型环境，持续优化现有架构，逐步融合多模态信息，如华为、蔚来、Momenta

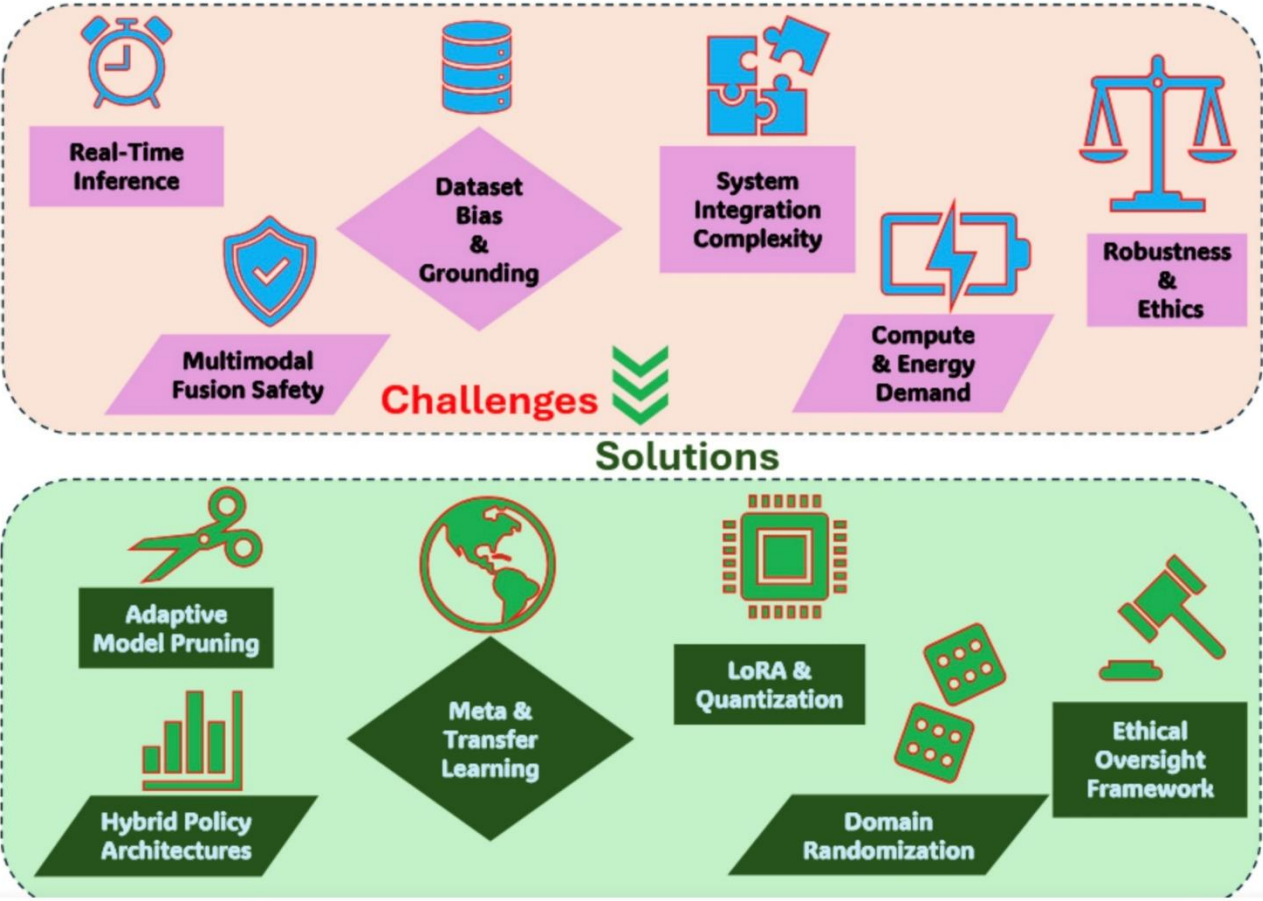
蔚来世界模型（NWM）-包含 VLA视觉语言行动模型的特征

在蔚来世界模型首个版本中全量推送的「停车场自主寻路」功能，就是蔚来世界模型对语言输入、图像及环境认知、理解及推理能力和执行动作的最直接体现。这也是行业首次量产VLA形态的类似功能。

华为（WEWA）靳玉志 - 更看重WA，中间省掉L这个环节

WA就是直接通过行为端，或者说直接通过vision这样的信息输入控车，而不是把各种各样的信息转成语言，再通过语言大模型来控制车。注：这里的vision只是一个代表，它可能来自于声音，可能来自于vision，也可能来自于触觉。

图、VLA模型面临一系列相互关联的挑战



应用大模型的过程中，

解决软硬件资源的效率问题

1) 系统集成复杂性：集成感知、推理、控制等各种模块会形成难以优化和维护的复杂架构。

2) 算力与能耗需求：大规模 VLA 系统的能耗与算力需求阻碍了其在嵌入式或移动平台上的部署。

3) 实时推理限制：由于自回归解码器的顺序特性以及多模态输入的高维度，实现资源受限硬件上的实时推断依然很困难。

保障模型稳健安全的能力

4) 多模态对齐：将视觉、语言和动作融为一体的控制策略，在遇到意外环境变化时会产生安全薄弱环节。

5) 泛化与评估：数据集偏差和定位错误损害了模型的泛化能力，常导致在分布外任务上失效。

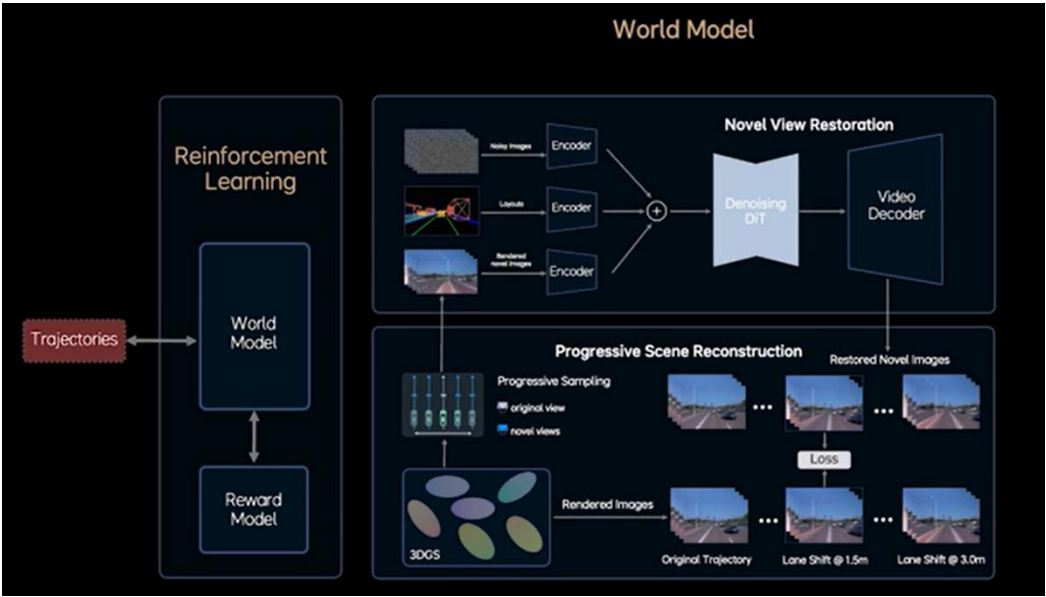
6) 鲁棒性与可靠性：对环境变化的鲁棒性以及隐私与偏见缓解等伦理考量，又为社会和法规层面增添了新的挑战。

3.3 世界模型&强化学习是VLA路径玩家的下一发力重心

理想MindVLA 依托自研的云端统一世界模型，该模型融合了重建与生成技术。

其中，重建模型具备三维场景还原能力，生成模型则可实现新视角补全以及未见视角预测。通过将这两种技术路径相结合，MindVLA 构建出贴近真实世界、符合物理世界规律的仿真环境，为解决训练偏差问题提供了有效方案。

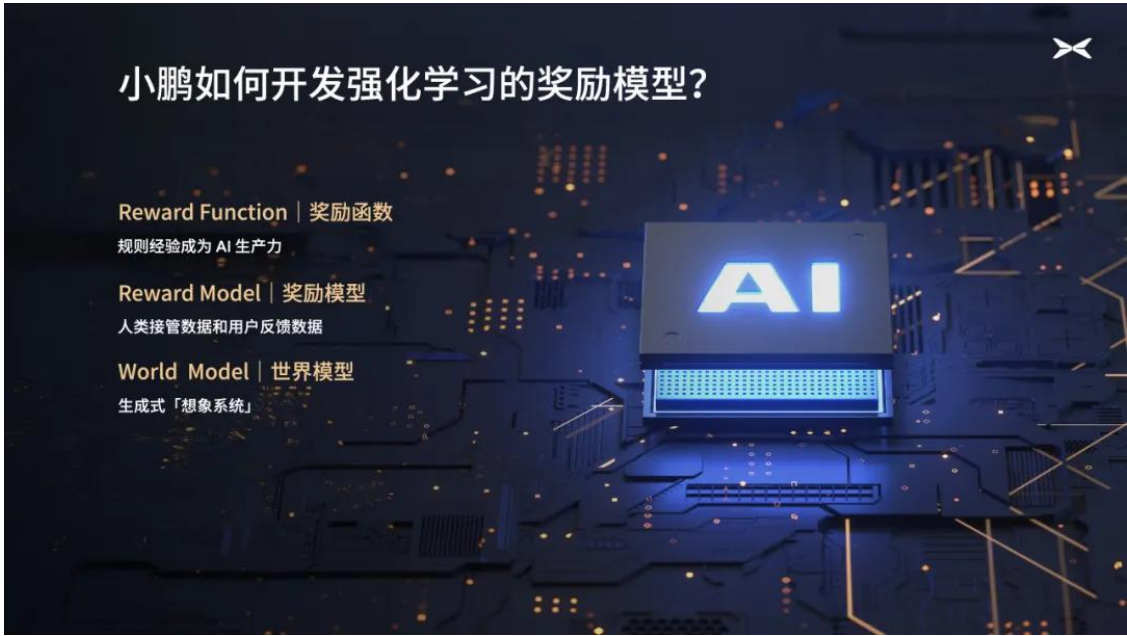
图、理想-世界模型中的强化学习



小鹏汽车正着手开发世界模型(World Model)，作为“云端模型工厂”的重要一环，支持基座模型的性能优化。

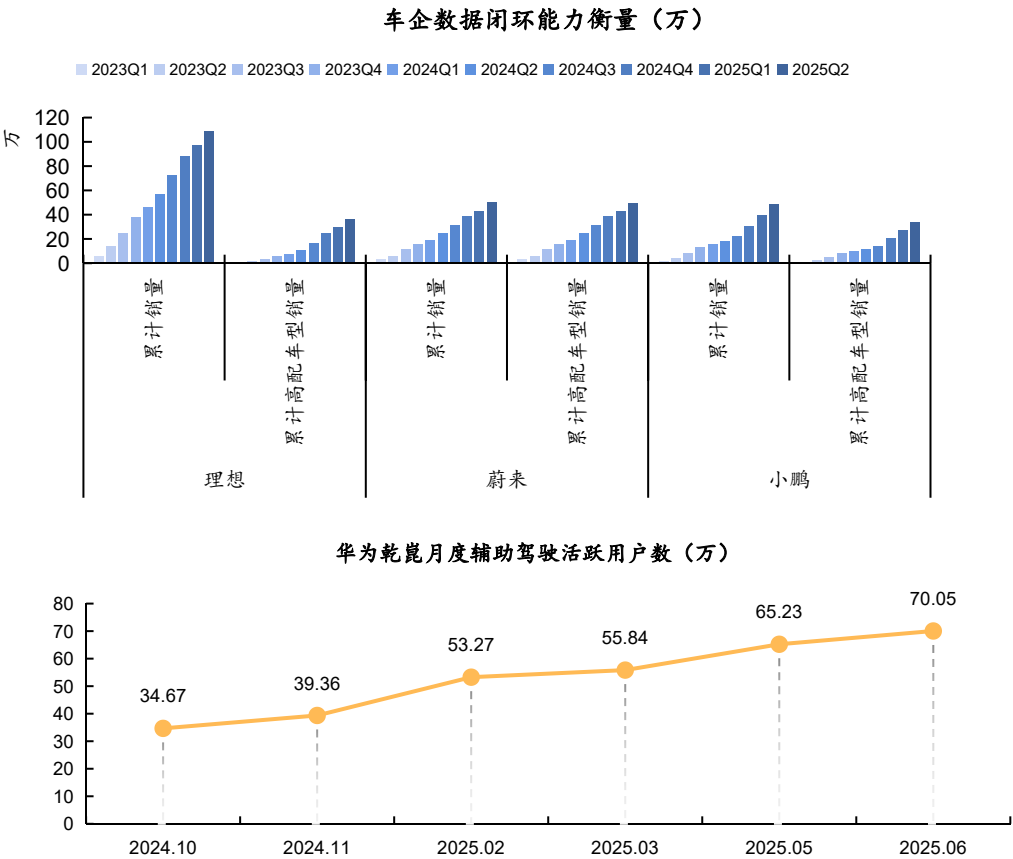
根据小鹏汽车XPENG公众号资讯可知，小鹏的世界模型是一种实时建模和反馈系统，能够基于动作信号模拟出真实环境状态，渲染场景，并生成场景内其他智能体（也即交通参与者）的响应，从而构建一个闭环的反馈网络，帮助基座模型不断进化，逐渐突破过去“模仿学习”的天花板。

图、小鹏-世界模型开发



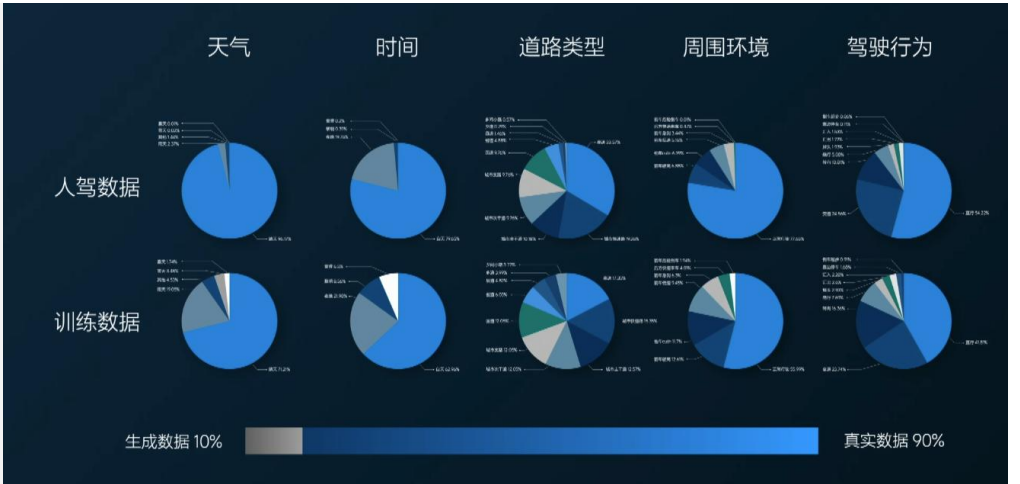
3.4 有效数据量决定模型迭代能力

我们认为数据闭环是企业构筑辅助驾驶能力领先性的核心，
累计销量：新势力车企普遍通过全量数据采集来构建竞争优势，即使低配车型也贡献数据（选自2023年以来累计零售销量，忽略衰减体量）
高配硬件车型累计销量：数据采集的前提条件是硬件基础，即使车辆没有激活或者使用智能驾驶软件（车企智驾版本销量占比为估算值）
用户活跃量：活跃用户数据是“更好的真实样本”，尤其对解决长尾问题至关重要

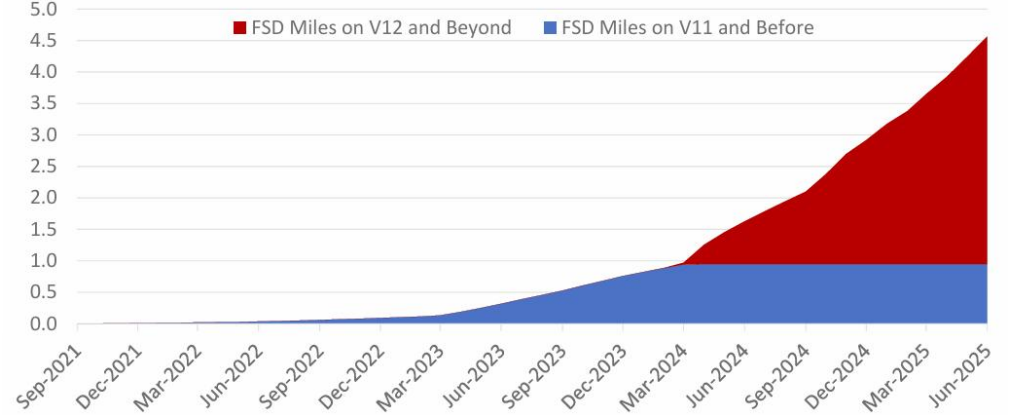


以真实数据为基，使用生成数据作为有效补充：累计辅助驾驶里程数据有效性高，数据规模越大，模型通过强化训练（RLHF）迭代升级的上限越高，形成“数据飞轮”效应。
 华为 - 智驾总里程超过35.41亿公里（2025.6），已实现6亿公里云端仿真验证（2025.4）

图、理想（2025.7）
 - 已累计采集12亿公里有效驾驶数据
 - 训练数据基于人驾数据的调整，训练数据是真实数据和生成数据的组合（9:1）



图、特斯拉-
 - FSD累计里程已超过45亿英里（2025Q2）
 - 影子模式（Shadow Mode）：特斯拉车辆在后台运行影子模式，通过模拟真实驾驶场景不断优化算法。



3.5 持续优化MPA&MPI向L2+演进

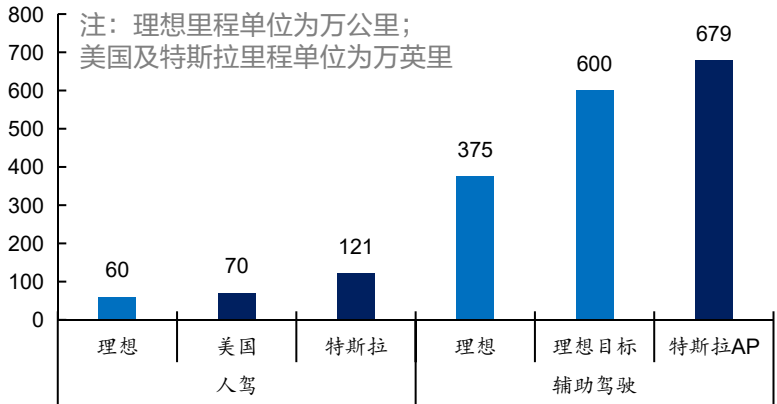
根据当前头部车企（如特斯拉、理想、Momenta）的驾驶数据表现，MPA能力已达到人驾安全水平的6倍以上，且仍在持续优化。
技术演进方向：在减少接管次数的同时，保障甚至提升安全能力，智驾体验逐步接近“全场景可用”。

MPA 平均事故里程-是底线，决定技术能否规模化应用

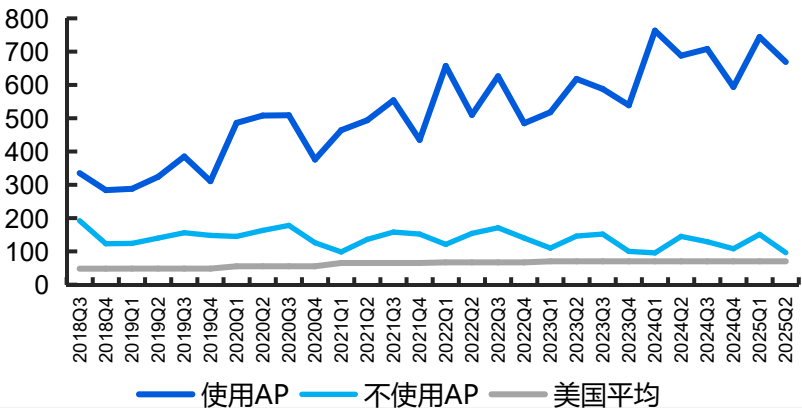
图、智己汽车辅助驾驶方案（Momenta配套）安全倍数



图、理想与特斯拉MPA数据

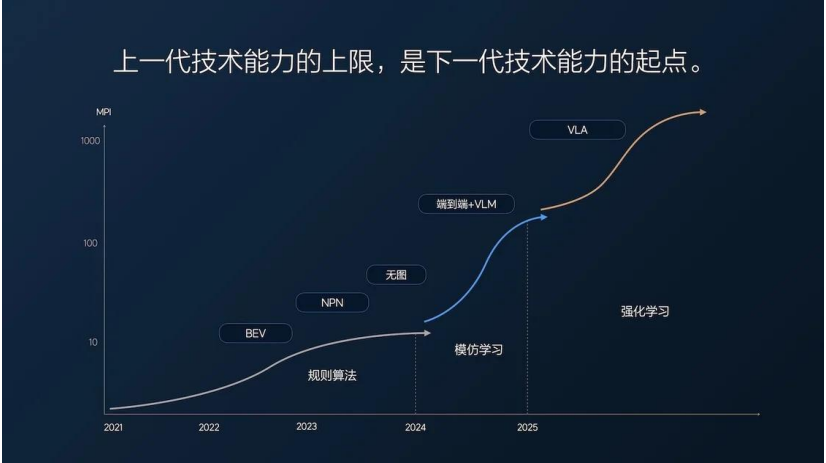


图、美国每发生一次事故行驶距离（万英里）



MPI 平均接管里程-是天花板，决定用户体验和商业效率

图、理想-基于端到端+VLM已实现约MPI225公里，目标通过VLA司机大模型达到MPI1000公里



图、特斯拉不同FSD版本下平均无接管高速里程（英里）
基于FSD13.2实现MPI213英里



VLA与世界模型在技术上并非同级或对立关系。我们将发展路径分为两派，实质上是产业玩家在实现端到端能力之后，在能力优化侧重点上出现了分化。

- 先重塑至VLA架构，融合语言模块，再提升强化学习的能力，如理想、小鹏、元戎启行
- 先搭建世界模型环境，持续优化现有架构，逐步融合多模态信息，如华为、蔚来、Momenta

比较与结论

- ◆ **通往高阶智驾的互补路径：**VLA 侧重语义融合与交互控制，通过语言增强可解释性与用户协同；世界模型侧重环境预测与仿真可靠性，通过生成技术提升安全冗余。
- ◆ **技术融合趋势明显：**双方均在向对方领域渗透：VLA 引入强化学习与仿真优化动作生成，世界模型扩展语言交互与多模态理解。
- ◆ **数据与安全为核心壁垒：**数据飞轮效应成为竞争关键-头部车企凭借高活跃用户积累海量里程数据，通过RLHF驱动模型高效迭代；安全性能量化用户体验：当前第一梯队MPA（平均事故里程）已达人类驾驶安全水平的6倍以上，MPI（平均接管里程）持续优化。

行业评级与投资建议：多模态VLA模型引入长序思维能力，世界模型提供仿真与强化训练环境，有效数据闭环+安全能力提升（MPA&MPI），共同构筑智能驾驶长期壁垒。考虑到L3辅助驾驶商用标准落地在即，多款L3级架构车型陆续上市，场景切割的阶段性功能实现，并向全场景覆盖的最终愿景发力，维持辅助驾驶行业“推荐”评级。智驾能力成为车企估值分化关键因素，短期关注**数据规模大、算法迭代快**的车企：**理想汽车、小鹏汽车、蔚来汽车、鸿蒙智行合作车企，辅助驾驶解决方案供应商Momenta、元戎启行等。**

- ◆ 辅助驾驶政策法规推进不及预期：辅助驾驶政策法规若推进缓慢，将延缓高级别功能的商业化落地进程。
- ◆ 数据安全与隐私保护监管趋严：数据安全监管趋严则会直接增加企业合规成本并限制数据使用。
- ◆ 技术可靠性与长尾场景安全问题导致的信任与合规风险：技术可靠性不足引发的安全事故，更会严重打击市场信心。
- ◆ 车企销量与技术搭载不及预期：若车企销量不及预期，将直接影响技术供应商的订单与收入。
- ◆ 行业竞争加剧导致技术研发投入承压：行业竞争白热化会持续推高研发投入，挤压企业利润空间。

汽车小组介绍

戴畅，首席分析师，上海交通大学本硕，9年汽车卖方工作经验，全行业覆盖，深耕一线，主攻汽车智能化和电动化，善于把握行业周期拐点，技术突破节奏，以及个股经营变化。
胡惠民，汽车行业分析师，华威大学金融硕士，4年汽车卖方研究经验，擅长对乘用车行业整体需求的判断，以及自下而上对个股的挖掘与跟踪。
吴铭杰，汽车行业分析师，上海财经大学金融专业硕士，2年汽车市场研究经验，擅长发现个股边际变化，从底部挖掘潜力个股，目前主要覆盖汽车热管理及机器人产业链。
陈飞宇，汽车行业分析师，墨尔本大学金融硕士，3年汽车市场研究经验，擅长把握产业技术变革，深度跟踪公司经营变化，主要覆盖汽车零部件智能化软硬件。
徐鸣爽，研究助理，复旦大学文学学士、波士顿大学经济学/东北大学数据分析双硕士，2年卖方研究经验，对搭建行业研究与数据跟踪框架有心得，目前主要覆盖商用车。

分析师承诺

戴畅, 陈飞宇, 本报告中的分析师均具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立，客观的出具本报告。本报告清晰准确的反映了分析师本人的研究观点。分析师本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收取到任何形式的补偿。

国海证券投资评级标准

行业投资评级

推荐：行业基本面向好，行业指数领先沪深300指数；
中性：行业基本面稳定，行业指数跟随沪深300指数；
回避：行业基本面向淡，行业指数落后沪深300指数。

股票投资评级

买入：相对沪深300 指数涨幅20%以上；
增持：相对沪深300 指数涨幅介于10%～20%之间；
中性：相对沪深300 指数涨幅介于-10%～10%之间；
卖出：相对沪深300 指数跌幅10%以上。

免责声明

本报告的风险等级定级为R3，仅供符合国海证券股份有限公司（简称“本公司”）投资者适当性管理要求的客户（简称“客户”）使用。本公司不会因接收人收到本报告而视其为客户。客户及/或投资者应当认识到有关本报告的短信提示、电话推荐等只是研究观点的简要沟通，需以本公司的完整报告为准，本公司接受客户的后续问询。

本公司具有中国证监会许可的证券投资咨询业务资格。本报告中的信息均来源于公开资料及合法获得的相关内部外部报告资料，本公司对这些信息的准确性及完整性不作任何保证，也不保证其中的信息已做最新变更，也不保证相关的建议不会发生任何变更。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。报告中的内容和意见仅供参考，在任何情况下，本报告中所表达的意见并不构成对所述证券买卖的出价和征价。本公司及其本公司员工对使用本报告及其内容所引发的任何直接或间接损失概不负责。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或者金融产品等服务。本公司在知晓范围内依法合规地履行披露义务。

风险提示

市场有风险，投资需谨慎。投资者不应将本报告为作出投资决策的唯一参考因素，亦不应认为本报告可以取代自己的判断。在决定投资前，如有需要，投资者务必向本公司或其他专业人士咨询并谨慎决策。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。投资者务必注意，其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。

若本公司以外的其他机构（以下简称“该机构”）发送本报告，则由该机构独自为此发送行为负责。通过此途径获得本报告的投资者应自行联系该机构以要求获悉更详细信息。本报告不构成本公司向该机构之客户提供的投资建议。

任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。本公司、本公司员工或者关联机构亦不为该机构之客户因使用本报告或报告所载内容引起的任何损失承担任何责任。

郑重声明

本报告版权归国海证券所有。未经本公司的明确书面特别授权或协议约定，除法律规定的情况外，任何人不得对本报告的任何内容进行发布、复制、编辑、改编、转载、播放、展示或以其他方式非法使用本报告的部分或者全部内容，否则均构成对本公司版权的侵害，本公司有权依法追究其法律责任。

国海证券 · 研究所 · 汽车研究团队

心怀家国，洞悉四海



国海研究上海

上海市黄浦区绿地外滩中心C1栋
国海证券大厦

邮编：200023

电话：021-61981300

国海研究深圳

深圳市福田区竹子林四路光大银
行大厦28F

邮编：518041

电话：0755-83706353

国海研究北京

北京市海淀区西直门外大街168
号腾达大厦25F

邮编：100044

电话：010-88576597