端到端自动驾驶行业研究报告



辰韬资本
CHERISH CAPITAL

南京大学上海校友会
自动驾驶分会

九章智驾
智能驾驶 产业智库

Copyright © 2024 辰韬资本, 南京大学上海校友会自动驾驶分会, 九章智驾. 保留所有权利。

报告摘要

端到端逐渐成为自动驾驶行业的共识

2023 年以来，在行业龙头特斯拉的标杆作用、大模型代表的 AGI 技术范式、以及自动驾驶拟人化和安全性需求的共同推动下，自动驾驶行业对于端到端的关注度一路升温。产业界、学术界和资本市场在端到端自动驾驶领域都有里程碑事件发生。

端到端自动驾驶已经成为明确的行业共识。本行研报告访谈了 30 余位自动驾驶行业一线专家，其中 90% 表示自己所供职的公司已投入研发端到端技术。

作为新兴技术领域，端到端的基本概念存在误区，本报告提出一套可供参考的概念定义和术语体系

早期，端到端的核心定义是“从传感器输入到控制输出的单一神经网络模型”，近年来，端到端的概念有了更大范围的延展。本报告认为，端到端的核心定义标准应为：感知信息无损传递、可以实现自动驾驶系统的全局优化。

基于以上定义标准，并结合自动驾驶系统中 AI 的应用，本报告将自动驾驶技术架构分为四个阶段，分别是感知“端到端”、决策规划模型化、模块化端到端、以及 One Model 端到端，其中，后两个阶段符合前述端到端的定义标准。

同时，端到端与大模型、世界模型、纯视觉传感器方案等相关概念常常被混淆，本报告也对其区别与联系做了辨析。

端到端的落地面临诸多挑战，包括技术路线、数据和算力需求、测试验证、组织资源投入等

技术路线：端到端技术路线还未形成最佳实践，技术路线存在分歧。

数据：在端到端技术架构下，训练数据的重要性得到前所未有的提升，其中，数据量、数据标注、数据质量和数据分布的相关问题都可能成为限制端到端应用的挑战。

训练算力：端到端训练算力需求急剧提升，行业头部玩家均储备千卡～万卡级训练算力。

测试验证：现有测试验证方法不适用于端到端自动驾驶，行业亟需新的测试验证方法论和工具链。

Copyright © 2024 辰韬资本, 南京大学上海校友会自动驾驶分会, 九章智驾. 保留所有权利。

组织资源投入：端到端需要组织架构重塑，也需要将资源投入倾斜到数据侧，对现有模式提出挑战。

另一方面，虽然有观点认为车端算力不足和可解释性问题是端到端落地的限制因素，但本报告提出相反结论并进行辨析。

展望：端到端将于 2025 年开始上车，带动上游技术进步、市场和产业格局演变

自动驾驶行业头部玩家纷纷提出端到端量产规划，预计模块化端到端系统将于 2025 年开始上车。

技术上，端到端的落地会推动其依赖的上游工具链、芯片等加速进步。市场端，端到端带来的自动驾驶体验提升，将会带来高阶辅助驾驶渗透率的提升；由于其强泛化性，端到端也可能驱动自动驾驶跨地理区域、跨国家、跨场景的应用。产业格局方面，端到端使数据和 AI 人才的重要性进一步提升，可能催生新的产业分工和商业模式。

端到端自动驾驶与通用人形机器人相互启发，竞逐物理世界 AGI 落地之路

自动驾驶在发展的早期借鉴了很多机器人行业的积累，包括感知算法、规划算法、中间件和传感器等多个层面。近年来，自动驾驶技术和产业成熟度提高，其中端到端自动驾驶提供了一套已验证、可量产的基于数据驱动的 AGI 技术范式，对于通用人形机器人的有较强的借鉴作用。

目前，自动驾驶和通用人形机器人已经成为物理世界 AGI 发展最重要的两个应用领域，比较二者在落地之路上遇到问题和解决路径，可以给另一个领域更大的启发。

Copyright © 2024 辰韬资本, 南京大学上海校友会自动驾驶分会, 九章智驾. 保留所有权利。

内容目录

前言

端到端开启新一轮自动驾驶产业革命 …… 01
端到端的“战略共识”和“战术分歧” …… 02

01
端到端的基本概念

1.1 传统自动驾驶架构 …… 04
1.2 端到端自动驾驶的定义 …… 05
1.3 端到端相关概念的联系与区别 …… 08
1.4 早期端到端自动驾驶的关键成果 …… 11
1.5 近期端到端自动驾驶的重要进展 …… 14

02
端到端主要玩家分析

2.1 主机厂 …… 17
2.2 自动驾驶算法和系统公司 …… 19
2.3 自动驾驶生成式 AI 公司 …… 23
2.4 学术研究机构 …… 24

03
端到端的发展驱动力

3.1 对标驱动 …… 26
3.2 用户体验驱动 …… 27
3.3 组织价值驱动 …… 27
3.4 技术驱动 …… 29

04
端到端落地之路的挑战

4.1 技术路线分歧 …… 31
4.2 对训练数据的需求量空前提升 …… 31
4.3 对训练算力的需求越来越高 …… 32

Copyright © 2024 辰韬资本, 南京大学上海校友会自动驾驶分会, 九章智驾. 保留所有权利。

4.4 测试验证方法尚不成熟 …… 34
4.5 组织资源投入的挑战 …… 34
4.6 车载芯片算力会不会成为瓶颈? …… 34
4.7 模型缺乏“可解释性”是否构成障碍? …… 35

05
展望端到端

5.1 量产进度 …… 36
5.2 技术发展趋势 …… 36
5.3 产业发展趋势 …… 40

06
端到端自动驾驶与通用机器人

6.1 阶段一：机器人赋能自动驾驶早期的发展 …… 43
6.2 阶段二：自动驾驶产业化加速，端到端技术可反哺机器人 …… 44
6.3 阶段三：自动驾驶和机器人竞逐物理世界 AGI …… 44

本报告访谈和编写项目组

特别声明

Copyright © 2024 辰韬资本, 南京大学上海校友会自动驾驶分会, 九章智驾. 保留所有权利。

图示目录

图示 1：2017-2023 年中国无人驾驶汽车行业投资整体情况 …… 01
图示 2：行业研究问卷受访者结构及其对端到端大模型的态度 …… 02
图示 3：对于端到端落地时间、技术终局和行业格局的预测 …… 03
图示 4：传统自动驾驶架构的典型代表——Apollo 3.0 软件架构 …… 04
图示 5：自动驾驶架构演进示意图 …… 06
图示 6：典型的模块化端到端 UniAD 架构图 …… 07
图示 7：DAVE-2 的网络结构 …… 11
图示 8：Wayve 基于强化学习的端到端神经网络 …… 12
图示 9：传统自动驾驶解决方案与 OpenPilot 方案对比 …… 13
图示 10：UniAD 架构图 …… 14
图示 11：LINGO-2 架构 …… 15
图示 12：小鹏发布端到端大模型 …… 17
图示 13：华为发布 ADS3.0 的端到端架构 …… 18
图示 14：零一汽车发布端到端自动驾驶系统 …… 19
图示 15：元戎启行端到端演进路线 …… 19
图示 16：商汤绝影发布 UniAD 端到端解决方案 …… 20
图示 17：鉴智机器人 GraphAD 端到端网络 …… 21
图示 18：地平线 SuperDrive 技术架构向端到端演进 …… 22
图示 19：DriveDreamer 模型架构 …… 24
图示 20：理想汽车在产品发布会上分享与 MARS Lab 合作技术的落地应用情况 …… 25
图示 21：特斯拉 FSD 平均接管里程和无接管行程次数占比的演变 …… 26
图示 22：大语言模型将传统自然语言处理的多个任务统一为一个基础模型 …… 28
图示 23：MLOps 工具链全景图 …… 29
图示 24：自然语言处理启发自动驾驶从传统范式到 AGI 新范式的演变 …… 30
图示 25：英伟达 H100 GPU 头部客户排名预测 …… 33
图示 26：视觉感知算法领域开源项目与闭源项目的性能提升对比 …… 37
图示 27：基于世界模型的仿真器和传统仿真器对比 …… 39
图示 28：在端到端量产前和量产后的高速 NOA 和城市 NOA 的渗透率预测 …… 41
图示 29：对通用人工智能实现时间的预测 …… 45

Copyright © 2024 辰韬资本, 南京大学上海校友会自动驾驶分会, 九章智驾. 保留所有权利。

Copyright © 2024 辰韬资本, 南京大学上海校友会自动驾驶分会, 九章智驾. 保留所有权利。

前言

端到端开启新一轮自动驾驶产业革命

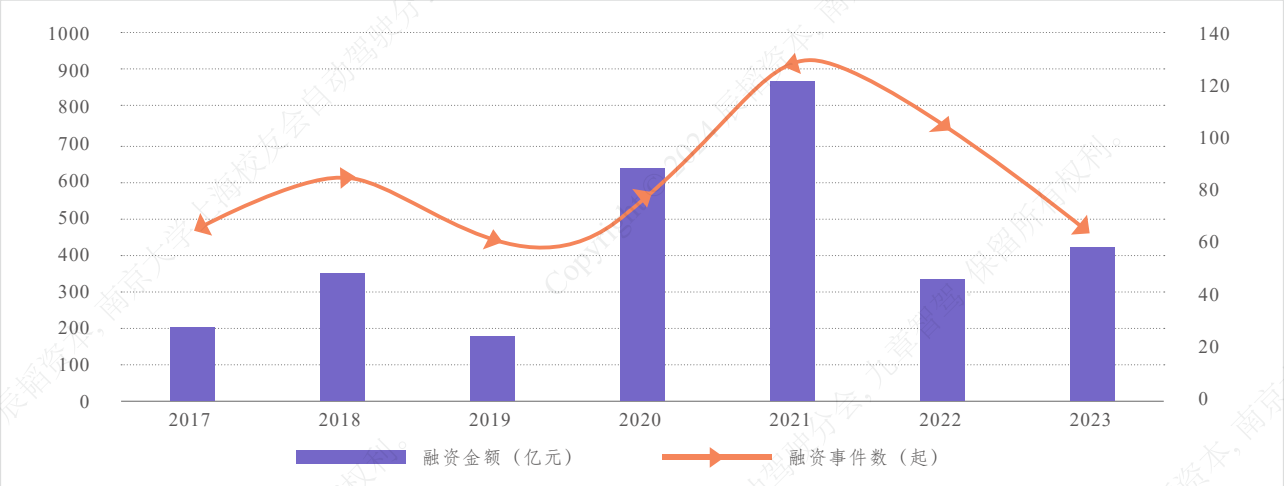
2023 年以来，端到端自动驾驶在学术界、产业界的热度逐渐升高。

学术研究方面，2023 年 6 月，上海人工智能实验室提出的 UniAD（Unified Autonomous Driving）获 CVPR 2023 最佳论文，大大提升了这一领域的关注度。紧随其后的是，一系列端到端自动驾驶的研究成果涌现，这一热度一直持续至今。

产业落地方面，特斯拉自 FSD (Full Self Driving) v12 更新以来切换至端到端技术架构，国内以鸿蒙智行、小鹏汽车、元戎启行、商汤绝影、零一汽车为代表的头部汽车主机厂和智能驾驶技术公司纷纷投入研发端到端系统，并于最近半年陆续对外披露上车量产规划。

端到端的升温还体现在资本市场对相关公司的高度关注上。2024 年 5 月，聚焦研发端到端自动驾驶以及自动驾驶大模型的英国初创公司 Wayve.AI 官宣获得了 10.5 亿美元融资，其股东包括微软、软银、英伟达等。自 2021 年上半年以来，全球自动驾驶投融资市场进入低迷期，已经鲜有资本在自动驾驶领域投入巨额资金（上一次有超过 10 亿美元的融资事件是 Waymo 于 2021 年上半年完成的 25 亿美元融资），但 Wavye 的融资可能意味着端到端技术将引领自动驾驶进入下一波产业发展和资本涌入的浪潮。

图示 1：2017-2023 年中国无人驾驶汽车行业投资整体情况（单位：亿元，起）



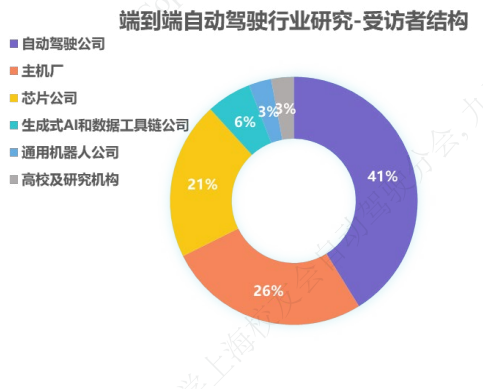
数据来源：IT 桔子 前瞻产业研究院

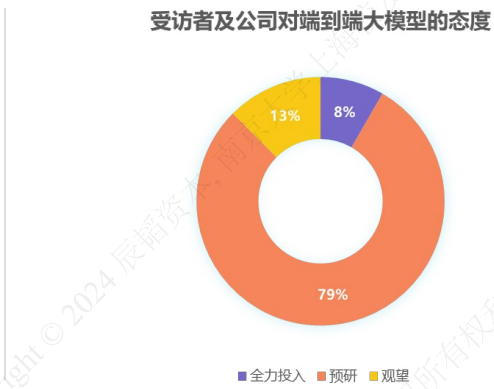
无论是学术界、产业界在端到端自动驾驶领域的快速进展，还是资本市场对相关公司的关注，都指向一个趋势：端到端开启了新一轮的自动驾驶产业革命。

端到端的“战略共识”和“战术分歧”

为了深入理解端到端大模型在自动驾驶产业中的发展现状和未来趋势，我们开展了一项全面的调研，访谈了 30 余位自动驾驶行业一线专家、研究人员和公司决策者，覆盖了主机厂、算法公司、芯片公司、数据服务和工具链公司等与端到端自动驾驶密切相关的产业链环节。

图示 2：行业研究问卷受访者结构及其对端到端大模型的态度





数据来源：本报告编写组，访谈专家

我们首先初步调研了受访者及其所供职公司对端到端自动驾驶的整体态度。与上节所述的现象类似，受访者普遍认可端到端这一技术方向，其所在公司，要么已经投入重要资源全面拥抱端到端，要么已经开启端到端自动驾驶的前瞻预研探索，持观望态度的公司极少。

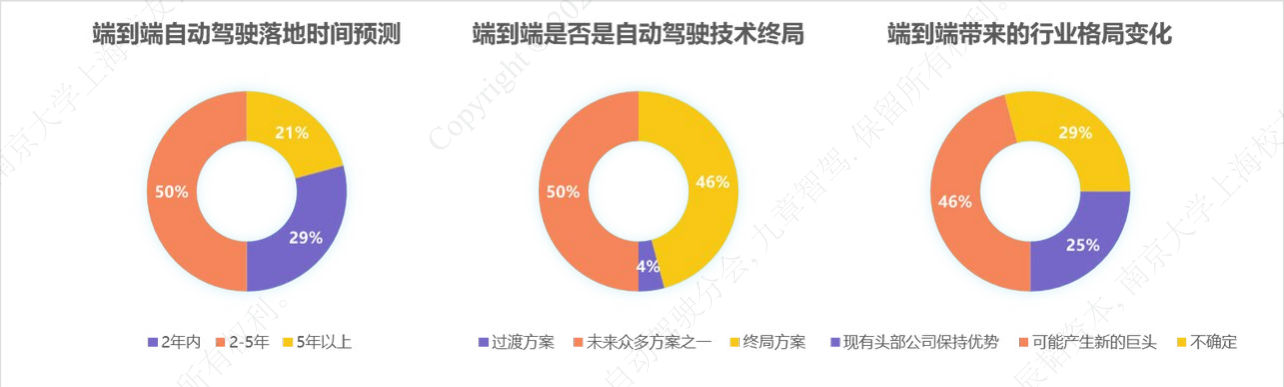
然而，在对端到端发展的具体问题的预判上，受访者的分歧较大：

a. **落地时间预判**：一半左右受访者认为端到端自动驾驶会在 2~5 年间落地，态度非常激进（2 年内落地）和态度非常保守（5 年以上落地）的观点也同时存在。

b. **技术终局预判**：一半受访者认为端到端是自动驾驶的最终解决方案；一半受访者认为端到端只是未来众多技术方案之一。

c. **行业格局预判**：一半受访者认为端到端会催生出新的巨头公司，而现有巨头的优势将不再明显；另一部分受访者则认为现有的巨头公司仍然能够保持竞争优势。这反映了行业对于“端到端是否是一种颠覆式创新”存在的分歧。

图示 3：对于端到端落地时间、技术终局和行业格局的预测



数据来源：本报告编写组，访谈专家

这一问卷调研反映了目前自动驾驶行业的现状：一方面端到端受到广泛关注，大部分技术公司都难以承受错过这一次技术革命的后果；另一方面，由于行业对端到端的探索刚刚起步，对许多基础问题缺乏讨论，对其未来的发展方向还难以达成共识。

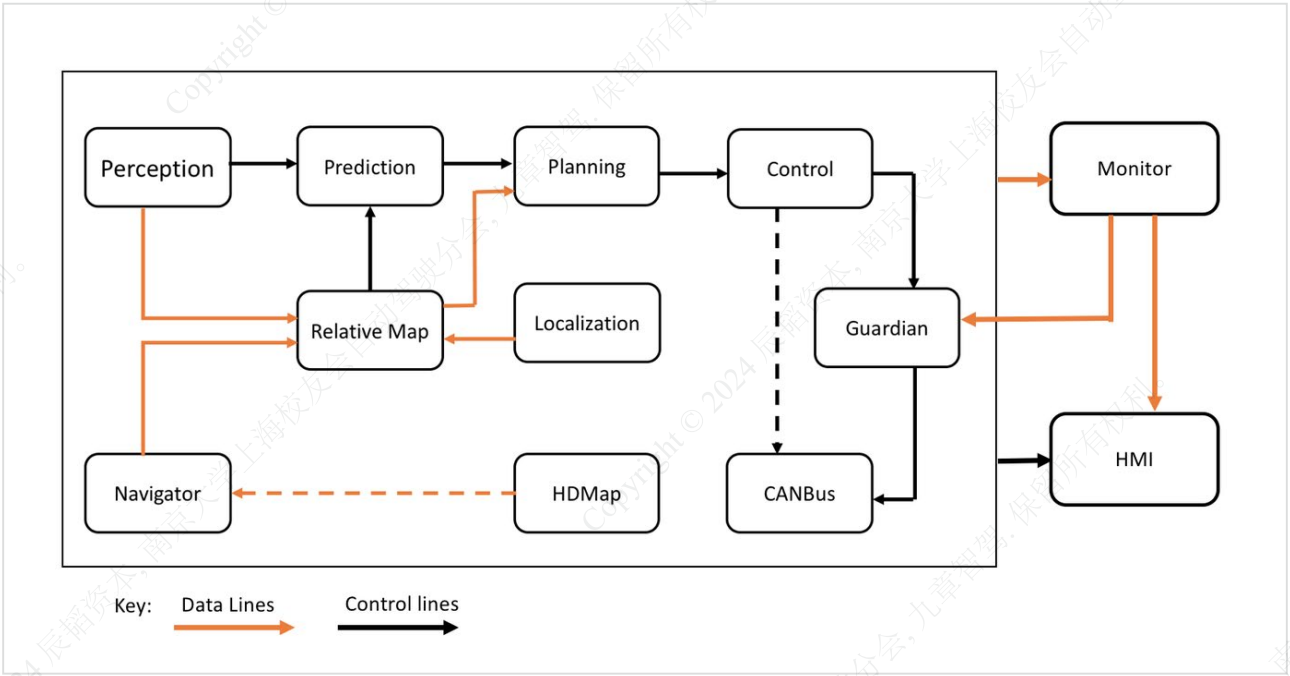
鉴于以上背景，本报告希望全面深入地探讨端到端自动驾驶的基本概念、参与者、发展驱动力、落地挑战和未来展望，并尝试为行业构建基础共识的认知和讨论基础，促进端到端自动驾驶行业更加蓬勃地发展。报告尝试在客观中立的前提下提取共识和有可供参考的观点和信息，并厘清行业面临的主要分歧，为相关企业和研究者提供参考。

Copyright © 2024 辰韬资本, 南京大学上海校友会自动驾驶分会, 九章智驾. 保留所有权利。

第一章 端到端的基本概念

1.1 传统自动驾驶架构

图示 4：传统自动驾驶架构的典型代表——Apollo 3.0 软件架构



图源：Apollo Auto

传统的自动驾驶架构更多衍生于机器人架构，所以在机器人领域的感知 - 定位 - 规划三大模块以及基于此进行的功能衍生在很长时间内成为了自动驾驶架构发展的基调。

典型的自动驾驶系统通常包含以下几个核心模块：

- 感知模块（Perception）

感知模块负责收集和解释车辆周围环境的信息。这包括使用各种传感器，如摄像头、激光雷达（LiDAR）、雷达和超声波传感器来检测和识别周围的物体，如其他车辆、行人、交通信号和道路标志。感知模块需要处理来自这些传感器的数据，并将其转化为车辆可以理解的环境模型。传统的感知模块的输出更多是基于人类的定义，如对于周边障碍物的检测，对于边界以及区域的分割等等。传统的感知系统要保证其输出能够为人所理解，代表了人类对于环境的一种抽象。

- **定位模块（Localization）**

定位模块的任务是确定车辆在环境中的精确位置。这通常涉及到使用全球定位系统（GPS）、惯性测量单元（IMU）和传感器数据来估计车辆的全局位置。此外，还可以使用基于地图的定位技术，如将车辆的传感器数据与预先制作的详细地图进行匹配。

- **预测模块（Prediction）**

预测模块用于预测其他道路使用者的行为和意图，如预测其他车辆的行驶轨迹和行人的移动。这有助于自动驾驶系统提前做出决策，以避免潜在的碰撞和冲突。

- **决策与规划模块（Decision and Planning）**

决策模块负责根据感知和预测的信息来制定车辆的行驶策略。这包括选择最佳的行驶路径、决定何时变道或超车，以及如何应对复杂的交通情况。规划模块则负责生成详细的行驶轨迹，确保车辆能够安全、高效地从当前位置行驶到目的地。

- **控制模块（Control）**

控制模块是执行决策规划模块输出的实际车辆控制命令的系统。它负责精确控制车辆的油门、刹车和转向，以实现平滑和安全的驾驶。

传统方案在模块上的划分使得每个模块的独立开发更加容易，在进行问题追溯时也更加简单快速。然而，这种方案的代价是限制了模块及整个系统所能达到的灵活性和泛化性的上限。

1.2 端到端自动驾驶的定义

围绕端到端自动驾驶的很多争议都来自于对其概念定义的不清晰，行业中看待这一技术的观点呈现出两极分化：“技术原教旨主义者”会认为，市场上很多公司所宣传的宣传“端到端”并不是真正的端到端；“实用主义者”则会认为只要基本原理符合，并且产品性能提升，端到端的准确内涵并不重要。

无论采信哪种观点，对于端到端技术的讨论语境始终无法拉齐。本章节中，我们希望对端到端自动驾驶技术做一个系统性的概念梳理，并提出一套可供参考的术语体系。

端到端架构演进

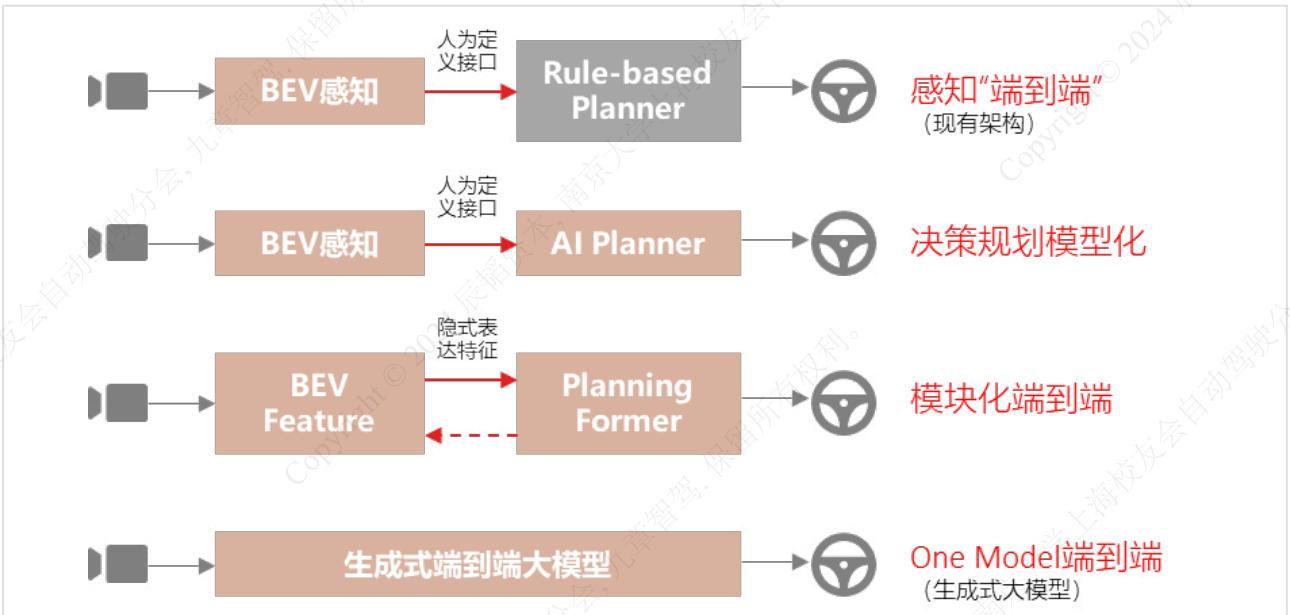
目前，自动驾驶架构的演进可以分为四个主要阶段：

第一阶段：感知“端到端”。这一阶段，整个自动驾驶架构被拆分成了感知和预测决策规划两个主要模块，其中，感知模块已经通过基于多传输器融合的 BEV (Bird Eye View，鸟瞰图视角融合) 技术实现了模块级别的“端到端”。通过引入 transformer 以及跨传感器的 cross attention 方案，感知输出检测结果的精度及稳定性相对之前的感知方案都有比较大的提升，不过，规划决策模块仍然以 Rule-based 为主。

第二阶段：决策规划模型化。这个阶段，整个自动驾驶架构被仍然分为感知和预测决策规划两个主要模块，其中，感知端仍保持上一代的解决方案，但预测决策规划模块的变动比较大 —— 从预测到决策到规划的功能模块已经被集成到同一个神经网络当中。值得注意的是，虽然感知和预测规划决策都是通过深度学习实现，但是这两个主要模块之间的接口仍然基于人类的理解定义（如障碍物位置，道路边界等）；另外，在这一阶段，各模块仍然会进行独立训练。

第三阶段：模块化端到端。从结构上来讲，这一阶段的结构和上一阶段比较类似，但是在网络结构的细节及训练方案上有很大不同。首先，感知模块不再输出基于人类理解定义的结果，而更多给出的是特征向量。相应地，预测决策规划模块的综合模型基于特征向量输出运动规划的结果。除了两个模块之间的输出从

图示 5：自动驾驶架构演进示意图



数据来源：辰韬资本

基于人类可理解的抽象输出变为特征向量，在训练方式上，这个阶段的模型必须支持跨模块的梯度传导——两个模块均无法独立进行训练，训练必须通过梯度传导的方式同时进行。

第四阶段：One Model/ 单一模型端到端。在这一阶段，就不再有感知、决策规划等功能的明确划分。从原始信号输入到最终规划轨迹的输出直接采用同一个深度学习模型。基于实现方案的不同，这一阶段的 One Model 可以是基于强化学习（Reinforcement Learning, RL）或模仿学习（Imitation Learning, IL）的端到端模型，也可以通过世界模型这类生成式模型衍生而来。

其中，模块化端到端以及 One Model 端到端都是基于全局优化的视角进行设计，且都满足梯度反向传播的特性，接下来，我们详细阐述下两个阶段的来龙去脉及技术特点。

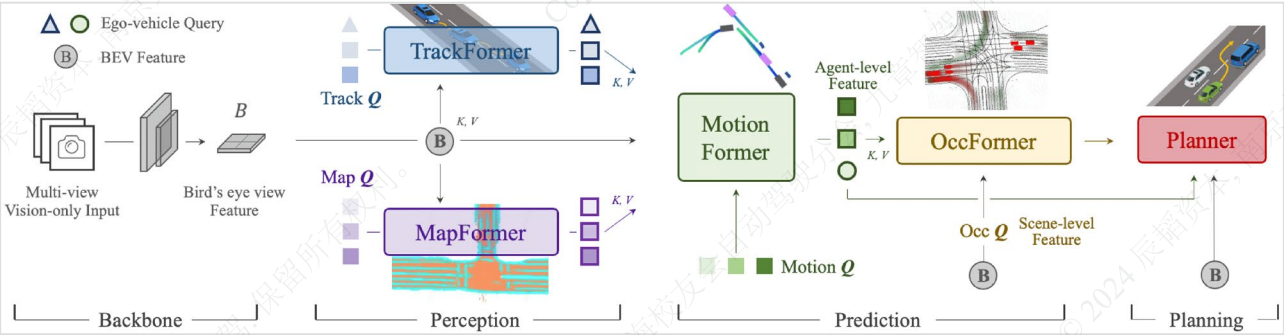
模块化端到端自动驾驶

相比于第一阶段 Rule-based 的决策方案，第二阶段的决策规划模块化很大程度地提升了决策规划应对复杂外界状况的上限。一方面，基于模型的方案能够最大程度地利用数据提升最终效果；另一方面，当现行模型大小不足以应对当前场景的复杂度时，扩大决策规划模型并重新进行训练在工程实现上也要比持续更新和维护规则库或状态空间模型容易很多。

在各功能模块都完成了“神经网络模型化”后，接下来的技术发展方向更多地体现在如何通过改进各功能模块间的互联方式获得更好的效果。而以这个标准来看，第二阶段的方案仍然存在着固有缺陷。

一方面，为了进行每个模块的独立训练和验证，接口的设计需统一抽象为人类的理解形式，这种方案在带来训练方便及验证便利性的同时，会不可避免地以

图示 6：典型的模块化端到端 UniAD 架构图



Yihan Hu et al. Planning-oriented Autonomous Driving

信息的损失为代价；另一方面，由于各模块之间无法进行全量有效的梯度传导——对每个模型的训练和优化更多地局限在模块内部，因而，在系统层面看，这更多是一种局部优化而非全局优化的方案。

而第三阶段的模块化端到端则通过避免对接口信号的过度抽象保证了信息的完整性，而跨模块的梯度传导也保证了对端到端模型的所有训练都有助于最终达到全局优化的效果。

One Model 端到端自动驾驶

虽然 One Model 属于第四阶段的方案，但是这个概念被提出的时间比模块化端到端更早。在自动驾驶产业刚刚开始起步的 2016 年，英伟达就提出采用单个神经网络（卷积 + 全连的简单架构）来实现端到端的自动驾驶，输入和输出就是最原始的传感器信号、方向盘角度及油门开度。

但由于结构设计过于简单，模型的规模也过小，这种方案仅能支持高速或者简单道路状况下的自动驾驶，且仅仅完成了小规模的 demo 验证，与量产需要的可靠性差距较大。

然而，随着 Transformer 网络架构和车端算力（逐步可支持 0.1B~1B 级参数量网络运行）的提升，One Model 的端到端方案又重新回到人们的视野中，甚至很可能成为端到端的终局解决方案。

相比于模块化方案，One Model 端到端方案虽然在训练以及调试上更为复杂，但在理论上，其最终效果具有更高的天花板。主要原因是，无论是采用基于 RL/IL 的学习类架构，还是采用以世界模型作为基底的衍生架构，为了保证能够拥有对世界更全面的理解，One Model 端到端方案的训练过程能够涵盖更广范围的数据，这就使得对真实世界的所有知识和认知都可以完整地应用于自动驾驶，因而模型可以实现更好的泛化效果。

此外，这类架构对于环境以及和其他物体交互的理解具有很强通用性，因此，它就不单单为自动驾驶，也为如机器人等其他领域的应用打好了基础，从而最终实现跨领域模型的统一。

1.3 端到端相关概念的联系与区别

端到端与决策规划模型化

行业里一种常见的误解是“通过模型而非规则实现决策规划模块即为端到端方案”，即将上面提到的第二阶段也归类为端到端。根据我们之前的分析，决策

规划模型化应该被理解为端到端的必要而非充分条件，做出这种判断的根据是，第二阶段的决策规划模型化与第三阶段的模块化端到端在以下两点上体现出明显区别：1）它与感知模块的接口是通过人为定义的目标而非底层特征，导致方案不支持梯度跨模块传导；2）各模块的训练及优化只局限在局部而无法扩展到全局。

端到端与大模型

端到端与大模型这两个概念经常被混淆在一起。但实际上，两者并不必然相关——大模型更多关注模型的参数数量以及涌现能力，而端到端更多强调的是结构上的梯度可传导以及全局优化。目前的大模型为端到端实现提供了很好的可选方案，但是端到端并非必然基于大模型实现。

从目标上来讲，大模型则像是一个多面手，它们拥有大量的参数，能够处理包括自动驾驶在内的多种复杂任务，如自然语言处理、图像识别等。大模型训练所需的数据类型结构更广，而且由于其广泛的应用范围，对于可解释性以及可靠性的要求并不高。相比之下，端到端的最终目的是让车辆能够自主导航并安全行驶。它需要学习和适应各种驾驶场景，以便做出准确的驾驶决策，更多基于标注数据进行训练，对于系统的可靠性要求更高。

实际上，自动驾驶及具身智能领域的“大模型”，往往不是传统意义上的“大”模型。这类大模型由于更多考虑了车端算力以及实时性要求的性质，很难达到和NLP (Natural Language Processing) 或者通用 AI 领域同等的标准。如果不考虑Nvidia 规划 2025-2026 年量产的 Thor，目前车端算力较大的也只有几百 TOPS (Tera Operations Per Second)，这种算力水平远远无法满足动辄 10B 甚至上百 B 的大模型需求，所以，自动驾驶领域的大模型都是小于 1B 的模型，这里所谓的“大”的定义更多的是相对于对原来感知系统采用的几百万参数的小模型而言。

当然，值得注意的一点是，目前自动驾驶产业采用端到端方案的诉求之一也是希望这种架构能够解决传统自动驾驶架构在决策和规划领域束手无策的一些复杂 case，而这些问题的解决从智能的角度来讲很依赖网络的推理能力——与GPT3 及其后继的大模型涌现出来的能力相似，这也是大家将二者联系起来的直接原因。在未来，车端模型的大型化以及 Transformer（或者类似变体）化是一个重要趋势，虽然二者在定义上存在着明显不同，但是在未来端侧算力持续增加的情况下，两者在自动驾驶领域的交叉会越来越多。

端到端与世界模型

世界模型的概念是指能够学习及揭示真实世界物理及数学定律的模型，例如能够学习并理解重力、力的相互作用等常见特性，并基于这些理解给出对未来一段时间的预测视频作为模型输出。

在世界模型与端到端自动驾驶的关系上，基本存在以下两种不同的观点：

第一，世界模型在整个自动驾驶算法的开发体系中以数据生成器的角色存在。

因为世界模型先天地具备根据外部输入条件生成视频以及与其他车辆的交互情况输出视频的能力，所以，通过世界模型生成的视频，可以作为端到端模型训练中的数据源，源源不断地合成数据，从而支持模型训练的全流程。这种合成数据可以覆盖大量采集车无法采集到的极端工况（如车祸或长尾的障碍物类型等），对于训练以及测试覆盖度的提升有很大帮助。同时，由于数据是通过模型在线生成，几乎没有任何离线工作，其成本优势也极为明显。

第二，只要对世界模型进行微小的调整或者增加一些输出链路及模块，就可以实现 One Model 端到端自动驾驶。

世界模型自身具备理解周围环境以及交互情况，并对其他道路交通的参与者的行为做预测的能力。这些感知和预测的时序信息就已经提供了做出决策和规划的基础，同时，因为世界模型本身在推理和理解上具有很好的基础（否则，世界模型就无法抽取到物理规律），它也具备了基于所有已掌握信息进行推理和最佳决策及规划的能力。因而，只要对世界模型进行微小的调整或者增加一些输出链路及模块，就可以很快实现端到端自动驾驶。

目前，这种方案已经有公司及研究机构进行了初步的探索，可行性得到了初步验证。不过，短期内制约这种方案上车的难点在于车端算力尚无法支持大的世界模型运行，至于后续是否可通过蒸馏或者其他降秩的方式在保持对真实世界理解的能力下最大程度地裁剪模型，还需要等待端侧硬件算力的持续迭代。

端到端与纯视觉传感器方案

目前的端到端自动驾驶方案对于车端传感器没有特殊要求，所以采用端到端的方案对于车端的感知传感器架构几乎不造成任何影响。

然而，部分从业者认为，端到端方案“只能基于纯视觉”。产生这种误解的主要原因在于：目前基于视觉 BEV 的感知及跟踪架构已经是业界主流，为保证更好的预测效果以及传递更多的感知层面信息，当前的端到端更多是继承了原来的视觉 BEV 方案，并将其作为感知模块，在此基础上进行了一定的延伸。

其实，无论是 BEV 还是端到端，从传感器处理上讲都属于前融合方案，在传感器选择上具有很强的灵活性。不排除未来在激光雷达性价比得到大幅提升，或 4D 毫米波的点云数量及质量得到快速发展的情况下引入其他类型传感器的输入作为整个网络的感知基础。

1.4 早期端到端自动驾驶的关键成果

本章第二节中提到，简单的 One Model 端到端方案在自动驾驶发展早期就已经出现，本节将具体展现其中三个关键的成果，以具体案例呈现早期端到端自动驾驶的技术方案。

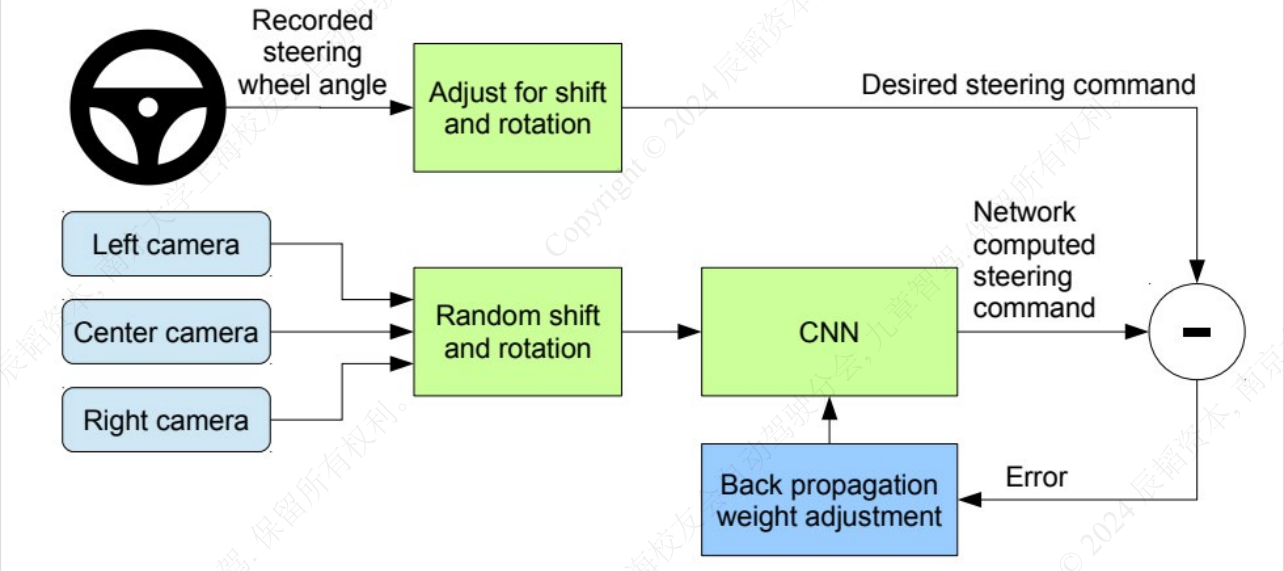
NVIDIA: DAVE-2 (2016)

2016 年 4 月，英伟达团队发表了一篇名为 *End to End Learning for Self-Driving Cars* 的论文，展示了基于 CNN 的端到端自动驾驶系统 DAVE-2。

该系统通过一个卷积神经网络（Convolutional Neural Network, CNN）处理车辆前方的摄像头图像，并直接输出转向角度。训练过程中，模型通过模拟驾驶数据进行学习。该系统在不同类型的道路上展示了出色的驾驶性能，包括城市道路和乡村公路。Demo 展示了车辆在复杂的道路环境中能够平稳驾驶，并对不同的交通状况进行适应。

DAVE-2 系统展示了端到端神经网络在自动驾驶中的巨大潜力，打破了传统的模块化自动驾驶系统框架，是近年来端到端自动驾驶领域的开创性工作。这项研究证明了通过直接从数据中学习驾驶策略，可以大幅简化系统设计，并提高适应性和鲁棒性。DAVE-2 为后续端到端自动驾驶研究奠定了基础，推动了该领域的快速发展。

图示 7：DAVE-2 的网络结构

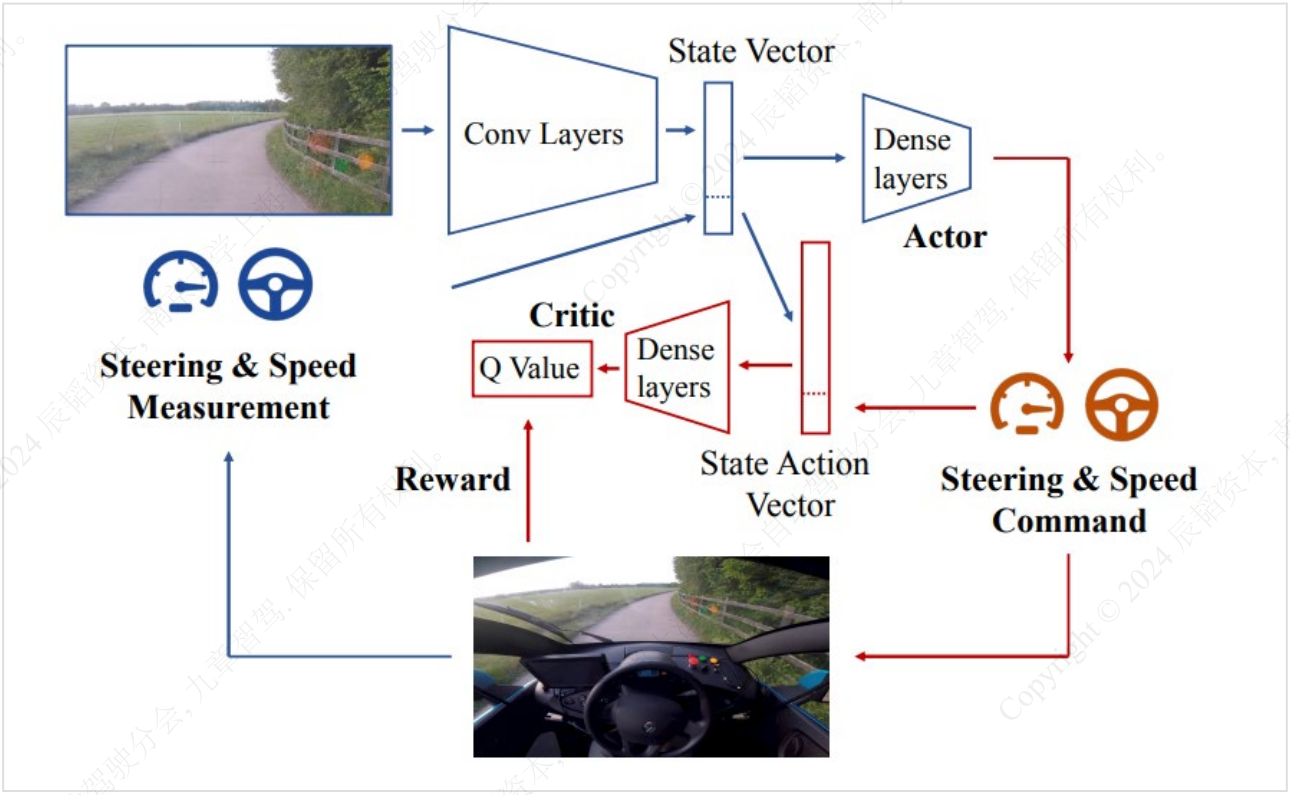


Mariusz Bojarski et al. End to End Learning for Self-Driving Cars

Wayve: Learning to Drive in a Day (2018)

Wayve.AI 是一家英国的自动驾驶公司，2017 年成立于剑桥。2018 年 9 月，Wayve 发表了一篇名为 *Learning to Drive in a Day* 的论文。Wayve 展示了其自动驾驶系统在短时间内学习驾驶的能力。与 DAVE-2 主要基于 CNN 的模仿学习方法不同，Wayve 结合了强化学习和深度学习算法，强化学习能够让系统通过试错过程自我改进，适应不同的驾驶环境和情况，而不仅仅依赖于预先收集的大量数据。该系统能够在仅一天的训练时间内，实现在复杂的城市环境中进行驾驶的能力。Wayve 的 Demo 展示了车辆在伦敦的开放道路上成功应对各种交通状况的表现，包括红绿灯、行人和其他车辆。

图示 8：Wayve 基于强化学习的端到端神经网络



Alex Kendall et al. Learning to Drive in a Day

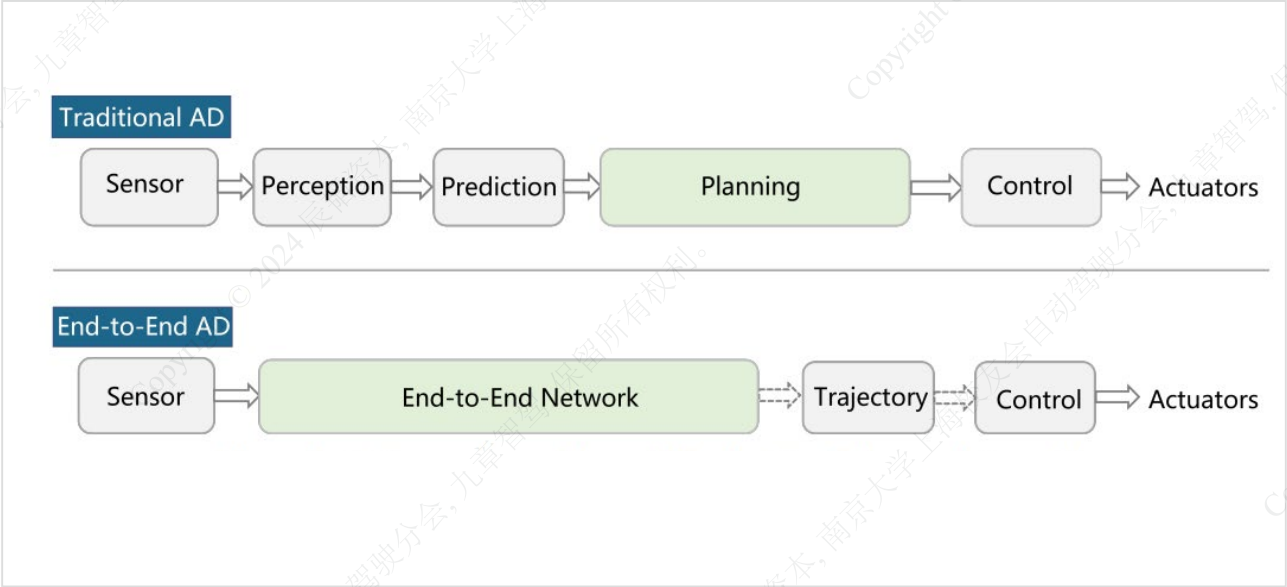
Wayve 的研究展示了基于深度强化学习的端到端系统的快速适应能力，证明了端到端学习在处理复杂城市道路驾驶任务中的有效性。这项研究为自动驾驶系统的灵活性和效率提供了新的视角，推动了快速部署和适应的可能性。

comma.ai: OpenPilot (2017)

comma.ai 是一家成立于 2015 年的美国自动驾驶公司，早期 comma.ai 通过手机的后置摄像头和手机计算芯片（骁龙系列），以及 comma.ai 研发的 Openpilot 自动驾驶软件（2017 年首次发布，持续迭代），实现了性能优异的 L2 级自动驾驶，2020 年曾被 Consumer Report 评为性能最佳的辅助驾驶系统，超越特斯拉 Autopilot 和凯迪拉克 SuperCruise。其与多达 200 余种车型的适配性，为 L2 级辅助驾驶的大规模推广提供了一种后装的解决思路。

2020 年后，OpenPilot 逐渐转向端到端神经网络模型，被认为是第一个大规模商业化的端到端自动驾驶产品。

图示 9：传统自动驾驶解决方案与 OpenPilot 方案对比



Li Chen et al. Level 2 Autonomous Driving on a Single Device: Diving into the Devils of Openpilot, (https://arxiv.org/abs/2206.08176)

同时，OpenPilot 是一个开源项目，通过开源和社区驱动的方式，comma.ai 在推动自动驾驶技术的普及和创新方面发挥了重要作用。其开放的开发平台和工具使得更多开发者能够参与到自动驾驶技术的开发和改进中 。

遗憾的是，OpenPilot 并未公布其训练数据和训练细节。上海人工智能实验室深度解析了 OpenPilot 的 Supercombo 端到端神经网络模型，并给出了其参考架构和改进点。

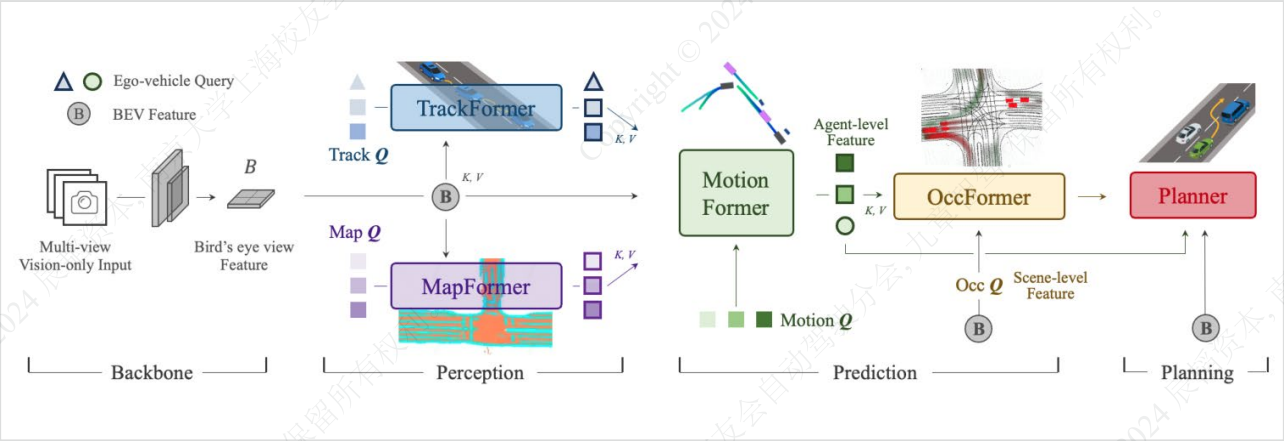
1.5 近期端到端自动驾驶的重要进展

2023 年以来，模块化端到端和 One Model 端到端都有了重大的进步。UniAD 提出了革命性的模块化端到端方案，成为这一技术路线的基准范式；Wavye 的生成式世界模型 GAIA-1，以及视觉 - 语言 - 动作模型 LINGO-2 可能是未来 One Model 端到端的重要基础；FSD v12 作为一个可以感受的产品让端到端技术路线出现在更多人的视野中。本节将呈现这几个重要进展的技术方案。

OpenDriveLab: UniAD (2023)

UniAD 代表了一种创新的全栈 Transformer 端到端模型设计，它通过集成多个查询组（query groups）来实现。在该模型的架构中，我们可以观察到两个关键的感知模块和两个核心的预测模块，以及一个至关重要的规划模块。这些模块的设计遵循了 Transformer 架构的先进理念，确保了高效的信息流和处理能力。

图示 10：UniAD 架构图



Yihan Hu et al. Planning-oriented Autonomous Driving

在 UniAD 中，TrackFormer 模块通过 query 与 BEV 特征的交互，实现了对周围环境的精确感知。MapFormer 模块则通过 Map query 更新，进一步丰富了环境特征。MotionFormer 模块利用 Motion query 与环境特征和 BEV 特征的交互，预测了未来轨迹，为决策提供了重要信息。而在 TrackFormer 中，特定的 ego-vehicle query 用于表示自车属性，为规划任务提供了关键信息。规划模块将 MotionFormer 更新后的 ego-vehicle query 与 BEV 特征进行深度交互，实现了对环境的全面感知和预测，从而优化规划任务。

Wayve: GAIA-1 (2023)

在人工智能领域，GAIA 模型以其创新的生成式世界模拟技术，为自动驾驶技术带来了革命性的突破。GAIA-1，作为该系列的最新版本，通过整合视频、文本和动作输入，显著提升了生成逼真驾驶视频的能力。它不仅能够精确控制自动

驾驶车辆的行为和场景特征，而且其多模态特性使得 GAIA-1 能够根据各种提示模态和组合生成多样化的视频内容。

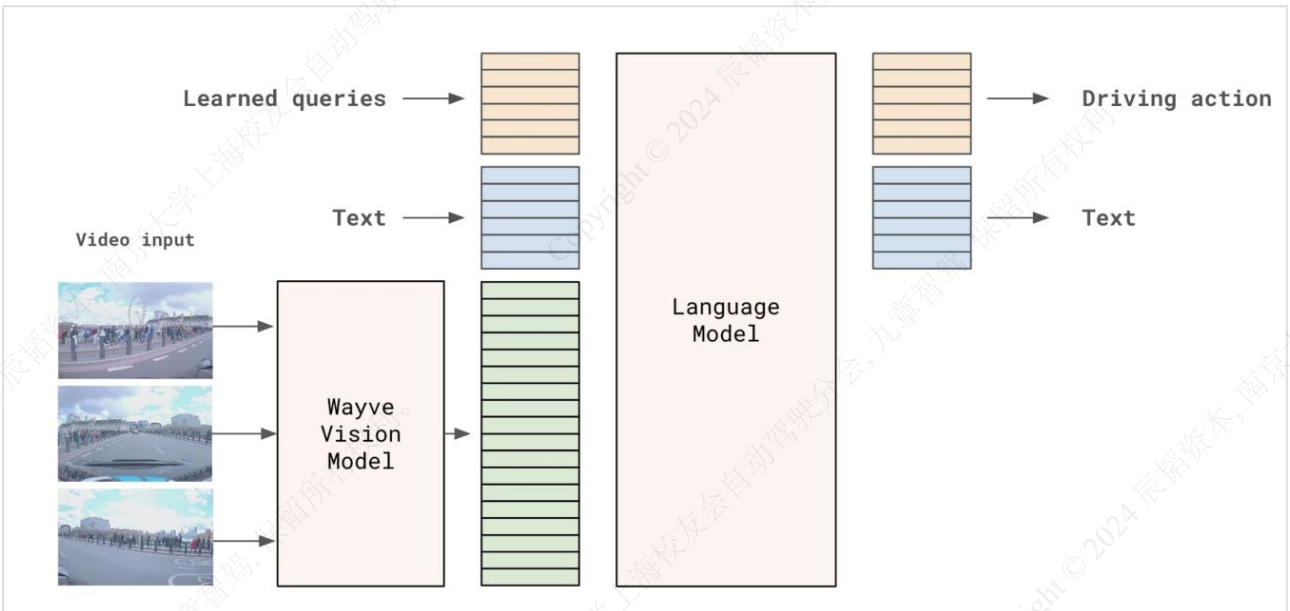
GAIA 模型的问世，标志着人工智能在模拟物理世界方面的重大进步。它能够生成长达数分钟的驾驶视频，这些视频细节丰富，严格遵循场景特征和车辆行为的预设规则。这一技术突破极大地提升了自动驾驶技术的决策力和安全性，有效解决了人工智能在预测和导航复杂现实世界交互时面临的诸多挑战。

Wayve: LINGO-2 (2024)

Wayve 公司最新推出的 LINGO-2 模型，为自动驾驶技术带来了重大突破。这项工作首次将视觉 - 语言 - 动作融合的大模型搭载上车，并开始商业化测试。视觉 - 语言 - 动作大模型构建了自动驾驶 AGI 的新范式，显著提升了 AI 驾驶系统的可解释性，使 AI 的决策过程更加透明，也将 AI 用于自动驾驶的能力提升到新的水平。LINGO-2 模型不仅能够执行驾驶任务，还能用自然语言与用户进行沟通，解释其决策背后的原因。例如，当 AI 决定减速时，它会向用户解释是因为检测到前方有行人过马路，从而增强用户对自动驾驶系统的信任感。

此外，LINGO-2 模型还具备根据用户的自然语言指令调整驾驶行为的能力。用户只需发出简单的指令，如"靠边停车"或"右转"，AI 就能理解并执行，同时解释其决策依据，展现出高度的互动性和智能性。LINGO-2 模型还支持视觉问答功能，能够回答用户关于场景和驾驶行为的问题，展示其对周围环境的深入理解和安全导航的能力。例如，当用户询问为什么左转时，AI 会用自然语言解释“左转可以更快到达目的地”。

图示 11：LINGO-2 架构



图源：Wayve

Tesla: FSD v12.3（2023）

Tesla 宣称从 FSD12.3 版本开始采用端到端自动驾驶方案。相比之前的非端到端版本，其在复杂场景上体现出更强的泛化能力，与其他车辆和行人的交互体现出更多的灵活性，驾驶风格也更加贴近人的习惯，其总体表现已经远远超过 Rule-based 决策规划模块的技术方案。业内普遍认为这些表现的巨大提升很大程度地建立在决策规划模块化基础方案之上，但是目前尚无法具体判断出 Tesla 的实现方案处于本报告定义的“决策规划模型化”“模块化端到端”或者“One Model 端到端”中的哪一个阶段。

特斯拉自动驾驶总监 Ashok Elluswamy 在 CVPR 2023 上分享了团队在 World Model 方向的研究进展。该研究很可能成为 FSD 下一阶段“One Model 端到端”自动驾驶方案的基础，并且有潜力进一步发展成为自动驾驶以及人形机器人通用的底层模型。

第二章 端到端主要玩家分析

2.1 主机厂

小鹏汽车

2024 年 5 月 20 日，小鹏汽车举办 AI Day 发布会，董事长、CEO 何小鹏宣布端到端大模型上车。小鹏的端到端大模型有三个组成部分：神经网络 XNet+ 规控大模型 XPlanner+ 大语言模型 XBrain。小鹏汽车在发布会上表示，端到端大模型上车后，18 个月内小鹏智能驾驶能力将提高 30 倍，每 2 天内部将做一次智驾模型的迭代。

图示 12：小鹏发布端到端大模型



图源：小鹏汽车

鸿蒙智行（类主机厂）

2024 年 4 月 24 日，华为智能汽车解决方案发布会上，华为发布了以智能驾驶为核心的全新智能汽车解决方案品牌——乾崑，并发布了并发布了 ADS 3.0。乾崑 ADS 3.0 的技术架构，感知部分采用 GOD（General Object Detection，通用障碍物识别）的大感知网络，决策规划部分采用 PDP（Prediction-Decision-Planning，预测决策规控）网络实现预决策和规划一张网。

ADS 3.0 在 ADS 2.0 基础上实现了决策规划的模型化，为端到端架构的持续演进奠定了基础。

图示 13：华为发布 ADS3.0 的端到端架构



图源：华为乾崑

蔚来汽车

蔚来在高阶辅助驾驶研发领域一直保持领先。据晚点 Auto 报道，自 2023 年下半年开始，蔚来已经投入几十人团队研发端到端自动驾驶，并计划于 2024 年上半年上线基于端到端的主动安全功能。蔚来智能驾驶研发副总裁任少卿认为，自动驾驶的大模型需要拆解成若干个层级，第一步是模型化，行业基本已经完成了感知的模型化，但是规控的模型化方面头部公司也没有完全做好，第二步是端到端，去掉不同模块间人为定义的接口，第三步是大模型。

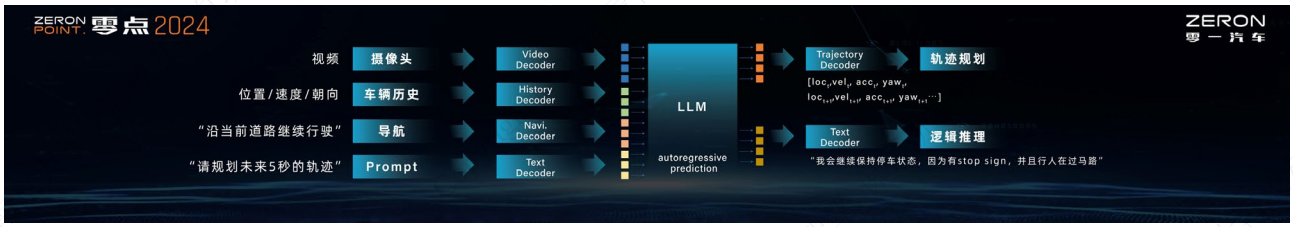
零一汽车

零一汽车是一家新能源重卡科技公司。零一致力于在核心动力总成、集成式热管理、自动驾驶技术等核心系统上实现全栈自研，并通过软件定义硬件，利用数据和技术重构供应链，实现自主安全可控与结构性成本优化。2024 年 5 月，零一汽车成功发布了两款量产纯电牵引车 " 惊蛰 " 与 " 小满 "。

近期，零一也推出了基于大模型的端到端自动驾驶系统。整个系统使用摄像头和导航信息作为输入，经过多模态大语言模型的解码产生规控信号和逻辑推理信息，将系统复杂度降低 90%。通过模拟人类的驾驶行为与思考过程，模型在仅使用视觉信息的情况下展现了非常强的泛化能力，并在多个数据集中获得世界第一的成绩。

零一计划在 2024 年底实现端到端自动驾驶的部署上车，2025 年在商用车与乘用车平台上同时实现量产，并计划于 2026 年实现高阶自动驾驶的大规模商业化运营。

图示 14：零一汽车发布端到端自动驾驶系统



图源：零一汽车

2.2 自动驾驶算法和系统公司

元戎启行

2023 年 3 月，元戎启行推出国内首款不依赖高精度地图、可实现全域点到点功能的高阶智能驾驶解决方案 DeepRoute-Driver 3.0。DeepRoute-Driver 3.0 发布的同时，元戎启行已经在进行端到端模型研发。2023 年 8 月，元戎启行运用端到端模型完成了道路测试。在道路测试中端到端模型表现惊艳，该车会顾虑后车需求主动礼让后车。在路况复杂的城中村狭窄路段，搭载端到端模型的车辆行驶流畅，无顿挫感；在 2024 年 4 月 25 日的北京车展上，元戎启行对外展示了即将量产的高阶智驾平台 DeepRoute IO 以及基于 DeepRoute IO 的端到端解决方案。该方案采用 NVIDIA DRIVE Orin-X 系统级芯片，200+TOPS 算力，并配备 1 颗固态激光雷达，11 颗摄像头。

元戎启行 CEO 周光在 2024 年 GTC 大会上表示："未来人工智能技术将在物理世界得到广泛应用。端到端模型会重塑物理世界的人工智能技术，终结一个以'规则驱动'为主导的原始人工智能时代，开启一个以'深度学习'为引擎的通用人工智能时代。

图示 15：元戎启行端到端演进路线



图源：元戎启行

商汤绝影

2024 年 4 月举办的北京车展上，商汤绝影推出面向量产的端到端自动驾驶解决方案“UniAD”。本次车展上，商汤绝影展示的端到端自动驾驶系统，无需高精地图，通过数据学习和驱动就可以像人一样观察并理解外部环境，然后基于足够丰富的感知信息，UniAD 能够自己思考并作出决策，像人一样开车，流畅进行无保护左转、快速通行人车混行的红绿灯路口，自主解决各种高难度的城市复杂驾驶场景。

图示 16：商汤绝影发布 UniAD 端到端解决方案



图源：商汤绝影

基于本报告第一章中的定义，商汤的的“端到端 UniAD”归属于“模块化端到端”类型，与决策规划模型化的技术架构（即“两段式端到端网络”）相比，不需要对感知数据进行抽象和逐级传递，实现了感知决策一体化和系统的联合优化。

同时，商汤绝影也发布了其下一代自动驾驶技术 DriveAGI，即基于多模态大模型对端到端智驾方案进行改进和升级的“One Model 端到端”。

小马智行

基于此前在模块化自动驾驶技术上的深远积累和技术优势，小马智行开始研发端到端自动驾驶模型，为自动驾驶更大范围使用、更快扩区速度做准备。2023 年 8 月，小马智行将感知、预测、规控三大传统模块打通，统一成端到端自动驾驶模型，目前已同步搭载到 L4 级自动驾驶出租车和 L2 级辅助驾驶乘用车。小马

智行端到端自动驾驶模型既可作为 L4 级车辆的冗余系统，也可作为 L2 级车辆的解决方案。

小马智行端到端自动驾驶模型具有四大优势：一、多维度的数据来源：L4 级自动驾驶车辆行驶数据，L2 级量产车中人类驾驶员数据，V2X 路段摄像头数据，日常生活中的数据等都可作为数据来源；二、全面的数据处理工具：基于在 L4 级自动驾驶上的多年积累，小马智行已拥有一套完整的数据评估体系，包含前期高质量数据挖掘清洗，测试使用的大规模仿真系统等；三、模型具有可解释性，不再是黑盒状态：小马智行结合驾驶意图、应用场景融入规则性指令，例如交通法规、驾驶偏好等；四、出色安全的驾驶技术：小马智行自动驾驶测试里程已达 3500 万公里，安全性比人类司机高 10 倍；不仅向不同场景中的优秀人类驾驶员学习，还帮助其他驾驶员减少错误，从而提高我们自身的安全性。

鉴智机器人

在 2024 北京车展期间，鉴智机器人联合创始人、CTO 都大龙表示，鉴智机器人原创的自动驾驶端到端模型 GraphAD 已经可量产部署，并正与头部车企进行联合开发。在主流端到端开环规划评测上，GraphAD 在各项任务上均达到领域最佳性能，规划任务上拟合误差和避障指标更是远超此前的最佳方案。

谈及端到端自动驾驶，都大龙说：“之所以将端到端范式称为 GraphAD，是因为鉴智使用了图形结构来显示建模目标，包括动态和静态障碍物之间的关系——这使得端到端模型训练起来更容易也进一步减少对数据量的需求。”都大龙指出，如果只使用积累下来的专用数据，是无法做到快速上车的，而鉴智目前在做的，是将端到端与世界模型相结合。构建生成式的仿真模型，使得 Agent 不断与仿真模型交互，以此将世界的规律和人类的知识迁移到 Agent 里。“我们将真实世界的信息模拟压缩成一个生成式模型，同时要不断对其进行几个层面的优化：真实度、可控度以及可交互度。其中，交互要无限逼近现实感。”

图示 17：鉴智机器人 GraphAD 端到端网络



图源：鉴智机器人

英伟达

作为 AI 生态赋能者，英伟达可以提供从芯片、工具链到智驾解决方案的全栈产品，将 AI 领域最前端的技术赋能至智驾。

2023 年夏季，吴新宙加盟英伟达成为汽车业务负责人。此后，英伟达加大对智驾业务的投入，强化英伟达对于自身全栈软硬件方案提供商的定义。

2024 北京车展前夕，吴新宙展示了英伟达自动驾驶业务从 L2 到 L3 的发展规划，其中提到规划的第二步为“在 L2++ 系统上达成新突破，LLM（Large Language Model，大语言模型）和 VLM（Visual Language Model，视觉语言模型）大模型上车，实现端到端的自动驾驶”。吴新宙认为，端到端是自动驾驶的最终一步，接下来几年端到端模型和原有模型会在自动驾驶中相辅相成，端到端模型提供更拟人且灵活的处理，而原来的模型和方法则可以保证安全性。

地平线

地平线早在 2016 年便率先提出了自动驾驶端到端的演进理念，并持续取得技术创新与突破：在 2022 年提出行业领先的自动驾驶感知端到端算法 Sparse4D；2023 年，由地平线学者一作发表的业界首个公开发表的端到端自动驾驶大模型 UniAD，荣获 CVPR 2023 最佳论文。

同时，地平线积累了基于交互博弈的端到端深度学习算法，大幅提升智驾系统在复杂交通环境中的通过率、安全性和舒适度。在硬件技术上，地平线专为大参数 Transformer 而生的新一代智能计算架构 BPU 纳什，能够以高度的软硬协同打造业界领先的计算效率，为自动驾驶端到端和交互博弈提供智能计算最优解。

图示 18：地平线 SuperDrive 技术架构向端到端演进



图源：地平线

2.3 自动驾驶生成式 AI 公司

光轮智能

光轮智能致力于为企业落地 AI 提供合成数据解决方案，结合生成式 AI 和仿真技术，为行业提供多模态、高真实度、可泛化、全链路的合成数据。解决自动驾驶、具身智能行业中真实数据采集难、Corner Case 数据缺乏、标注成本高、回环周期长、利用率低等问题。

光轮智能由谢晨博士创立，结合生成式AI和仿真技术提供合成数据解决方案。谢晨曾在英伟达（美国）、Cruise（美国）、蔚来汽车等企业担任自动驾驶仿真负责人，国际首创将生成式 AI 融入仿真。清华大学智能产业研究院助理教授赵昊担任光轮智能首席科学家。赵昊深耕基于生成式 AI 的仿真、自动驾驶与具身智能算法，曾主导研发全球首个开源的模块化真实感自动驾驶仿真器 MARS。光轮团队拥有国内外最多次从 0-1 合成数据开发和落地经验，汇聚国际顶级生成式 AI 算法专家、英伟达传感器仿真负责人、国内领先自动驾驶感知负责人、28 岁阿里 P8、机器人国际大赛冠军、清华特奖、多次创业者等。员工来自英伟达、华为、Cruise、蔚来、百度、达摩院等，拥有多项国内外技术专利。公司 2023 年初成立，已完成种子轮、天使轮、天使 +、Pre A 轮四轮融资，累计融资上亿元。

面对市场上迅猛增长的合成数据需求，光轮智能坚持高质量高效率地交付合成数据，现已交付多家国内外头部主机厂、Tier 1、自动驾驶公司数千万商业订单，服务量产落地以及端到端算法预研。在自动驾驶端到端方面，光轮作为引领者，开发了自研的端到端数据与仿真的全链路解决方案，积累了大量自动驾驶端到端实战经验。2024 年 3 月，光轮智能与上海人工智能实验室联合推出并开源自动驾驶 3D Occupancy 合成数据集 "LightwheelOcc"，用于 CVPR 2024 自动驾驶挑战赛，本届挑战赛包含了以端到端为代表的众多自动驾驶领域关键技术赛题的比赛。除自动驾驶领域外，光轮智能也开始服务具身智能、多模态大模型等领域，目前国内合成数据领域市场份额稳居第一。

极佳科技

极佳科技是一家专注于世界模型技术和视频生成应用的公司。2023 年 9 月，极佳科技推出了全球首个物理世界驱动的自动驾驶世界模型 DriveDreamer，在业界引起了广泛的关注。

DriveDreamer 是一个生成与理解统一的世界模型架构，基于其高真实度、高效率以及高可扩展性的特点，首先能够实现自动驾驶场景的高效数据生成，用于解决自动驾驶训练的数据短缺，特别是 Corner Case 难以收集的问题；其次通过与驾驶控制信号的结合，DriveDreamer 可以实现高效的数据生成、编辑与交互，从而用于实现端到端自动驾驶的全链路闭环仿真；同时 DriveDreamer 所具备的场景理解能力，可以扩展实现直接输出端到端的动作指令，成为新一代端到端方案的重要环节。

目前，基于 DriveDreamer 的产品与方案已经在多个主机厂和自动驾驶科技公司实现定点落地，成为广受认可的自动驾驶世界模型方案。

图示 19：DriveDreamer 模型架构



图源：极佳科技

2.4 学术研究机构

上海人工智能实验室

上海人工智能实验室近年来为自动驾驶技术的发展做出了很大的贡献。

2022 年，上海人工智能实验室开源了 BEVFormer 架构，时至今日依然是自动驾驶行业内最通用的视觉感知算法架构。2023 年 6 月，上海人工智能实验室、武汉大学及商汤科技联合提出的端到端自动驾驶算法 UniAD，获得 CVPR 2023 最佳论文，是近十年来计算机视觉顶级会议中第一篇以中国学术机构作为第一单位的最佳论文。受到 BEVFormer 和 UniAD 的启发，自动驾驶行业在 BEV 感知、端到端自动驾驶方面的研究大大加速。

近半年来，上海人工智能实验室还推出利用大语言模型进行闭环端到端自动驾驶的工作 LMDrive、自动驾驶视频生成模型 GenAD 等，上海人工智能实验室从多个技术维度对自动驾驶技术进行探索，从而多方位提升其智能性。

同时，上海人工智能实验室还主导了 DriveLM（语言 + 自动驾驶数据集）、GenAD（驾驶视频数据集）、OpenLane（车道线数据集）、OpenScene（3D 占用空间预测数据集）多个自动驾驶开源数据集建设，主办了 CVPR 2023 自动驾驶挑战赛、CVPR 2024 自动驾驶挑战赛，对自动驾驶研究生态的发展起到重要推动作用。

清华大学 MARS Lab

清华大学 MARS Lab 由前 Waymo 科学家赵行成立和主导。2021 年初，MARS Lab 提出了视觉为中心的自动驾驶框架 VCAD，发表了一系列 BEV 和端到端自动驾驶的基石研究论文和工作：首个基于 Transformer 的视觉 BEV 检测模型 DETR3D、首个视觉 BEV 3D 物体跟踪模型 MUTR3D、首个基于 Transformer 的多传感器前融合 3D 物体检测模型 FUTR3D、首个端到端轨迹预测模型 ViP3D、首个 3D 占用网络的评测基准数据集 Occ3D 等。

MARS Lab 也是最早发表“无图”自动驾驶方案的团队：2021 年初，发布首个在线高精度地图模型 HDMapNet；2022 年初，发布首个矢量高精度地图模型 VectorMapNet；2023 年初，开创性地提出了基于众包的先验神经网络的地图模型，实现了自动驾驶地图的记忆、更新、感知一体化。该系列工作为行业指明了技术落地方向，其中合作企业理想汽车将相关技术在其新能源车产品中进行了广泛应用落地。

图示 20：理想汽车在产品发布会上分享与 MARS Lab 合作技术的落地应用情况



图源：理想汽车

近期，MARS Lab 提出了自动驾驶快慢双系统 DriveVLM-Dual，一套基于大模型的可落地的自动驾驶方案。DriveVLM-Dual 将大模型和传统自动驾驶方案有机结合，大幅提升了大模型的空间理解能力，并且规避了大模型的推理速度问题。

第三章 端到端的发展驱动力

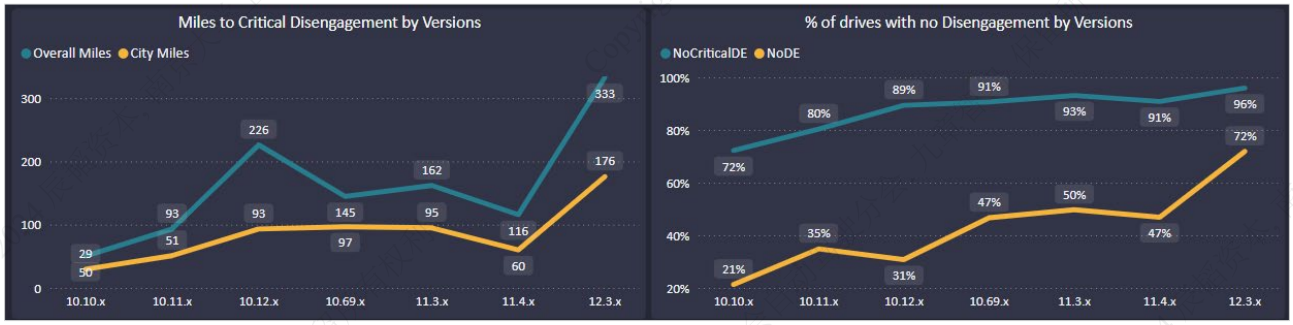
本章节将讨论端到端自动驾驶近一年来快速发展的原因，从对标、用户体验、组织价值、技术演进角度分析端到端的发展驱动力。

3.1 对标驱动：特斯拉 FSD 的标杆作用吸引行业关注

大部分行业专家表示，特斯拉 FSDv12 的优秀表现，是端到端自动驾驶这一技术路线快速形成大范围共识的最重要的推动力；而在此之前，从来没有一个自动驾驶产品可以让从业者和用户如此便捷地感受到技术带来的体验提升。

2024 年 3 月，特斯拉开始在北美地区大范围推送 FSD v12，舆论对其性能的正面评价占据了主流。许多用户表示，特斯拉 FSD v12 在施工路段、无保护左转、拥挤的环岛以及为后方车辆挪动空间时表现优异。

图示 21：特斯拉 FSD 平均接管里程和无接管行程次数占比的演变



图源：Tesla FSD Tracker

由第三方网站 FSD Tracker 统计的特斯拉车辆接管里程数据也表明了 FSD v12 的巨大性能提升。在此前很长一段时间，FSD 的版本迭代处于瓶颈期，自 2022 年初 FSD v10 更新以来，其接管里程数据保持在稳态波动，行业普遍认为这是传统架构的工程优化陷入瓶颈的表现；但 FSD v12 更新后，较之此前版本，用户完全无接管的行程次数占比从 47% 提升到了 72%，平均接管里程（Miles Per Intervention, MPI）从 116 英里提高到了 333 英里，性能的大幅提升代表着端到端技术突破了原有的技术瓶颈，推动自动驾驶系统的能力再上台阶。

作为自动驾驶行业最重要的标杆企业之一，特斯拉的技术路线一直备受关注。从 Elon Musk 在 2023 年 5 月首次公开提出“特斯拉 FSD v12 是一个端到端 AI”，到 2024 年 3 月特斯拉开始大范围推送 v12，特斯拉端到端技术不断演进和成熟的过程，也是中国自动驾驶行业逐渐凝聚共识的过程。很多业内专家表示，公司下决心投入端到端自动驾驶，与特斯拉的进展密不可分。

3.2 用户体验驱动：端到端解决长尾场景的安全性和动态博弈的拟人化程度

端到端作为一种技术路径，能够得到车企客户和消费者的关注，是因为其能带来用户体验的显著提升。

第一是提升了系统的安全性。自动驾驶的很多长尾场景是“只可意会，不可言传”的，意即对于人类驾驶来说可以靠直觉“搞定”，但很难用规则化的语言来表述，这类问题是传统技术路径依赖的专家系统很难解决的。某自动驾驶芯片公司 AI 负责人提到，诸如“道路上一滩正在起火的油”与“道路上的积水”、“正面飘来的空塑料袋”和“前车落下的钢筋”这类需要常识推理的场景，以及“不同地区的不同红绿灯外观和路口等待规则”这类需要复杂环境理解能力的场景，要么很难用规则准确描述，要么其开发工程量太大（因为专家无法总结所有规律）。这类场景对于端到端系统来说，可以被训练为隐式的中间表示，从实践上表现出很强的应对能力。

第二是驾驶风格的拟人化。当前的自动驾驶功能在车道保持、定速巡航这类简单场景下的表现已经和人类司机相当，但是在拥堵跟车、加塞响应、变道超车等需要博弈的场景，其驾驶风格通常与乘客的预期不一致。究其原因，是因为基于规则的决策规划系统需要海量的工程优化才能拟合出人类的高动态驾驶策略。相比之下，端到端系统能够表现得更像人类司机，这有利于自动驾驶系统与用户建立信任。

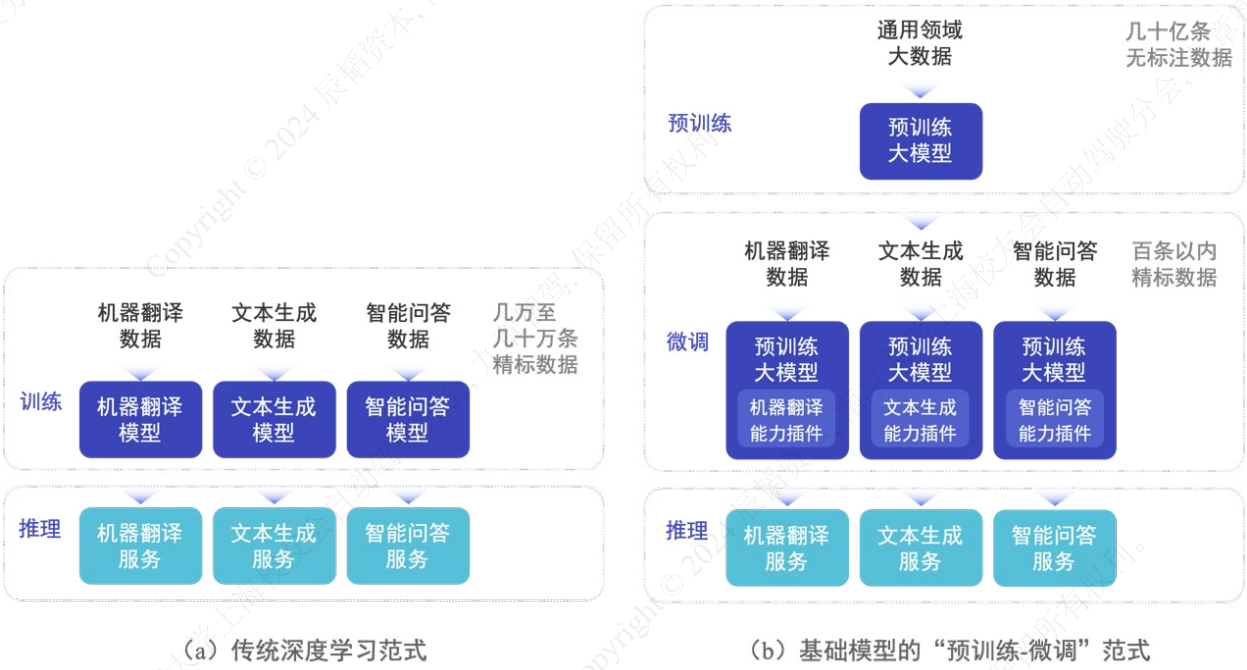
3.3 组织价值驱动：简化研发流程，优化组织效率

目前，大多数自动驾驶公司在量产落地阶段都会面临大量的研发人员投入的挑战。传统技术架构下，要实现高阶驾驶功能的交付，头部公司基本都需要投入近千人或上千人的资源。而多位专家认为，端到端会简化自动驾驶的研发流程，因而有很大概率可以降低自动驾驶研发的人力成本。

端到端对研发流程体系的简化可以体现在两个方面：

第一是功能系统层面，端到端意味着原来以横向控制和纵向控制为基础、从底盘执行单元推演而来的自动驾驶功能体系可以被彻底革新。将浩如烟海的“Use Case 定义和分析”的工作转向“场景提取”和“数据挖掘”方向。类比 ChatGPT 的出现将原来细分的 NLP 任务统一，端到端自动驾驶也可能意味着自动驾驶任务的统一。

图示 22：大语言模型将传统自然语言处理的多个任务统一为一个基础模型



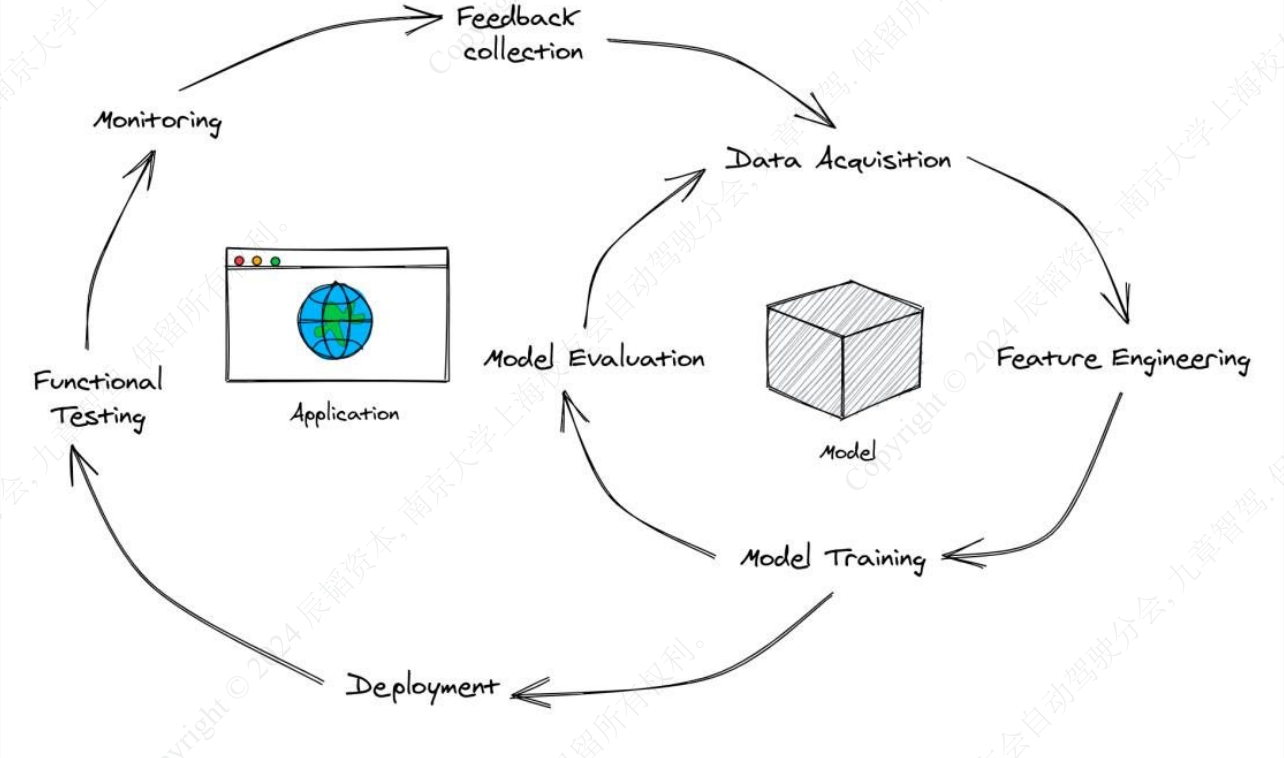
图源：清华大学官网

第二是算法层面，传统的感知 - 融合 - 预测 - 决策 - 规划架构可能涉及到十几个子系统和更多的软件模块，而端到端则可以将与之相关的子系统缩减为一个。子系统简化意味着研发团队的分工简化，并可以大大减少部门墙对组织效率的影响。

许多专家都表示，很多时候，自动驾驶测试中遇到的问题，很难准确界定属于感知部门还是决策规划部门，也许无论哪个子系统的优化都可以解决这个问题，因而，在实际工作中这两个部门的接口也最容易遇到责权不清的问题，这就间接地增加了系统工程、各个算法方向上的人力投入。而端到端的核心技术原理是全局统一优化，这同时也意味着组织目标的统一对齐。

除了功能、系统、算法团队的简化外，端到端对公司组织架构的影响还体现在 AI Infra 的重要性将持续提升。由于核心算法的开发范式全部更迭为数据驱动的开发范式，数据处理（包括数据采集、场景挖掘和定义、合成数据等）以及训练平台基础设施将会变得越来越重要，自动驾驶公司的组织架构可能演变为 AI 算法 + AI Infra 的双核心。

图示 23：MLOps 工具链全景图



图源：Lj Miranda 个人网站

3.4 技术驱动：端到端是实现自动驾驶 AGI 新范式的基础

以 ChatGPT 为代表的生成式 AI 大模型彻底颠覆了传统的自然语言处理，而 Elon Musk 表态“特斯拉即将迎来 ChatGPT 时刻”和对外公开端到端的时间都在 ChatGPT 推出半年后，因此业界普遍猜测，ChatGPT 的出现是影响特斯拉投入端到端路线的重要驱动因素。

虽然 GPT、Sora 等很难直接用于自动驾驶任务，但其代表的 AGI (Artificial General Intelligence, 通用人工智能）新范式的思路同样可以应用于自动驾驶。根据鉴智机器人联合创始人、CTO 都大龙博士等专家的观点，AGI 新范式对自动驾驶的启发包含以下几点：

a. 生成式的训练框架，特点是自回归、模型可与环境交互，减少了数据标注的巨额成本。

b. Transformer 作为基础模型结构的强大能力。

c. 通过增加模型规模和数据量提升模型性能（Scaling Law）。

图示 24：自然语言处理启发自动驾驶从传统范式到 AGI 新范式的演变

	自然语言处理		自动驾驶	
	传统范式	AGI 范式	传统范式	One Model 范式
训练框架	监督学习	生成式框架	监督学习	生成式框架
数据需求	标注数据	无标注 + 人类对齐	标注数据	无标注 + 人类对齐
模型规模	千万级参数	十亿级参数	千万级参数	十亿级参数
基础模型架构	CNN / LSTM	Transformer	CNN + Transformer	Transformer

来源：本报告编写组

从图表中可以看出，以上提到的 a、b、c 三个特点，现有的模块化自动驾驶架构是无法实现的，而端到端正好是实现自动驾驶 AGI 范式的基础。

第四章 端到端落地面临的挑战及辨析

虽然行业内对于端到端的发展趋势已经形成普遍共识，但目前已经或者接近量产的方案还停留在“决策规划模型化”阶段，距离模块化端到端或 One Model 端到端自动驾驶量产落地还有距离。本章将集中讨论端到端方案在落地中面临的挑战。

4.1 技术路线分歧

正如本报告第一章所提到的，在端到端的概念下，存在不同结构的端到端神经网络，不同公司和研究机构可能采用不同的技术路线和模型设计，例如“模块化端到端”采用监督学习的训练范式，“One Model 端到端”可能更加侧重自回归和生成式的训练范式，以上两种路线都有公司押注。

虽然在理论上，每种技术路线都有其优劣势，但其最终对用户体验的价值如何、工程开发代价如何等关键问题，在当前还没有共识。造成当前技术路线分歧的原因，除了行业还处于发展早期以外，最关键的原因是：行业内尚未出现可供参考的最佳实践案例。

特斯拉在 2021 年和 2022 年举办的两场 AI Day 基本确定了 BEV 和 Occupancy Network 的网络结构的标准，而其端到端的网络结构至今为止还没有对外披露。开源社区的标志性成果 UniAD 尽管提供了一个公开的技术架构，但 UniAD 尚未真正经过量产交付的验证，特别是与其配套的模型验证和评测方案、数据采集和处理方式还没有明确，目前，UniAD 也无法作为确定的终局技术架构。

我们预判，在未来 1~2 年内，随着更多公司和研究机构在端到端领域加大投入并推出产品而逐渐明晰，技术路线将会逐渐收敛。

4.2 对训练数据的需求量空前提升

端到端自动驾驶是数据驱动的模型，因此，训练数据的重要性前得到所未有的提升。端到端数据的挑战可以分为数据量、数据标注、数据质量和数据分布几方面。

首先是训练端到端所需的数据量。特斯拉在多个场合中提到，其 FSD 训练需要用到上千万个视频片段，假设每个视频片段时长为 30~60 秒，以此为参考，训练端到端模型最起码需要几万小时的视频数据。目前规模最大的公开数据集包含大约 1200 小时数据，这意味着，要拿到端到端研发的入场券，自动驾驶公司还必须使用更大规模的非公开数据。

Wayve 研发的生成式世界模型 GAIA-1 使用了该公司在过去 4 年内采集的 4700 小时驾驶数据训练；许多国内的自动驾驶公司表示，端到端第一阶段研发用

到的数据量约为几千小时。这些数字似乎表明，端到端所需要数据量并不是大到难以想象，几乎所有拥有车队的自动驾驶公司都可以达到这一水平。然而，真正的挑战来自于为提高模型性能所需的海量数据。小马智行 CTO 楼天城表示，现阶段，训练出一个一般性能的端到端模型，自动驾驶的数据量已经不是一个问题；但是要训练出一个高性能的端到端模型，对数据的质量要求可能是几个量级的提升，这是自动驾驶行业都会面临的挑战。

第二是数据标注需求的变化。长期来看，端到端由于没有感知 - 决策规划的中间接口，面向感知的标注需求将会转向面向规划的标注需求，对现有的包括 3D 目标检测、车道线检测、路面标识语义分割等感知模块的真值的标注需求将大大减少。但短期内，大部分公司，特别是采用“模块化端到端”架构的公司，依然会对感知的中间结果做标注和监督，提高训练的效率。

第三是对数据质量的要求。一位自动驾驶工程师提到，他们在训练端到端模型时发现，原本积累的路测数据只有 2% 可以用。特斯拉用来训练端到端模型的几万小时自动驾驶数据，是从超过 20 亿英里（截至发稿）的 FSD 里程数据中挖掘出来的。对于端到端自动驾驶来说，其数据质量问题还叠加了更复杂的因素——对人类驾驶员能力的要求。自动驾驶系统的理想目标是像老司机一样开车，而很多驾驶员的驾驶行为并能达到“老司机”的水平，这就需要一套有效的数据管理和处理流程，仅提取老司机所驾驶的车辆在特定场景下的高质量数据，这也增加了数据获取的难度。

第四是数据分布问题。全面和多样化的数据对于端到端自动驾驶模型至关重要，但多样性很难用量化方法来掌握。数据的分布需要考虑诸多因素，例如环境因素包含各种天气和光线、道路情况。与特斯拉不同，许多公司面临的最大难题是缺乏数据采集能力。即使是车企采集的数据也往往是在受控环境中获得的，这些数据可能无法完全代表真实世界中的复杂情况。真正能够反映真实世界的复杂性的数据，应该是由大量用户在无意识中采集的。更有挑战的问题是，如何调整长尾场景（Corner Case）在训练数据中的分布比例，目前行业还没有一套行之有效的方法论，有专家认为，在解决数据分布的问题上，行业在模型化决策规划方面的探索会有很大价值。

合成数据可以解决上述部分问题。合成数据可以低成本增加训练数据的规模、泛化场景增加多样性、低成本生成长尾场景。目前，很多企业已经开始通过混合使用合成数据和真实数据来提高系统性能。

另一个亟待探索的方向是建立数据共享平台。目前，数据共享仍处于起步阶段，真正产生价值的数据共享平台可能需要一个极其强势的第三方来发起。尽管部分企业已经开始尝试内部数据共享，但整个行业内尚未形成统一的数据共享标准和惯例。一些地方政府已经开始出台相关规定，以促进数据共享和智能交通的发展。

4.3 对训练算力的需求越来越高

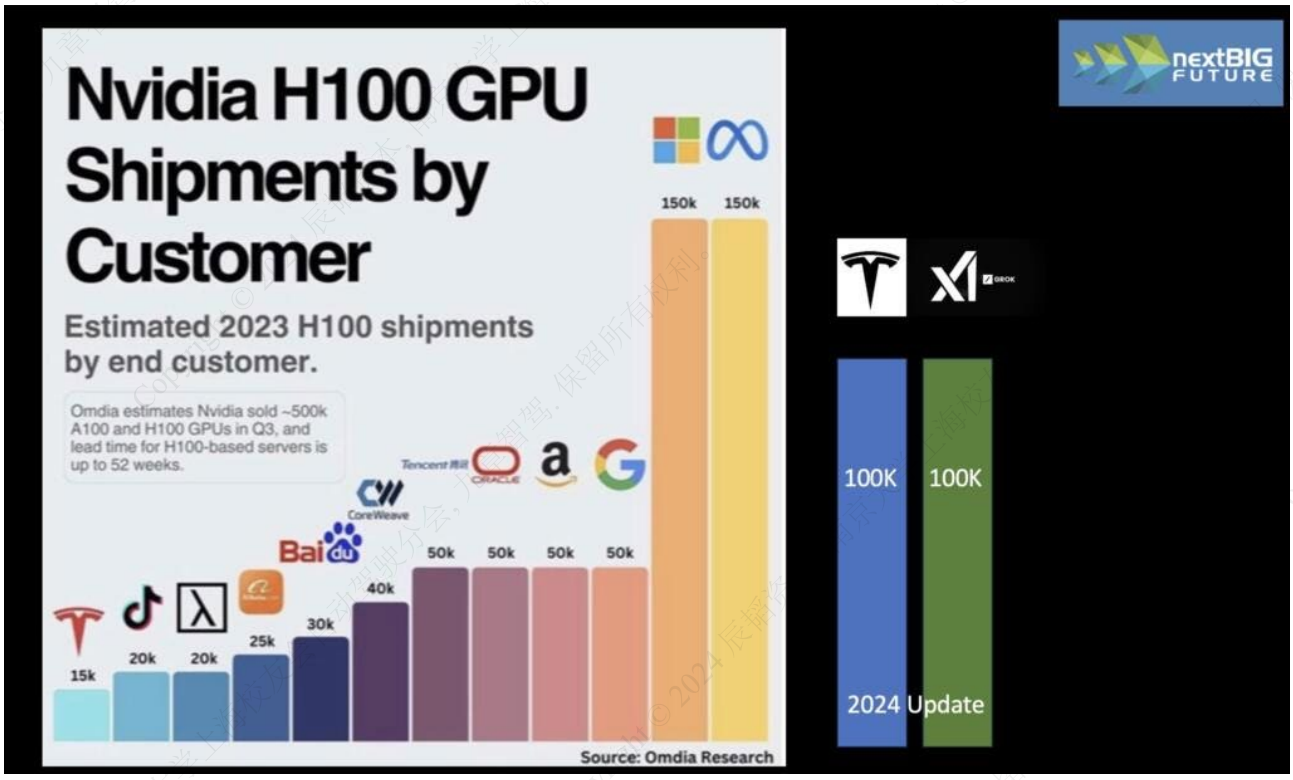
随着自动驾驶系统的 AI 模型化程度越来越高，其对训练算力资源的需求越来越大。尽管大部分公司表示 100 张大算力 GPU 可以支持一次端到端模型的训练，

但这并不意味着端到端进入量产阶段只需要这一数量级的训练资源。

量产研发阶段更需要考虑团队分工和模型迭代效率问题，企业所拥有的训练算力越大，完成端到端模型训练的时间就越短，越能抢占市场先机。另外，当企业拥有更大训练算力时，更有可能研发出性能更强（大多数时候意味着云端模型的参数量更大）的自动驾驶端到端模型，也会提升部署在车端的模型的能力。

特斯拉近年来不断增加训练算力投入，2023 年底已经跻身英伟达 H100 头部客户之列。在 2024Q1 财报电话会上，特斯拉表示，公司已经有 35000 张 H100 GPU，并计划在 2024 年内增加到 85000 张 H100 以上，达到和谷歌、亚马逊同一梯队。此前，特斯拉还部署了规模更大的 A100 GPU 训练集群，其实际训练算力投入在自动驾驶行业中遥遥领先。

图示 25：英伟达 H100 GPU 头部客户排名预测



图源：Nextbigfuture.com

国内主机厂和自动驾驶公司的训练算力很少能达到特斯拉的规模。小鹏是在训练算力建设方面比较领先的公司，其于 2023 年 8 月宣布建成了“扶摇”自动驾驶智算中心，算力可达 600PFLOPS（以英伟达 A100 GPU 的 FP32 算力推算，约等于 3 万张 A100 GPU）。商汤大装置已经布局全国一体化的智算网络，拥有 4.5 万块 GPU，总体算力规模达 12000PFLOPS，2024 年底将达到 18000PFLOPS。大部分研发端到端自动驾驶的公司目前的训练算力规模在千卡级别，随着端到端逐渐走向大模型，训练算力将显得捉襟见肘。

4.4 测试验证方法尚不成熟

实车测试验证的成本高昂，针对经典的自动驾驶架构，行业已经有一套行之有效的方案进行模型上车前的测试验证，即感知算法使用回灌数据进行离线开环测试，规控算法基于模拟器进行闭环测试验证。

然而，以上方法无法满足端到端自动驾驶的测试验证需求。一方面，在基于数据回灌的开环测试条件下，端到端系统无法与环境交互，系统一旦出现偏离采集路径的操作，后续系统的响应将无法评估。例如，商汤绝影智驾副总裁石建萍表示，公司之前基于开环测试验证后的模型，部署在实车上后差异很大，可见，离线数据开环验证的结果基本无法等效于实车的表现。另一方面，基于模拟器可以实现模型的闭环测试验证，但现有的模拟器在传感器一致性和保真度上离真实世界尚有较大差距。

4.5 组织资源投入的挑战

如第三章所述，端到端将带来自动驾驶团队的组织重塑。大部分 All in 端到端的公司都在削减原有团队规模，将团队重心调整到 AI 大模型和数据基建方向。参考特斯拉从 FSD v11 到 v12 经历的“性能爬坡”，端到端模型上车后，其早期性能可能不会强于极致优化的经典技术方案，这也对管理层进行技术路线转型的决心提出考验。

同时，削减团队规模不意味着总体投入的减少，一方面，端到端转型会使现有智驾团队的人员规模减少，另一方面，端到端对数据、数据基础设施的投入增加。对部分公司管理层来说，扩张团队的决策是容易的，但对数据、工具链等“看不见、摸不着”的资源的投入更需要认知模式的切换。

4.6 车载芯片算力会不会成为瓶颈?

无论开发何种自动驾驶技术，算法工程师总是希望车载芯片算力越大越好。因此，当提到端到端模型部署的挑战时，“端到端需要更大算力芯片”是一个容易达成的推论。然而，与直觉不同的是，部分受访者认为，车载芯片算力限制并不是端到端落地的主要矛盾，而这样的观点也许更具有现实意义。

一方面，当前以英伟达 Orin 和地平线 J6 为代表的大算力车载芯片已经可以支持 BEV + Transformer 模型的部署；端到端的基础思想是整个算法处理流程的模型化，从经典架构到端到端，总的代码数量会显著降低，端到端神经网络带来的计算资源消耗相比 BEV 模型并不一定会有显著的提升。对更高算力的渴望更多来自于模型参数量和模型性能的提升，而不是来自于经典算法架构到端到端的转变。

另一方面，软件与芯片一直是动态演化的，与其思考“端到端模型需要多少算力”,不如考虑“基于当前的芯片,应该如何优化端到端模型以实现高效的部署”。智驾芯片的性能上限依然在不断提升，几乎每 3 年其算力就会上一个台阶，可以支持自动驾驶公司部署性能更强大的端到端模型。同时，行业头部玩家如英伟达、华为、地平线、蔚来、Momenta 都在逐渐走向软硬一体，针对自研模型定制化开发更匹配的计算芯片，可以做到最大程度的优化。

4.7 模型缺乏“可解释性”是否构成障碍?

行业中有一种观点认为，端到端系统由于没有模块间显示表达的中间结果，无法保障各个模块的确定性和安全性，这增加了错误发生的风险和调试的难度。这样的担忧对于追求安全性和确定性的 L4 级无人驾驶系统更加显著。

然而，大部分受访专家认为，缺乏“可解释性”并不会成为限制端到端模型应用的问题。主要原因有二:

第一，尽管端到端没有显示表达的中间结果，但是模块化端到端依然保留了各个主要功能模块，中间的特征输出可以被进一步提取为可解释的数据，帮助分析和理解模型是如何工作的。

第二，对可解释性的担忧从 AI 在计算机视觉领域应用以来一直存在，但与其性能较传统算法的显著提升相比，可解释性成为一个次要考量因素。目前的自动驾驶技术方案中，基于 AI 模型的 BEV 感知也存在可解释性问题，但是 BEV 模型由于其性能的显著性能提升，已经成为自动驾驶感知的主流技术方案。

第五章 展望端到端

5.1 量产进度

2025 年模块化端到端开始上车，2026 年至 2027 年 One Model 端到端开始上车

目前，国内头部乘用车自动驾驶企业已经公开端到端自动驾驶方案在 2024-2025 年上车的规划，包括华为、小鹏、元戎启行、商汤绝影等。

部分专家认为，这一技术演进的时间进度可以参考国内企业追赶特斯拉 BEV/Occupancy Network 的进度。特斯拉在 2021 和 2022 年年底的 AI Day 上分别公布了 BEV 和 Ocuupancy Network 的技术架构，而国内车企开始 OTA 基于 BEV/Occupancy Network 的功能普遍在 2023-2024 年，与特斯拉的研发进度差大概在 1.5~2 年。虽然行业内关于特斯拉的具体技术方案的猜测有一些分歧，但大部分专家认为当前 FSD v12 的端到端方案属于模块化端到端方案，参考前述追赶进度，国内自动驾驶公司的模块化端到端方案上车量产时间可能会在 2025 年。

同时，行业中很多公司也在积极布局下一代的 One Model 端到端系统。该技术方案需要更多模型、数据、训练算力和芯片层面的升级，中性预期其落地时间会晚于模块化端到端 1~2 年时间，从 2026 年至 2027 年开始上车量产。

商用车自动驾驶方面，零一汽车在 5 月 16 日举办的“零点”新品发布会上首次披露了其端到端自动驾驶系统的进展，零一汽车也因此而成为商用车领域第一家公开投入端到端技术的公司。零一汽车认为，由于商用车自动驾驶的应用场景比乘用车更加可控，公司规划在 2025 年开始测试 One Model 的端到端系统，2026 年开始在部分应用场景开始稳定运营，并实现常态无人化。

5.2 技术发展趋势

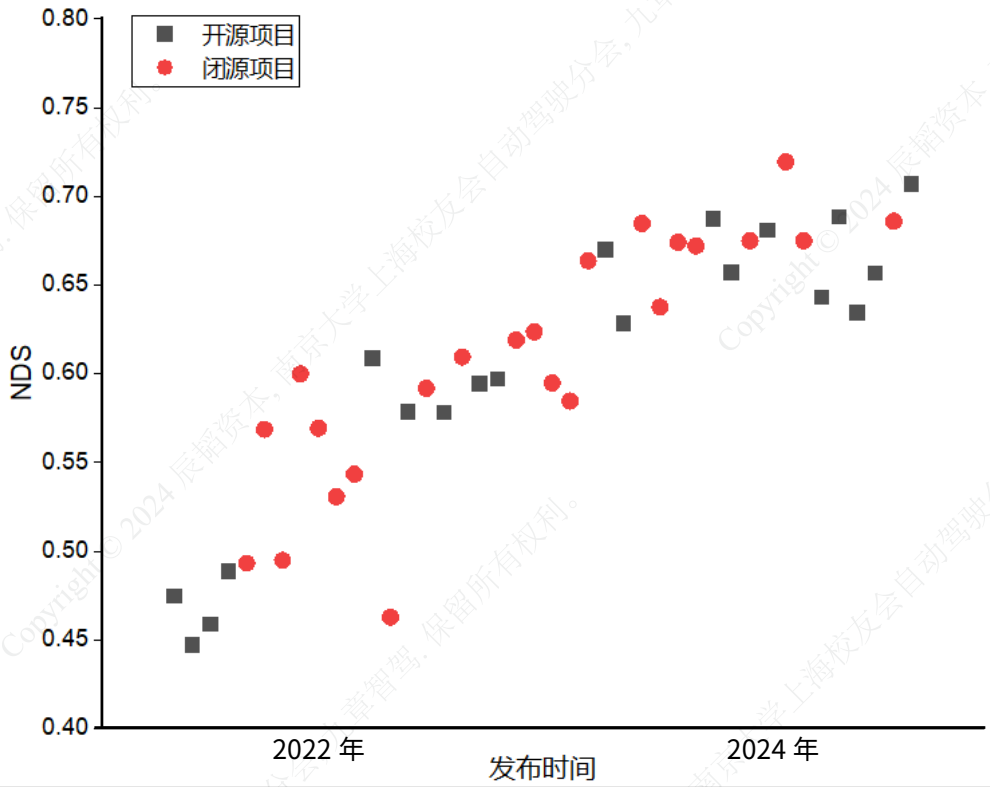
开源生态在端到端算法的发展中扮演重要角色

开源项目和社区在资源共享、社区协作、人才培养方面的优势在 AI 模型日趋复杂、迭代速度加快的当下显得越来越重要。

自动驾驶领域，复盘 2021 年以来基于 BEV 的视觉感知算法的发展历程，开源社区在其中扮演了举足轻重的作用。从自动驾驶视觉感知领域最权威的

nuScenes 目标感知天梯榜上，可以一窥开源与闭源生态的发展规律：2022 年以来视觉感知算法的性能（以 nuScenes 的综合评分 NDS 衡量）接近翻倍，在这一不断进步的过程中，一半以上的优秀项目是开源的；同时，开源项目在这一榜单上的表现并不弱于闭源项目。开源社区的贡献不仅在算法本身的提升上表现突出，还在数据集的共享、代码的开放和社区的协作上体现出强大的推动力。

图示 26：视觉感知算法领域开源项目与闭源项目的性能提升对比



来源：nuScenes 目标检测天梯榜，本报告编写组整理

2022年以来大模型和生成式 AI领域的蓬勃发展，也离不开开源社区的贡献。开源的 AI 项目正在显著增长，贡献者数量也在不断增加。尽管闭源的技术巨头为 AI 技术开发投入了大量金钱、计算资源、数据等成本，但开源社区的发展速度和成果表现总能迎头赶上。

端到端自动驾驶算法较 BEV自动驾驶感知算法的复杂度进一步提升，可以推测，其开发者生态更接近大模型的现状，即走闭源路线的巨头公司可能掌握先机，但是开源生态也会蓬勃发展，迎头赶上。

另一方面，虽然开源看似会带来“技术平权”，但这并不意味着自动驾驶公司或主机厂丧失壁垒。长期来看，模型本身收敛后，数据将会成为模型性能的护城河。另外，如何把 AI模型转变为具有优秀用户体验的产品将成为自动驾驶公司和主机厂的差异化竞争力来源。

闭环仿真将成为端到端量产的重要基础

实车测试的成本高昂，在自动驾驶系统上车前，往往需要进行大量的仿真测试来帮助算法或系统低成本地快速迭代。现有的模块化自动驾驶算法已经形成了一套行之有效的测试方法：感知算法的评测较多依赖真实道路的数据采集并进行开环的回灌测试，决策规划算法的评测主要依赖模拟器和场景库。由于感知算法不涉及系统与环境的交互，所以开环测试是可以接受的。同时决策规划算法的评测大都基于理想的环境输入，因此模拟器与物理世界的一致性不会成为一个主要问题。

端到端自动驾驶不再区分这样的模块，因此其测试评估方法需要同时具备真实性和可交互性，开环测试方法无法评估系统偏离人类驾驶员的操作后的场景变化，闭环仿真成为实车测试以外最重要的基础工具。多位专家认为，闭环仿真工具的研发是端到端上车的必要条件，而目前在研发端到端的团队几乎也都在同步开展闭环仿真工具的探索。

用于端到端的闭环仿真模拟器需要构建几方面重要的能力：

- 场景定义和生成
- 传感器仿真
- 交通流仿真
- 车辆动力学仿真

其中，车辆动力学仿真已经发展多年，成熟度较高。而前三者随着近年来计算机图形学、计算机视觉、生成式 AI 的发展，已经有了长足的进步。

目前，学术界普遍采用 CARLA 作为端到端开发的闭环仿真模拟器，但与真实数据相比，其生成的数据的真实性、复杂性、真值可获取性、实时性无法满足量产的需求。

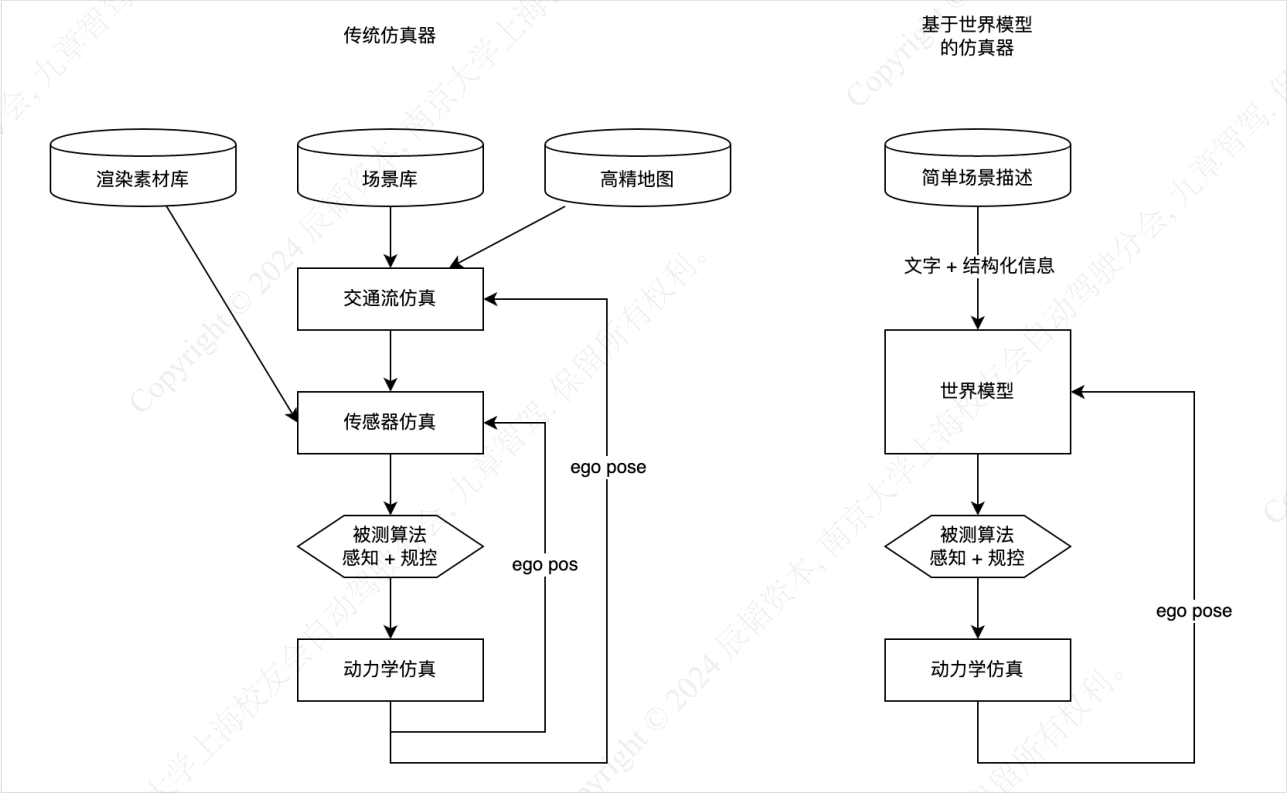
光轮智能团队针对以上问题，利用在 Nvidia、Cruise 和蔚来汽车内部成功的 0-1 闭环仿真技术和工程经验，结合生成式 AI，开发出针对端到端算法研发的数据与仿真全链路解决方案。

针对传感器仿真的真实性问题，光轮智能通过从 0 到 1 的技术创新，并与国内外顶尖实验室和高校合作，基于 3D Gaussian Splatting（3DGS）技术研发准确度领先于 SOTA 的面向开放世界街道场景的三维重建技术（其中 3DGS 真实动目标重建部分的代码已开源），并实现了车道级偏移后新视角图像的准确生成。同时，光轮独立研发了行业上首套 AI 与仿真混合渲染引擎，首次实现结合 3DGS 和实时 Ray Tracing，可以将 3DGS 的静态环境与传统 CG 渲染的动态物体有机融合渲染，实时渲染多路摄像头和激光雷达，具备深度、语义分割在内的各类真值标注的能力。目前，光轮智能的三维重建混合渲染自动驾驶仿真器已实现重打光技术，使得白天 / 晴天采集的数据可以在重建后进行黑夜 / 雨雪天 / 沙尘暴等恶劣情况的模拟。

针对动态交通流仿真真实性和复杂性问题，光轮智能自主训练开发了基于 LLM + Diffusion 的智能体 Agent 技术，利用大量真实交通流数据进行训练，可以为每一辆交通流车辆分配不同的驾驶风格，显著提升动态交通流的多样性、生成 Corner Case 的效率和覆盖度。光轮智能大幅优化了该算法的运行效率，可以在实时模拟中，同时闭环控制相对大规模数量的交互智能体，实现闭环仿真中智能体与端到端自动驾驶算法的高交互甚至对抗仿真。

最后，针对仿真支持端到端算法训练实时性的需求，光轮智能深度优化了仿真器模拟和渲染的性能，使得多路传感器的闭环模拟可以达到超实时运行。同时该仿真器提供 Gym-Style API，用户可以利用光轮智能仿真器实现端到端算法闭环评测与训练。目前，光轮的解决方案正在服务多家头部端到端算法预研。

图示 27：基于世界模型的仿真器和传统仿真器对比



图源：极佳科技

极佳科技的闭环仿真的核心是一套被称为世界模型的多模态视觉生成大模型。在其训练过程中引入了数十亿图像数据进行预训练的 Diffusion Model 作为基础模型，在其上引入了 ActionFormer 模块、经自动驾驶场景特别优化的大语言模型、以及统一多视角视频模型（UniMVM）。由此支持以多模态数据（文本、图像、 HDMap，3D 框、历史驾驶动作）作为控制条件输入，预测生成：符合输入条件的、支持规控算法闭环控制，并且能够在一个统一模型架构内，同时支持环境生成、交通流生成、传感器数据生成。

车载芯片架构创新将更好地支持端到端模型部署

自动驾驶端到端大模型等高阶智驾技术的发展，离不开车载芯片架构的持续创新，智能驾驶算法的高速演进，也对车载芯片的AI计算能力提出了更高的要求，需要在满足智能驾驶场景AI性能的同时，更好地兼顾整体端到端的数据流处理分析，并能够有效保障汽车功能安全。

端到端模型倚赖Transformer作为基础模型结构的强大能力，需要原生支持Transformer的芯片架构。一个原生支持Transformer的计算架构应包含高浮点算力、特殊算子优化加速、高带宽存储等特点。以地平线为例，其新一代智能计算架构BPU Nash支持线程并发的SIMT Vector Processing Unit (VPU)，支持BF16/FP16/FP32多种浮点数据类型，在性能和精度之间取得更好的平衡。BPU Nash对transformer中的特殊算子做了特别优化，能够支持Layer-norm, Softmax, Transpose, Reshape等算子的硬件加速。BPU Nash还设计了全新的存储系统，用于数据的缓冲和交换，以先进的总线架构配合高带宽的DDR，使得数据模型的参数能够实现高效的吞吐，有效缓解内存墙问题。

芯片架构的创新更离不开芯片设计的原材料——IP的创新。安谋科技(Arm China）坚持在Arm全球标准的基础上开展本土创新，将Arm技术与自研业务产品进行异构融合，打造面向智能汽车的高性能异构计算芯片IP平台，这不仅涵盖了基于Arm平台的Cortex®-M系列、Cortex-R系列、Cortex-A系列等不同处理器IP产品，并包含带有功能安全的汽车增强（AE）产品系列，还囊括了“周易”NPU、“星辰”CPU、“山海”SPU、“玲珑”VPU等安谋科技自研业务产品，通过各个计算单元之间的协同与互补，能够有效兼顾通用性与专用性，充分满足汽车芯片厂商的个性化定制需求。2024年，全新发布的Arm汽车增强（AE）处理器首次为汽车应用引入最新的Arm®v9架构技术和服务器级性能，其中包括聚焦自动驾驶应用的Arm Neoverse V3AE，更强的服务器级性能可以应对后续更高强度的负载需求，为端到端自动驾驶模型的持续迭代至更强性能打下核心技术基础。

另外，从BEV到端到端，自动驾驶算法的迭代周期相对较短，而传统的芯片设计周期受限于“一环扣一环”的设计流程，动辄需要3至5年才能完成一次迭代，这种硬件先行的思路在算法技术架构高速迭代的今天正面临着极大挑战，行业迫切需要一套加速汽车整体开发周期的方法。以Arm平台为例，通过与自动驾驶生态合作伙伴更早期的协同，可以实现更无缝、更快速的软件开发和芯片设计，全方位加速智能汽车整体开发周期。

5.3 产业发展趋势

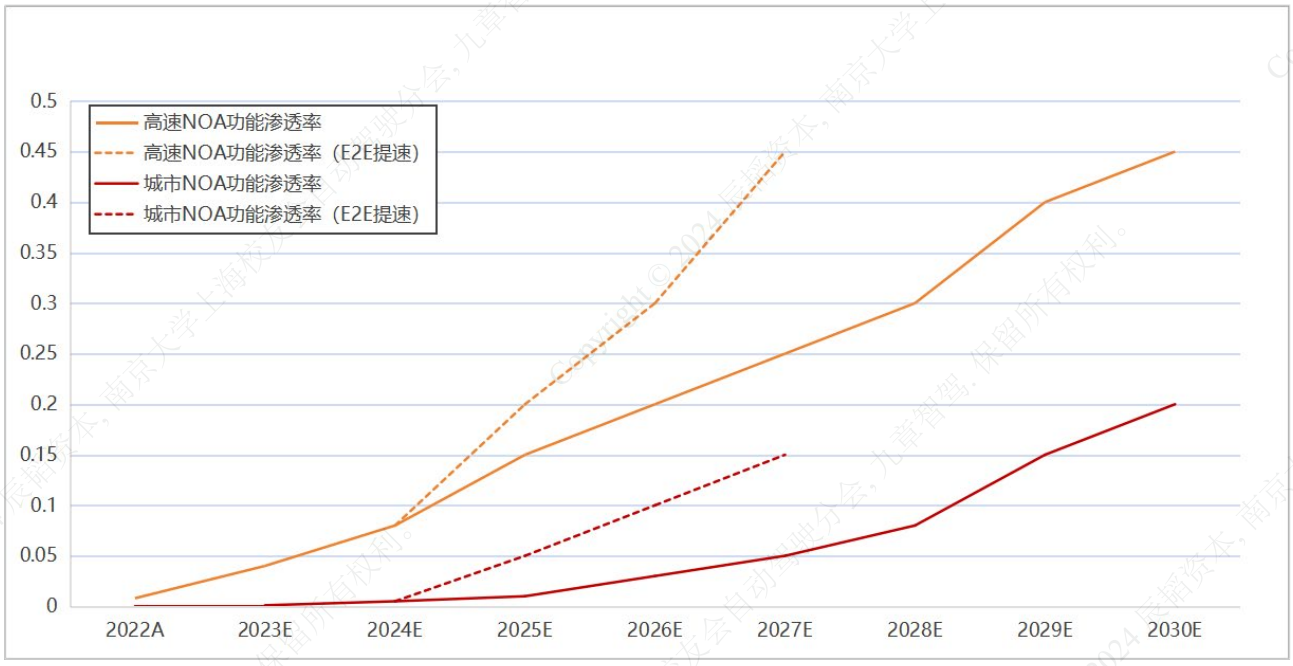
端到端加速自动驾驶整体渗透率提升

自动驾驶的商业模式大致分为两类，其一是高速 NOA/ 城市 NOA 等辅助驾驶功能作为整车的附加软件面向消费者收费，其二是无人驾驶（L4）作为生产力工具向 B 端客户收费。

端到端是创新的自动驾驶技术方案，其带来的自动驾驶性能提升和用户体验提升会让更多消费者愿意接受自动驾驶，从而加速自动驾驶渗透率的提升。以特斯拉为例，其城市 NOA 软件 FSD 在以前版本只有 2% 左右的订阅率，但是自 2024 年 3 月特斯拉开启 FSD v12 大规模推送和免费试用后，Elon Musk 表示其订阅率有大幅度提升 —— 业界普遍猜测，其订阅用户比例可能提升至 20% 左右。

基于特斯拉的案例，可以合理推测，当自动驾驶行业进入端到端时代后，其整体的渗透率提升将大大加速。2023 年 9 月中信证券发布的研报中预测，至 2026 年高速 NOA/ 城市 NOA 的自动驾驶渗透率预计将达到 20%/3%。2025 年起，端到端发展的提速将催化各级别自动驾驶功能渗透率大幅提升，我们据此更加乐观地预测，高速 NOA 渗透率至 2026 年将超过 30%，城市 NOA 渗透率超过 10%。

图示 28：在端到端量产前和量产后的高速 NOA 和城市 NOA 的渗透率预测



数据来源：中信证券，辰韬资本预测整理

数据和 AI 人才的重要性提升，产业分工格局演变

端到端进一步提升了数据及 AI 能力在整个产业链上的重要性。

何小鹏在小鹏汽车 520 AI DAY 上提出，端到端会加速完全无人驾驶的落地，同时提高了智驾的数据门槛。而数据闭环所需的算力费用高昂，智驾产业的头部效应会越来越明显，少数几家企业会脱颖而出，并且，与采用传统的模块化自动驾驶方案的企业差距会越拉越大。

从生产要素角度考虑，目前智能驾驶的研发成熟度与“工程师人数”高度相关，头部主机厂和自动驾驶公司的智能驾驶团队规模均在 1000 人以上。核心原因是，基于现有技术架构，城市 NOA 的开城过程中，会遇到大量 Corner Case，需要大量的规控程序员和测试工程师处理。

行业切换至端到端架构后，“高质量数据”和“顶尖 AI 人才”可能成为更加重要的资源要素。

其中，数据是主机厂的核心资产，能够持续量产交付的主机厂并不会缺少数据；而对于顶尖 AI 人才来说，只有头部主机厂和部分头部自动驾驶供应商才有吸引力。在这种背景下，部分拥有数据、但自研能力较弱的主机厂将与 AI 人才集聚的自动驾驶技术公司产生更强的合作动力，这给了一些 AI 能力强的自动驾驶供应商新的商业机会。

自动驾驶国际化和场景迁移进程提速

端到端的强泛化能力可能突破不同场景、不同车型的壁垒。

比如，智能驾驶公司在出海时遇到的一个普遍的问题是，由于目前不同国家地区的驾驶场景、交通规则都有较大差异，因而，场景适配所需的工作量极大。而端到端的强泛化能力可能大大减少这一差异带来的资源投入，进而加速智驾公司在不同场景、不同国家或地区的应用。今后，在全球智能驾驶产业发展中，基于端到端技术的自动驾驶公司有更强的出海开展业务的能力。

另外，现有的智能驾驶方案对车型的依赖度较高，特别是，模型需要针对不同车型的传感器位置和传感器参数的变化做额外的训练和适配。端到端不仅可以实现车型尺寸间的泛化，甚至有可能突破乘用车和商用车之间的数据壁垒，让辅助驾驶行业和无人驾驶行业产生新的化学反应，这也许会加速辅助驾驶和不同商用场景无人驾驶能力的迁移。

第六章 端到端自动驾驶与通用机器人

一直以来，自动驾驶和机器人是紧密相关、相互借鉴和启发的两个行业，本章将集中讨论自动驾驶早期发展与机器人的历史渊源，以及近年来自动驾驶端到端技术可能对机器人行业的启发；最后，我们将呈现关于物理世界 AGI 的实现路径的不同观点。

6.1 阶段一：机器人赋能自动驾驶早期的发展

自动驾驶的早期发展过程中，借鉴了大量机器人技术的积累。这些借鉴不仅体现在传感器的使用上，还包括感知算法、规划算法、中间件等多个层面。

自动驾驶系统大量依赖激光雷达、摄像头和超声波传感器，这些传感器技术最初在机器人领域得到了广泛应用和成熟。机器人使用这些传感器来感知环境、避障和定位。例如，激光雷达在机器人导航中的应用可以追溯到 20 世纪 80 年代末和 90 年代初，麻省理工学院（MIT）和卡内基梅隆大学（CMU）的研究人员在那个时期开始将激光雷达应用于机器人系统中。2004 年美国国防部举办的 DARPA 无人驾驶挑战赛上，激光雷达首次被应用于自动驾驶汽车。

感知算法方面，SLAM（Simultaneous Localization and Mapping）技术是机器人自主导航的核心技术之一，允许机器人在未知环境中实时构建地图并确定自身位置。自动驾驶系统借鉴 SLAM 技术，实现高精度地图构建和车辆定位。例如，百度的 Apollo 自动驾驶平台利用 SLAM 技术构建高精度地图，提供车辆定位和导航支持。这种技术的应用提高了自动驾驶系统在复杂城市环境中的导航能力。

路径规划算法方面，目前广泛应用于自动驾驶行业的经典算法 A* 和 Dijkstra 算法均来自于机器人行业。例如，Waymo 的自动驾驶系统使用了一种改进的 A* 算法，能够在复杂城市环境中计算出最优驾驶路径，同时避开交通拥堵和事故区域。

中间件在机器人系统中用于实现不同模块之间的数据通信和协调。自动驾驶系统借鉴这一概念，通过中间件实现各子系统的高效协同。ROS（Robot Operating System）是机器人领域的标准中间件，许多自动驾驶系统在开发初期都使用 ROS 进行原型设计和测试。

6.2 阶段二：自动驾驶产业化加速，端到端技术可反哺机器人

2015 年以来自动驾驶技术快速迭代，叠加近几年来智能电动汽车在乘用车消费市场的渗透率跨越鸿沟，带动了自动驾驶产品真正迈向大规模应用，而这一阶段机器人行业没有出现类似的规模化应用。

除成本因素外，机器人最大的问题是智能化、通用性不足。传统的机器人受限于算法能力，并不具备智能化的理解、交互能力。其中智能化和通用化程度最低的代表是工业机械臂，其操作基本都基于固定运动轨迹编程，每次切换应用场景和操作的产品都需要重新编程；随着机器人算法的进步，也逐渐发展出一些“清洁机器人”、“物流机器人”等细分赛道，但目前这类产品有明确的功能属性和应用场景范围，智能化和通用化程度有待提升。

不过，机器人和自动驾驶技术的基本架构有一定通用性，自动驾驶端到端代表的数据驱动学习范式同样可以应用于机器人。端到端相比基于规则的经典算法的转变，在机器人领域的重要性甚至超过自动驾驶。

清华大学交叉信息研究院赵行教授表示：“传统的自动驾驶技术方案很难解决机器人的问题，以感知能力为例，自动驾驶系统运行的道路环境最主要的元素是车辆、行人、车道等，在经典算法架构下以目标检测、车道线检测等任务的形式表达可以解决大部分问题；而机器人可能产生互动的环境要素过于多样，例如家用机器人可能需要与数据线等特殊几何形态物体、流体等难以简单表征的元素交互。因此传统感知和规控方案无法用于通用机器人。相比之下，数据驱动和 AI 是对机器人更重要的思路。”

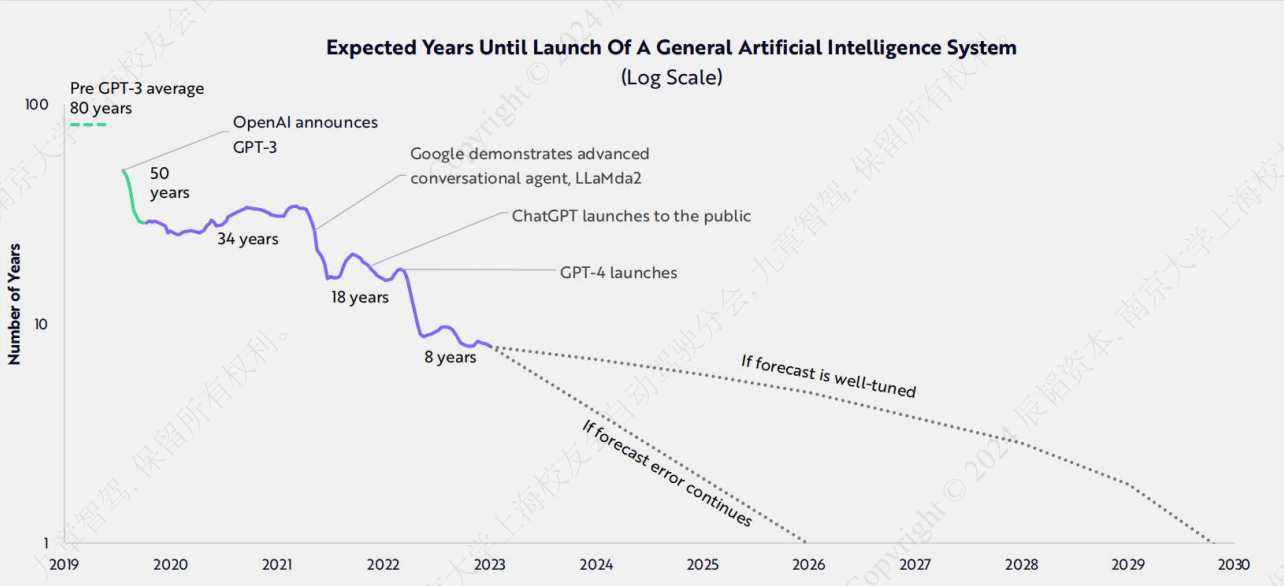
基于以上背景，自动驾驶端到端已经提供了一种可供参考、可量产的数据驱动技术范式。自动驾驶端到端的探索不仅会从算法和模型架构上启发机器人采用数据驱动技术范式，同样也为通用机器人如何构建数据获取路径和算力等数据基建提供了参考。

6.3 阶段三：自动驾驶和机器人竞逐物理世界 AGI

随着大模型在 2022 年以来的快速发展，人们对通用人工智能的到来时间有了更乐观的预期。方舟资本发布的《Big Ideas for 2024》报告中指出，目前，通用人工智能预期实现的时间已经提前到 2027-2031 年。在这一语境下，AGI 主要指的是语言大模型、多模态大模型等虚拟世界中的 AGI。

另一方面，当虚拟世界的 AI 已经展现出足够的智能化，人们自然而然想尝试将 AI 拓展到物理世界，从而引发了 2023 年以来“具身智能”领域的研究热潮。

图示 29：对通用人工智能实现时间的预测



图源：Ark Investment, Big Ideas for 2024

在物理世界 AGI/ 具身智能的语境下，端到端自动驾驶和通用机器人是这一研究范畴最重要的两个应用领域。阿里云智能集团副总裁李强在 2024 年 3 月的电动汽车百人会上提到“具车智能”概念，将具身智能与自动驾驶间做了概念上的关联。

虽然讨论“自动驾驶和通用人形机器人谁会率先实现物理世界 AGI”这一问题并无现实价值，但无论哪个领域率先实现，支撑这一判断的考虑因素却可以给另一个领域带来启发。我们将在报告的下文呈现行业专家对于这一问题的正反两方面观点。

支持“通用人形机器人将会率先实现物理世界 AGI”这一观点的事实主要是其安全性要求低于自动驾驶汽车，因而可能更容易达到可用状态——在大多数应用场景中，机器人主要面对的是静态的障碍物而非动态变化的环境，这种稳定的工作条件使得机器人有更多时间处理信息和进行深入推理，从而提高操作效率和减少意外。

这一观点对自动驾驶的启发是：反过来说，提高计算和响应的速度能够显著提高安全性，因此，更高效的计算平台、更灵敏的底盘执行器是加速自动驾驶应用的关键环节。

支持“自动驾驶将会率先实现物理世界 AGI”这样观点的主要事实是：

a. 自动驾驶的场景相对结构化。自动驾驶汽车主要操作在相对可预测和规范化的环境中，如公路和城市街道。这些场景虽然复杂，但遵循一套明确的交通规则，

车辆的任务目标是明确的。相比之下，通用人形机器人面临的挑战则更为多样和复杂。人形机器人需要应用于多样的场景，如工厂、家庭、办公室或公共场所等，其对泛化能力的要求远高于自动驾驶汽车。鉴智机器人联合创始人、CTO 都大龙提到，通用机器人必然将落地自动驾驶，因为自动驾驶有着最统一的场景需求，有目前最成熟的传感器及执行器。

这一观点对机器人的启发是：追求“一定范围的下的通用”可能是机器人更务实的落地路径，参考 L4 级无人驾驶发展的经验教训，通用机器人也应当从特定场景开始应用，逐渐拓展至更广泛的场景。

b. 通过影子模式获取海量数据的路径已经打通。Wayve.AI CEO Alex Kendall 在一篇文章中表示，数以百万辆计的汽车都配备了训练和部署具身智能所需的传感器和计算能力，这些汽车每天都在路上跑，这意味着与其他机器人应用不同，数据收集可以以较低的成本实现。这是 Wayve 认为自动驾驶将助推具身智能的发展的主要原因。另外，驾驶场景的数据同质性很高，这意味着每一辆车采集的数据在端到端的范式下是能够为训练后的模型应用起到重要作用的，但是机器人的应用场景多样，比如从工厂里拧螺丝到餐馆里洗盘子，每一个任务都可能需要不同的数据和数据处理方法，这样的数据异质性显著增加了技术实现的复杂度。

这一观点对人形机器人的启发是：需要找到一个可以快速形成数据反馈的场景和机制。目前，机器人的数据采集除了使用机器人直接采集外，还可以通过遥操作、动作捕捉等方式进行，但是其获取海量高质量数据的能力依然非常匮乏。长期来看，仿真和合成数据是机器人获取数据的重要手段。

总的来说，随着 AI 技术的不断进步，自动驾驶和机器人技术的相互融合、相互借鉴将更加深入，共同推动 AGI 走向物理世界，创造更大社会价值。

本报告访谈和编写项目组

本行业研究报告由辰韬资本、南京大学上海校友会自动驾驶分会和九章智驾三方联合发起及撰写，本报告的成功编撰得到了众多行业专家、组织和企业的鼎力支持，衷心感谢所有对本次行业研究报告做出贡献的人员和组织。

* 以下名单均按拼音首字母排序

策划 / 指导团队

辰韬资本： 萧伊婷
九章智驾： 苏清涛
南京大学上海校友会自动驾驶分会： 罗雪霜

采编团队

主笔
辰韬资本： 刘煜冬、赵磊
责任编辑
九章智驾： 苏清涛
专家访谈
南京大学上海校友会自动驾驶分会： 陈璐、董婷婷、姜丽丽、Molly、宋怡、肖恩、曾恺翔、章清泉、张子杭
编纂支持
南京大学上海校友会自动驾驶分会： 董婷婷、Molly、宋怡
辰韬资本： 彭轶宁

支持团队

辰韬资本： 孙箫笛、孙　欣、王　琴

参与访谈及审阅的专家

戴文鼎　路特斯机器人
都大龙　鉴智机器人
何　俊　柏川数据
侯　聪　轻舟智航
黄泽铧　零一汽车
李弘扬　上海人工智能实验室
楼天城　小马智行
罗　恒　地平线
毛继明　极佳科技
石建萍　商汤绝影
孙鲁毅　安霸半导体
孙韶言　极佳科技
孙又晗　前擎天智卡
王　弢　前小鹏汽车
谢　晨　光轮智能
徐泽林　辉羲智能
杨继峰　行业专家
姚卯青　前蔚来汽车
殷　鹏　香港城市大学
张　鹏　行业专家
赵　行　清华大学 MARS Lab

在此，也向所有不便署名的采编团队成员及访谈专家表示衷心感谢。

此外，也特别感谢柏川数据（南京大学校友企业）和侨银数智在人员方面提供了重要支持。以及南京大学上海校友会苏州科创中心（云帆乘风）、宝通科技在此次组织及协调外部合作中发挥的重要作用。

最后，再次向所有参与和支持本次自动驾驶行业报告编撰的专家和团队表示衷心的感谢！

特别声明

分析师声明

作者具有专业胜任能力，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解。本报告清晰准确地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响，特此声明。

免责声明

本报告编写组不会因为接收人接受本报告而将其视为客户。本报告仅在相关法律许可的情况下发放，并仅为提供信息而发放，概不构成任何广告。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。在法律许可的情况下，辰韬资本及其所属关联机构可能会持有报告中提到的公司的股权，也可能为这些公司提供或者争取提供筹资或财务顾问等相关服务。

本报告的信息来源于已公开的资料，本报告编写组对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见及推测仅反映报告编写组于发布本报告当日的判断，本报告所指的公司或投资标的的价格、价值及投资收入可升可跌。过往表现不应作为日后的表现依据。在不同时期，报告编写单位可发出与本报告所载资料、意见及推测不一致的报告。报告编写单位不保证本报告所含信息保持在最新状态。同时，报告编写单位对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。本报告中涉及企业产品的相关图片均整理自公开渠道，或经企业授权使用。

版权声明

本报告版权仅为报告编写单位所有。未经报告编写单位书面许可，任何机构和 / 或个人不得以任何形式翻版、复制和发布本报告。任何机构和个人如引用、刊发本报告，须同时注明出处，且不得对本报告进行任何有悖原意的引用、删节和 / 或修改。

Copyright © 2024 辰韬资本, 南京大学上海校友会自动驾驶分会, 九章智驾. 保留所有权利。

辰韬资本
CHERISH CAPITAL

南京大学上海校友会
自动驾驶分会

九章智驾
智能驾驶 产业智库

2024 年 6 月