

## 汽车行业深度报告

# AI 系列深度 25 期：从模仿学习到强化学习的范式跃迁

增持（维持）

2026 年 08 月 07 日

证券分析师 黄细里

执业证书：S0600520010001

021-60199793

huangxl@dwzq.com.cn

研究助理 童明祺

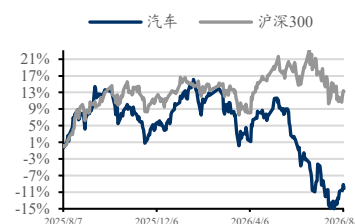
执业证书：S0600125080005

tongmq@dwzq.com.cn

### 投资要点

- **模仿学习的能力上限在数字 AI 与物理 AI 之间存在结构性分野。**大语言模型预训练本质上是对互联网级人类知识的规模化模仿，学习对象覆盖公开语料中的多领域知识，广度上限高于单一专家，因此可形成较高能力底座；自动驾驶端到端模型主要学习人类驾驶轨迹，而 L4 安全目标要求显著超越平均人类表现，模仿学习天花板约束更强。
- **闭环驾驶会放大模仿学习的结构性缺陷。**行为克隆在训练分布内表现稳定，但一旦进入示范数据未覆盖状态，模型输出偏差会推动车辆进入新的状态分布，形成协变量漂移；长尾、危险和事故场景也难以依靠真实道路主动采集。强化学习因而不只是后训练补强，而是驾驶模型突破模仿边界的重要环节。
- **强化学习与世界模型在物理 AI 中形成双向共生。**真实道路无法承担反复试错、失败重置和危险事件注入，世界模型因此成为策略训练、奖励设计与安全验证的闭环试错场；同时，真实数据回流与策略训练也会暴露仿真偏差，推动世界模型持续校准。
- **产业实践正在向“端到端+世界模型仿真+强化学习后训练”收敛，但具体路线仍有分化。**特斯拉、小鹏、理想、Momenta、地平线、千里科技、卓驭、元戎启行等均已公开将强化学习纳入训练链路；小鹏、理想、元戎、千里更偏向 VLA 或多模态路线，华为、Momenta、地平线相对弱化语言中介，更强调世界模型构建与强化学习闭环。
- **L4 Robotaxi 路线更关注大规模闭环仿真、安全冗余和超越人类驾驶水平。**小马智行 PonyWorld 等公开表述突出强化学习范式和世界模型仿真，Waymo、文远知行则更多强调基础模型、大规模仿真和端到端能力建设。行业共性是以仿真环境承载闭环训练，差异集中在语言中介、强化学习披露程度和量产辅助驾驶/L4 运营定位。
- **我们认为，数字 AI 底座模型发展路径已逐渐清晰、价值在于确定性，但其难以实现物理世界的建模闭环；我们认为物理 AI 将成为物理世界的 AI 解决方案，从而在 AI 的当前发展阶段成为增量更大的方向，智能驾驶作为最先跑通闭环的代表场景，我们认为应重点关注。**
- **风险提示：**技术迭代不及预期的风险；大模型厂商 ARR 增速放缓的风险；云服务商资本开支不及预期的风险；智能驾驶及机器人商业化落地不及预期的风险。

### 行业走势



### 相关研究

《AI 系列深度 17 期：视觉智能为何引入 Transformer? 》

2026-08-06

《AI 系列深度 16 期：语言智能为何从 RNN 转向 Transformer? 》

2026-08-06

## 内容目录

|                                        |           |
|----------------------------------------|-----------|
| <b>1. 强化学习成为高物理闭环任务关键环节</b>            | <b>4</b>  |
| 1.1. 数字 AI: 互联网级知识语料抬高模仿学习上限           | 4         |
| 1.2. 物理 AI: 人类驾驶演示约束模仿学习上限             | 4         |
| 1.3. 模仿学习的结构性缺陷在闭环驾驶中被放大               | 4         |
| 1.4. 强化学习在驾驶中的定位由事后增强转向关键环节            | 5         |
| <b>2. 强化学习与世界模型: 高物理闭环下的工程共生</b>       | <b>6</b>  |
| 2.1. 强化学习需要可试错的环境, 真实道路无法承担这一角色        | 6         |
| 2.2. 世界模型为强化学习提供高保真的闭环试错场              | 6         |
| 2.3. 真实数据与强化学习反过来校准世界模型的保真度            | 7         |
| 2.4. 生成式与重建式世界模型的结合是当前主流工程取向           | 8         |
| <b>3. 厂商实践: 端到端之上叠加强化学习后训练</b>         | <b>8</b>  |
| 3.1. 量产乘用车辅助驾驶: 端到端之上的强化学习后训练正在收敛为共识   | 8         |
| 3.2. L4 Robotaxi: 以世界模型驱动的闭环仿真, 安全冗余优先 | 9         |
| 3.3. 不同路线的共性与分歧                        | 10        |
| <b>4. 结论: 强化学习从可选增强走向关键环节</b>          | <b>11</b> |
| <b>5. 风险提示:</b>                        | <b>12</b> |

图表目录

图 1: 模仿学习与强化学习的能力天花板（概念示意） .....5

图 2: 强化学习与世界模型的双向共生闭环（概念示意） .....8

表 1: 模仿学习与强化学习在数字 AI 与物理 AI 中的角色对比 .....6

表 2: 代表性厂商“端到端+世界模型+强化学习”技术布局梳理 .....11

## 1. 强化学习成为高物理闭环任务关键环节

数字 AI 与物理 AI 均以模仿作为重要训练范式，但模仿对象差异决定了强化学习在两类系统中的角色分化。理解这一分化，是判断强化学习为何在自动驾驶中由后训练工具走向关键环节的前提。

### 1.1. 数字 AI：互联网级知识语料抬高模仿学习上限

大语言模型的预训练在机制上可视为一种模仿。其训练目标是预测下一个 token，学习对象是经过筛选的互联网级人类文本语料。由于相关语料汇集了大量人类在不同领域的知识与表达，预训练可使模型在知识覆盖广度上显著高于单一个体，模仿由此可以支撑较高能力底座。

强化学习在大语言模型中更多承担事后对齐与推理增强的角色。基于人类反馈的强化学习用于让模型输出更符合人类偏好，而面向可验证任务的强化学习则用于提升推理能力。DeepSeek 团队 2025 年发表于《自然》的研究显示，在不使用监督微调的前提下，纯强化学习可以诱发自我反思、验证与策略调整等推理行为，并在数学、编程等可验证任务上超过以人类示范做监督学习的同类模型。这一结果说明，在数字 AI 中强化学习能够在模仿底座之上进一步突破，但模仿本身已提供了足够高的起点。

### 1.2. 物理 AI：人类驾驶演示约束模仿学习上限

自动驾驶端到端模型的主流训练方式同样具有模仿属性，但模仿对象主要是人类驾驶演示数据。Momenta 在阐述其技术路线时，将模仿学习概括为模型从人类司机驾驶行为中学习，其能力上限更容易受到示范数据质量、覆盖范围与人类驾驶分布约束；同时，人类驾驶并不完美，复杂低频场景样本稀缺，模型可学习空间因此受到约束。卓驭科技 AI 负责人陈晓智的表述与之相近，其认为预训练阶段本质上是模仿学习，可将系统能力推升至约 80 分水平，而从 80 分提升至 95 分则需要强化学习。

驾驶任务对能力上限的要求与模仿学习所能提供的上限之间存在落差。L4 自动驾驶的目标是在安全性上超越人类，而非趋近人类平均水平。小马智行联合创始人楼天城在阐述 PonyWorld 世界模型时表示，即使完美复刻人类顶尖司机的驾驶行为，也无法满足 L4 级自动驾驶对安全性的苛刻要求，因为机器需要理解“为什么”，而不仅是“怎么做”。地平线创始人余凯亦认为，用户行为数据价值有限，行业头部玩家不应仅依赖司机数据学习，未来在仿真平台中通过强化学习训练可能成为关键增量。

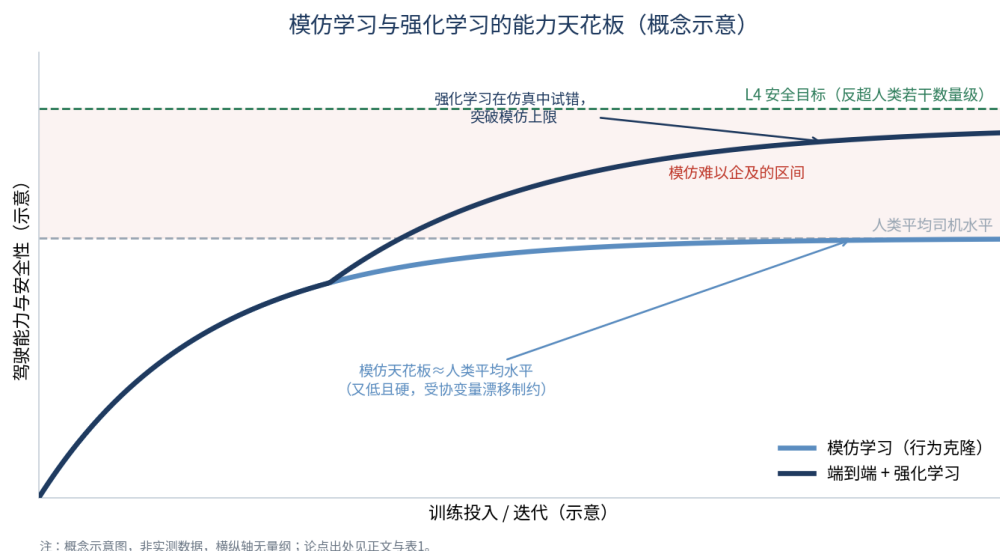
### 1.3. 模仿学习的结构性缺陷在闭环驾驶中被放大

行为克隆式模仿学习在闭环控制中存在协变量漂移问题。学术界对此已有长期讨论，Ross 等人在 2011 年提出的 DAgger 及后续研究指出，策略在执行早期产生的微小动作偏差会使车辆进入训练分布之外的状态，误差随时间累积放大，最终导致策略在新状态下失效。该问题在自动驾驶闭环场景中尤为关键，因为驾驶是动作影响后续观测的序贯

决策过程，而非单步预测。

**模仿学习从人类数据中习得的行为并非天然最优。**端到端模型可能同时学习有效驾驶行为与低质量驾驶行为，强化学习则可通过奖励设置保留安全、高效、舒适的动作，并抑制不良驾驶行为。由此可见，模仿范式的另一局限在于难以自发区分被模仿行为的优劣，而强化学习可通过奖惩信号提供行为筛选机制。

图1：模仿学习与强化学习的能力天花板（概念示意）



数据来源：东吴证券研究所绘制；为概念示意，论点出处见正文与表1

**特斯拉公开口径提供了量化参照。**关键干预间隔里程在FSD v12.5较前代改善约100倍，并预期v13改善约1000倍。该数据为公司口径，v13为预期目标，属前瞻信息，存在不确定性。即使最终兑现程度仍待验证，目标量级仍反映行业对“远超人类干预频率”的安全目标诉求，而这一目标难以仅依赖模仿平均人类司机达成。

#### 1.4. 强化学习在驾驶中的定位由事后增强转向关键环节

**强化学习在自动驾驶中的定位不同于大语言模型。**在大语言模型中，强化学习主要服务于后训练、偏好对齐与可验证推理增强；在自动驾驶中，模仿底座受人类演示数据和协变量漂移制约，若要达到超越人类的安全目标，强化学习在高物理闭环、高安全约束场景中更接近突破上限的关键环节。

**表1：模仿学习与强化学习在数字 AI 与物理 AI 中的角色对比**

| 维度          | 数字 AI，大语言模型            | 物理 AI，自动驾驶            |
|-------------|------------------------|-----------------------|
| 预训练或模仿的对象   | 互联网级人类文本语料，下一 token 预测 | 以人类驾驶轨迹为主的演示数据，行为克隆   |
| 模仿可达到的能力上限  | 在广度上逼近并可超越单一个体         | 趋近平均人类司机，难以自发超越人类安全水平 |
| 强化学习的定位     | 后训练、偏好对齐与可验证推理增强       | 在高物理闭环任务中突破模仿上限、筛选行为  |
| 奖励或反馈来源     | 人类偏好，或数学、编程等可验证答案      | 安全、效率、舒适等驾驶目标的奖惩信号    |
| 对环境或世界模型的依赖 | 较低，文本可直接评测             | 较高，需高保真闭环环境以避免奖励钻空    |
| 典型失败模式      | 幻觉、对齐偏差                | 协变量漂移、误差累积、长尾场景失效     |

数据来源：《DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning》，《A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning》等其他文献，东吴证券研究所整理

## 2. 强化学习与世界模型：高物理闭环下的工程共生

强化学习要在驾驶中发挥作用，离不开可反复试错的训练环境，而真实道路难以承担这一角色。世界模型的意义在于提供可控、可重置、可注入危险事件的闭环环境，从而支撑策略训练、安全验证和模型迭代。

### 2.1. 强化学习需要可试错的环境，真实道路无法承担这一角色

强化学习的核心机制是在与环境交互中通过奖惩信号优化策略，这要求环境能够反复复现、注入危险事件，并在失败后重置。真实道路不满足上述条件，碰撞类数据在真实道路上极难且不应主动获取，危险场景也无法随意复现。相关驾驶模仿学习研究已将世界模型作为神经模拟器，用于在训练中生成新状态，从而在降低对额外真实道路主动采集依赖的同时缓解协变量漂移，并为强化学习提供可行试错载体。

物理 AI 对交互环境真实度的要求高于传统游戏引擎式仿真。已有研究和产业实践显示，过去基于 3D 游戏引擎的尝试存在场景真实度不足、规模较小等问题，模型也可能利用仿真与真实世界之间的偏差形成不可迁移策略，导致训练结果在真实场景中失效。该问题在驾驶中体现为奖励钻空，即策略利用仿真环境的失真之处获得高奖励，却无法迁移到真实道路。

### 2.2. 世界模型为强化学习提供高保真的闭环试错场

世界模型被多家厂商定位为强化学习的训练场。理想汽车在阐述 MindVLA 时表示，强化学习若要避免奖励钻空，必须与足够逼真的世界模型交互；其采用的“重建+生成”统一世界模型，将三维场景重建与新视角生成结合，构建接近真实世界的仿真环境，并

在其上开展大规模闭环强化学习。据理想汽车公开表述，相关重建与生成方法中有四篇论文入选 CVPR2025，并将 3D 高斯渲染训练速度提升至七倍以上；该类数据仍需按公司口径处理。在下一代 MindVLA-o1 框架中，世界模型被命名为 MindSim，并与数据引擎、VLA 模型、强化学习基础设施共同构成闭环。

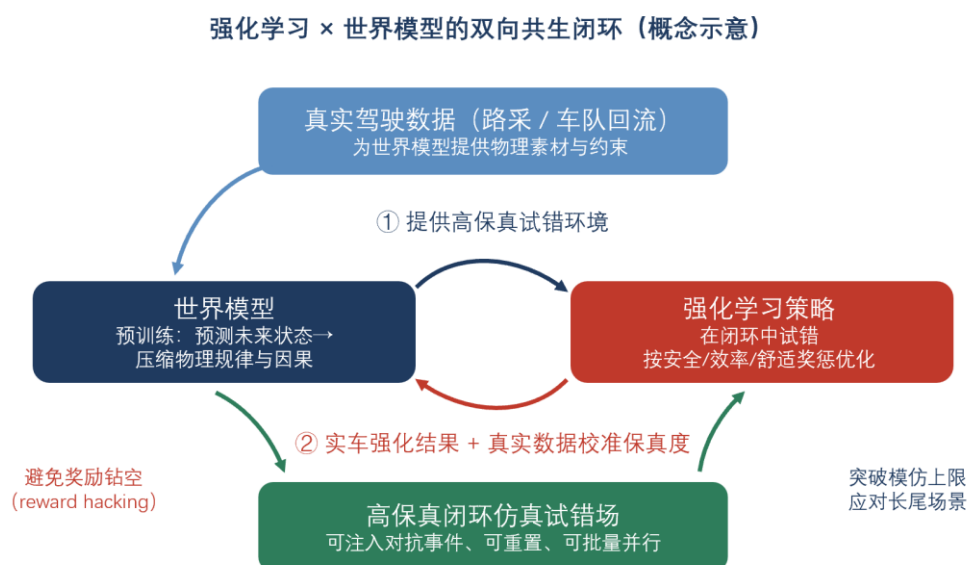
**华为乾崮的世界引擎亦承担类似训练场职能。**其 ADS4 采用世界引擎与世界行为模型的双层架构，云端世界引擎利用难例扩散模型生成高质量场景，难例场景密度可达真实世界的约 1000 倍，用以训练车端模型。华为相关负责人将 VLM 与 VLA 定位为能力增益项，同时强调能力提升核心仍在强化学习，即将端到端大模型置于由海量真实数据构建的仿真环境中探索。据 Momenta 公开表述，其 R7 世界模型将这一逻辑分为三层，即世界模型预训练、世界模型仿真、以及在世界模型中进行强化学习，并将前两层构建的虚拟世界作为强化学习训练场。

### 2.3. 真实数据与强化学习反过来校准世界模型的保真度

**世界模型的保真度依赖真实数据持续校准，由此形成双向共生。**理想汽车相关表述指出，重建式方法基于真实数据重建，符合物理约束但新视角泛化受限；生成式方法泛化能力较强，但物理约束不足。两者结合，才有望得到既符合物理规律又具备泛化能力的世界模型。这一结合本身依赖真实路采数据为重建提供素材、为生成提供约束，世界模型并非脱离真实数据的纯生成系统。

**Momenta 对世界模型的解释呼应了大语言模型的预测范式。**其首席执行官曹旭东在发布 R7 时表示，预测是智能进化的核心基石，大语言模型通过预测下一个词来压缩数字世界的常识，世界模型则通过预测物理世界未来状态来理解物体属性与因果关系。我们认为，这一表述指向世界模型与强化学习绑定的深层原因：世界模型先通过对真实数据的预测学习物理规律，再以此为基础充当强化学习环境；真实数据质量直接决定环境保真度，进而影响强化学习有效性。

图2：强化学习与世界模型的双向共生闭环（概念示意）



数据来源：Momenta、理想等公司公开表述，东吴证券研究所整合绘制

## 2.4. 生成式与重建式世界模型的结合是当前主流工程取向

世界模型作为基础设施的工程化趋势，在国际前沿实验室中同样较为明确。Wayve 自 2023 年起发布 GAIA 系列生成式世界模型，GAIA-1 为早期版本，GAIA-2 为多视角可控的潜在扩散世界模型，可生成包含突然加塞、行人异常行为、极端天气等安全关键场景的合成数据；2025 年 12 月发布的 GAIA-3 参数规模约 150 亿，定位转向自动驾驶系统的评估与验证。英伟达于 2025 年初发布 Cosmos 世界基础模型平台，面向自动驾驶与机器人等物理 AI 系统，Wayve 为其早期合作方之一。因此，无论量产车企还是研究机构，均在将高保真世界模型作为支撑数据生成、仿真训练与闭环评估的共用基础设施。

## 3. 厂商实践：端到端之上叠加强化学习后训练

代表性厂商的强化学习实践主要围绕量产辅助驾驶与 L4 Robotaxi 两类场景展开。

### 3.1. 量产乘用车辅助驾驶：端到端之上的强化学习后训练正在收敛为共识

特斯拉路线被多家厂商作为参照。FSD 自 v12 起强化端到端神经网络路线，后续公开技术分享进一步强调世界模拟器、车队回流难例和强化学习在训练链路中的作用，训练目标包括避免碰撞、减少急刹、保持合理车距等安全相关结果。Elluswamy 在 2025 年 ICCV 主旨演讲中介绍了一个学习型世界模拟器，可同时合成 8 路相机画面、注入行人或加塞等对抗事件，并重放历史失败案例用于训练与强化学习。

小鹏与理想是偏向视觉-语言-动作的量产路线代表。小鹏建设“云端模型工厂”，在

云端训练参数规模约 720 亿的世界基座模型，并通过后训练、强化学习和模型蒸馏生产车端模型 XVLA 后部署上车；用于训练的视频数据已达 2000 万 clips 并计划增至 2 亿，云端算力约 10EFLOPS。理想的 MindVLA 以“重建+生成”世界模型作为强化学习训练场，下一代 MindVLA-o1 进一步将世界模型、数据引擎、VLA 模型与强化学习基础设施构成闭环，并称将显著降低渲染成本与训练成本。两者均采用“云端训练大模型、蒸馏部署至车端”的路径，强化学习主要位于后训练环节。

**Momenta 与地平线代表相对弱化语言中介的量产路线。**Momenta 的一段式端到端于 2024 年落地，2025 年在端到端基础上引入强化学习，推出 R6 飞轮大模型；公开披露显示，该模型基于约 30 亿公里实战里程提炼出的约 7000 万条数据，在仿真环境中通过奖惩机制开展闭环训练，首搭于别克至境 L7，并被称为国内首个在端到端基础上量产落地的强化学习大模型。地平线 HSD 采用“一段式端到端+强化学习”架构，并以“端到端世界模型+交互博弈”为核心，通过仿真平台中的强化学习提升长尾场景处理能力，首搭于星途星纪元 E05。

**千里科技、卓驭与元戎启行呈现更明显的路线分化。**千里科技与阶跃星辰联合研发千里智驾 RLM 强化学习多模态大模型，分为基础大模型训练、智驾数据特训、强化学习三段，强化学习阶段通过设置奖励模型降低对数据分布的强依赖、抑制不良驾驶行为、降低兜底规则接管频率；其与吉利联合的千里浩瀚 G-ASD 引入基于 VLA 框架的世界行为模型，公开披露算力约 23.5EFLOPS。千里科技董事长印奇对相关名词持审慎态度，认为 VLA 与世界模型即便在大模型领域也尚未成熟，在智驾中的应用成熟度仍需验证。卓驭科技将强化学习定位为把系统从约 80 分提升至 95 分的环节。

**华为乾崮路线具有一定独立性。**ADS4 的 WEWA 架构由云端世界引擎与车端世界行为模型组成，世界引擎以难例扩散模型生成高密度难例场景，实现“用 AI 训练 AI”；车端世界行为模型融合全模态感知与混合专家能力。华为相关负责人将 VLA 定位为能力增益项，并把强化学习与世界引擎中的探索作为能力提升核心。ADS4 云端 AI 算力约 45EFLOPS，高质量训练数据约 10 亿公里。我们认为，华为与小鹏、理想的差异主要在于是否以语言为中介，而非是否使用强化学习与世界模型。

### 3.2. L4 Robotaxi: 以世界模型驱动的闭环仿真，安全冗余优先

**Waymo 公开口径侧重基础模型与大规模仿真。**其 2024 年发布的端到端多模态研究模型 EMMA 基于 Gemini，将导航指令、车辆状态等非传感输入与轨迹、三维位置等输出统一表示为自然语言文本，在运动规划等任务上取得领先结果；Waymo 公开口径将其视为探索多模态基础模型在自动驾驶中的应用，并强调大语言模型与视觉语言模型的运用。

**小马智行对“超越而非复刻人类”的判断表述较为明确。**第七代 Robotaxi 方案采用多传感器冗余配置，并以影子模式持续优化算法；PonyWorld 世界模型基于强化学习范式，

具备生成场景与传感器数据、高保真仿真、驾驶行为多维评估三项能力；据公开口径，其世界模型可按周生成超百亿级别的仿真测试里程与高风险场景变体，具体单位与统计口径需以公司披露为准。

**文远知行公开口径侧重端到端与大规模仿真。**文远知行自研端到端 AI 模型突破传统模块化框架，并以神经网络与博弈论实现类人驾驶行为规划，嵌入 WeRide One 平台形成持续学习与优化的飞轮效应；其云端模拟器每日可运行数万公里仿真里程。文远知行与博世合作推出一段式端到端方案 WePilot AiDrive 并计划量产。

### 3.3. 不同路线的共性与分歧

无论量产车企还是 Robotaxi 厂商，代表性厂商均在不同程度上围绕“端到端模型、世界模型仿真、强化学习/仿真训练”构建提升长尾安全性的技术组合；其中，明确披露强化学习的厂商普遍需要以世界模型或仿真环境承载闭环训练。分歧主要体现在三个方面：一是是否以语言为中介，小鹏、理想、元戎、千里浩瀚 G-ASD 偏向 VLA，华为、地平线、Momenta 相对弱化语言中介；二是强化学习披露的明确程度，特斯拉、小鹏、理想、Momenta、地平线、千里、小马已明确披露强化学习环节，Waymo 与文远知行的公开口径相对侧重端到端与仿真；三是落地形态，量产辅助驾驶更强调车端部署与成本约束，L4 Robotaxi 更强调安全冗余与超越人类。

表2: 代表性厂商“端到端+世界模型+强化学习”技术布局梳理

| 厂商      | 定位              | 主要技术标签                  | 强化学习的角色                    | 世界模型/仿真环境                  | 落地状态                     |
|---------|-----------------|-------------------------|----------------------------|----------------------------|--------------------------|
| 特斯拉     | 量产辅助驾驶/Robotaxi | 端到端神经网络, FSD V12→V14    | V14 在模仿外引入更激进 RL, 优化安全关键结果 | 学习型世界模拟器, 可合成 8 路相机、注入对抗事件 | FSD V14 已推送, Robotaxi 试点 |
| 小鹏      | 量产辅助驾驶          | 世界基座模型 72B → 车端 XVLA    | 云端后训练含 RL, 再蒸馏上车           | 在研世界模型作为基座训练场              | XVLA 量产路线                |
| 理想      | 量产辅助驾驶          | MindVLA/MindVLA-o1      | 重建+生成世界模型中做大规模闭环 RL        | 重建+生成统一世界模型, MindSim       | MindVLA 规划量产             |
| Momenta | 量产辅助驾驶供应商       | 一段式端到端→R6 飞轮→R7 世界模型    | R6 量产端到端+RL, R7 在世界模型中做 RL | R7 三层世界模型                  | R6 首搭别克至境 L7             |
| 地平线     | 芯片+量产方案         | HSD 一段式端到端+强化学习         | 交互式博弈强化学习, 仿真平台训练          | 端到端世界模型+交互博弈               | HSD 首搭星途星纪元 E05          |
| 千里科技    | 量产方案+大模型        | 千里智驾 RLM; 千里浩瀚 G-ASD    | 预训练+特训+RL 三段               | 仿真世界模型; G-ASD 世界行为模型       | 千里浩瀚多车型                  |
| 华为乾崮    | 量产方案 L2-L4      | WEWA: 世界引擎+世界行为模型       | 以仿真环境中 RL 探索为核心            | 云端世界引擎生成高密度难例              | ADS4 已推送                 |
| 卓驭      | 量产方案供应商         | 端到端+强化学习                | 把系统从 80 分提升至 95 分          | 公开披露有限                     | 多车型量产合作                  |
| 元戎启行    | 量产方案供应商         | 车端 VLA                  | 公开侧重 VLA, RL 细节有限          | 探索车外具身场景                   | VLA 方案推进量产               |
| Waymo   | L4Robotaxi      | Waymo 基础模型; EMMA 研究     | 公开口径侧重基础模型与仿真              | 大规模仿真, EMMA 基于 Gemini      | Robotaxi 已商业化            |
| 小马智行    | L4Robotaxi      | 第七代 Robotaxi; PonyWorld | 基于 RL 范式的 PonyWorld        | 生成场景、高保真仿真、行为评估            | 第七代量产、车队扩张               |
| 文远知行    | L4Robotaxi/ 通用  | WeRide One; 一段式端到端      | 公开侧重端到端、博弈论、飞轮             | 大规模云端仿真                    | Robotaxi 多城运营            |

数据来源: 特斯拉、小鹏、理想等公司官网、公告及公开报道, 东吴证券研究所整理

#### 4. 结论: 强化学习从可选增强走向关键环节

我们认为, 强化学习从可选增强走向关键环节, 核心源于模仿学习在物理闭环任务中的能力边界。第一, 模仿学习的能力上限在数字 AI 与物理 AI 之间存在结构性分野。大语言模型因模仿对象是互联网级人类知识而具备较高能力底座; 自动驾驶因模仿对象趋近平均人类司机, 且受协变量漂移制约, 模仿天花板更低、约束更强。这使强化学习在驾驶等高物理闭环任务中, 由可选的事后增强转向突破上限的关键环节。

**第二，在物理 AI 中，强化学习与世界模型的绑定更多源于工程约束。**真实道路难以承担反复试错的角色，世界模型被用作高保真闭环试错场以降低奖励钻空风险；世界模型的保真度又依赖真实数据与实车强化结果持续校准，二者由此形成双向共生。这一绑定已在国内量产车企与国际前沿实验室同时出现，世界模型正成为数据生成、仿真训练与闭环评估的共用基础设施。

**第三，产业落地呈现收敛中有分化的格局。**多数厂商正在围绕“端到端+世界模型仿真+强化学习/仿真训练”方向演进，但在是否以语言为中介、强化学习披露明确程度，以及面向量产辅助驾驶还是 L4 运营等方面仍存在分化。需要强调的是，相关概念的产业成熟度仍有争议，部分厂商对 VLA 与世界模型等名词持审慎态度；算力与数据规模多为公司口径，性能改善与量产时间多为前瞻信息，宜结合后续披露持续跟踪。上述判断为技术综述，不构成投资建议。

## 5. 风险提示：

**技术迭代不及预期的风险：**数字 AI、物理 AI 技术进展低于预期，任务或应用场景拓宽速度低于预期。

**大模型厂商 ARR 增速放缓的风险：**若客户增长、续费率或客单价不及预期，模型厂商 ARR 增速可能放缓。

**云服务商资本开支不及预期的风险：**若自由现金流承压，或 AI 算力需求及投资回报低于预期，云服务商可能下调资本开支。

**智能驾驶及机器人商业化落地不及预期的风险：**若产品可靠性、成本下降或用户需求低于预期，相关产品商业化进程可能放缓。

## 免责声明

东吴证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。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。

在法律许可的情况下，东吴证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，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。

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的，应当注明出处为东吴证券研究所，并注明本报告发布人和发布日期，提示使用本报告的风险，且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

## 东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的），北交所基准指数为北证 50 指数），具体如下：

公司投资评级：

买入：预期未来 6 个月个股涨跌幅相对基准在 15%以上；

增持：预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间；

中性：预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间；

减持：预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间；

卖出：预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级：

增持：预期未来 6 个月内，行业指数相对强于基准 5%以上；

中性：预期未来 6 个月内，行业指数相对基准-5%与 5%；

减持：预期未来 6 个月内，行业指数相对弱于基准 5%以上。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所  
苏州工业园区星阳街 5 号  
邮政编码：215021

传真：（0512）62938527

公司网址：<http://www.dwzq.com.cn>