下载日志已记录，仅供内部参考

推演世间万千变局，重塑智能底层范式

——世界模型行业专题研究

推荐|维持

相关研究报告

报告作者

分析师 耿军军
执业证书编号 S0020519070002
电话 021-51097188-1856
邮箱 gengjunjun@gyzq.com.cn

报告要点：

- **世界模型有望加速 AGI 的落地，长期发展空间广阔**

世界模型作为人工智能迈向通用人工智能 AGI 的重要方向，涉及到人工智能、心理学、计算机视觉、机器人、认知科学等多个学科，近年来受到广泛关注。世界模型的发展历史很短，定义还未完全统一。通常来说，世界模型是智能体通过数据驱动的方式学习得到的、对其所处环境的时空动力学规律、状态-动作因果转移关系、奖励机制与不确定性进行参数化建模的可计算模型。

- **VLA 与世界模型赋能智能驾驶，世界模型的长期发展值得期待**

从全球范围来看，智能驾驶技术逐步演进为两大核心路线共同发展的态势：第一条路线以“视觉-语言-动作”(Vision-Language-Action, VLA)链路为核心架构，第二条路线是以物理推演为核心驱动力的世界模型(World Model)路线。VLA 和世界模型各具优势：VLA 适合快速迭代，同时兼容现有量产平台，短期内易于落地；世界模型则代表了更底层的认知方式，强调物理规律和空间理解力，适合长期演进。世界模型通过统一时空表征、环境动力学模拟与反事实推演能力，打通智能驾驶感知、预测、决策、规划、控制、仿真及数据闭环全流程，推动智能驾驶从被动式响应向主动预判式智能升级，长期空间广阔。

- **世界模型赋能具身智能，助力产业规模化落地**

从产业实践来看，具身智能的算法架构正经历从视觉-语言-动作(Vision-Language-Action, VLA)模型向世界-动作模型(World-Action Model, WAM)的范式跃迁。2026 年初发布的 NVIDIA Cosmos Policy，进一步验证了 WAM 范式替代传统 VLA 模型的技术可行性。世界模型通过构建可交互的物理环境仿真器，使智能体具备预测未来状态、评估动作后果的能力，为具身智能的认知升级提供了核心技术支撑。

- **各大机构纷纷投入世界模型，技术路线多样化**

目前，世界模型在学术界和产业界热度较高，各大机构纷纷投入，根据自己的理解和需求，开拓了不同的技术路线，呈现百花齐放的态势。

WorldIncept：在 IROS 2026 WorldArena 2.0 Challenge 取得第一。

极佳视界 GigaWorld-1：专为具身智能设计的动作控制世界模型。

小鹏 X-World：被大量用于环境仿真与模型评估。

Xiaomi Auto World Model：三大核心场景完成业务落地。

腾讯混元 3D 世界模型 2.0：提出新一代的多模态世界模型架构。

英伟达 Cosmos 3：面向物理 AI 的开放世界基础模型。

谷歌 Genie 3：能够生成前所未有的多样化交互式环境。

杨立昆(Yann LeCun)AMI Labs：技术路线基于 JEPA，核心思路是在抽象的表征空间中预测事物的未来状态。

李飞飞 World Labs：Atlas 最核心的设计是空间上下文。

- **风险提示**

世界模型商业化进展低于预期的风险；世界模型技术迭代较快，研发投入不足导致竞争力下降的风险；人才流失风险。

下载日志已记录，仅供内部参考

目 录

1. 世界模型有望加速 AGI 落地 4
1.1 世界模型发展概况 4
1.2 世界模型与智能驾驶 6
1.3 世界模型与具身智能 8
2. 世界模型研究与产业进展 12
2.1 视启未来 WorldIncept 12
2.2 极佳视界 GigaWorld-1 13
2.3 小鹏 X-World 14
2.4 小米汽车 Xiaomi Auto World Model 16
2.5 腾讯混元 3D 世界模型 2.0 17
2.6 英伟达 Cosmos 3 18
2.7 谷歌 Genie 3 19
2.8 杨立昆(Yann LeCun)AMI Labs 20
2.9 李飞飞 World Labs 21
2.10 字节跳动 21
3. 风险提示 23

图表目录

图 1：世界模型发展概况 4
图 2：全球世界模型的主流技术架构流派 5
图 3：世界模型赋能智能辅助驾驶演进历程 6
图 4：世界模型在智能驾驶中的应用场景 7
图 5：高阶智能辅助驾驶算法洞察 7
图 6：中国具身智能产业发展阶段 8
图 7：英伟达 Cosmos Policy 10
图 8：EVAC：全球首个基于机器人动作序列驱动的具身世界模型 10
图 9：WorldArena 2.0 排名 12
图 10：视启未来 WorldIncept 模型获奖证明（2026 年 9 月） 13
图 11：极佳视界与合作伙伴落地解决方案 14
图 12：基于世界模型的仿真测试 15
图 13：X-World 模型框架 15
图 14：WorldRec 架构 17
图 15：混元 3D 世界模型 2.0 核心技术架构 18
图 16：Cosmos 3：面向物理人工智能的通用世界模型 19
图 17：Google Genie 3 与类似产品对比 19
图 18：JEPA（联合嵌入预测架构） 20
图 19：Atlas 核心设计 21

下载日志已记录，仅供内部参考

表 1：大语言模型 LLM 与世界模型对比 5

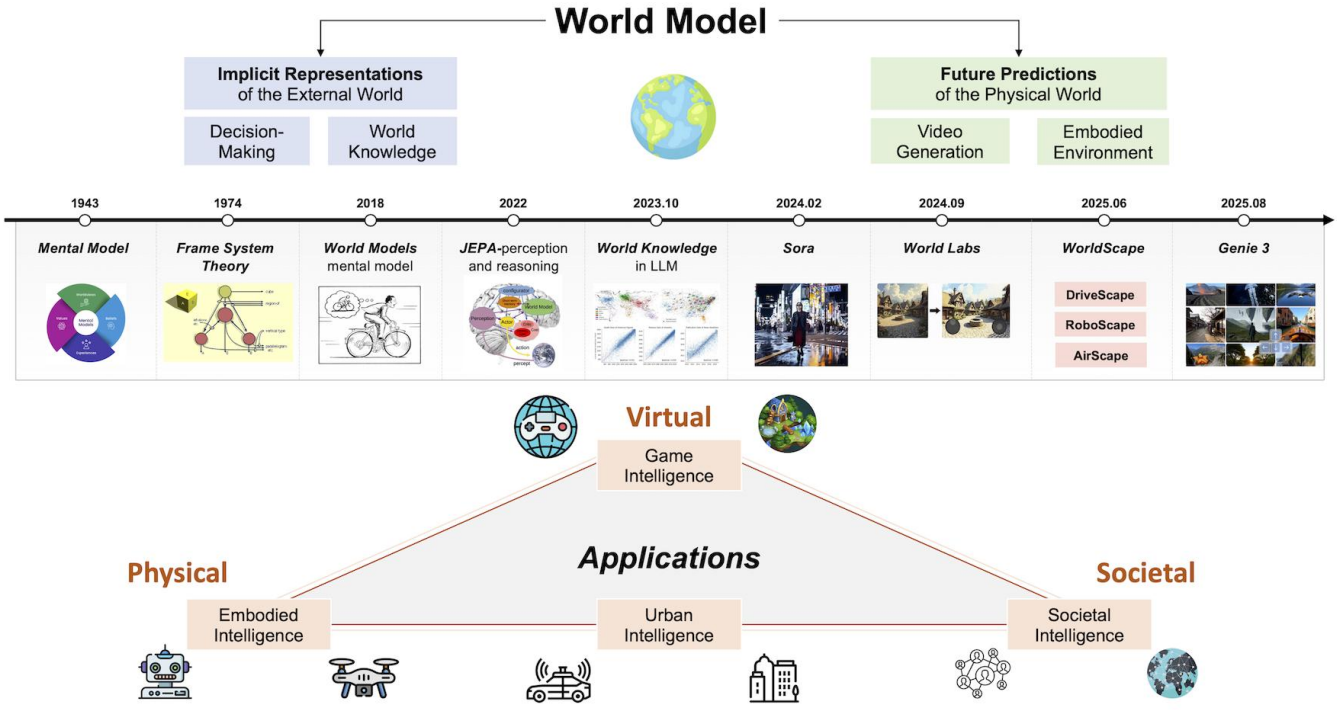
下载日志已记录，仅供内部参考

1. 世界模型有望加速 AGI 落地

1.1 世界模型发展概况

世界模型作为人工智能迈向通用人工智能 AGI 的重要方向，涉及人工智能、心理学、计算机视觉、机器人技术、认知科学等多个学科，近年来受到广泛关注，资本市场逐步成为推动世界模型行业发展的重要驱动力。世界模型的发展历史很短，定义还未完全统一。通常来说，世界模型是智能体通过数据驱动的方式学习得到的、对其所处环境的时空动力学规律、状态-动作因果转移关系、奖励机制与不确定性进行参数化建模的可计算模型。不同领域对世界模型的理解和应用方式虽各有差异，但都在推动这一概念的不断深化与拓展。从历史脉络来看，世界模型的思想最早可追溯到 20 世纪 60 年代 Minsky 提出的框架理论(Frame System Theory)，以及强化学习中用于建模环境动态的方法。进入深度学习时代，2018 年 Ha 等人提出的循环神经网络隐式模型，重新点燃了学界对世界模型的研究热情。近年来，随着大语言模型(LLM)、自监督学习（JEPA 架构）、视频生成与交互式三维环境等技术的快速发展，世界模型逐渐成为人工智能跨模态、跨任务的重要研究方向。

图 1：世界模型发展概况



资料来源：清华大学 Future Intelligence laB，国元证券研究所

随着技术的逐步成熟和成本持续下降，世界模型有望从汽车、具身智能机器人等高端领域下沉至更多细分场景，市场规模有望迎来爆发式增长。根据工信头条的数据，预计 2030 年全球世界模型产业规模将突破 3000 亿美元，中国市场占比有望超过 40%。其中，基础设施层将成为增长核心：高精度数据采集领域，高精地图、空间扫描需求将持续爆发；算力服务领域，针对世界模型的专用云服务将不断涌现，降低中小企业

下载日志已记录，仅供内部参考

使用门槛；工具链领域，适配中小企业的轻量化 SaaS 工具将快速普及，形成“基础功能免费、高级功能付费”的商业模式。

根据清华大学 Future Intelligence laB 的研究：世界模型主要承担两大核心功能：理解外部世界的机制(internal representation)；预测未来状态并辅助决策(future prediction)。从模型类别来看，主要包括：

- **隐式表示类世界模型**：代表性方法包括基于模型的强化学习(MBRL)、侧重于存储世界知识的大语言模型和多模态模型，强调通过构建内在表示来理解外部世界的机制；
- **预测类世界模型**：典型方向包括视频生成与交互式三维环境建模，更侧重于对未来状态的模拟与推演。

世界模型与大语言模型 LLM 有本质的区别。

表 1：大语言模型 LLM 与世界模型对比

维度	大语言模型 LLM	世界模型
本质	文字生成器	空间逻辑生成器
目标	预测下一个 token	预测下一帧物理世界
理解	语言模式	重力、因果、空间结构
应用	写作、对话、代码	机器人、自动驾驶、手术

资料来源：科技咖啡馆 TECHCAFE 微信公众号，国元证券研究所

目前，世界模型的创新持续推进，形成了几套不同的技术架构，各有特色，满足不同场景多样化的应用需求。

图 2：全球世界模型的主流技术架构流派



资料来源：毕马威《具身智能行业跨越奇点：世界模型如何重塑具身智能产业与商业新范式》，国元证券研究所

目前，世界模型的主要应用场景包括：

下载日志已记录，仅供内部参考

- **智能驾驶**：通过对交通环境与驾驶行为的建模，显著提升车辆在复杂环境下的感知与预测能力；
- **具身智能**：借助世界模型，支持各类型的机器人在真实环境中实现更高效的规划与操作；
- **社会模拟**：利用世界模型模拟群体行为和社会动态，助力理解复杂社会系统及其决策过程。

1.2 世界模型与智能驾驶

根据亿欧智库的研究：世界模型通过统一时空表征、环境动力学模拟与反事实推演能力，打通智能驾驶感知、预测、决策、规划、控制、仿真及数据闭环全流程，打破传统模块化架构信息割裂、误差累积的瓶颈，推动智能驾驶从被动式响应向主动预判式智能升级。

图 3：世界模型赋能智能辅助驾驶演进历程



资料来源：亿欧智库，国元证券研究所

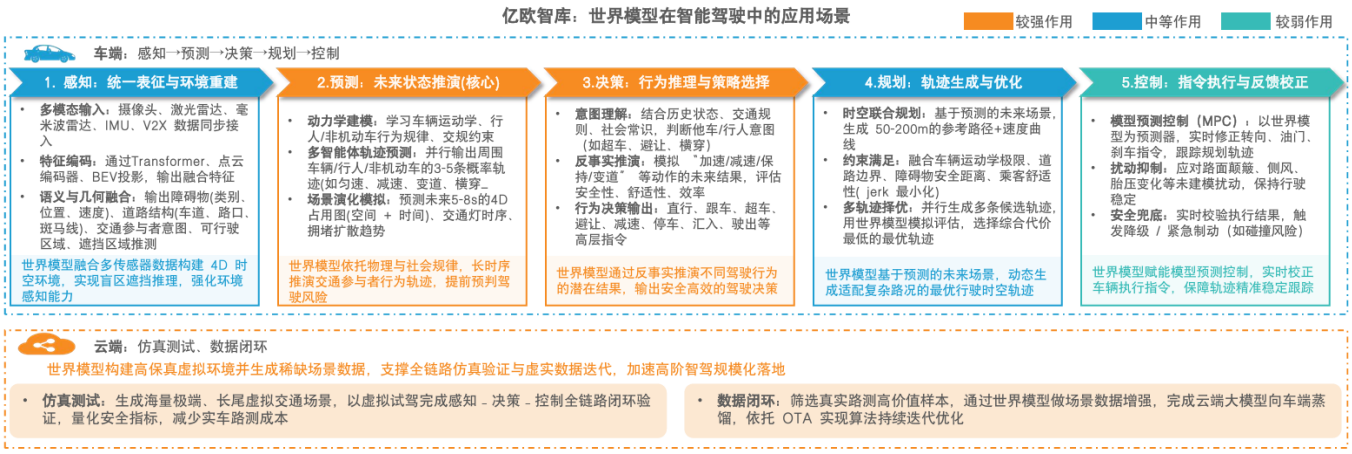
从技术角度来看，世界模型在车端和云端将发挥不同的作用，推动智能驾驶技术快速进步。在车端，世界模型能够实现交通参与者长时序行为推演与类人化博弈决策，补齐长尾风险预判短板；同时赋能感知盲区推理、复杂场景规划与精准控制，对预测和决策的影响最大。在云端，世界模型依托虚拟场景生成与虚实数据迭代，大幅降低实车测试成本，成为推动高阶智能驾驶规模化落地的关键技术。

从产业角度来看，世界模型重构了智能驾驶技术的迭代范式，可以有效提升行驶的安全性、驾乘的舒适性与复杂场景的通行能力，尽管面临算力、实时性等诸多挑战，但仍是目前实现 L3 及以上高阶智能驾驶、推动技术从感知智能迈向认知智能的核心突破口。

根据亿欧智库的研究，世界模型将在：感知、预测、决策、规划、控制等环节发挥不同的作用。

下载日志已记录，仅供内部参考

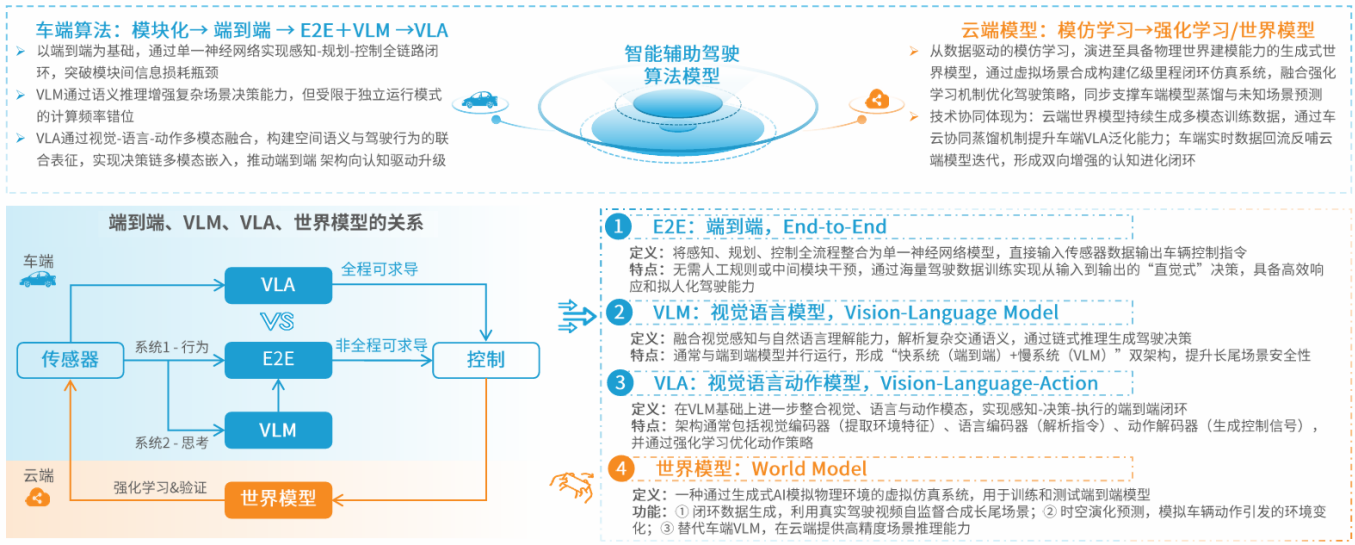
图 4：世界模型在智能驾驶中的应用场景



资料来源：亿欧智库，国元证券研究所

从全球范围来看，智能驾驶技术逐步演进为两大核心路线并行发展的态势：第一条路线为以“视觉-语言-动作”(Vision-Language-Action, VLA)链路为核心的架构体系，第二条路线则是以物理推演为核心驱动力的世界模型(World Model)路线。VLA 和世界模型各有优势：VLA 适合快速迭代、兼容现有量产平台，短期内易于落地；世界模型则代表了更底层的认知方式，强调物理规律和空间理解力，适合长期演进。

图 5：高阶智能辅助驾驶算法洞察



资料来源：亿欧智库，国元证券研究所

- VLA：VLA 通过“视觉-语言-动作”多模态融合，构建空间语义与驾驶行为的联合表征，实现决策链多模态嵌入，推动端到端架构向认知驱动升级。
- 世界模型：世界模型路线通过绕开语言转化环节，直接将 3D 高斯表征、点云等空间感知数据输入大模型，在模型空间内完成物理规律推演，并直接输出车辆控制指令，形成“云端世界引擎+车端世界行为模型”的技术架构-云端世界引擎负

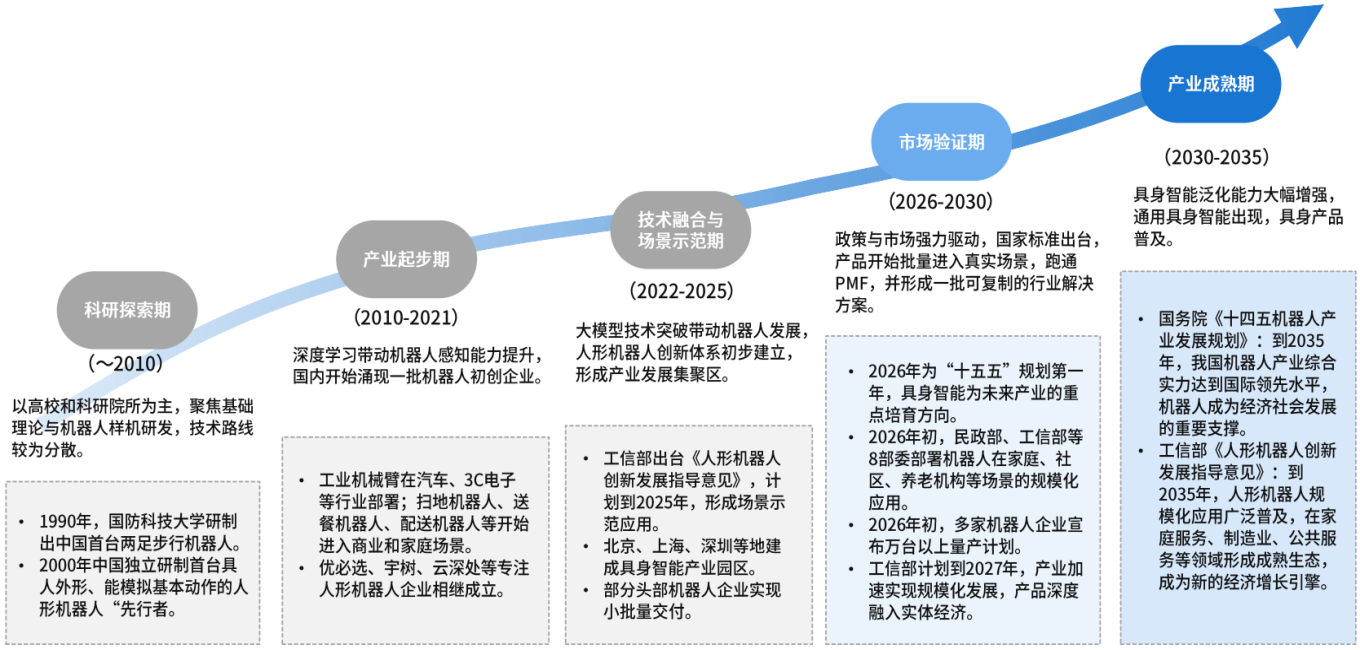
下载日志已记录，仅供内部参考

责大规模场景的物理规律建模与预训练，车端世界行为模型则基于实时感知数据进行快速决策。

1.3 世界模型与具身智能

根据中国人工智能学会《中国人工智能系列白皮书-具身智能(2026)》：具身智能作为人工智能领域的重要研究方向，专注于物理本体与真实外界环境的互动并实现智能行为的理论与技术研究，涵盖了环境感知、记忆与推理、对话交互、自主学习、决策规划、动作执行等多项复杂技术，最终在真实的物理世界中展示出类人的智能行为。相对于静态的、离身的人工智能，具身智能具有涉身性、情境性、主动性和交互性等多种特点。具身智能的思想和研究跨越了多个学科，主要包括：哲学、认知与神经科学、计算机科学、机器人学等，体现了多学科交叉的特性。

图 6：中国具身智能产业发展阶段



资料来源：易观分析，国元证券研究所

根据毕马威的研究，目前具身智能产业化面临四大核心瓶颈：

- **数据瓶颈：真实机器人数据难以支撑规模化训练**
  - 真实机器人数据采集成本高、场景覆盖窄，难以穷尽不同物体、动作、环境和机器人形态。
  - 开放世界视频、人类行为与机器人数据虽分别提供了物理规律、任务逻辑和动作锚定，但三者天然不对齐。例如同样是“拿杯子”，视频呈现的是画面，人类数据记录的是关节配合，而机器人数据则是具体的机械臂角度与力度。
- **任务瓶颈：短动作能力难以自然延展为长链路任务能力**
  - 当前机器人已能完成抓取、放置、开门等相对明确的短动作，但真实产业场景更多是连续、多步骤、强上下文依赖的长链路任务，例如清洁房间、仓库

下载日志已记录，仅供内部参考

分拣和工厂装配等。

- 长任务要求模型持续记住物体位置、遮挡状态、任务进度和前序动作带来的后续影响。一旦缺乏长期状态保持能力，机器人容易在多步骤执行中出现状态漂移、动作断裂或任务遗忘。

- **控制瓶颈：现实世界的理解难以稳定转化为可执行动作**
  - 视觉模型可以识别环境，生成式模型可以预测未来画面，但机器人真正需要的是判断“如果我现在这样行动，世界会如何变化”。
  - 开放世界视频通常缺少动作条件，人类示范难以直接映射到机器人执行机构，机器人交互数据又过于稀缺和狭窄，导致模型即使能够理解真实场景，也未必能够形成稳定的感知-动作映射。

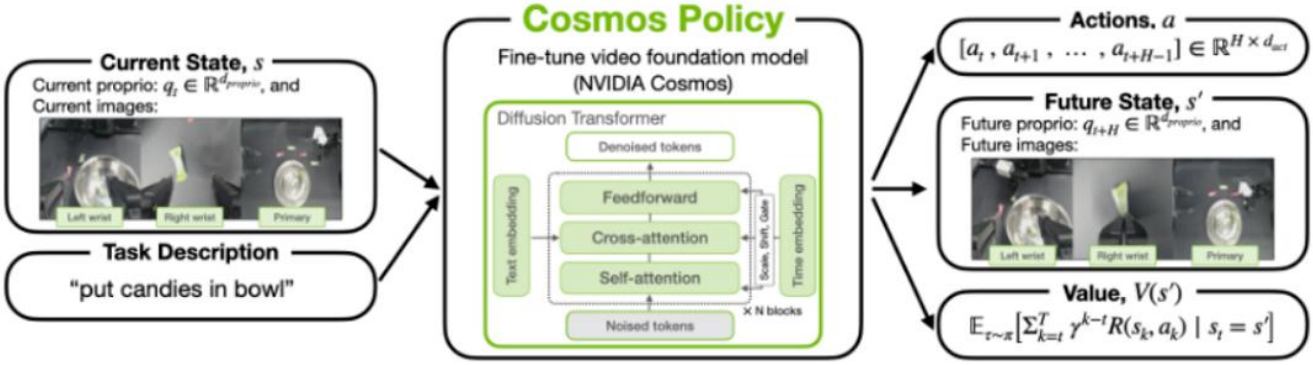
- **部署瓶颈：离线模型能力难以直接进入真实闭环系统**
  - 机器人行动具有强实时性，抓取、避障、移动和协作都需要低延迟反馈。如果模型推理速度慢、显存占用高、通信成本大，就难以嵌入真实机器人系统，也无法在观察-行动-反馈过程中持续修正策略。
  - 具身智能要从实验室演示走向产业化部署，需要世界模型具备可运行、可部署、可闭环的工程能力，而不仅是离线评测中的生成或预测能力。

根据中国人工智能学会《中国人工智能系列白皮书-具身智能（2026 版）》，具身智能的算法架构正经历从视觉-语言-动作(Vision-Language-Action, VLA)模型向世界-动作模型(World-Action Model, WAM)的范式跃迁。2026 年初以来，这一趋势愈发显著：虽然 VLA 模型仍是当前主流技术路线，以 OpenPI、OpenVLA 为代表的系列工作表现稳定，但其固有架构在动态环境适应与长程规划方面已暴露出结构性瓶颈。世界模型通过构建可交互的物理环境仿真器，使智能体具备预测未来状态、评估动作后果的能力，为具身智能的认知升级提供了核心技术支撑。

2026 年初发布的 NVIDIA Cosmos Policy，进一步验证了 WAM 范式替代传统 VLA 模型的技术可行性，推动 WAM 逐步成为学术界与产业界的共识性技术路线。Cosmos Policy 的核心思想是将视频扩散模型从“生成视频”转变为“生成动作”，即利用其对时序变化的建模能力，直接输出机器人控制信号。

下载日志已记录，仅供内部参考

图 7：英伟达 Cosmos Policy

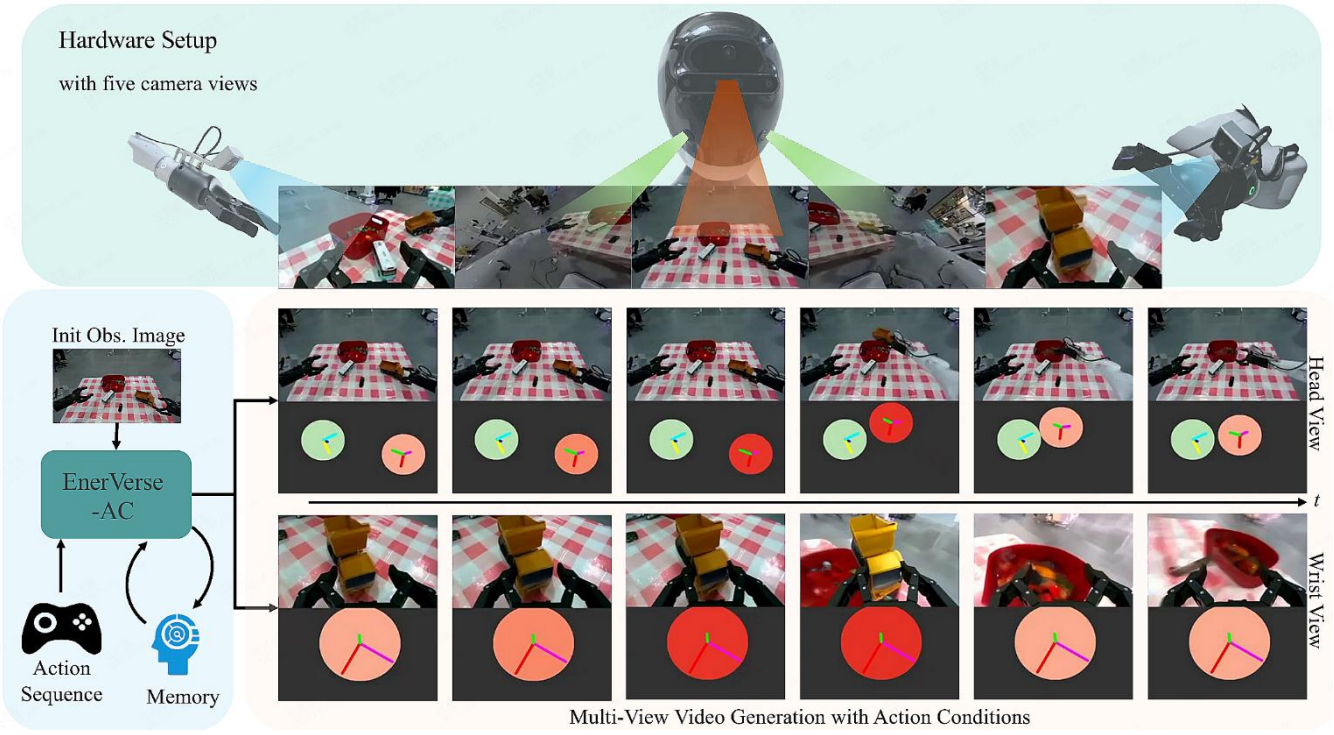


我们提出了 Cosmos Policy，这是一种基于 NVIDIA Cosmos-Predict2-2B 视频基础模型进行微调的先进机器人策略。Cosmos Policy 能够处理多模态输入和多视角摄像头图像，并预测：（1）机器人动作片段；（2）未来状态（由机器人本体感觉和图像观测结果表示）；以及（3）价值（未来状态下的预期奖励）。基础视频模型架构未做任何更改，所有模态均通过视频扩散学习目标进行联合建模。

资料来源：PnP 机器人微信公众号，国元证券研究所

2025 年 5 月，智元机器人发布具身智能领域双重里程碑式突破：全球首个基于机器人动作序列驱动的具身世界模型 EVAC (EnerVerse-AC)、具身世界模型评测基准 EWMBench。这两大创新成果现已全面开源，旨在构建“低成本模拟-标准化评测-高效迭代”的全新开发范式。

图 8：EVAC：全球首个基于机器人动作序列驱动的具身世界模型



资料来源：智元 AGIBOT 微信公众号，国元证券研究所

下载日志已记录，仅供内部参考

EVAC 是一个能够动态复现机器人与环境复杂交互的世界模型，标志着从传统仿真到生成式模拟的跃迁。其核心能力是：从“物理执行”到“像素空间”的精准映射。EVAC 基于前序工作 EnerVerse 架构持续演进，创新引入多级动作条件注入机制，实现“物理动作-视觉动态”的端到端生成。

目前来看，VLA 向 WAM 的跃迁本质是具身智能从“模仿人类指令”到“理解物理因果”的认知升级。世界模型的实时交互生成能力打破了仿真与现实的边界，为闭环规划与长程决策提供了统一框架。在下一阶段，WAM 将逐渐成为具身智能系统的核心组件。

下载日志已记录，仅供内部参考

2. 世界模型研究与产业进展

目前，世界模型的研究加速推进，全球各大机构纷纷推出自己的模型，WorldArena 2.0 提供了该领域较为权威的评测基准，排名也在快速更新。

图 9：WorldArena 2.0 排名

Model	EWMScore-P (Difficulty/OOD Adjusted)	Rank	Visual Quality	Motion Quality	Content Consistency	Physics Adherence	3D Accuracy	Controllability
WorldIncept	72.33	1	74.42	31.42	77.05	78.12	99.24	90.76
BWM-Turbo	71.49	2	68.05	32.18	80.96	75.91	98.48	90.02
DWM	71.15	3	80.94	30.98	74.86	67.26	98.22	87.98
Waterloo	71.05	4	67.80	30.57	80.30	73.48	98.96	92.42
ActiFlow	70.83	5	68.29	32.02	79.30	73.95	98.41	89.43
PrismWorld	70.70	6	81.80	32.37	69.44	67.09	98.50	89.27
sonata-wm	70.62	7	68.86	29.90	81.07	70.62	98.83	90.47
Concerto-AWM	70.56	8	68.93	29.88	81.13	70.47	98.74	90.12
Snowl-WM	70.55	9	68.12	30.58	80.57	72.24	98.65	89.35
K1-World	70.48	10	81.84	32.28	69.82	65.57	98.34	88.84
Nora-WM	70.46	11	68.94	29.94	81.03	69.83	98.72	90.04
Kairo-WM	70.39	12	68.13	30.56	80.49	71.56	98.58	89.04
Scope-WM	70.39	13	68.14	31.39	79.22	71.77	98.72	89.33
Hera-WM	70.29	14	68.60	30.89	79.11	71.81	98.54	88.94
NexWorld	70.28	15	81.45	32.36	69.25	66.08	98.16	88.25

资料来源：WorldArena 2.0，国元证券研究所

注：排名日期：2026-10-14

WorldArena 2.0: Extending Embodied World Model Benchmarking on Modality, Functionality and Platform

2.1 视启未来 WorldIncept

WorldIncept 是视启未来面向真实世界理解、预测与交互研发的视觉原生世界模型。2026 年 9 月，在仿真视频生成这一直接考验世界模型“能否构建可信视觉世界”的核心场景中，WorldIncept 已具备领先的综合表现。尤其是在经过测试难度与分布外(OOD)情况调整后的 EWMScore-P 评估中仍取得最高分，说明模型不仅能够生成高质量画面，也展现出较强的复杂场景适应能力和跨分布泛化能力。

WorldArena 2.0 是 IROS 2026 WorldArena 2.0 Challenge 的官方评测基准，覆盖视觉质量、运动质量、内容一致性、物理遵循性、3D 准确性和可控性等多个维度。Track 1 的领先成绩，是 WorldIncept 在当前视觉世界建模能力的一次集中验证，也为其进一步服务于具身智能、仿真训练和机器人学习等应用奠定了基础。

下载日志已记录，仅供内部参考

图 10：视启未来 WorldIncept 模型获奖证明（2026 年 9 月）



资料来源：IDEA 数字经济研究院微信公众号，国元证券研究所

视启未来（深圳）科技有限公司是由粤港澳大湾区数字经济研究院（IDEA 研究院）核心孵化的人工智能科技企业，成立后完成近亿元天使轮融资，是国内聚焦具身智能通用视觉大模型研发与产业化的标杆企业。公司由 IEEE Fellow 张磊博士领衔创办，核心团队全程承接 IDEA 研究院 DINO-X 系列通用视觉大模型的核心技术、研发团队及全套知识产权，深耕智能视觉与物理 AI 融合赛道，核心定位为具身机器人、智能终端打造专属视觉中枢，破解传统机器视觉场景适配性弱、泛化能力不足的行业痛点。

2.2 极佳视界 GigaWorld-1

2026 年 3 月，极佳视界(GigaAI)发布专为具身智能设计的动作控制世界模型 GigaWorld-1。在当时的世界模型权威评测基准 WorldArena 中，以 62.34 的综合得分刷新世界纪录，并在 3D 准确度、物理遵循和视觉质量等核心维度上展现出领先优势。

从技术路线上看，GigaWorld-1 是一款专为具身场景打造的 AC-WM（Action-Conditioned World Model，动作控制世界模型）。相较于传统的世界模型，GigaWorld-1 深度继承并发展了极佳视界在 2025 年 7 月发布的 EmbodieDreamer 核心架构。该方案不仅引入了显式的动作建模机制，从根本上保证了视频生成过程中的几何一致性；更创新性地融入了可微分物理引擎，从而获取精准的机械臂物理参数，以实现

下载日志已记录，仅供内部参考

对复杂物理交互过程的真实模拟与严格遵循。在此前沿架构的基础上，极佳视界进一步引入了团队长期积累的上万小时高质量真实机器人操作视频数据进行训练，极大地增强了模型在开放场景下的泛化能力与高精度的动作遵循表现。

2026 年 4 月，极佳视界携手一汽模具、阿里云，完成了具身智能机器人在真实工业制造场景的全流程解决方案落地。这是国内首个以真实汽车工业制造场景为核心，围绕箱体拆垛、跨区域搬运、动态避障、精准操作等高频作业任务，构建了包含数据采集及标注、模型训练、应用部署、升级迭代的全流程方案。

图 11：极佳视界与合作伙伴落地解决方案



资料来源：极佳视界微信公众号，国元证券研究所

根据公司官方微信公众号，极佳视界 2026 年 4 月完成近 15 亿元 B1 轮融资，该轮融资由某知名科技巨头、多个顶级国家队基金、伊利集团 CVC 健瓴资本、普华资本、华福投资、毅达资本、新型工业化基金、盛景嘉成、图灵资管、凯洋资本、武汉高科、贵阳金投、山东产投等顶尖国资平台、产业资本、双币财务机构等共同投资，华控基金、华民投、一村资本、领阳投资等多个老股东持续超额加注。此前 2026 年 3 月，公司完成近 10 亿元 Pre-B 轮融资，1 个月内融资 25 亿元，估值超百亿元。

2.3 小鹏 X-World

2026 年 3 月，小鹏汽车正式发布世界模型 X-World 技术报告，从数据、模型、训练、验证及应用等多层面详解 X-World 的构建与使用。X-World 是一个基于视频扩散生成技术(Video Diffusion)构建的、服务于自动驾驶的可控多视角生成式世界模型，具备多视角下即时响应与持续生成的能力。

技术报告展示了 X-World 在小鹏汽车自动驾驶中的实际应用价值，该模型已投入到闭环仿真测试、在线强化学习、数据生成等实际生产环节。在向用户推送的第二代 VLA 的研发与验证过程中，X-World 也已被大量用于环境仿真与模型评估。

下载日志已记录，仅供内部参考

图 12：基于世界模型的仿真测试

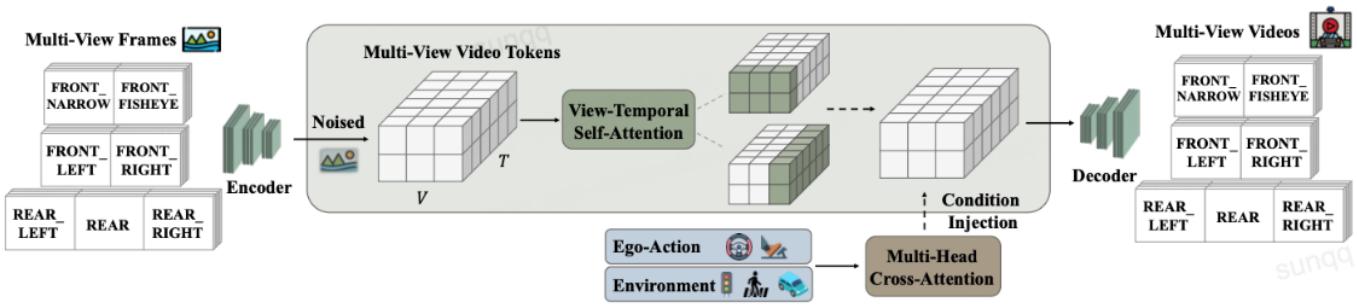


资料来源：小鹏汽车微信公众号，国元证券研究所

小鹏生成世界模型团队试图构建一个“现实世界模拟器”，让它能够在给定动作条件下生成符合物理约束的未来视频，同时在持续生成过程中保持良好的可控性与稳定性。在这一背景下，X-World 应运而生，通过输入多摄像头的历史视频流和待执行的驾驶动作（或动作序列），便可生成对应的未来多摄像头视频流。X-World 可以被视为一个会“思考”驾驶场景的物理 AI 系统，它能够根据当前的路况和驾驶操作，想象出数秒后路况的变化。

架构层面，X-World 基于当前领先的视频生成模型 WAN 2.2 构建，沿用其潜空间视频生成范式，将视频 VAE 与基于 DiT 的潜空间去噪器相结合。底层采用高压缩比 3D 因果自编码器(VAE)，大幅降低计算与内存开销，支持长时序视频建模，从而更好地捕捉丰富的时空依赖关系，同时降低延迟、加快推理速度。模型骨干为定制化的 DiT 骨干网络，通过视角-时间自注意力机制，实现时间维度和视角维度的联合建模，从而确保 7 路摄像头视角间的一致性。X-World 还提供了一套全面的条件控制接口，包括自车动作、动态交通参与者、静态道路元素（如车道线、道路边界）、摄像头内外参，可对驾驶场景的生成过程进行细粒度调控。这些设计共同实现了多输入条件下的可控多视角生成。

图 13：X-World 模型框架



资料来源：小鹏汽车微信公众号，国元证券研究所

下载日志已记录，仅供内部参考

2.4 小米汽车 Xiaomi Auto World Model

2026 年 5 月，小米汽车正式发布 Xiaomi Auto World Model 全新框架，为业界辅助驾驶世界模型提供了新的框架路径，推动行业从“场景感知”向“认知推演、场景进化”的高阶形态跃迁。这是小米首次将三维重建与视频生成深度耦合的一体化架构，以“重建锚定几何、生成填补想象”的新范式，打破行业长期采用的重建、生成独立拆分路线。在 Waymo、nuScenes 等主流基准测试中全面取得 SOTA，并已在小米汽车合成数据、仿真测试、智能座舱三大核心场景完成业务落地。

Xiaomi Auto World Model 提出一个全新的整合框架，将重建模块(WorldRec)与生成模块(WorldGen)深度耦合，让两者在结构上互相约束：

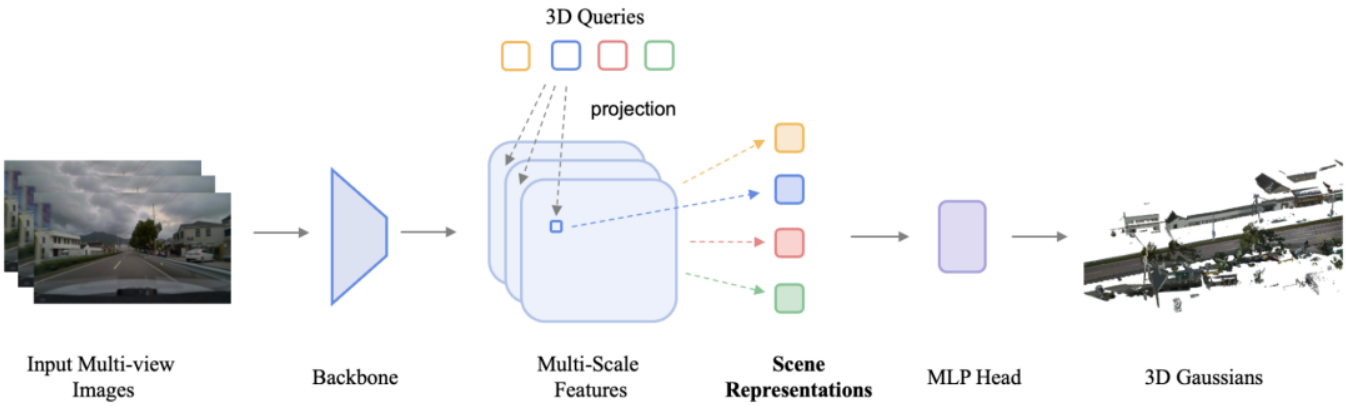
- **重建侧给生成“打地基”**：WorldRec 维护一个随观测增量扩展的 4D Gaussian 全局表示，把这个 3D 几何投影到自车视角后作为渲染先验喂给生成模型。这意味着生成模型在已观测区域不再“自由发挥”，而是被几何约束锁住“车道结构、建筑位置、相机”间的一致性都由重建来兜底，生成只负责补全光照、纹理和未观测区域。
- **生成侧给重建“扩边界”**：在重建覆盖不到的时空区域（未来帧、未观测视角、遮挡区），WorldGen 用生成能力把内容补出来，让整个世界模型不再受限于“开过的路”。
- **两者共同压制长时序漂移**：重建提供的确定性几何先验持续校正生成过程，从根源上抑制曝光偏差带来的累积误差，让一分钟级别的长视频生成依然保持稳定。

重建提供 3D 几何作为结构化锚点，约束生成过程的稳定性；生成则把预测能力延伸到观测之外，弥补重建的边界。两者形成闭环、互相增益，从三个关键维度实现了“1+1>2”的协同效应：

- **高稳定性**：WorldRec 的确定性几何约束，有效抑制长时序自回归中的误差累积与内容漂移。
- **高一致性**：4D 场景表征作为跨帧共享记忆，确保不同时刻、不同视角下场景内容全局一致。
- **高真实性**：WorldGen 以 WorldRec 渲染的 RGB 图像为几何骨架，使合成内容既符合物理布局，又贴近真实传感器观测，显著缩小了“仿真-现实”的领域鸿沟。

下载日志已记录，仅供内部参考

图 14：WorldRec 架构



资料来源：小米技术微信公众号，国元证券研究所

在重建领域，Xiaomi Auto World Model WorldRec 在公开的 Waymo 数据集上，全面领先此前 SOTA 方法 DGGT，PSNR（衡量重建精度的核心指标，越高越好）达到 28.48，超出 DGGT 约 1 个点。在 nuScenes 零样本泛化测试中，对比 SOTA 方法，依然保持领先，说明它不只在特定的数据集上表现好，对新场景也有很强的适应能力。在生成领域，Xiaomi Auto World Model 在 H20 GPU 上，单视角生成速度达到 0.19 秒/帧，三视角为 0.46 秒/帧，对比同为自回归方法的 Epona（1.06 秒/帧）快了 5.6 倍。并且，WorldGen 还支持高达 81 帧的连续生成（10Hz/30Hz，最长可达 1 分钟），而大多数公开基线模型仅能生成 8-16 帧。在权威的 nuScenes 数据集上，WorldGen 取得了 FVD 64.97 和 FID 7.04 的成绩，FVD 指标超越了所有对比的同类双向与自回归方法模型，保持了极具竞争力的 FID。

2.5 腾讯混元 3D 世界模型 2.0

2026 年 4 月 16 日，腾讯正式发布并开源混元 3D 世界模型 2.0(HY-World 2.0)。混元 3D 世界模型 2.0 是一个多模态世界模型，能够理解文字、图片、视频等不同类型输入，自动生成、重建和模拟 3D 世界，同时支持多格式 3D 资产（Mesh/3DGS/点云等）导出，可以与现有的游戏工作流无缝对接，用于快速生成游戏地图和关卡原型。

混元 3D 世界模型 2.0 让 AI 从“对话”、“画图”、“捏物体”，进化到了“造世界”-包含人、物、景的完整世界。不同于谷歌的 Genie 3 和混元世界模型 1.5 仅能生成视频文件，世界模型 2.0 的实用性更进一步：它可以直接生成可二次编辑的 3D 资产，这些生成的 3D 资产可以直接导入到游戏制作或者具身仿真引擎，实现真正可玩、可用。

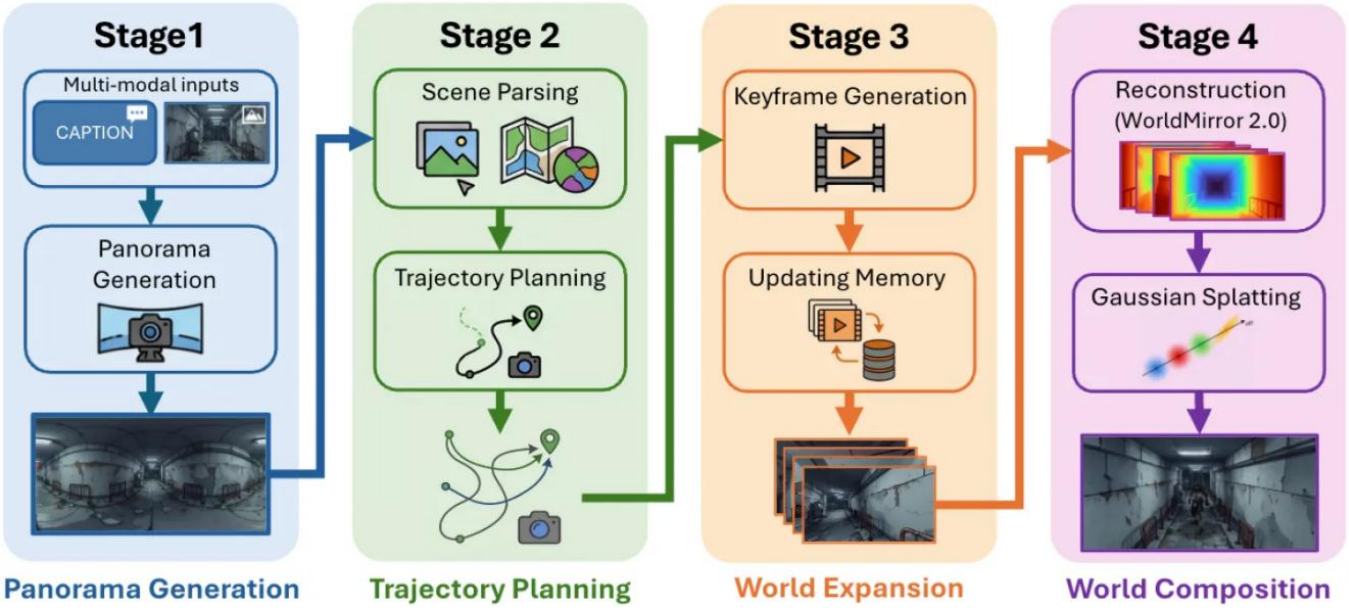
同时，混元 3D 世界模型 2.0 支持复刻真实 3D 场景，用户输入一段真实空间的视频或者多视角图片，模型就能构建出高精度的数字孪生空间。基于升级后的 WorldMirror 2.0 架构，模型支持任意尺寸图像与视频输入，一次性预测密集点云、多视角深度图、表面法线及相机参数，一次生成永久可复用。未来，室内装修预览、城市规划、文化遗产保护等场景，都可以基于这一能力实现快速构建和还原。

混元 3D 世界模型 2.0 提出了新一代的多模态世界模型架构：以 3D 为核心，整合空间理解、生成、与重建能力，可将多模态输入转化为三维世界：(1)通过全景图生成完

下载日志已记录，仅供内部参考

成世界初始化；(2)通过空间理解与规划得出探索路径；(3)通过基于记忆的世界扩展来扩充世界观测范围；(4)利用 3D 重建构建最终三维资产。

图 15：混元 3D 世界模型 2.0 核心技术架构



资料来源：混元世界模型官网，国元证券研究所

2.6 英伟达 Cosmos 3

2026 年 6 月，NVIDIA 推出 NVIDIA Cosmos™ 3，这是一款面向物理 AI 的开放世界基础模型，基于突破性的混合 Transformer (Mixture-of-Transformers)架构所构建，将视觉推理、世界生成和动作预测集成至单一系统中。Cosmos 3 是全球首款完全开放的全模态模型，能够以领先的物理精度原生理解并生成文本、图像、视频、环境音和动作，将物理 AI 的训练和评估周期从数月缩短至数天。NVIDIA 同时宣布成立 NVIDIA Cosmos Coalition，这是世界模型构建者与 AI 开发者的全球协作组织（成员包括 Agile Robots、Black Forest Labs、Generalist、LTX、Runway 和 Skild AI）旨在共同推动下一代世界模型发展。

NVIDIA 创始人兼首席执行官黄仁勋表示：“得益于多模态推理语言、视觉和世界模型领域的多项突破，物理 AI 爆发的时代已近在眼前。Cosmos 3 系列开放前沿全模态模型，助力开发者在构建机器人、智能汽车和视觉 AI 方面实现了代际飞跃，使其能够在物理世界中完成感知、推理、规划并采取行动。”

开发者可将 Cosmos 3 用作：

- 具备跨模态理解和推理的视觉语言模型。
- 可用于模拟物理环境并预测未来世界状态以进行训练和评估的世界模型或视频基础模型。
- 帮助训练机器人执行特定任务的世界动作模型的骨干网络。



下载日志已记录，仅供内部参考

图 16：Cosmos 3：面向物理人工智能的通用世界模型



资料来源：NVIDIA 官网，国元证券研究所

2.7 谷歌 Genie 3

2025 年 8 月，Google 正式发布 Genie 3，这是一款通用型世界模型，能够生成前所未有的多样化交互式环境。Genie 3 是首个支持实时交互的世界模型，用户只需输入文本提示，Genie 3 就能以每秒 24 帧的速度实时生成可供自由探索的动态世界，并在 720p 分辨率下保持数分钟的画面一致性。相较于前两代产品，它不仅能在 720p、24fps 的规格下即时生成可交互的三维场景，还把单次交互时长从几十秒延长到数分钟，并首次加入“可提示的世界事件”和“视觉记忆”等关键特性，使得同一房间里的涂鸦、家具位置等细节在玩家多次往返时依然保持一致，真实感大幅提升。

图 17：Google Genie 3 与类似产品对比

	GameNGen	Genie 2	Veo	Genie 3
Resolution	320p	360p	720p to 4K	720p
Domain	Game-specific	3D Environments	General	General
Control	Game-specific	Limited keyboard / mouse actions	Video-level description*	Navigation; Promptable world events
Interaction Horizon	A few seconds	10-20 seconds	8 seconds	Multiple minutes
Interaction Latency	Real time	Not real time	N/A	Real time

资料来源：智东西微信公众号，国元证券研究所

Genie 3 目前的实用场景包括：

- **虚拟训练场**：Genie 3 能在几秒内拼装出仓库、雪道等场景，为机器人或自动驾驶算法提供“无限刷题”的沙盒，比昂贵的真实试验场更灵活。

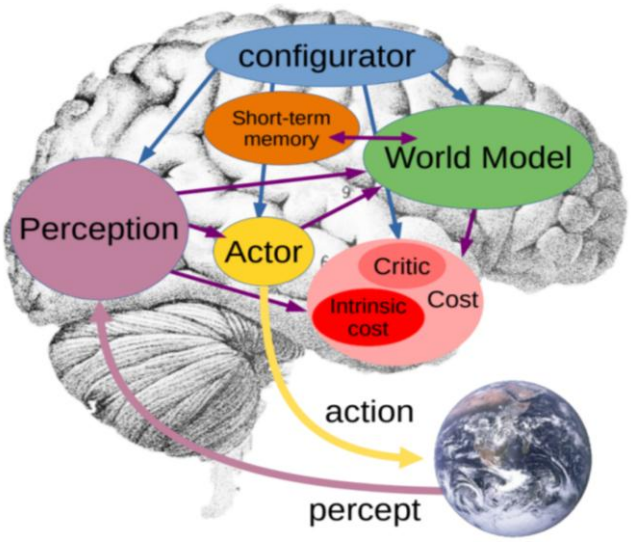
下载日志已记录，仅供内部参考

- **游戏与内容创作**：设计师只需一句话就能生成可游玩的关卡，降低独立游戏开发门槛；影视前期预演、数字孪生亦可受益。
- **AI 智能体研究**：更长、更稳定的交互回合让强化学习与规划算法有机会在同一世界中持续试错、形成长期策略，而不仅仅是“小试牛刀”的短局。

2.8 杨立昆(Yann LeCun)AMI Labs

图灵奖得主杨立昆认为“世界模型”是一个能够构建现实的因果模型，并能在此模型内部进行规划的系统。AMI Labs 的技术路线基于杨立昆提出的 JEPA（Joint Embedding Predictive Architecture，联合嵌入预测架构）。与当前主流大模型逐字预测下一个词不同，JEPA 的核心思路是在抽象的表征空间中预测事物的未来状态。它不预测下一个像素或下一个词，而是预测“接下来会发生什么”的整体模式。

图 18：JEPA（联合嵌入预测架构）



资料来源：智东西微信公众号，国元证券研究所

2026 年 10 月 5 日，杨立昆创立的 AMI 公司发布分层世界模型 H-JEPA，并开源全部代码与权重。模型堆叠多层 JEPA 架构，高层负责长远粗规划，低层细化动作，在表征空间预测未来状态，无需重建像素。在迷宫仿真任务中分层模型大幅提升规划成功率，计算开销更低，为智能体、机器人视觉规划提供可复现开源方案。

H-JEPA 的长时序视觉规划能力，天然适配需要与物理/虚拟环境交互的智能场景，核心落地方向集中在四大领域：

- **具身机器人**：机械臂柔性操作、室内外自主导航等场景中，机器人可以通过 H-JEPA 在表征空间快速规划运动路径，无需逐帧重建环境画面，响应速度与复杂环境鲁棒性显著提升。
- **虚拟智能体**：游戏 AI、数字人等虚拟角色可依托分层规划能力，完成长流程的任务交互，比如复杂关卡探索、多步骤人机协作任务。
- **自动驾驶**：用于行车场景的长时序轨迹预测与决策规划，提前预判多步交通态势，支撑高速、城区等复杂场景的安全决策。

下载日志已记录，仅供内部参考

- **工业自动化**：产线机器人的柔性作业调度、物流 AGV 的动态路径规划等，通过分层规划适配多变的工业生产场景。

2.9 李飞飞 World Labs

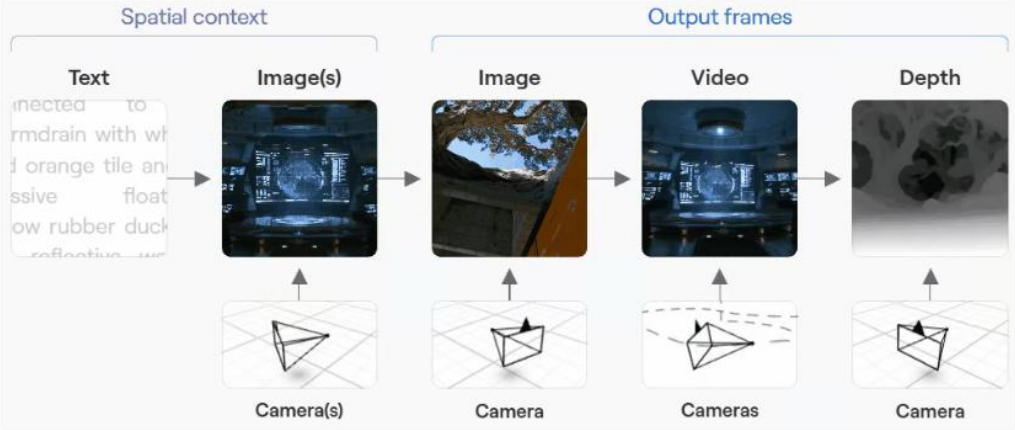
2026 年 9 月，李飞飞的创业公司 World Labs 发布新一代世界模型 Atlas。World Labs 把 Atlas 称作 omni model（全模态模型），从零开始预训练，原生处理文本、图像、视频和 3D 四种数据。官方𝕏上的表述是“世界上第一个能以像素级相机控制生成图像与视频帧、并将其重建为 3D 的多模态世界模型”。

Atlas 的四大能力包括：

- **相机可控生成**：从一到六张参考图出发，按用户指定的机位和角度生成新视角，最高输出 1440p、时长最长 1 分钟的视频。
- **空间重建**：从一张到几十张输入图重建真实场景，既输出新视角的图像帧，也输出点云和 3D 高斯泼溅这样的显式 3D 结果。
- **时空模拟**：从输入视频中同时建模空间与时间，用于视频重构图和机器人的 Real-to-Sim 流程。
- **图像生成**：包括文生图和 360 度全景图。

Atlas 最核心的设计是空间上下文(spatial context)。

图 19：Atlas 核心设计



Atlas 是一种多模态自回归扩散 Transformer。其输入在三维空间中进行 grounded，以形成空间上下文，并基于该上下文生成多模态输出。

资料来源：机器之心微信公众号，国元证券研究所

2026 年 9 月，AMD 正式官宣：已达成最终协议，收购李飞飞领导的 World Labs。交易完成后，李飞飞将出任 AMD 执行副总裁兼首席科学家，直接向 AMD 董事长兼 CEO 苏姿丰汇报。

2.10 字节跳动

根据网易科技微信公众号，字节跳动想用世界模型，把视频变成一个能“走进去”的

下载日志已记录，仅供内部参考

世界。9 月 7 日，彭博社援引知情人士称，字节跳动正在基于 Seedance 开发实时空间视频生成模型，报道称该项目涉及跨部门及算力资源协调，最快可能于 10 月推出，计划仍可能调整。它瞄准的不只是一段生成好的视频，而是能随用户声音和动作变化的虚拟环境，应用方向包括直播、短剧和游戏，并计划服务 Pico 头显。

下载日志已记录，仅供内部参考

3. 风险提示

1) **世界模型商业化进展低于预期的风险**：目前，世界模型仍处于早期研发阶段，受技术瓶颈、算力成本、场景验证难度等因素的影响，行业规模化落地、商业闭环构建进度存在较大不确定性，收入与利润的兑现可能慢于预期；

2) **世界模型技术迭代较快，研发投入不足导致竞争力下降的风险**：世界模型技术迭代迅速，技术路线持续快速演进；若研发资金、核心人才等投入不及预期，世界模型迭代速度滞后，技术优势可能削弱，产品市场竞争力可能下滑；

3) **人才流失风险**：世界模型是人工智能的前沿领域，行业仍处于早期探索阶段，兼具多模态、物理仿真、机器人虚实迁移等跨学科属性，兼具理论与工程落地能力的复合型高端人才供给稀缺。行业头部企业持续高薪争夺核心研发人才，人才流动性较高。若无法维持有竞争力的激励体系、算力资源与科研环境，核心研究员、仿真工程骨干存在流失风险。

下载日志已记录，仅供内部参考

投资评级说明

（1）公司评级定义		（2）行业评级定义	
买入	股价涨幅优于基准指数 15%以上	推荐	行业指数表现优于基准指数 10%以上
增持	股价涨幅相对基准指数介于 5%与 15%之间	中性	行业指数表现相对基准指数介于-10%~10%之间
持有	股价涨幅相对基准指数介于-5%与 5%之间	回避	行业指数表现劣于基准指数 10%以上
卖出	股价涨幅劣于基准指数 5%以上		

备注：评级标准为报告发布日后的 6 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现，其中 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 指数或纳斯达克指数，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。

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告。本人承诺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操守和专业能力，本报告清晰准确地反映了本人的研究观点并通过合理判断得出结论，结论不受任何第三方的授意、影响，特此声明。

证券投资咨询业务的说明

根据中国证监会颁发的《经营证券业务许可证》（Z23834000），国元证券股份有限公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询业务是指取得监管部门颁发的相关资格的机构及其咨询人员为证券投资者或客户提供证券投资的相关信息、分析、预测或建议，并直接或间接收取服务费用的活动。证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

法律声明

本报告由国元证券股份有限公司（以下简称“本公司”）在中华人民共和国境内（台湾、香港、澳门地区除外）发布，仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。若国元证券以外的金融机构或任何第三方机构发送本报告，则由该金融机构或第三方机构独自为此发送行为负责。本报告不构成国元证券向发送本报告的金融机构或第三方机构之客户提供的投资建议，国元证券及其员工亦不为上述金融机构或第三方机构之客户因使用本报告或报告载述的内容引起的直接或连带损失承担任何责任。本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的信息、资料、分析工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他投资标的的投资建议或要约邀请。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。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。在法律许可的情况下，本公司及其所属关联机构可能会持有本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投资银行业务服务或其他服务，上述交易与服务可能与本报告中的意见与建议存在不一致的决策。

免责条款

本报告是为特定客户和其他专业人士提供的参考资料。文中所有内容均代表个人观点。本公司力求报告内容的准确可靠，但并不对报告内容及所引用资料的准确性和完整性作出任何承诺和保证。本公司不会承担因使用本报告而产生的法律责任。本报告版权归国元证券所有，未经授权不得复印、转发或向特定读者群以外的人士传阅，如需引用或转载本报告，务必与本公司研究所联系并获得许可。

网址：www.gyzq.com.cn

国元证券研究所

合肥	上海	北京
地址：安徽省合肥市梅山路 18 号安徽国际金融中心 A 座国元证券	地址：上海市浦东新区民生路 1199 号证大五道口广场 16 楼国元证券	地址：北京市朝阳区安定路 5 号院 3 号楼中建财富国际中心 5 层
邮编：230000	邮编：200135	邮编：100029