



F R O S T & S U L L I V A N

60 Years of Growth, Innovation and Leadership

2025年中国世界模型发展洞察

2025年8月

A Frost & Sullivan
White Paper

执行摘要

本报告聚焦“世界模型”（World Models）这一前沿人工智能技术，分析其发展现状、技术路径、市场格局及未来趋势。世界模型是理解现实世界动态（包括其物理和空间属性）的生成式 AI 模型。它们使用文本、图像、视频和运动等输入数据来生成视频。通过学习，它们能够理解现实世界环境的物理特性，从而对运动、应力以及感官数据中的空间关系等动态进行表示和预测，加速物理AI的虚拟世界生成，生成可扩展的增强型数据，从而消除数据瓶颈，实现更高效的基础模型训练。

研究目的

本白皮书的研究目的在于全面梳理世界模型的发展历程、现状、核心技术及其在智能驾驶中的应用，并通过对不同厂商能力的对比分析，探讨世界模型未来的发展趋势。世界模型作为模拟和预测现实世界动态的生成式AI框架，已在自动驾驶、机器人、数字孪生等领域展现出巨大潜力，我们期望为相关领域的研究者、开发者以及企业提供有价值的参考信息，促进技术进步和产业发展。

目录

◆ 报告摘要	
◆ 关键发现	
◆ 章节一：世界模型概述	6
• 世界模型定义与基本概念	7
• 世界模型核心要素	8
• 世界模型发展历程与现状	9
• 世界模型关键能力及主流场景	10
◆ 章节二：世界模型技术能力分析	11
• 世界模型关键技术与能力	12
• 世界模型技术指标概述	14
• 世界模型对比及评价	15
• 主流世界模型介绍	16
◆ 章节三：世界模型在自动驾驶中的应用	17
• 自动驾驶系统概述	18
• 自动驾驶发展现状及核心痛点	19
• 世界模型在自动驾驶中的生成场景	20
• 世界模型在自动驾驶中的关键能力	21
• 世界模型在自动驾驶中的价值分析	22
• 世界模型在自动驾驶中的未来方向	23

目录

◆ 章节四：世界模型在具身智能中的应用	24
• 具身智能基本概述	25
• 具身智能发展现状	26
• 具身智能核心痛点及需求分析	27
• 世界模型在具身智能中的关键能力	28
• 世界模型在具身智能中的价值分析	29
• 世界模型在具身智能中的未来方向	30
◆ 章节五：世界模型厂商竞争格局	31
• 世界模型厂商综合实力评估指标定义	32
• 中国世界模型厂商总览与简介	34
• 中国世界模型厂商综合实力评估表现	35
• 世界模型优秀案例分析	36



7 Key Findings 关键发现



多模态感知融合、真实物理世界构建与动态环境预测作为世界模型的核心能力正成为驱动智能系统向高阶感知与决策能力跃升的核心技术支柱。

当前在自动驾驶领域世界模型应用成熟度高于具身智能领域，未来，世界模型将成为具身智能的

“认知核心”



世界模型的本质是**视觉和空间**的认知，最重要的是保证4D数据重建的真实性

目前，中国世界模型领域玩家已经超过

10+

其中头部玩家正凭借其先发优势与技术能力，与世界领先的厂商同台竞技。

超过 80%

自动驾驶算法使用世界模型进行辅助训练

以商汤为代表的

平台型
世界模型提供商

通过构建开放通用的产业基座，避免了车企重复自研与科研模型的工程化瓶颈，有效赋能自动驾驶、具身智能等领域的规模化应用与跨场景迁移

商汤绝影「开悟」世界模型在技术产品、应用能力以及生态建设三方面综合评价位居榜单

第一

第一章

世界模型概述

世界模型概述

世界模型定义

世界模型（World Models, WMs）作为一种**生成式AI模型**，能够通过**构建内部表征理解真实世界的动态规律**（涵盖物理特性与空间属性），同时借助文本、图像、视频和运动数据等输入信息生成视频内容。通过从感官数据中学习表征和预测运动、力和空间关系等动态特性，**实现对现实环境物理属性的理解并通过生成环境及动作，从而模拟、指导及实施决策。**

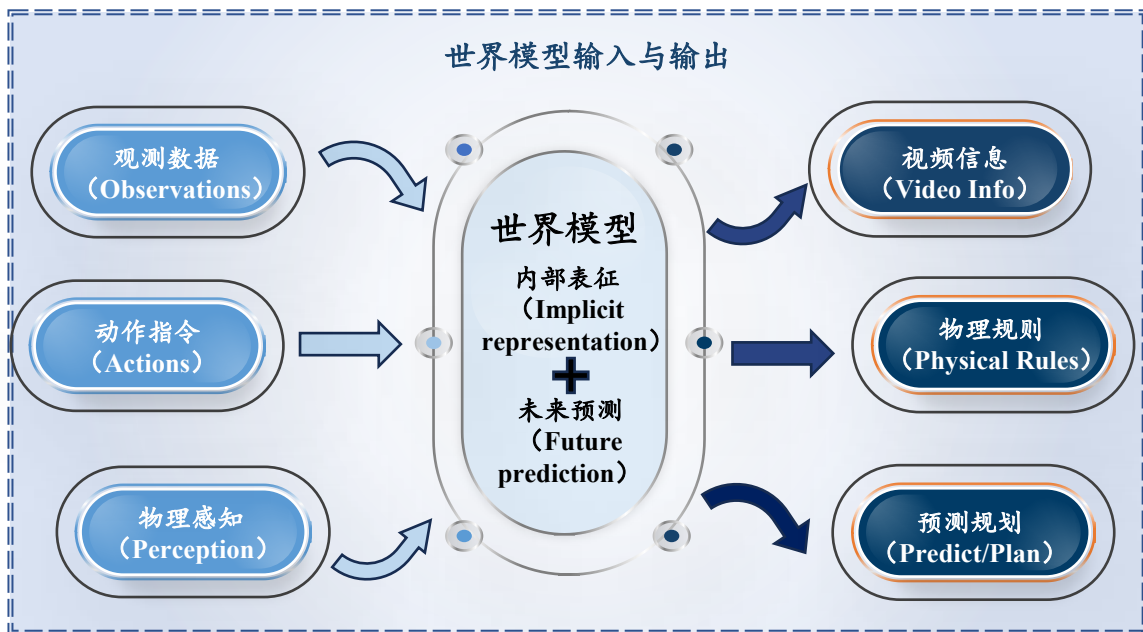
这类模型已在自动驾驶、机器人技术、数字孪生等前沿领域展现出巨大应用潜力，

正成为驱动智能系统向高阶感知与决策能力跃升的核心技术支柱。

世界模型基本概念

内部表征的世界模型可以通过概率模型、物理模型等形式存在，它们通过对历史数据的学习和理解，形成对未来事件和状态的预测。

未来预测的世界模型则依赖于生成模型，如扩散模型或视频生成模型，通过模拟连续的时间动态来生成真实场景。



世界模型核心要素

“世界模型不仅应当感知和建模现实世界，还应具备展望可能存在的未来状态的能力，从而为决策提供指导。”

——Worlds Labs 李飞飞

感知与建模

人类的感知能力是世界模型构建的第一步。当我们通过视觉、听觉、触觉等感官获取到外界信息时，大脑就会将这些信息处理并转化为对世界的认知。在机器学习中，这一过程通常通过传感器、图像识别、自然语言处理等技术来实现。

状态表示

一旦获取了感知数据，世界模型就需要将这些信息表示为某种形式的状态。比如，在强化学习中，状态通常是对环境某一时刻的抽象表示。在其他领域，状态可能是对事物属性或时间序列数据的集合。

学习与更新

世界模型的另一重要特征是能够通过学习不断更新自己的结构。无论是人类通过经验积累，还是人工智能通过训练数据，世界模型都能够随着新信息的加入而不断优化，以提高其对未来预测的准确性。

动态性与预测能力

世界模型不仅仅是被动地接收信息，它还具有根据现有数据进行预测和推理的能力。例如，基于过去的经验，世界模型可以预测未来的状态，帮助决策者在面对不确定性时做出更加理性和高效的选择。

世界模型发展历程与现状

- AI世界模型仍处早期阶段，以感知压缩和因果推理为主，闭环尚未稳定成型。
- 发展历程从理论提出到产业试点，包括MuZero发布、BEV模型落地和“升格”系统升级。
- 未来将突破三方面：多模态感知、因果可控生成、与具身智能系统深度融合。



从感知压缩到因果推理，世界模型正处于从早期探索走向复杂智能行为生成的关键过渡期

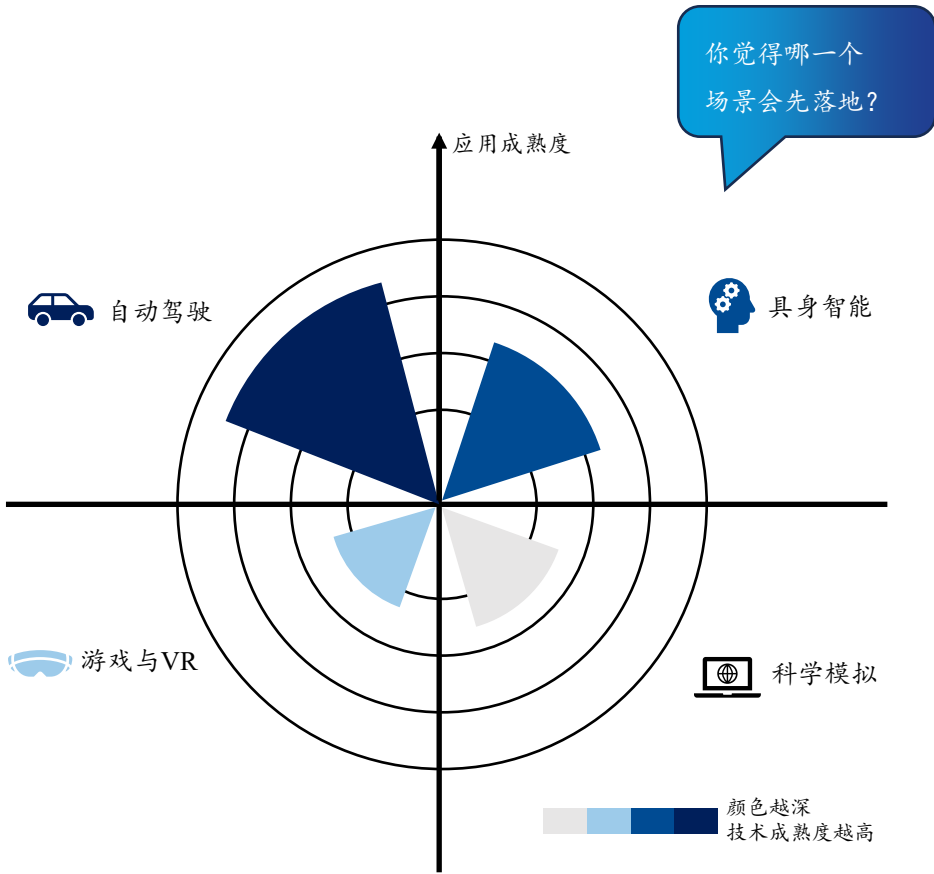
- 目前，世界模型整体仍处于早期探索阶段，大多集中在感知层面的模拟与压缩，尚未真正实现“感知-预测-决策”一体化的稳定闭环。尽管在自动驾驶领域已有试点应用，但多依赖特定环境与强先验，不具备通用性与长期泛化能力。
- 未来的发展方向将集中在三个方面：一是通过多模态输入增强对世界状态的理解；二是引入因果建模与可控生成机制，提升预测准确性与行为规划能力；三是将世界模型与具身智能系统深度融合，实现从“观察世界”到“理解并参与世界”的跃迁。



哪些领域将迎来变革？

——世界模型四大应用场景

世界模型主要场景成熟度



哪些领域将迎来变革？

——世界模型四大应用场景

一、自动驾驶

实时感知周围环境，世界模型可以分析其他车辆、行人、非机动车的运动状态和行为意图，**预测它们未来的行动轨迹**。例如，当识别到骑车人转头的动作时，世界模型能够判断骑车人可能有左拐的意图，从而提前提醒自动驾驶系统做好减速、避让等准备。此外，世界模型还可以**模拟不同天气、路况下的驾驶场景**，帮助车辆制定更合理的行驶策略，大幅提升自动驾驶的安全性和可靠性，推动自动驾驶技术从辅助驾驶向完全自动驾驶迈进。

二、具身智能

世界模型将彻底改变机器人的应用形态，使其**从只能执行简单重复动作的机器，转变为能够完成各种复杂任务的通用劳动力**。在家庭服务场景中，机器人可以利用世界模型理解家居环境和用户需求，自主完成打扫房间、整理物品等工作；在工业生产领域，机器人能够根据世界模型规划任务方案，灵活应对不同的生产需求，进行精密装配、质量检测、设备维护等操作。

三、游戏与虚拟现实

世界模型在游戏与虚拟现实中扮演着核心角色。它用于**构建虚拟环境的逻辑与物理规则**，使玩家能与动态场景互动。通过世界模型，系统可模拟重力、碰撞、光照等真实效果，提升沉浸感。同时，它还**支持AI角色的行为决策**，推动剧情发展。在VR中，世界模型确保**视觉、听觉与动作的实时同步**，增强用户体验的真实性与连贯性。

四、科学模拟

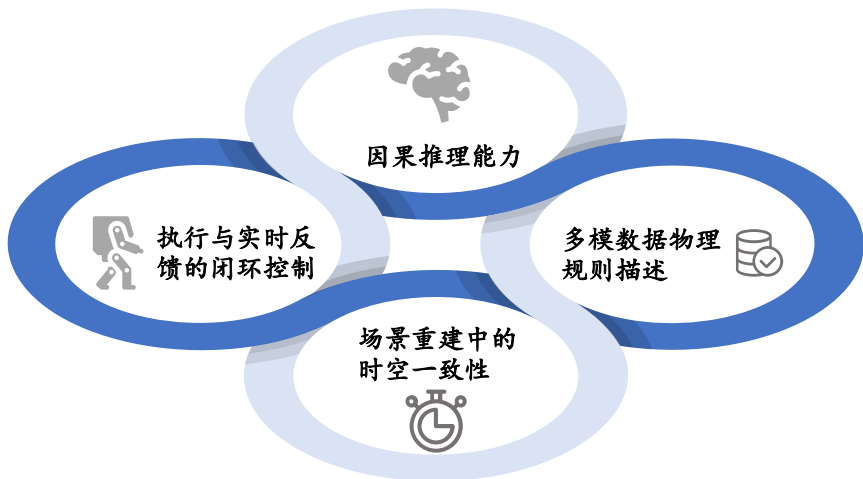
在科学研究和工程领域，世界模型为复杂系统的模拟和预测提供了强大的工具。通过跨尺度的推演，世界模型帮助科学家和工程师**更好地理解 and 掌控复杂系统**，推动科学技术的进步和社会的可持续发展。

注：图中圆形大小表示目前该领域的整体成熟度与投入情况

第二章

世界模型技术分析

世界模型的关键技术与能力



因果推理能力

因果推理要求 AI 能够回答“如果 A 发生，B 会怎样？”这类假设性问题，这对模型来说是一个巨大的挑战。因此人工智能不应该仅仅是对于已有知识的复制或检索，而应当具备一种更广泛、更深层次的理解周围环境的能力。

由于生活中大多数数据为动态视觉信息，如何利用图像、视频等视觉输入自发地理解现实概念并实现与人类的交互一直是一大难题，最先进的视觉问答模型在视觉因果推理方面与人类相比仍存在较大改进空间，世界模型通过构建视觉感知组件与记忆组件，所学习到的关于**环境的表征与对未来的预测**能有效地帮助模型理解现实

概念，构建强大的复杂因果推理能力，提升AI在动态环境中的自主决策能力。

例如：在自动驾驶场景中，需要模拟如果前方突然出现一个行人，车辆采取不同的避让策略会产生怎样的后果；在工业生产中，要预测如果设备某个部件出现故障，整个生产流程会受到哪些影响。

场景重建时空一致性

常见的视频生成通常侧重于以单一模态或单一视角生成数据，虽然其已经初步展现出对时空一致性、视觉因果链的建模能力，但受限于物体遮挡、视角信息固定、缺乏3D结构信息等因素，而难以实现真正的

时空一致性。

在3D世界中，我们的视角通常会随心所欲地沿着不同路径，向各种各样的方向移动，在这一场景下世界模型需要更强的时空一致性与动态3D环境生成能力。如保持空间结构稳定（如物体形状、位置、遮挡关系）和时间演化连贯（如运动轨迹、因果关系、物理规律）等。

当前世界模型通过长期记忆机制、潜在空间建模、对象中心表征等技术在更高维度空间利用潜变量向量代替像素重建，带来泛化能力和样本效率的提升，从而达成更优的时空一致性。

多模数据物理规则描述

模拟复杂的物理规则仍然是一项极具挑战性的任务，现实世界中的流体运动、物体碰撞等物理现象涉及到大量的参数和变量，具有高度的非线性和不确定性。以流体模拟为例，水的流动受到重力、粘度、表面张力等多种因素的影响，不同场景下的流体行为差异巨大，想要精确模拟十分困难。在物体碰撞模拟中，不仅要考虑物体的材质、形状、质量等属性，还需要准确计算碰撞瞬间的力的传递和能量转换，稍有偏差就会导致模拟结果与现实不符。



世界模型预测的是3D场景结构，而非简单的像素，这为后续的交互和真实感奠定了基础，因此生成的场景遵循基本的3D几何和物理规则（如物体的遮挡、深度、实体感），避免了传统模型生成视频常出现的扭曲、变形等“梦境般”的不真实感。且生成的世界是稳定的，不会因视角变化或短暂离开而改变，保证了体验的连贯性。

执行与实时反馈

在真实世界中，获取实时的交互数据面临诸多困难，收集数据需要耗费大量的时间、人力和物力，而且存在安全风险，难以获取足够丰富和多样化的数据。

而世界模型通过结合强化学习，可以实现从感知 → 建模 → 规划 → 执行 → 感知更新 → 模型修正的动态循环过程，甚至在过程中还可以根据主动干预来验证预测，并利用反馈信号持续优化自身。如在自动驾驶中世界模型可以根据感知建模信息预测行人轨迹，规划避让路径并执行转向操作，再通过更新的感知内容如雷达、视觉等信息修正预测模型。在这个过程中，低延迟响应的实时反馈是实际应用的基础，可以通过轻量化技术与潜空间状态生成等技术实现。

世界模型有哪些关键指标？

世界模型技术指标概述

FID

FID 是衡量生成图像与真实图像分布之间差异的指标，FID值越低，表示生成图像的质量和多样性越接近真实数据，用于评估世界模型生成的单帧图像质量

FVD

FVD 是FID的视频扩展版本，用于衡量生成视频与真实视频在时空特征上的相似性。评估世界模型生成的视频片段在内容和动态方面与真实视频的匹配程度

帧率

帧率指模型生成或预测的视频每秒包含的帧数，决定模型生成视频的的时间分辨率。高帧率可提供更平滑的视频，但计算成本更高。

时长

世界模型能够生成或预测的连续视频帧的数量或时间长度，衡量模型的长期预测能力。在相同帧率下，长时预测更难，易出现累积误差和模式崩溃。

一致性

指生成视频在时间维度上的连贯性，包括运动平滑性、物体身份保持、场景逻辑合理等。常考虑多模态内容一致性、物体一致性、时空一致性等。

模型	FID	FVD	帧率	时长	一致性
英伟达 COSMOS					
谷歌 Genie3					
Meta V-JEPA2					
商汤 绝影开悟					
Wayve GAIA2					

----- 全球世界模型主流模型技术能力指标均值线

不同世界模型厂商根据其自身战略与不同维度的技术优势，打造独特的世界模型能力及相关产品

生成式世界模型

生成式世界模型旨在学习一个能够生成或者预测的环境状态的概率模型。
给定当前状态和动作，生成式模型能够预测下一个可能的状态，能够生成逼真的环境观测，为智能体提供丰富的感知信息。但同时也需要处理高维数据，计算复杂度高。

非生成式世界模型

非生成式世界模型侧重于学习对控制有用的状态表示，避免预测所有的观测细节。
这类模型通常不显式地生成环境观测，而是通过其他方式来学习状态表示，例如预测价值函数、预测下一个状态的隐向量等。

谷歌 Genie3——世界模型前沿探索

1 摒弃传统硬编码物理引擎

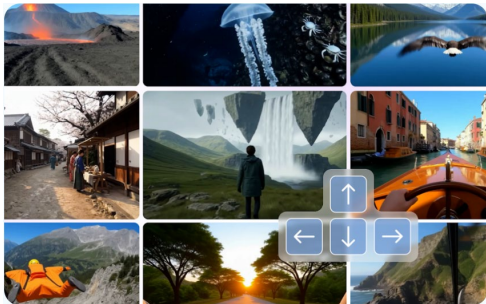
通过观察大量视频数据自主学习物理规律不需要硬编码的物理引擎，就能够理解重力、碰撞、水流动力学、光照效果等复杂物理现象。

2 高质量生成与实时交互

Genie3能够从简单文本提示生成720p、24fps的实时交互3D环境，远超传统AI视频生成质量，且对于交互的实时性支持更迅速。

3 多模态输入处理

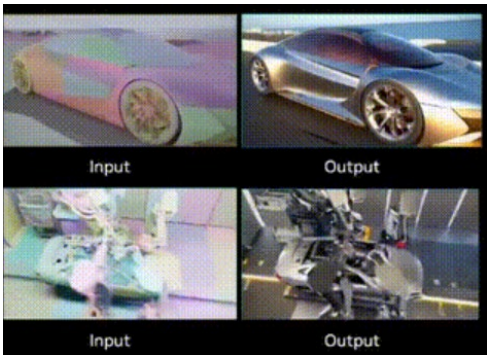
支持文本提示、图像、照片甚至手绘草图作为输入，理解并解析复杂的场景描述，同时支持键盘鼠标等标准交互设备，实现丰富的人机交互体验。



4 长期记忆一致性

通过引用历史轨迹信息维护长期一致性，能够记住数分钟前的环境状态，支持用户重新访问之前探索过的区域并保持完全一致性。

英伟达 COSMOS——世界模型应用先锋



1 基于真实物理的引擎

通过Omniverse、Genesis等，生成遵循现实物理规律的模拟视频数据。通过Cosmos Transfer来提升模拟数据的真实性和多样性

2 多种模型选择

Cosmos Reason、Transfer、Transfer精简版、Transfer2、Predict2等一系列预训练多模态模型提供了丰富的应用场景与功能选择。

3 数据工厂加速下游基础模型开发

从3D仿真场景中生成高保真度的合成数据，为下游用户提供合成数据生成和后训练能力，构建高效数据工厂，推动视觉 AI 和具身 AI 的发展。

Meta V-JEPA2——自监督基础世界模型

1 自监督训练

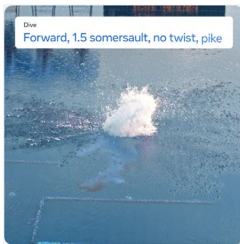
编码器和预测器通过视觉数据的自我监督学习进行预训练，利用丰富的自然视频来引导对物理世界的理解和预测。

2 解锁物理世界理解

与语言建模相结合时，提供了卓越的运动理解能力和领先的视觉推理能力。

3 预测未来动向

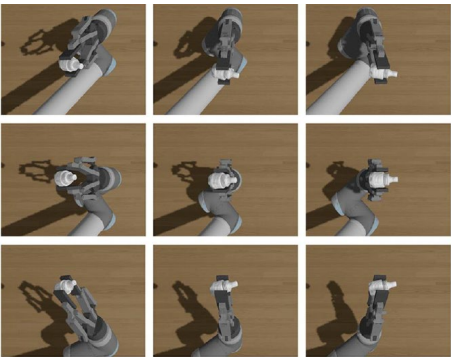
V-JEPA 2可以预测世界将如何演变，它利用世界模型来理解物理现实，预测结果，并计划有效的策略，并且只需要最少的监督。



4 应用于机器人与可穿戴设备

开启机器人技术的新时代，为在物理环境中导航以处理家务和复杂任务的AI Agent赋能，也可以帮助个体在嘈杂环境中导航，提供障碍物和危险的实时警报。

商汤 开悟世界模型——世界模型平台赋能具身智能



1 合成数据为核心

以合成数据为核心，支撑模型算法训练和仿真测试两翼发展。通过合成数据，实现与下游模型单帧级别实时交互。

2 具身智能训练场

提供从基础设施到模型再到平台的全套生态能力，使机器人能够在现实物理世界中自主探索，完成自我迭代与持续进化。Waymo Sim Agent 竞赛连续两年榜单第一名

4 全流程数据处理能力

为满足模型闭环高效实训，商汤世界模型提供了训练前的数据合成、编辑能力以及数据解析等能力；并且支持高质量4D数字资产重建与实时交互，从而推动具身模型在4D真实世界不断学习迭代。

3 首个高分辨率与稀疏控制的多视世界模型

视频分辨率满足机器人场景算法训练、稀疏控制更贴近具身数据集标注分布情况，二者加成有利于世界模型应用落地，结合多视角生产能力，突破了具身数据合成的技术瓶颈。

Wayve GAIA2——生成式AI世界模型

① 生成式世界模型

采用连续潜在空间，结合32倍空间压缩率与64通道高语义维度，通过增强的可控性、扩大的地理多样性，推动了合成数据生成的边界。

② 专为自动驾驶复杂场景构建

处理多个摄像头视角、不同的路况和关键的弯道情况。通过提供对关键驱动因素的精细控制，GAIA-2使工程师和研究人员能够创建更丰富、更真实的训练场景。

③ 多镜头高保真

GAIA-2利用潜在扩散框架，在多个摄像机视点上提供稳定、长时间的视频序列，具有时空一致性。同时提供对自我车辆行为和场景特征的精确控制。



④ 安全临界情况的模拟

GAIA-2可以系统地生成罕见和高风险的场景，如突然车道变换和紧急机动，以改善自动驾驶系统的测试和验证。

第三章

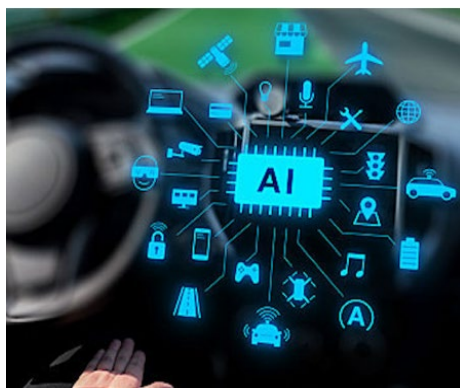
世界模型在 自动驾驶中的应用

自动驾驶发展概述

从L2提升到L3级自动驾驶需要跨越式的技术与成本，目前整体实际落地仍停留在L3以下。

按照SAE的标准,自动驾驶汽车可以依据智能化、自动化水平被分为6个等级,从L0到L5反映了车辆自动驾驶能力的不断提升以及人类驾驶员参与程度的逐渐减少。

近年来，中国自动驾驶渗透率整体呈现提升趋势，增长主要体现在L2及以上的车辆。其中，新能源车型的电气化和智能化优势为自动驾驶搭载提供了良好基础。



当前自动驾驶技术在L2级辅助驾驶已大规模落地。2024年上半年，中国乘用车L2及以上自动驾驶新车渗透率已达**55.7%**，预计2025年将达到**65%**。**L2级自动驾驶正从高端车型专属逐渐下放至更多中低端车型，向“全民智驾”跃迁。**L3级是辅助驾驶和自动驾驶的分水岭，从L2提升到L3级需要跨越式的技术与成本。目前，L3级自动驾驶已从前期验证阶段迈向商用初步落地阶段，**未来2-3年将成为L3级量产的关键期。**然而，对L4及更高级别的探索仍面临法规、成本与安全等限制，仍聚焦于Robotaxi、物流货运等限定场景。

自动驾驶当前发展障碍——6大核心痛点

1 环境理解能力不足



- **动态障碍物误判率高：**自动驾驶系统在恶劣天气（如雨、雾、雪）中使用的 LiDAR 等传感器表现明显下降
- **单一传感器可靠性受限：**多传感器融合复杂度高，时间同步误差过大会使感知融合失效造成误判

2 泛化能力差



- **长尾场景稀有：**需数亿甚至数十亿英里实测才能统计验证安全性，数据量需求不现实
- **仿真与真实场景鸿沟：**虚拟测试中动态物体行为建模误差较大，极端天气（如暴雨、大雾）的物理效应模拟失真度激增，导致实车失效率陡增

3 高度依赖高精地图



- **时效性难以保持：**流动环境中地图需频繁更新，否则造成定位与决策误差，但地图维护成本过高
- **跨区域适配成本高：**支持不同区域（如中/美/欧）的差异化交规需定制地图，开发周期延长，成本增加

4 决策预测不稳定



- **模块割裂导致响应延迟：**传统“感知-规划”架构缺乏整体协同理解，面对动态变换时容易产生保守或误判行为
- **长时序推理能力弱：**传统规则系统仅能预测1秒内路况，无法应对长时间未来场景预测需求

5 研发与运营成本高昂



- **硬件成本高：**高等级所需的 LiDAR、冗余感应器与冗余硬件显著提升单车成本，维护与地图标注、传感器校准带来长期经营压力
- **长期数据成本高：**长期训练还需庞大的数据团队和高成本数据标注流程，产业普遍负担重

6 安全信任度不足



- **安全验证时间长：**在统计意义上证明自动驾驶的致命事故率优于人类驾驶需耗时数十年甚至数百年，传统“行驶里程证明安全”策略几乎不可行
- **公众信任程度保守：**全球仅 38% 消费者信任 L3+ 系统，67% 担忧算法伦理决策不透明

世界模型在自动驾驶中生成了哪些场景？

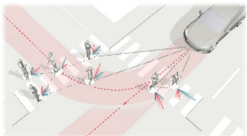
世界模型能通过生成多层复杂要素结合的场景，将传统算法难以覆盖的“高动态+高不确定性”场景转化为可控问题。

世界模型能够生成或应对的高复杂度、高风险的真实驾驶环境。这些场景通常具有高动态性、高不确定性的特点，传统算法难以覆盖。其核心价值在于解决自动驾驶的长尾问题，覆盖极端天气、突发风险、全域连续驾驶等不足0.1%但决定99.9%安全性的场景，推动L3级系统落地。

世界模型应用的生成场景是多层要素的结合：静态和动态参与者、道路环境、天气和光照条件、不同国家和地区的交通背景。

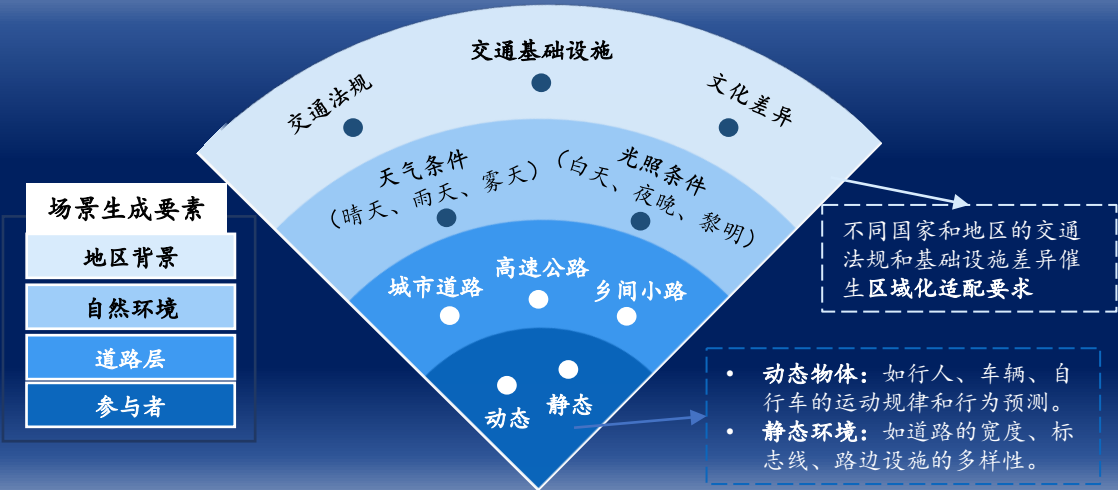


极端天气下动态避障



无规则路口群体博弈

世界模型在自动驾驶中的生成场景



世界模型如何赋能自动驾驶系统？

“超过80%的自动驾驶算法使用世界模型进行辅助训练”

世界模型通过构建“真实数据→模型训练→仿真场景验证→模型部署”的闭环反馈机制，并提供统一的潜在世界状态表征，从而为感知、预测、规划、控制等模块提供一致的认知语境。因此，世界模型能够推动自动驾驶系统持续学习、自主验证并快速迭代优化，从而显著提升端到端自动驾驶表现。

世界模型能够助力自动驾驶系统在产品性能和市场表现的双重升级。一方面，世界模型能够快速生成海量高保真场景，覆盖长尾与极端事件，显著增强系统鲁棒性与安全保障。另一方面，世界模型通过高效仿真替代真实路测，不再依赖昂贵标注和地图数据，在降低研发成本体系的同时推动产品快速迭代和市场拓展。

世界模型是突破L4（如Robotaxi等）规模化部署瓶颈的加速器，是构成自动驾驶智能体走向类人认知与判断模式的关键基础。

世界模型在自动驾驶中的能力分析



世界模型推动自动驾驶系统持续进化

进一步理解世界模型对自动驾驶的实际价值

“世界模型能够节省近50%的成本，提高约70%的效率”

世界模型能够自动生成自带标注的图像与视频数据，包括极端或罕见场景，**避免昂贵的真实数据采集成本**。同时，世界模型可在无需大量人工设计的情况下，生成多模态、跨时序一致的场景，**显著降低人工建模与标签标注的投入**。虽然构建世界模型初始需要大量训练与计算资源成本投入，但一旦训练完成，其生成额外高质量数据的边际成本极低，相较于采集真实数据和传统仿真重建场景更具**数据规模经济优势**。

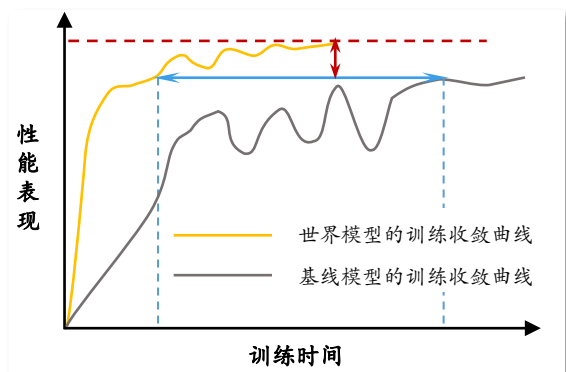
综合以下6个维度对比分析（A. 初始投入；B. 边际生成成本；C. 极端/长尾场景覆盖能力；D. 数据规模经济；E. 真实性；F. 一致性），世界模型拥有更优秀的成本表现，且生成的数据质量会随模型训练持续改进。

维度	真实数据采集	传统仿真	世界模型
A	高	中等	偏高
B	极高	中等偏高	低
C	有限	一般	优秀
D	弱	中等	强
E	高	较优	中等偏高
F	高	中等偏高	较优

由浅至深对应由低至高的表现优秀程度

世界模型通过模拟未来场景、提供更丰富、更高效的训练信号，显著加速了自动驾驶模型的收敛过程。这不仅提升训练效率，还增强了模型的性能和泛化能力。

下图中，横轴表示训练时间，纵轴表示模型在实际执行任务中的表现。图中两条曲线的对比显示，世界模型能在相同训练时间内，对比基线模型**更快达到更优性能**，提升自动驾驶感知、规划甚至端到端的学习效率。



- 世界模型的收敛曲线以更快的速度上升，其达到稳定性能所需训练时间或迭代次数远低于基线模型，表示模型能在更少的训练时间或更少的数据下，学到高质量的参数和策略。
- 世界模型的收敛曲线最终性能提升水平高于基线模型，基线模型的收敛曲线显得缓慢、震荡且总值较低。

未来世界模型在自动驾驶中将走向何处？

“世界模型未来预计1-2年内实现上车，引领自动驾驶从车端走向云端”

从“两段式”到“一段式”端到端，最终实现世界模型的应用，是自动驾驶**从车端到云端**的技术路线图。未来1-2年，世界模型的成熟将引领自动驾驶迎来真正的爆发。



通用世界模型

- 通用世界模型将具备更强的泛化和迁移能力，可以**跨城市、跨车辆平台**发挥作用，成为自动驾驶的“通用大脑”。有了通用世界模型，不同厂家和地区的自动驾驶系统或可共享同一套核心智能。
- 目前，多家主流厂商已将世界模型视为**下一代智能驾驶架构的核心**，计划训练能够全面表达和预测真实世界的通用模型。未来的通用世界模型或将融合数百亿参数、多模态输入，成为驱动L4/L5自动驾驶的基础设施。



多车协同建模

- 多辆自动驾驶车辆之间可以共享各自感知信息，在云端融合成更大范围、更精细的世界模型，实现车队层面的环境建模与决策优化，有助于解决单车视野受限、传感器冗余不足的问题。
- 未来，多车协同世界模型可应用于车队编队、智能调度等场景，通过**群体智能**优化决策与交通效率。
- 在**车路协同**上，路侧感知与车端世界模型融合也是典型方向，可扩展车辆对整体环境的认知。



跨模态世界建模

- 世界模型将综合视觉、雷达、地图、V2X通信、自然语言等多模态信息源，构建更全面的数字孪生，赋予模型“知识”和“常识”理解能力。
- 未来，世界模型将进一步深化与BEV架构和VLA（视-语-行动）模型的融合，实现向**统一语义认知与决策一体化**演进，使自动驾驶系统接近人类司机的直觉判断。



AI原生地图构建

- 世界模型将重塑高精地图的生成与使用模式，催生**AI原生地图**，支持车辆在行驶过程中实时构建环境模型，并同步至云端进行共享。
- 不同于静态预制地图，AI原生地图具备**动态更新**能力，深度融合车载感知，甚至能主动预测环境变化，支持陌生道路下的即时环境认知，实现“所见即所知”的能力。



与强化学习融合

- 未来自动驾驶系统或将大规模采用**世界模型辅助的强化学习**：先由世界模型模拟环境和车辆动力学，再让决策策略在其中通过强化学习不断试错优化。
- 世界模型与强化学习的结合有望使得自动驾驶成为真正具备**自主进化能力**的智能体系统，即通过不断地自我博弈和环境交互来提高。随着算力提升，自动驾驶系统可能探索出连人类都未想到的最优驾驶策略。

第四章

世界模型在 具身智能中的应用

从感知智能向具身智能的跃迁

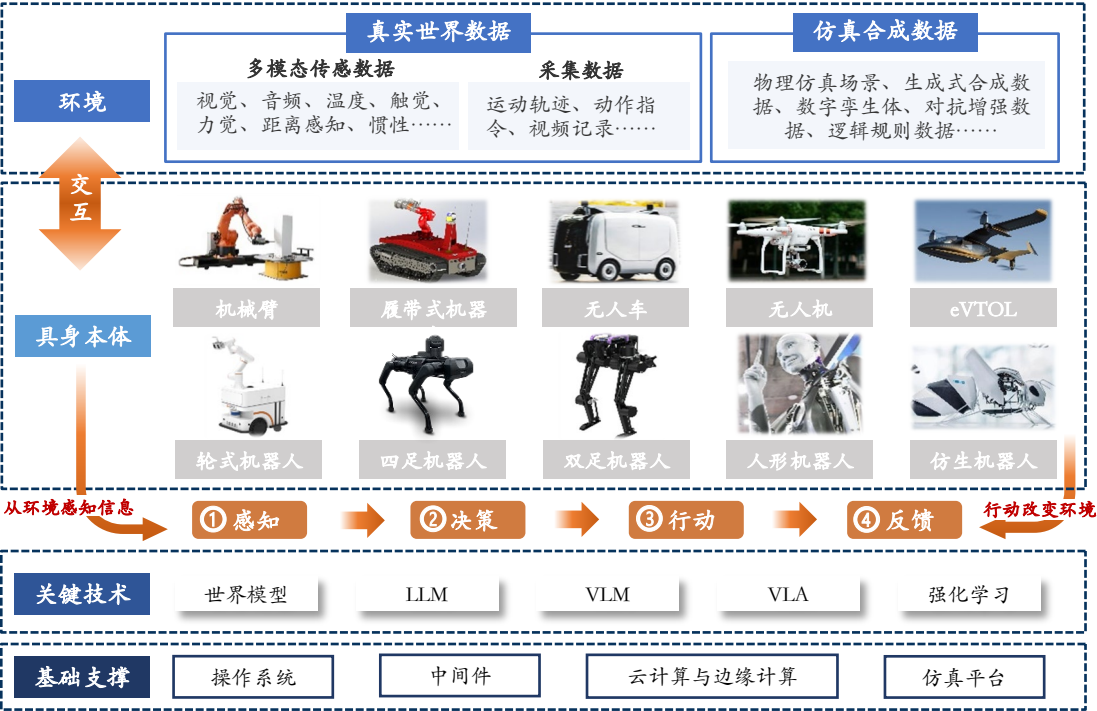
过去十年是感知智能快速发展的时期，而未来十年将是具身智能的黄金时代。

具身智能是指通过机器人等物体实体与环境交互，实现对世界的感知、决策、行动、反馈闭环，并能由此实现智能增长和行动自适应的智能系统。具身智能代表着AI从纯粹的信息处理转向物理世界的交互。

具身智能的基础架构由本体、数据和软硬件技术底座共同构成。

本体作为任务执行机构，其形态、传感器及核心零部件硬件直接影响感知与行动能力。其中，人形机器人凭借高自由度的操作移动能力、强环境适应性及良好接受度，被广泛视为最具潜力的通用形态。数据是关键驱动力，需要依赖高质量真实数据和高效低成本的仿真数据实现0到1的突破，并支撑1到N的深度复杂应用。软硬件底座提供三位一体的支撑：软件工具链支撑灵活开发与高效测试，通用计算平台保障复杂计算的可靠运行。

具身智能技术架构



来源：弗若斯特沙利文

具身智能代表了机器人发展的高级阶段，也是走向通用人工智能（AGI）的必经之路

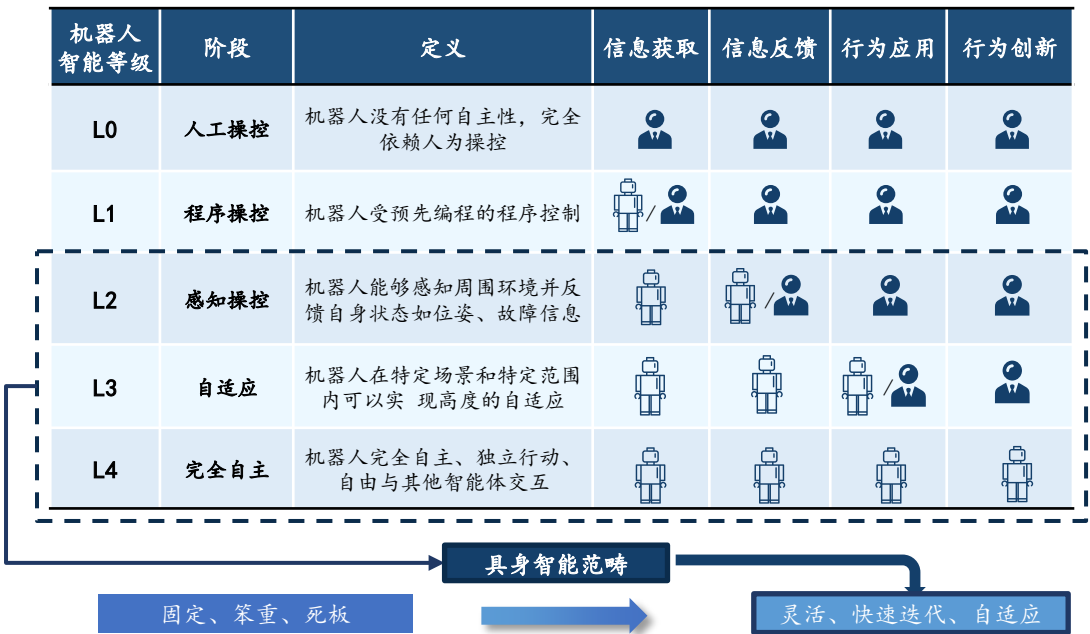
在机器人技术领域，智能机器人的自动化水平被划分为五个等级，从基础操作到完全自主，依次为L0（人工操控）、L1（程序操控）、L2（感知操控）、L3（自适应操控）和L4（完全自主）。

当前，尤其在工业制造领域，应用最为普遍的是L0和L1级别的机器人，如执行固定程序的机械臂和直角坐标机器人。尽管这些机器人能够高效地处理高精度的重复性制造任务，但它们在应对商业市场快速变化和迭代方面存在局限。

随着智能化、自动化水平的提升，L2至L4级别的机器人因其高级感知、决策和适应能力，对于满足工业制造、医疗健康、零售和服务等行业的复杂动态需求至关重要。

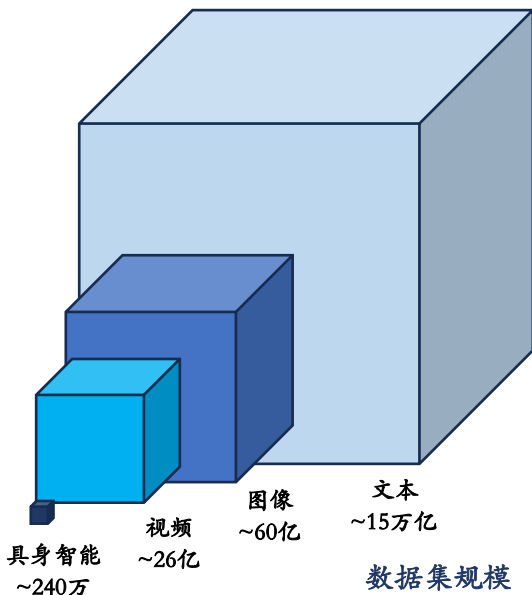
大模型技术的突破为具身智能带来了技术范式转移，具身智能产业目前正处于快速拓展期，预计未来将在更多行业实现规模化应用，成为推动社会智能化转型的重要力量。**长期来看，具身智能未来产业发展空间巨大，2050年潜在的全球市场规模将超过一万亿元。**

智能机器人分级



具身智能领域的核心痛点

——物理交互数据的“千倍级缺口”



具身智能的核心困境在于模型训练对数据的需求量大，但物理交互数据极度缺乏。在数据规模上，相比大语言模型的TB级存量和自动驾驶的PB级积累，具身智能只有约240万的积累但需要数百PB级的物理交互数据支撑，**数据缺口超过99%**。同时，真实数据的获取周期较长，无法满足具身智能爆发式的需求。

- 每家企业实现通用能力需约**400万小时**数据，国内全行业共需约**6000万小时**
- 数据采集时长：机器人本体企业约**6.5年**，模型企业约**32年**

具身智能所需的数据需要整合文本指令、多视角视觉、关节运动轨迹及物理交互等多维信号，复杂度远超纯文本或单一视觉模态。采集真实的物理交互数据则需要搭建各种工作生活场景，耗时长成本高，使得技术研发速度严重滞后。

数据获取困难、场景复杂多变、试错成本巨大、标注成本高昂、数据标准不一致等挑战严重阻碍了具身智能行业的发展。为了应对这些挑战，具身智能行业对**合成数据**的需求和认可逐渐增长，几乎所有头部的具身智能企业均在使用合成数据。



世界模型为具身智能提供大规模高质量合成数据

“世界模型为具身智能提供高质量、低成本、易扩展的合成数据生成路径，解决当前数据瓶颈”

□ 高质量合成数据

世界模型能生成视觉逼真、物理精确的合成数据，有效克服了传统仿真数据与真实世界之间的差异。

- **多模态融合**：先进的世界模型能够处理和理解多模态的输入信息
- **物理一致性**：通过整合物理引擎的约束，生成的合成数据不仅物理拥有精准性，行为和互动也更合理

□ 高成本效益与可扩展性

世界模型能够大幅降低数据获取的时间和经济成本，并轻松实现数据规模的扩展。

- **大幅降低数据成本**：合成数据生成主要依赖计算资源，能显著降低成本并缩短迭代周期
- **多样化数据扩展**：零风险生成在现实世界中稀少、危险的极端场景

□ 提升模型泛化能力

通过世界模型生成的海量、多样化合成数据进行训练，能显著提升具身智能模型在未知环境中的适应能力和任务执行成功率。

- **增强模型鲁棒性**：世界模型为智能体训练提供多样化的变体数据，显著增强模型对真实世界中的噪声和变化的鲁棒性
- **跨平台与跨任务迁移**：在大规模、多样化的合成数据上训练的基础模型，其学习到的表征和策略更可能捕捉到任务的本质和共性，从而更容易迁移到不同的机器人平台或执行相关的新任务上



如何量化合成数据的成本效益

- 数据获取成本降低约**60%**
- 缩短数据获取周期：本体企业缩短约**70%**，模型企业缩短约**95%**

(以真实数据：合成数据=1：9的比例保障训练数据有效性)

极少量真实样本



大量高质量合成数据

进一步理解世界模型如何赋能具身智能

“世界模型不仅为具身智能提供数据支持，更正在重塑其整个开发范式”

以世界模型为核心构建具身智能的一站式开发平台能够提供集成化的工具链和环境，重塑了具身智能的开发范式。通过将数据合成、算法训练和仿真验证无缝连接，形成闭环迭代系统，能够大幅提升研发效率，降低技术门槛，从而消除具身智能发展的工程障碍，让研究人员和工程师专注于算法和应用的创新，而不是基础设施的构建。

首先，世界模型为算法训练提供了安全、高效且可解释的仿真环境。

世界模型通过模拟智能体动作引发的精确状态变化和物理交互，支持对“感知-决策-执行”全流程进行闭环验证与迭代优化，从而系统性提升智能体在复杂动态环境中的适应性和可靠性。

其次，以世界模型为引擎的开发范式能够深度整合开发工具链，为具身智能研究提供端到端的高效支持。世界模型作为核心预测与生成引擎，能够无缝衔接从开发、训练到仿真部署的全流程，消除了工具链割裂带来的效率损耗。因此，世界模型能够大幅提升具身智能主流模型的开发与训练效率，并支持模型性能的优化。

以世界模型为核心的一站式闭环开发平台



来源：弗若斯特沙利文

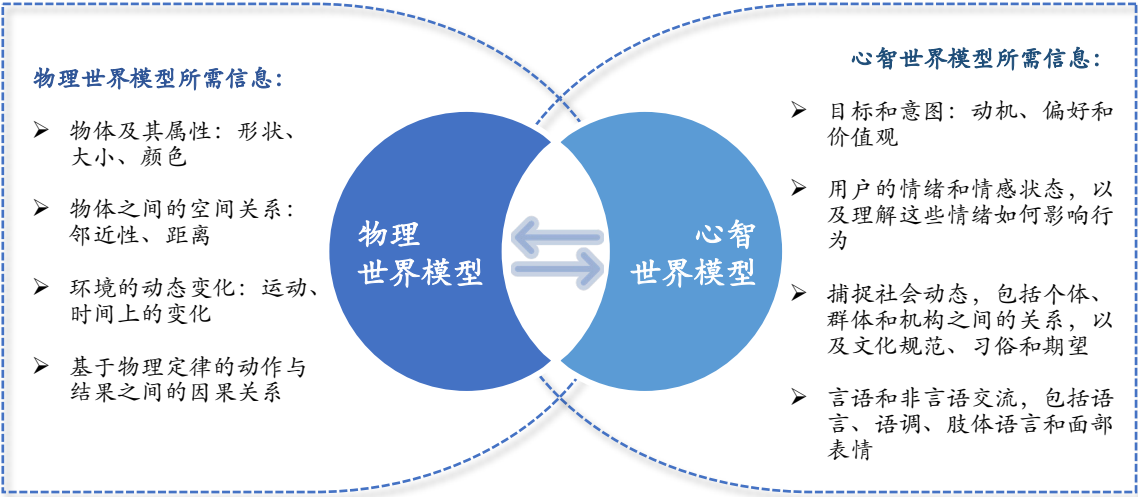
世界模型在具身智能中的下一步会走向何处？

“世界模型的未来框架是基于物理+心智的双轨建模”

相比于传统物理世界模型仅关注物体运动和机械因果的单一视角，未来的心智世界模型将人类的心理规律（如意图、情绪、信念、社会关系等）纳入建模范畴，形成“物理+心智”的双轨架构。这种架构使智能体能够表征并理解人类的心理状态、行为模式和文化惯例，进而实现对目标意图的预测、信念差异的推断以及情绪反应的模拟。借助这种能力，智能体可**从被动执行转向主动响应**用户需求，提供符合心理预期的协助，从而使人机交互从机械响应迈向具有情境感知与心理理解层次的协同模式。

同时，心智世界模型还为**多智能体协作**提供了“共识心智”的理论基础。各智能体不仅能识别共用的外部环境信息，还能够相互推断彼此的意图、信念和预期，通过心智建模实现高阶的行为对齐与策略协调。面对不确定或冲突情况下，心智世界模型帮助智能体识别他方内在目标，从而调整自身策略以寻找平衡，推动协作从简单的同步执行向情境感知和情绪兼顾的方向发展。这种协同机制将使未来的人机互动与智能体间协作更具**社会性与同理心**，更符合现实世界中多主体间合作的复杂性和柔性需求，推动具身智能向更复杂、更协调的群体协同演进。

未来双轨建模架构



第五章

世界模型厂商 竞争格局

世界模型厂商综合实力评估指标定义

应用能力	产品能力	生态建设
行业覆盖	环境建模与预测	数据生态
	决策与规划	软硬件协同生态
场景覆盖	模拟仿真预训练	合作伙伴生态
	研发投入与创新	开发者生态

应用能力

- **行业覆盖**考量是否已在主流车企或智能汽车品牌中实现技术部署；跨行业协同拓展能力：是否具备将世界模型技术延伸至智慧交通、智慧城市、工业互联网等领域的能力，其他行业如机器人、游戏等。
- **场景覆盖**考量是否能在不同复杂场景具备良好适应性与生成能力，如城市环境；高速与匝道；极端天气情况如雨雪、雾霾、夜间、强光等恶劣环境下的建模与预测；特殊区域覆盖；多模态交互场景兼容性等。

产品能力

- **环境建模与预测**考量对多源传感器数据的融合建模能力；动态目标行为预测：对行人、车辆、非机动车等交通参与者的未来轨迹预测准确性等。
- **决策与规划**考量在复杂环境中生成合规路径的能力；多目标优化能力；应对突发状况的反应速度，面对紧急情况的实时响应能力以及长周期任务规划的能力

- **模拟仿真与训练能力** 考量虚拟环境构建能力，能否构建高保真、可编程的模拟世界用于模型训练与验证；是否支持基于世界模型的端到端强化学习策略训练；
- **技术研发投入与创新能力** 考量企业在世界模型相关技术上的资金与人力投入情况；算法创新成果：是否提出具有原创性的世界模型架构或改进方案；专利与论文产出：在世界模型与智能驾驶交叉领域的专利申请数量及学术论文发表质量

生态建设能力

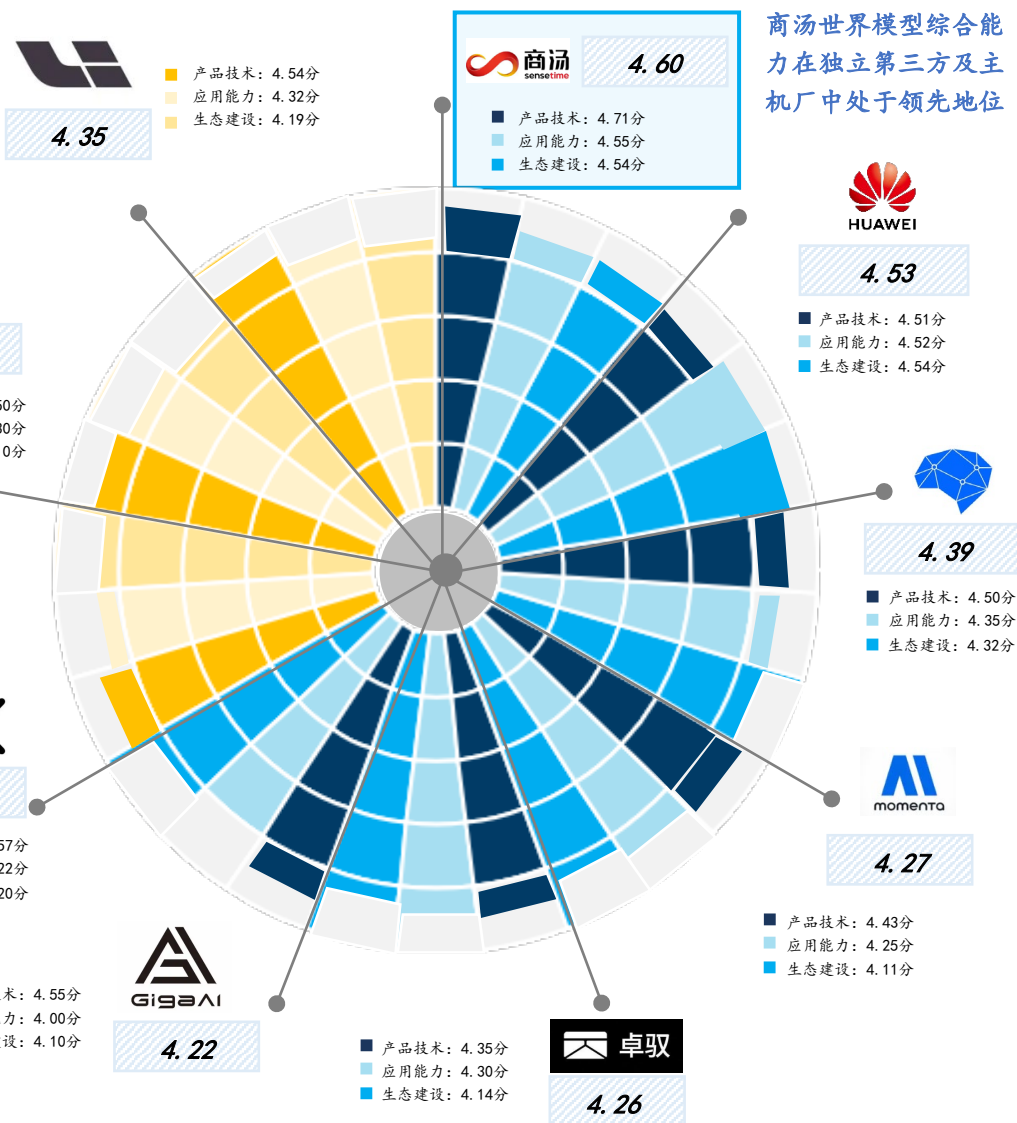
- **数据生态** 考量数据采集能力；数据标注与治理能力：数据清洗、标注、脱敏、版本管理等流程的完善程度；数据多样性与覆盖广度：涵盖不同地区、天气、光照、交通密度等场景的丰富度；数据安全与合规性。
- **软硬件协同生态** 考量芯片适配与算力优化能力：模型在不同计算平台上的部署效率；中间件与工具链完整性：是否提供从模型训练、部署、推理到监控的全流程开发工具；与其他系统集成能力：与现有自动驾驶系统的兼容性和耦合性等。
- **合作伙伴生态** 考量厂商的产学研合作深度，与高校、研究机构在世界模型方向的合作项目与成果转化；产业链协同能力：与车企、Tier1供应商、云服务商等上下游企业的联合开发案例等。
- **开发者生态** 考量厂商的开发者支持体系：是否提供SDK、API、文档、教程、示例代码等资源；是否具备开发者社区，社区活跃度以及开发者热度。是否建立闭环的用户体验反馈机制以持续优化模型性能及商业化落地能力。

中国世界模型厂商总览与简介



来源：弗若斯特沙利文

中国世界模型厂商竞争格局



*注：蓝色系厂商为第三方自动驾驶解决方案商世界模型能力，黄色厂商为造车新势力自研世界模型能力

Waymo Sim Agent 竞赛连续两年榜单第一名

开悟世界模型能力图谱



开悟世界模型成功案例

为自动驾驶厂商提供低成本海量仿真数据与极端场景覆盖，助力加速训练迭代与量产落地。

智能驾驶



商汤绝影与智己汽车共建端到端数据工厂，通过生成高风险长尾场景补足训练和验证数据，显著加快智驾量产落地



商汤绝影在上海自动驾驶实训场支撑数据到模型上车的全链路，规模化生成多视角仿真数据，降低数据成本并提升研发效率。

技术瓶颈突破

首个支持高分辨率与稀疏控制的多视世界模型，突破具身智能数据合成技术瓶颈

具身智能

合成数据能力领先

一链生成场景多样性可控泛化场景要素可控耦合、任意编辑
3D技术控制生成逼真轨迹



基础动作：拧螺丝、跳跃、精细放置、手指抓取……
场景轨迹：医疗护理、超市补货、流水线工作……

商汤绝影构建了以开悟世界模型为核心引擎的**悟能具身智能平台**，实现视觉感知、精准导航与多模态交互，并通过端侧与云侧算力支撑，实现智能体在真实环境中的自主理解与行动。



华为——盘古世界模型

功能介绍



盘古世界模型

动态生成可交互、可漫游的数字空间，构建智能驾驶和具身智能机器人训练所需环境。



图像理解

深度解析图像的场景、实体、属性、逻辑等，综合理解图像含义。



视频生成

加速视频创作，丰富感官体验，让创意动起来，带你跨越时空的旅行。



图像生成

利用大数据和深度学习，将灵感转化为高质量图片，重塑视觉艺术形态。

场景应用及成功案例

互动娱乐

图像理解识用户动作表情，图像生成造个性视觉，视频生成创动态剧情，盘古世界模型构交互场景，共推互动娱乐革新。

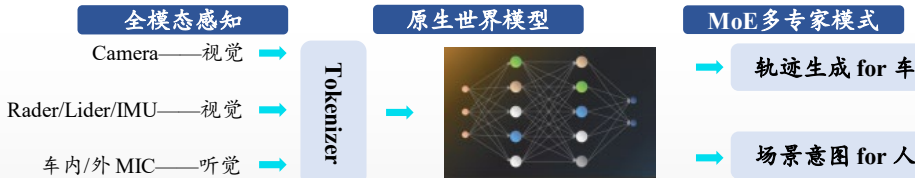
影视制作



图像生成与视频生成技术赋能影视制作，可智能生成场景、角色与特效，实现动态剧情生成，提升制作效率与创意灵活性。

万兴科技与华为云达成战略合作，共同推进AI大模型的技术和应用创新。以华为云盘古多模态大模型为基座，持续优化万兴天幕音视频大模型2.0，构建其音视频垂类能力。

智能驾驶



广汽借助华为云盘古多模态大模型，打造业界首个支持点云生成的大模型，为其端到端仿真高效迭代提供强有力支撑。



NWM世界模型

三大核心功能

理解世界

NWM 通过自回归学习传感器输入及物理规律，实现全量重建并具备强泛化能力。它能像人类一样聚焦关键要素，并在雨天、雪天、昼夜等环境中想象平行世界，从而学会复杂场景中的驾驶。



推演世界

NWM 通过自回归学习传感器输入及物理规律，实现全量重建并具备强泛化能力。它能像人类一样聚焦关键要素，并在雨天、雪天、昼夜等环境中想象平行世界，从而学会复杂场景中的驾驶。



仿真世界

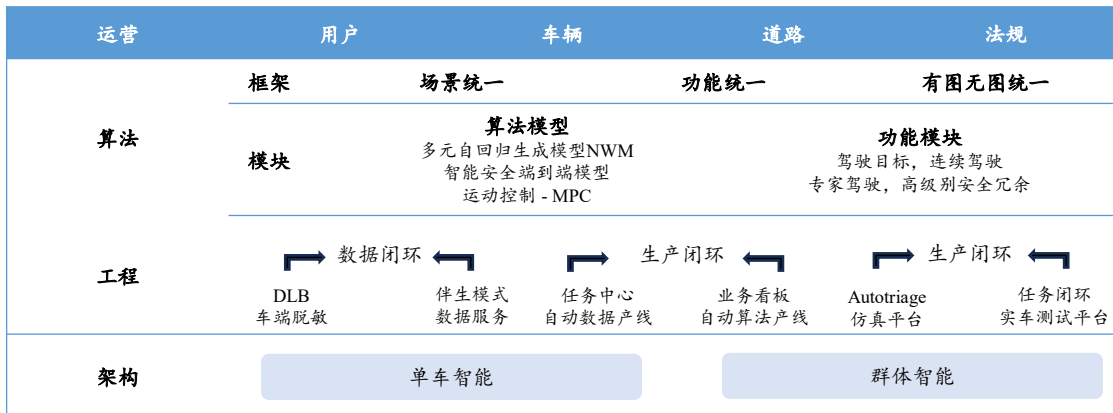
NWM 通过自回归学习传感器输入及物理规律，实现全量重建并具备强泛化能力。它能像人类一样聚焦关键要素，并在雨天、雪天、昼夜等环境中想象平行世界，从而学会复杂场景中的驾驶。



模型架构

点到点全域领航辅助2.0
覆盖高速、城区、停车场的全域智能驾驶

智能安全辅助2.0
突破传统主动安全局限，360度保障安全



关于我们

弗若斯特沙利文是一家能够协助企业实现高速增长，优化投融资渠道，解决企业发展瓶颈的跨国咨询公司。我们凭借遍布全球 45 个办事处的增长顾问团队，精准识别数百个行业的增长机遇，积累了对全球价值链运作方式的深刻洞察。弗若斯特沙利文创新的市场进入策略和经过验证的最佳实践实施方案，已通过我们的“增长即服务”管道 (Growth Pipeline as a Service)，助力众多世界领先企业成功实现商业模式转型，使客户能够轻松创建并实施源源不断的增长机遇。

弗若斯特沙利文处于一个独特生态系统的中心，该生态系统融合了最佳实践辅导、高管同行支持社群和以增长为导向的内容。我们始终一心一意致力于通过有管理的增长重塑世界。