

一文读懂世界模型（World Model）：AI的“数字大脑”，通往AGI的核心基石

发布于 2026-05-08 19:45:12 1.2K 0 举报

在当下AI领域，“具身智能”和“通用 人工智能 （AGI）”已成为热议话题，而支撑这些技术落地的核心基础，正是“世界模型（World Model）”。

世界模型并不是某个具体的模型，而是AI理解物理世界、预判未来并自主决策的“内在思维系统”。简单来说，它让AI在“脑海中”构建一个虚拟世界，在行动前先“脑补”后果，再实际执行。

本文将以通俗易懂的语言，结合干货内容，全面拆解世界模型的起源、原理、应用及最新进展，尽量做到即便是新手也能轻松理解。

一、开篇先搞懂：世界模型到底是什么？

很多人会将世界模型和大语言模型（LLM）混淆，但二者的核心定位完全不同：

ILLM（如GPT、文心一言）擅长“读万卷书”，理解语义、进行推理，但不掌握物理规律——它无法理解“推杯子会掉”“走路会受重力影响”。

I世界模型擅长“行万里路”，能够理解物理世界规则，预测动作后果，并支持AI在现实世界中实际行动。

通俗地说，世界模型是AI内置的“虚拟仿真系统”，它能抽象复刻物理世界的空间结构、运动规律和因果关系，让AI无需频繁在真实环境中试错，就能在虚拟空间中推演、规划和优化动作，从而大幅提升决策效率和安全性。

世界模型之所以必要，是为了突破传统AI“落地痛点”，让AI从“纸上谈兵”走向“实际行动”。具体来说：

I传统AI局限：纯语言模型或简单算法要么不懂物理规律，无法应对动态场景；要么必须在真实环境中反复试错，成本高昂（如机器人摔倒、自动驾驶 碰撞导致硬件损耗），且存在安全风险。

I复杂环境挑战：真实世界场景多变且充满不确定性，人类无法预设所有规则，AI必须具备自主理解和预判能力，才能适应不同环境。

世界模型解决了这些问题，为AI搭建了“虚拟练兵场”，让其在无风险、低成本环境中学会适应世界、预测结果，这也是其成为具身智能和通用人工智能核心基石的关键原因。

核心特征

I预测性：能预判“做某个动作后，世界会发生什么”（如“踩刹车，车会减速”）。

I模拟性：内置虚拟世界，AI可以在其中尝试各种可能性，无需冒险试错。

I压缩性：将摄像头、雷达等高维数据压缩为核心特征，降低计算压力。

I泛化性：掌握通用物理规律后，能适应陌生场景（如学会走路后，在草地或石板路上同样可行）。

霞姐聊IT

LV.0

这个人很懒，什么都没有留下~

文章 60

作者相关精选

NVIDIA具身智能版图初探

VLA：通用机器人时代的核A

为什么做车的公司，更容易什

目录

▶ 一、开篇先搞懂：世界模型到底

二、追溯根源：世界模型是怎

三、核心原理：AI的“虚拟大脑

四、主流流派：世界模型有哪

五、这些突破值得关注

六、落地场景：世界模型早已

七、现存瓶颈与未来趋势

交个朋友

加入AI工作实战群

学习大模型落地案例 实操行业场景应用

加入AI趋势交流群

领券

加入AI学习入门群

换一批

腾讯云 Hy Token Plan 全新发布

最强最新自研 Hy3 preview 模型

复杂推理、指令遵循、上下文学习等能力大幅提升

Lite 版仅售28元

相关产品与服务

轻量应用服务器

https://cloud.tencent.com/developer/article/2666102

1/9

轻量应用服务器（Tencent Cloud Li
代开箱即用、面向轻量应用场景的云

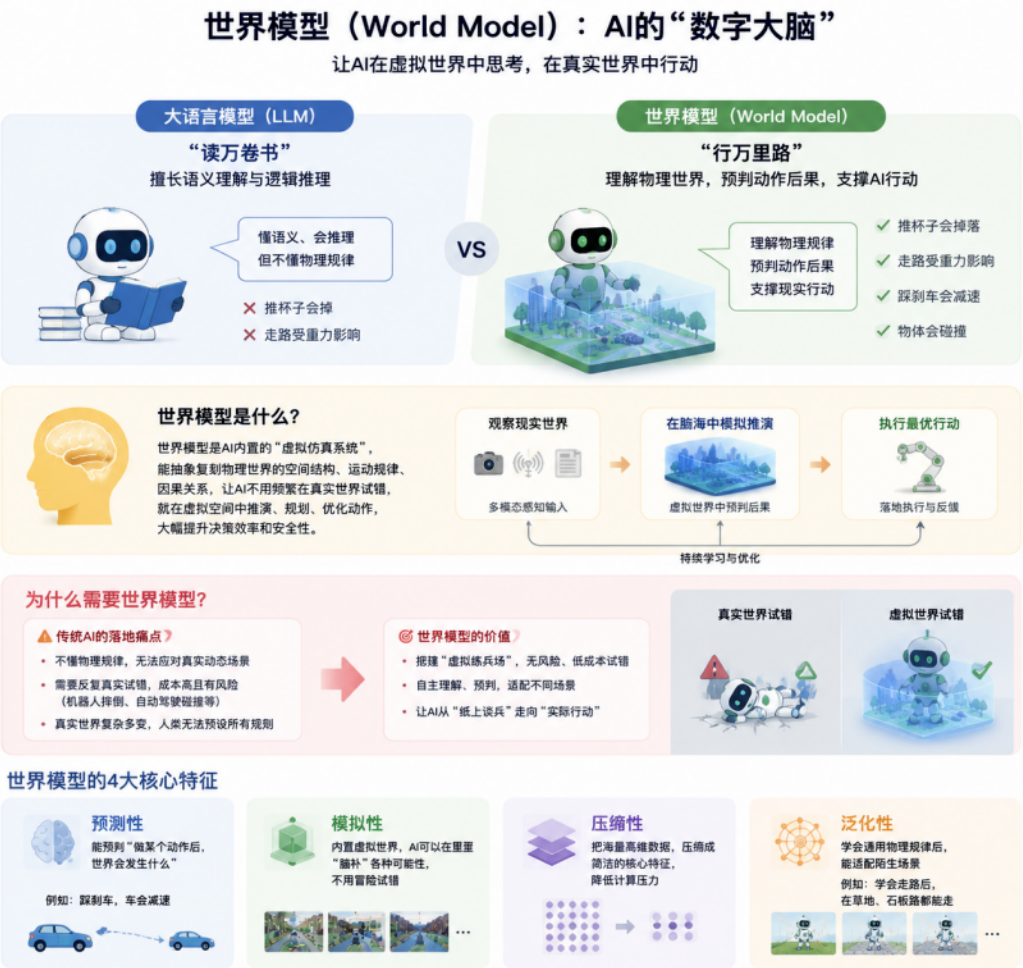
[产品介绍](#) [产品文档](#)

4核4G3M云服务器 新用户

加入讨论

用户12479753

使用AI学习新语言/能力，如
云+未来峰会上的超级大脑背
什么是AI Agent 安全？



二、追溯根源：世界模型是怎么来的？

世界模型的思想，本质是“模仿人类的认知方式”，其发展大致可以分为三个关键阶段：

1.早期思想：源于人类的“心理模型”

早在1943年，心理学家Kenneth Craik提出，人类在日常生活中依靠大脑构建“心理模型”来预测事件结果，从而进行高效决策。

例如，看到下雨会预判“不打伞会淋湿”，进而决定带伞。

这种内部模拟环境的理念，为后来的世界模型提供了思想基础。

到了20世纪80-90年代，AI研究者将心理模型思想引入机器人和 智能体 领域，提出智能体在自主行动时需要构建内部环境模型，以减少对真实环境的频繁试错。

这一阶段的研究为 强化学习 和模型预测控制的理论奠定了基础。

2.里程碑：2018年，现代世界模型正式成型

2018年，Ha和Schmidhuber提出了经典的世界模型架构——VAE+RNN。

该方法首先使用变分自编码器（VAE）将高维视觉输入压缩为潜在空间表示，然后利用 循环神经网络 （RNN）预测潜在状态的动态变化，实现“虚拟试错”。

这是世界模型首次从理论走向可落地应用，为后续所有基于潜在空间的世界模型奠定了基础。

3.爆发期：2022年后，自监督范式普及并深化

自2022年起，世界模型进入快速发展期。

其中最重要的突破是Yann LeCun团队提出的JEPA（联合嵌入预测架构）。

JEPA通过自监督学习优化潜在空间预测，使模型在高维、复杂环境中更加高效和稳定。

相比传统方法，它减少了计算冗余，提高了泛化能力，能够处理更多不确定性场景。

在这一阶段，世界模型研究呈现出几个显著趋势：

自监督学习成为主流：不依赖人工标注，能从大量无标签视频数据中自主学习物理规律。

多模态融合：结合视觉、文本和传感器数据，实现对复杂环境的综合理解。

端侧部署和轻量化：针对机器人和自动驾驶场景，世界模型开始在低算力设备上实时运行。

具身智能与交互增强：模型不仅能够预测状态，还能在虚拟环境中进行动作推演，支撑机器人、人形智能体等实际操作。

这一时期奠定了世界模型在具身智能、自动驾驶、工业仿真等领域快速落地的基础，同时为未来AGI的发展提供了核心技术支撑。

领券

三、核心原理：AI的“虚拟大脑”，到底怎么工作？

世界模型的工作逻辑与人类“观察-思考-行动”的流程高度相似，其核心由三个模块组成，形成闭环联动：

1.感知编码器：AI的“眼睛和耳朵”

负责接收外界的多模态数据，例如摄像头画面、雷达信息或人类文本指令。

通过CNN、VAE等技术，将这些高维、杂乱的数据压缩为简洁、有语义的低维特征，例如“杯子”“桌子”“距离1米”。

2.动态模型：AI的“大脑核心”

作为世界模型的核心模块，动态模型负责学习状态变化规律，

例如：“推杯子→杯子滑动→掉落”“踩油门→车加速”，

用于建模当前状态与动作对未来状态的因果关系。

常用技术包括RNN、Transformer和概率模型，既可进行确定性预测（如“踩刹车一定减速”），也可进行不确定性预测（如“雨天刹车减速距离延长”）。

3.决策控制器：AI的“手脚”

基于动态模型的未來状态预测，结合强化学习或路径规划算法，选择最优动作。

例如，“预判行人可能横穿马路，提前减速”“预判杯子会掉，伸手接住”，并将动作指令下发给执行单元（机器人或自动驾驶汽车）。

整个流程形成闭环：AI观察环境 → 编码特征 → 预测未来 → 执行动作 → 接收反馈 → 优化模型，使系统随着使用不断提高精度和适应性

四、主流流派：世界模型有哪些“技术路线”？

依据建模范式、技术核心与应用导向的差异，结合中科院、MBZ 未来实验室、新加坡南洋理工大学（NTU）、牛津大学（Oxford）联合发表的世界模型综述（2026），当前世界模型可划分为四大核心分支，各分支定位清晰、侧重不同，适配多样化场景需求：

1.潜空间世界模型（Latent-space WMs）

以低维潜在表征为核心，通过VAE、JEPA等技术将高维观测数据（如图像、雷达）压缩为简洁、有语义的潜在特征，并在该空间内建模环境动态变化与因果关系。

其核心优势是计算效率高、算力开销低，无需冗余的高维数据处理，适合机器人、车载设备等端侧实时部署场景。

典型代表模型有Dreamer系列和JEPA（杨立昆团队提出）

2.观测层生成式世界模型（Observation-level generative WMs）

不进行中间特征压缩，直接对原始高维观测数据（如视频帧）进行时序建模与场景生成，追求画面细节还原和物理仿真精度。

该分支保留高维数据中的完整细节，适合长时序视频生成、高保真虚拟场景仿真、数字孪生 等应用场景。

代表模型包括OpenAI Sora和英伟达Cosmos。

3.强化学习驱动世界模型（RL-driven WMs）

该分支以模型基强化学习（MBRL）为核心目标与落地导向，将世界模型作为智能体的专属虚拟仿真环境，通过内部无风险想象推演生成仿真交互轨迹，大幅降低真实环境试错成本、提升策略训练样本效率。

分支核心聚焦“模拟推演 - 策略训练 - 迭代优化”的闭环决策流程，可兼容概率动态建模能力，应对真实场景的随机扰动与环境不确定性，完美适配自动驾驶、人形机器人、工业智能控制等高安全要求的复杂决策场景。

典型代表方法包括PlaNet、MBPO（Model-Based Policy Optimization）、SLAC，这类方法以世界模型为核心支撑完成策略优化，不局限于固定的潜在空间建模架构，可灵活适配潜空间表征、直接观测生成等多种底层建模方式。

4.对象中心世界模型（Object-centric WMs）

以“对象”为核心建模单元，先对环境中独立对象（如车辆、行人、物体）进行识别与分割，再分别建模各对象的属性、运动规律及交互关系，最终整合为完整世界模型。

该方法在空间关联和因果推理上优势明显，可精准捕捉多对象交互的动态变化，适配复杂交通场景和多智能体协同任务。结合Transformer架构，可进一步提升长时序建模能力，

例如Google Genie 3可归入该类的延伸应用。

五、这些突破值得关注

近几年，世界模型技术发展迅速，核心突破可概括为七个方向，每个方向都在推动AI向“通用智能”迈进：

1.基础世界模型大型化与通用化

以 Genie 2（Google DeepMind）、HWM 2.0（DeepMind）为代表，模型能够从特定任务（如视频生成）扩展到通用世界建模任务，支持复杂场景的高保真可交互模拟。

2.自监督范式成为主流

以 JEPA 为代表，世界模型可以从大量无标注视频中自主学习环境动态和物理规律，大幅降低训练成本，并提升泛化能力，减少对人工标注的依赖。

3.大模型+世界模型“双脑架构”



LLM 负责理解任务和指令（如任务分解），世界模型负责动作推演与落地执行（如机器人如何抓取物体），二者协同实现端到端自主决策。

4.端侧轻量化与实时部署

地平线、黑芝麻、英伟达等推出专用芯片，使世界模型能够在机器人、自动驾驶和其他低算力设备上实时运行，无需依赖云端算力，从而推动量产落地。

5.物理引擎融合数据驱动，提升建模精度

通过将 Motive、HWM 等物理引擎的规则融入世界模型，可精准模拟物理现象，如流体、柔性物体、交通动态等，实现物理一致性和高保真预测。

6.少样本/零样本适应能力

借助大规模预训练模型，世界模型能够在只提供少量样本的情况下快速适配新场景，例如首次遇到新类型物体时也能预测动作结果。

7.国内落地加速，商业化验证

国内人形机器人、自动驾驶和仿真平台已开始量产搭载轻量化世界模型，在具身控制、交通预测和仿真规划等场景进行验证，逐步形成自主技术体系。



技术演进时间线（2018—2026）

- 1.2018 年：Ha & Schmidhuber 提出的经典 World Model，开启现代世界模型时代。
- 2.2022—2023：JEPA、Smallville、DayDreamer 等，强调潜在空间表示与自监督学习。
- 3.2024—2025：Genie 2、Meta AI、Sora、Vista、Voyager、HWM 2.0 等大模型，推动高保真、交互式和长时序建模。
- 4.2025—2026（预测）：Holodeck、RoboTwin、CitySim、DriveWorld 等，标志着可交互、长时序、多智能体、分层与生态建设方向的突破。

图表显示的技术趋势包括：

从特定任务生成向通用世界模型扩展、从高维生成到可交互操作、从监督学习向自监督学习迁移。

场景落地覆盖机器人、自动驾驶、城市仿真、游戏生成、科学发现等五大核心方向。

六、落地场景：世界模型早已不是“实验室技术”

现在，世界模型已经渗透到生活和工业的多个领域，其核心落地场景主要有五个，每一个都与我们日常和生产密切相关：

1. 人形机器人与具身智能（最核心场景）

特斯拉Optimus、优必选Walker X等人形机器人能够保持平衡、灵活抓取物体，依靠世界模型在虚拟空间中反复推演步态和抓取动作，再迁移至真实环境，从而大幅降低摔倒和操作失误的概率。

2. 高阶自动驾驶

L2–L4级自动驾驶系统离不开世界模型的预测能力：

它可以实时预测行人和车辆的未来轨迹，预判路口交通变化，提前规划避障路径，并在虚拟空间中模拟极端工况（如突发横穿），显著提升行车安全性。

3. 生成式AI（长视频与虚拟场景）

OpenAI Sora、Google Genie等生成式AI能够制作逼真长视频，世界模型保证视频中物体运动遵循物理规律（如苹果掉落、水流流动），避免出现“悬浮”或“穿模”等不合理现象。

4. 工业智能

工业机械臂的装配、焊接及巡检机器人的路径规划都可通过世界模型优化：在虚拟空间中模拟和推演轨迹，避免碰撞、提升效率，尤其适用于高危或复杂的工业环境。

5. 数字孪生与仿真推演

城市交通、能源系统和 [智能制造](#) 的数字孪生依赖世界模型进行未来态势模拟，例如预测高峰期交通拥堵、推演工厂生产流程优化方案，为决策提供可靠参考。

七、现存瓶颈与未来趋势

尽管世界模型近年来实现了快速迭代与规模化落地，但行业发展仍面临诸多技术瓶颈亟待突破：长 [时序预测](#) 过程中易出现误差累积导致的推演“跑偏”问题，复杂物理场景（如流体运动、柔性物体形变）的建模精度仍有待提升，同时端侧设备的算力约束尚未完全解决，一定程度上制约了其规模化普及。

结合当前行业技术趋势与前沿研究方向，未来世界模型将重点朝着以下五大方向稳步推进：

1.高维多模态统一建模：打通视觉、雷达、文本、传感等多源异构数据的表征壁垒，实现对物理世界更全面、更精准的环境理解与状态感知；

2.长时序无漂移预测：攻克多步预测误差累积难题，突破时序推演的稳定性瓶颈，实现小时级甚至更长周期的精准、稳定推演；

3.自监督/ 少样本学习深度普及：进一步强化模型的自主学习能力，持续降低对人工标注数据的依赖，大幅提升模型在陌生场景下的快速适配能力；

4.大模型与世界模型深度融合：持续深化“双脑协同”架构，推动二者在语义理解、物理推演、动作执行上的无缝衔接，成为通用具身智能体的标准核心架构；

5.安全与可解释性升级：将安全约束与伦理框架深度植入模型设计，强化决策过程的可追溯性与可管控性，破解AI“黑箱”难题，保障其在高安全需求场景的可靠应用。

八、最后总结

世界模型的核心价值，在于推动AI 从“单纯的认知思考”向“认知与行动兼备”跨越 —— 它如同智能体内置的“数字大脑”，赋能AI 真正理解物理世界的运行规律、精准预判未来状态、自主规划最优决策路径。

从2018 年经典 World Model 架构奠定基础，到如今大模型融合、端侧轻量化落地、多场景商业化验证，世界模型已明确成为通往通用人工智能（AGI）与具身智能的核心技术基石。

未来，随着各类技术瓶颈的逐步突破，世界模型将持续渗透到自动驾驶、人形机器人、工业仿真、数字孪生等更多领域，深刻改变人类生产生活方式，开启物理智能的全新发展阶段。

本文参与 [腾讯云自媒体同步曝光计划](#)，分享自微信公众号。

原始发表：2026–05–08，如有侵权请联系 cloudcommunity@tencent.com 删除

[自动驾驶](#) [机器人](#) [model](#) [模型](#) [数据](#)

领券

评论

期待你的精彩评论

0/300

发表

推荐阅读

编辑精选文章			换一批
万字详解高可用架构设计	15613	Go 开发者必备：Protocol Buffers ...	10451
10分钟带你彻底搞懂分布式链路跟踪	9216	多租户的 4 种常用方案	14215
亿级月活的社交 APP，陌陌如何做...	11360	60页PPT全解：DeepSeek系列论文...	12688

世界模型，大语言模型之后的第二场革命

基础 模型 视频 系统 机器人

在过去两年里，大语言模型（LLM）在语言理解、文本生成、推理、指令执行等方向取得突破，使AI在理解语义层面变得越来越可靠。



数据猿 2025/10/21 671 0

世界模型：不止是SeedDance 2.0的内核，更是AGI的核心基石

内核 视频 数据 机器人 模型

最近随着seeddance 2.0的出现，世界模型逐步走入大众的视野。当传统的大语言模型逐渐消耗完互联网的实时数据，性能走进瓶颈的时期，世界模型则作为一个核心概念，被无数行业和学术界所关...



赛博解生 2026/04/16 514 0

如何构建物理AI？NVIDIA给出了自己的答案

nvidia 部署 模型 数据 机器人

从ChatGPT到DeepSeek，从Sora到Cursor，多模态与大模型一路狂飙，不断刷新我们对“智能”的认知上限。但一个更现实的问题也日益显现——当AI面对的不再只是“文字、图像与代码”，而是现实...



数据猿 2025/10/21 485 0

为什么做车的公司，更容易做出人形机器人？

模型 数据 系统 自动驾驶 机器人

在过去的几年里，我们习惯性地将自动驾驶汽车和人形机器人视为两个不同的领域——讨论自动驾驶时，核心聚焦于车辆自主行驶能力；提及人形机器人时，则更侧重人工智能与物理世界的交互能力...

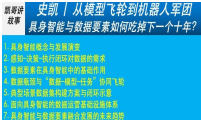


霞姐聊IT 2026/04/20 161 0

史凯 | 从模型飞轮到机器人军团：具身智能与数据要素如何吃掉下一个十年？

部署 服务 模型 数据 机器人

具身智能（Embodied Intelligence）是一种让“算法长在设备上”的前沿人工智能范式：智能体被置于机器人、自动驾驶车辆等物理载体之中，通过与环境持续互动，形成感知—认知—决策—执行的闭环...



凯哥 2025/12/24 825 0

NVIDIA具身智能版图初探

具身智能 机器人 nvidia 模型 数据

在具身智能（Embodied AI）和物理智能（Physical AI）的前沿技术中，NVIDIA无疑占据了一个重要的位置。



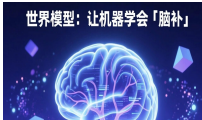
霞姐聊IT 2026/04/28 215 0

世界模型：让机器学会「脑补」

自动驾驶 游戏 机器人 模型 视频

这件事对人类来说太自然了，自然到我们从来不觉得这是一种“能力”。但对机器人来说，这几乎是最难的事情之一——在脑子里模拟一下还没发生的事，然后据此做出决策。

陆业聪 2026/03/10 241 0

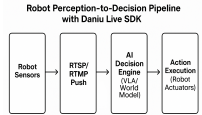


从「行走」到「思考」：机器人进化之路与感知—决策链路的工程化实践

机器学习平台 机器人视觉 机器学习 人工智能 机器人

近年来，人形机器人和具身智能频频登上舞台，从马拉松式的步态演示到竞技擂台上的“对打”，一次次制造着社交媒体上的话题热度。然而，炫技过后留下的核心问题是：机器人能否真正走出实验室…

音视频牛哥 2025/08/18 682 0



一文读懂具身智能（AI+机器人）

模型 数据 具身智能 机器人 行业

前段时间，具身智能领域出现了一个标志性事件——北京人形机器人创新中心开源的XR-1模型，首次通过国家具身智能标准测试。这个模型背后的UVMC（多模态视动统一表征）技术，堪称是AI+机…

臻成AI大模型 2026/02/02 714 0



10亿美金融资背后的真相：为什么世界模型是物理AI的“氧气”？

数据 机器人 架构 论文 模型

AI教父杨立昆离开Meta创办AMI Labs时撂下一句话：“整个行业都陷入了LLM思维的泥潭。”

AI 男神说 2026/03/02 278 0



具身智能明日之境：肉眼可见的科幻世界

机器人 汽车 模型 数据 具身智能

科幻正在成为现实。具身智能——有身体、会思考、能行动的机器人——不再是实验室里的演示，而是一批公司正在推向量产的东西。本文系统梳理了具身智能的定义与发展脉络，展示了当前 Tesla…

Jack. 2026/04/21 595 0

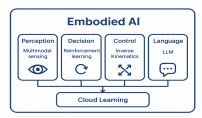


机器人 / 仿生 / 机械人身化 AI（Embodied AI）-----制造业智能化转型的关键引擎

机器人 AI 互动体验展

人工智能（AI）的早期发展主要集中在认知层面——如自然语言处理（NLP）、计算机视觉（CV）、推荐算法等，它们“思考”但不“行动”。

知孤云出岫 2025/10/14 940 0



浅谈世界模型（下篇）

人工智能

● 视频生成流派：Sora（OpenAI）、Genie（Google DeepMind）

languageX 2026/02/11 1.3K 0

深度学习进阶（六）——世界模型与具身智能：AI的下一跃迁

模型 数据 深度学习 机器人 架构

在过去的几年中，AI的发展速度令所有人目不暇接：从卷积网络征服图像识别，到Transformer统治自然语言处理；从ChatGPT点燃语言智能的时代，到Diffusion模型开启内容生成的革命。

海棠未眠 2025/10/22 1.7K 0

通通3.0、通脑发布！国家级平台定调AGI未来

行业 架构 系统 人工智能 机器人

“本次论坛系统阐述中国AGI原创路线，并展示了“通通 3.0”与“通脑”两大标杆成果

数据猿 2026/04/13 253 0



人形机器人的进化之路 | 2.5万字圆桌实录

机器人

本期为《仲夏六日谈》第四季第五期节目文字内容，主题为《从「行走」到「思考」机器人进化之路》。

小腾资讯君 2025/08/06 517 0

2026年AI五大硬核方向解析：从技术门槛到产业落地，开发者必看指南

人工智能 产业分析

随着大模型技术的持续迭代，AI已经从实验室走向产业深处，不再是单纯的算法博弈，而是与物理世界、科学研究、工程制造、数据服务深度融合的综合性赛道。对于腾讯云开发者而言，选对方向、找准定位，才能在AI浪潮中抓住核心机遇。本文将拆解...

math chen 2026/03/15 2.4K 0

具身人工智能全面综述

机器人 迁移 模型 数据 人工智能

Aligning Cyber Space with Physical World: A Comprehensive Survey on Embodied AI

CreateAMind 2026/03/11 793 0



从开普勒到牛顿：如何改进Transformer读懂物理世界

模型 数据 机器人 解决方案 论文

我们总在聊AI的“智能”，以transformer为底座，现在的AI模型已经写上万行严丝合缝的代码，能通过最难的律考和数学；能背下整本《经典力学》，一字不差地默写万有引力公式，能解复杂的天体运动...

赛博解生 2026/04/09 151 0



千寻智能解浚源：具身智能十年时间覆盖全球10%的人口 | 数据猿专访

部署 模型 数据 系统 机器人

如今但凡是标榜高科技的展会，总少不了人形机器人“来帮帮场子”。它们或是一记左勾拳右勾拳，或是一个漂亮的回旋踢，动作虽带着机械特有的笨拙感，但绝对是会场的顶流。作为暖场嘉宾，它们...

数据猿 2025/10/21 402 0



领券

社区

- 技术文章
- 技术问答
- 技术沙龙
- 技术视频
- 学习中心
- 技术百科
- 技术专区

活动

- 自媒体同步曝光计划
- 邀请作者入驻
- 自荐上首页
- 技术竞赛

圈层

- 腾讯云最具价值专家
- 腾讯云架构师技术同盟
- 腾讯云创作之星
- 腾讯云TDP

关于

- 社区规范
- 免责声明
- 联系我们
- 友情链接
- MCP广场开源版权声明

册

扫码

热门产品	域名注册 云存储	云服务器 视频直播	区块链服务	消息队列	网络加速	云数据库	域名解析
热门推荐	人脸识别 SSL 证书	腾讯会议 语音识别	企业云	CDN加速	视频通话	图像分析	MySQL 数
更多推荐	数据安全 网站监控	负载均衡 数据迁移	短信	文字识别	云点播	大数据	小程序开发

