具身智能如何迈过高质量数据“卡点”

▶ 本报记者 刘琴

作为人工智能（AI）与机器人技术结合的前沿领域，具身智能产业正在从实验室技术验证迈向实际场景规模化应用，但高质量数据的短缺成为制约其快速发展的核心瓶颈。如何破解这一难题，成为业内热议的话题。记者就此采访了具身智能领域多位专家。

在湖北人形机器人创新中心，工作人员通过训练机器人性能收集数据。 本报记者 刘琴/摄

数据采集为何难

“数据是具身智能产业的核心生产要素与‘燃料’，是算法、算力之外决定智能系统能力上限的关键变量。”湖北人形机器人创新中心首席运营官刘传厚介绍说，具身智能模型需海量多模态数据（视觉、力觉、触觉、运动轨迹、物理交互等）训练，数据的规模、质量、多样性直接决定模型感知、决策、控制能力与泛化性。当前，具身智能机器人与人类对话聊天时，有时会“答非所问”；跑步时，会因自身不稳而摔倒……这些都与缺乏高质量数据有关。

“对于具身智能而言，数据是连接虚拟与物理现实的‘桥梁’。”城市之间（北京）科技有限公司副总经理兼具身事业部总经理薛清恒表示，高质量数据决定了具身智能的泛化能力，只有丰富、高质量的数据才能让机器人在面对从未见过的场景、物体或干扰时，依然能做出正确的物理操作。

高质量的数据对提升具身智能机器人性能至关重要，但采集高质量数据面临多重挑战。

乐聚（深圳）机器人技术有限公司常务副总裁柯真东介绍道，采集高质量具身智能数据有五大难点。首先，采集成本高，需要专门的硬件设备和人员；其次，采集效率很低，采集和训练流程需要大量时间；第三，不同传感器有不同的数据格式和标准，导致数据难以共享和使用；第四，具身智能涉及多种感官和信息，采集和处理这些数据非常复杂；第五，行业标准不统一，导致采集到的数据质量很难保障。

“物理世界交互的复杂性、高成本与异构性等特征，导致具身智能数据采集难度极大。”刘传厚说，数据采集需部署人形机器人、机械臂、多模态传感器等专用硬件，而单台人形机器人成本高达数十万元，采集1万小时数据需上百万元软硬件投入。此外，真实场景环境复杂，具身智能数据采集需改造场地适配多场景，规模化场景部署成本高与难度大。

在北京人形机器人创新中心具身天工事业部负责人、具身智能机器人数据与训练基地负责人蒋未来看来，高质量数据采集主要面临三大难点：一是场景碎片化，真实环境千差万别，每个变量都在考验算法的泛化能力；二是不同构型的机器人传感器布局、关节自由度、控制接口各不相同，导致不同构型的数据难以迁移；三是数据质量参差不齐，数据采集涉及动作捕捉、多模态同步、人工标注等环节，任何环节偏差都可能产出“低质”数据。

采集方式各有千秋

据了解，目前主流的具身智能采集方式主要有3种：真实物理遥操作、仿真合成数据以及互联网视频学习/人体动作捕捉。

薛清恒介绍说，真实物理遥操作采集方式数据质量最高，但采集效率极低、成本极高，难以在短时间内获取百万级数据；仿真合成数据采集速度快、成本低，但仿真中的物理参数往往与现实不符，导致在仿真中训练好的模型在现实中往往“失灵”；互联网视频学习/人体动捕采集方式成本相对适中，但视频数据缺少“动作标签”和“力觉信息”，数据利用率较低。

薛清恒表示，现阶段，没有单一的有效数据采集方式，需要采用“虚实结合，预训练+微调”的复合模式。

蒋未来认为，人形机器人要想真正走进千行百业，就需要海量、多样、高质量的数据“原料”。真机数据是机器人智能从虚拟走向现实的必经之路，能够精准还原力觉反馈、触觉信息、环境干扰等仿真难以复制的细节。此外，真机数据还能有效解决“分布偏移”问题，机器人可以学习适应各种环境特征与突发情况，从而真正实现从实验室走向现实世界的跨越。

柯真东认为，更有效获取高质量数据的策略是分级训练和分层数据，即用低成本的数据做预训练，用高质量的真机数据做精调和落地微调。

在刘传厚看来，“仿真数据打底扩量+便携式轻量化采集+真机少量精调+Ego-centric（第一人称视角）多模态数据采集补全场景理解”的融合数据采集路径最有效，可平衡数据质量、成本与规模，是当前行业主流演进方向。

“三位一体”破解采集难题

如何破解高质量数据短缺难题，推动具身智能产业发展？

刘传厚表示，在技术层面，需要攻关仿真与迁移技术、发展数据生成与增强技术、创新轻量化采集技术等，从而提升数据供给效率；在产业生态层面，通过建设国家级数据基础设施、构建数据共享与交易生态、推动场景开放与协同采集等方式，实现共建共享，打破数据孤岛；在政策与资本层面，通过出台专项扶持政策、引导资本精准投入、设立产业基金等方式，支撑和引导产业发展。

刘传厚认为，各类企业要进行差异化布局。头部企业要聚焦仿真技术、世界模型研发，建设自有数据采集平台，主导行业标准制定，构建数据壁垒；中小企业要聚焦细分场景数据采集、标注服务，或参与开源数据社区，通过差异化竞争推动产业发展。

薛清恒表示，破解具身智能数据采集难题，需要构建数据基础设施、算法范式革新、产业协作机制“三位一体”解决方案。一是要构建标准化与开放的数据底座；二是在算法层面，大力发展“世界模型”与仿真技术，让数据采集从“劳动密集型”彻底转向“算力密集型”，还要加强机器人的“自监督”学习，让机器人通过自主探索生成数据，在与环境交互中自动学习物理规律；三是推动硬件层面的成本下降与一体化设计。例如在硬件出厂时就预置标准化数据采集接口和传感器套件，让每一台机器人都能成为“数据采集员”，形成“部署—应用—数据回流—迭代”的商业闭环。

“破解具身智能数据短缺难题的核心逻辑是，开源打破数据孤岛，普惠降低行业门槛，生态聚集开发者，闭环实现可持续发展，共同推动具身智能产业发展。”柯真东说。

湖南智能网联产业服务中心年内投用

本报讯 近日，湖南智能网联产业服务中心项目建设已完成联合验收。该项目计划今年年内投用。

作为湘江智能网联产业园核心产业载体，湖南智能网联产业服务中心项目由湖南湘江智能科技创新中心有限公司（以下简称“湘江智能”）承建，总建筑面积约3.1万平方米。该中心以智能网联大厦、产业展示中心等为核心构建智能网联产业集群载体，致力于打造全国有影响力的智能网联创新研发总部基地。待项目建成后，将集聚产业链上下游企业，提供研发、办公、展示等一体化服务。

此外，湘江智能网联产业园重点建设项目——湘江智能网联交通产业园项目一期8栋主体结构顺利封顶，砌体和外墙工程同步推进，已完成总工程量的20%。

据介绍，湘江智能网联产业园首开区安置房项目主体已于今年3月31日封顶，预计10月份完工。

配套道路建设方面，望雷大道箱涵工程已完成，路基完成85%；莲浦大道桥梁主体完工，路基完成60%，格宾式挡墙完成50%。道路建成后，将优化湘江智能网联产业园的路网结构，提升区域通达性，有力支撑园区开发与招商。

据悉，湘江智能网联产业园今年共安排41个项目，片区建设稳步推进。“我们将确保重点项目稳步落地，推动功能性项目在2026年如期完工并发挥实效。”湘江智能负责人表示。 李文晴

海口高新区AI创客训练营开办

本报讯 4月8日，海口高新区“火山智汇”人工智能创客训练营（第二期），在海南大学科技园举办。该活动以“AI养虾”为实战场景，聚焦人工智能（AI）技术在创业中的落地应用，吸引近百位创客参与探索“一人公司”（OPC）轻量化创业新模式。

此次训练营摒弃理论灌输方式，采用全程实战模式。人工智能专家邓锦宏以《AI养虾——带你手搓OPC一人公司》为题，手把手指导学员运用AI智能体搭建可落地的商业模型。

作为活动核心配套硬支持，海口高新区联合中国移动海口公司，为50位学员免费赠送价值千元的“AI龙虾”专属大礼包，涵盖三大核心权益：云电脑/云主机+OpenClaw（免费使用3个月）、移动云大模型CodingPlan（免费使用3个月）、2500万词元（Token）算力资源。所有礼包权益均在现场完成账号开通与“一对一”使用指导，确保学员即领即用、快速上手，助力创客们零门槛开启AI创业实践。

在交流答疑与资源对接环节，学员与讲师、中国移动技术专家“面对面”交流，有针对性地解决工具适配、算力配置、路径设计等实际问题。

据介绍，作为海南自贸港重点园区，当前海口高新区正通过“火山智汇”系列品牌构建课程+工具+资源“三位一体”的AI创业基础设施。未来该高新区将持续联合高校与龙头企业，提供更多前沿技术与实战机会，助力更多创客在AI领域实现创业梦想。 徐步遥