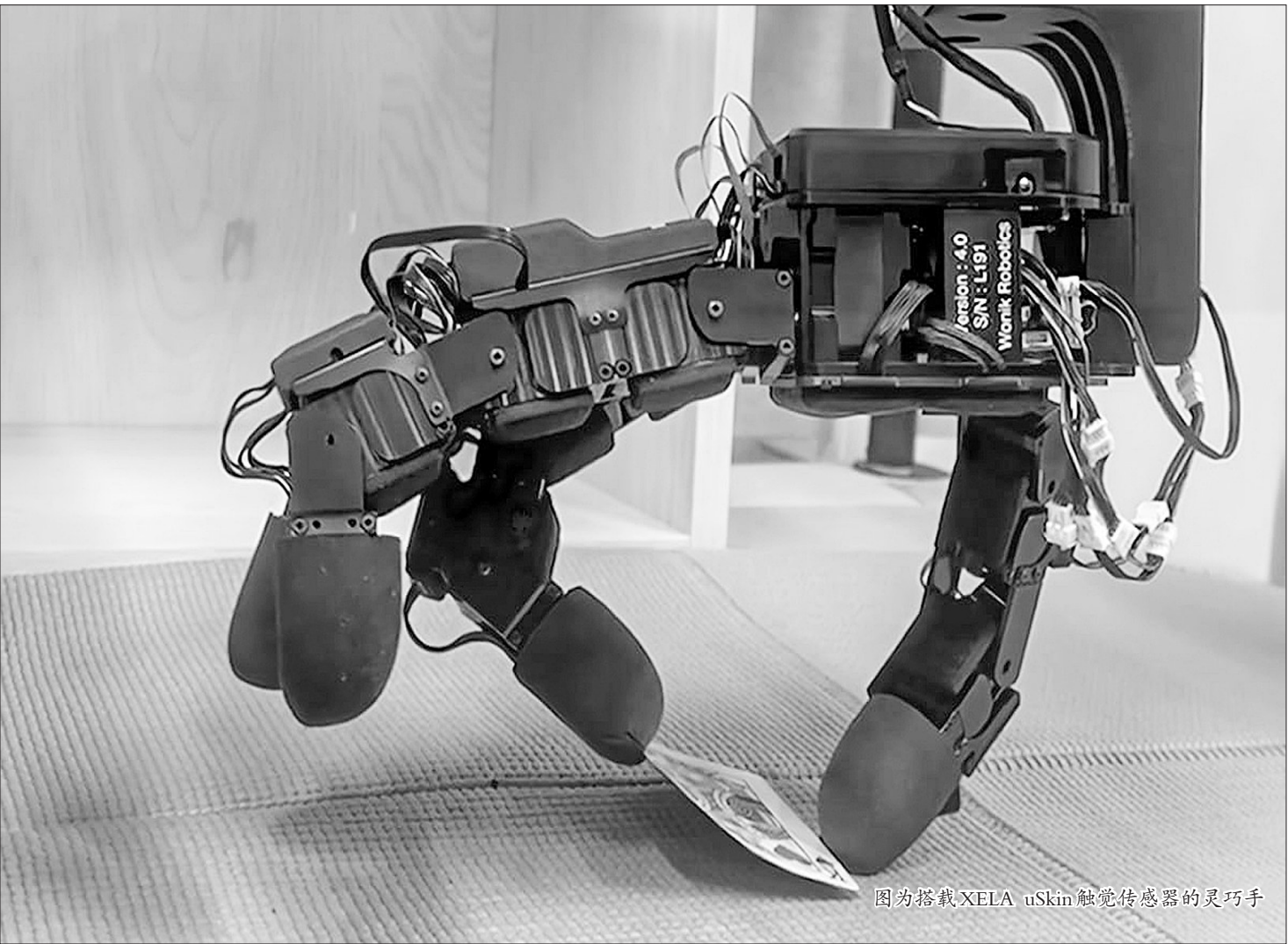


微观察

世界模型，卡在哪里？



图为搭载XELA uSkin触觉传感器的灵巧手

本报记者 姬晓婷 实习生 刘璐

8月21日，2026世界机器人大会主论坛上，进行了一场关于“世界模型到服务人类的现实距离”圆桌对话。就在这场对话开始前，中国电子学会世界模型专家委员会宣布正式成立，中国工程院院士、之江实验室主任王坚出任主任委员。这一动作强化了一个信号：世界模型——这个被寄望于让AI真正理解物理世界的核心技术，正从学术概念走向产业共识。世界模型被公认为具身智能走向落地的“核心基座”，但它究竟卡在哪里？正如浙江人形机器人创新中心首席科学家熊蓉在演讲中指出的——“物理世界具有不可约简性”，这就是世界模型当前面临的困境。结合大会多位嘉宾的观点，世界模型从实验室走向现实必须跨越三道坎。

世界模型不是大模型的延伸

要理解世界模型的卡点，首先要厘清一个认知误区：世界模型不是大语言模型的延伸。大语言模型基于互联网文本数据训练，擅长语言理解和生成，但它不理解物理世界。它知道“杯子”这个词，但不知道杯子是圆的、易碎的、会被打翻的。世界模型则要让机器人理解物理世界，理解空间关系、物体属性、因果推理和行动后果。奥比中光创始人黄源浩提出，大语言模型只能做语音交互，GPT指向“诗和远方”，世界模型才能落地干活；“干活”这件事需要空间智能，也需要视觉、触觉等高质量同步数据作为训练养料。大晓机器人董事长、商汤科技联合创始人王晓刚将世界模型的发展划分为三个阶段：第一阶段是数据增强，通过世界模型生成多样化训练数据；第二阶段是仿真闭环，结合机器人控制在虚拟环境中试错；第三阶段是具身智能，让机器人真正感知、理解物理世界并指导行动。从嘉宾发言看，行业目前仍处于第一阶段向第二阶段过渡的进程中。要达到第三阶段，至少面临三大卡点。

卡点一 真实世界数据的获取成本

世界模型落地的第一道坎是数据。世界模型需要大量真实世界数据来训练——不是互联网上的图片和文本，而是来自传感器、摄像头、机器人本体的多模态物理数据。问题在于，这类数据的获取成本极高。数据问题有多棘手？熊蓉直言——需兼具低成本、大规模、高质量。她指出，目前已有多渠道的数据采集方法，但如何提升数据质量，仍需持续优化。英特尔物理AI与机器人业务总经理Nagesh Puppala也提出了类似观点：“物理AI

需要传感器数据、真实世界数据，而不仅仅是互联网数据。”这与大模型的训练逻辑截然不同。大模型可以从互联网上抓取海量文本，但世界模型的数据必须来自真实物理环境——机器人在工厂里抓取物体的力反馈、在仓库里行走的视觉流、在装配线上操作的运动轨迹……这些数据集成成本高、标注难度大、隐私约束多。如何低成本获取大规模高质量的真实世界数据，是世界模型面临的第一道坎。

卡点二 Sim2Real——仿真与真实的鸿沟

但有了数据，不等于模型就能直接使用。第二个卡点，来自仿真与真实之间的鸿沟。在仿真环境中训练的模型，部署到真实世界后往往表现不佳。这就是业界常说的Sim2Real（即simulation to reality）问题——在仿真里面去学习，就会面临仿真和现实的鸿沟。为什么仿真训练出来的模型，到真实世界会“水土不服”？熊蓉分析了深层原因：“物理世界具有不可约简性。我们的专家模型无法用一套总结学习出来的规则去描述物理世界的复杂性。同样，现在的数据和模型能力，也没办法让它涵盖物理世界的丰富性和复杂性。”具体而言，仿真环境很难完整复刻真实世界的物理规律——摩擦力、弹性、流体动力学的模拟都存在近似误差；现实里

的光照变化、物体遮挡、环境反射又会给传感器带来噪声和不确定性；现实中大量长尾、罕见场景，也很难在仿真环境中全部穷举。为解决这一问题，业界正在探索多条路径。浙江人形机器人创新中心提出了“Real→Sim→Real数据飞轮”方案：先在真实环境中采集数据，通过高保真重构导入仿真环境，在仿真中增强训练，再部署到真实环境进行校正。熊蓉也提出“噪声注入训练”的思路——在仿真环境中主动向传感器数据注入光照变化、运动模糊、物体遮挡等各种噪声，使模型学会在不确定的环境中做出鲁棒决策。但这些方法仍处于探索阶段。Sim2Real的鸿沟，是世界模型走向落地的核心障碍。

卡点三 感知与表征——从看世界到懂世界

世界模型的第三道坎，在于“理解”。即便拿到了数据，跨过了仿真与真实的鸿沟，模型仍要真正“懂”物理世界——而这恰恰是当前世界模型表征层与感知层共同的短板。今天的多数世界模型，本质上是“看世界”而非“感受世界”。它们靠视觉表征学习，能识别物体的形状和位置，却难以表达物理交互的本质。熊蓉在主题演讲中指出，具身智能的核心在于与环境发生交互，这种交互“并不仅仅体现在视觉上”，“交互一定涉及力触觉，并且这种力触觉会跟形状、重量、材质、行为都是强相关的”，仅靠视觉很难还原这种物理交互的本质。XELA Robotics的Alexander Schmitz也强调：“视觉数据不足以抓取和操作物体，需要触觉数据来解决这个问题。”这意味着，抓取、装配、插入这类物理交互，依赖的恰恰不是“看见”而是“感受”。业界正在探索将视觉、力觉、触觉融合的通用表征方法，但距离成熟落地仍有距离。更进一步，模型即便“看见”了环境，也未必“理解”。熊蓉直言，语言视觉大模型“缺乏

人形机器人何时摆脱“遥控”

（上接第1版）

更大的看点在于，今年新增了乒乓球、自由搏击等高对抗类竞赛项目。这类项目的外部环境实时变化，机器人需要根据球体或对手的运动状态，在毫秒级时间内完成感知、预测、决策和动作执行，对系统级协同能力提出了极高要求。以乒乓球为例，机器人首先需要通过视觉传感器捕捉高速飞行的球体，估算其速度和运动轨迹，并迅速判断落点、选择回球策略，进而调动手臂、手腕和躯干在合适的时机完成击球。北京信息科技大学BiStrike战队队员李天乐在接受记者采访时表示：“完全自主决策，正是这个项目最大的难点。”

据运动会乒乓球项目技术代表吴飞介绍，今年乒乓球赛事的考验范围进一步扩大，由视觉识别、算法运算能力延伸到双足快速移动、全身重心动态调节、多关节协同发力等系统协同能力。参赛的人形机器人相互对抗，除挥拍外，还要完成走位、重心调节和多关节协同等运动控制动作，所有判断和执行均由机器人自主完成。支撑全自主对抗的，是机器人在高速感知、轨迹预测、实时决策和全身控制等环节的协同进步。开幕式上，与奥运冠军丁宁对打的智元远征A3机器人，搭载了由香港大学与超维动力联合开发的乒乓球算法。香港大学计算机与数据科学学院科研副院长、超维动力联合创始人罗平在接受记者采访时表示，团队采集了超过100小时的专业运动员对打数据，用以训练模型，并通过全身控制策略，使机器人能够在毫秒级内完成感知、预测、决策和动作执行。在比赛过程中，机器人的“大脑”“小脑”与机器人本体高效协同，从而快速完成击球。

在网球场地上，机器人面临的考验更加严峻。运动会开幕式上，北京银河通用机器人股份有限公司（以下简称“银河通用”）的人形机器人与网球冠军郑洁挥拍对弈，进行了多轮拉扯。面对高速来球，这台机器人可自主完成位置移动、落点研判与回击动作，实现灵活变向与网前截击，支撑起多拍连续回合。其间，机器人曾在摔倒后立刻站起，并重新进入对战状态，引起现场一阵惊叹。据银河通用方面介绍，其自研的“银河星脑”具身智能大模型首次将“大脑”“小脑”、神经网络集于统一模型中，可将对战策略直接转化为躯体运动，输出更拟人、更流畅的回击动作。从径赛到需要大小脑高度协同的乒乓球、网球展示等项目，本届赛事所呈现的能力维度，已经由单纯的运动控制扩展到动态感知、实时决策和全身协同。深圳市人工智能与机器人研究院具身智能中心主任刘少山认为，这代表机器人已经“开始从运动控制慢慢走向智能化”。这些变化背后，是视觉感知、实时决策、全身运动控制和本体硬件等多项技术的协同进步，也是机器人“大脑”与“小脑”协作能力不断增强的结果。

“全自主”走到了哪一步？

丢掉遥控设备是否意味着机器人已经能够摆脱对固定场景和针对性训练的依赖，直接适应全新环境？目前答案仍是否定的。赛场上的自主运行，建立在规则明确、环境相对标准化且经过反复训练的基础上，一旦场地条件、任务流程或外部干扰超出训练范围，机器人的感知、决策与控制能力仍可能明显下降。正如宇树科技创始人王兴兴给出的判断，在固定场景充分训练的情况下，机器人完成任务的成功率可以接近100%，但一旦物体、环境发生微小的变化，机器人的性能就会大幅衰减，这是当前具身智能模型普遍面临的现实。有记者在运动会现场注意到，在“打包入库”和“线缆连接”两个项目中，能够完整达成任务要求的几乎全部是遥控队伍，其间曾有全自主机器人不小心碰歪了打包箱，但未能即时调整纠错，导致后续所有药盒都放歪，暴露出机器人在突发情况应对和反馈环节上的不足。“线缆连接”项目的灵巧手赛队队员何伟介绍，团队在机器人本体到位后进行了约一个月的数据采集和训练，然而，赛场提供的线材比团队训练时使用的线材更软，这大大增加了机器人抓取、识别和对准接口的难度。线材软硬度的变化看似微小，却改变了物体受力后的形态，也使原有操作策略难以直接迁移。这些现象进一步说明，机器人已经开始从依靠人类实时控制，转向自主完成感知、决策和执行，但这种自主仍有较为清晰的能力边界。面对物体位置、材质和形态变化时，机器人能否发现原有策略已经失效，并在缺少人工干预的情况下重新规划任务，仍是其从“按程序干活”走向“根据现场变化干活”需要跨越的关键一道坎。与此同时，在一些项目上，“全自主”

尚未贯穿任务执行的完整流程。在径赛方面，机器人还无法做到自主起跑和平稳制动。并且在高速奔跑状态下，即便操作人员及时发出停止指令，部分机器人仍会因减速不及时而撞上终点缓冲挡板，进而失去平衡，重重摔倒在地。这意味着，目前径赛中的“全自主”主要覆盖定位、循迹和奔跑过程，距离准确捕捉并理解发令信号、瞬时状态切换以及根据实时速度、剩余距离和地面摩擦条件自主规划减速动作的理想状态，仍有一些技术环节需要补齐。

刘少山认为，“跑得快”体现出机器人在关节功率密度、机身轻量化、动态平衡、全身控制和高速运动能力等本体能力方面的提升。而判断什么时候起跑、如何停下，则揭示出机器人在自主研判能力、系统能力与安全性能方面的短板。

距离真正自主还有几道坎？

从本次运动会的赛场状况来看，场景泛化能力、自主纠错能力、长时间稳定可靠运行的能力，是目前人形机器人走向真正自主和真实场景应用需要跨越的几道坎。人形机器人继续走向更高水平的自主，首先需要练就的是跨越不同场景的泛化能力。王兴兴在2026世界机器人大会上指出，泛化能力不足，是当前全球具身智能面临的最大瓶颈。经过针对性的训练，机器人可以在标准化环境中高效执行任务，而一旦操作物体或临场环境发生细微的变化，模型输出与真实物理世界之间的偏差便会导致任务成功率大幅下降，这也是目前机器人在场景赛中完成度较低的关键原因。

而解决这一问题的核心之一，仍在数据。王兴兴表示：“行业需要海量高质量的人类与真机数据，并建立由大模型编程、仿真训练、实机测试和人机评价组成的自进化闭环。”随着具身智能逐步进入规模化应用阶段，模型训练对数据的规模、质量及场景覆盖面提出了更高的要求。然而，当前具身数据的采集、处理与流通体系仍处于发展初期，能够直接用于训练的有效数据供给有限，与行业千万小时量级的需求之间仍存在较大缺口。而真机数据采集成本高、积累速度慢，不仅需要机器人在真实环境中反复执行任务，还涉及设备损耗、场地搭建、人员操作和安全保障等投入，难以在短期内大量获取。北京智源人工智能研究院院长王仲远指出，行业正在仿真合成数据、互联网视频数据、穿戴设备采集人类行为数据等多条数据收集路线上并行探索。银河通用机器人股份有限公司创始人王鹤认为：“走向通用的关键在于融合VLA与世界模型的世界动作模型，以及仿真强化学习、人类数据、真机回流数据共同构成的数据体系；再通过只需普通视频、无须动作标签的低门槛训练，让各行业能够快速把通用模型训练成高成功率的场景模型。”

数据积累能够拓展机器人适应不同场景的能力范围，但现实环境中的异常情况难以被训练数据完全覆盖，机器人还需要建立能够发现偏差、判断原因并动态调整动作的自主纠错闭环。以本次场景赛中打包箱被碰歪为例，机器人需要重新感知箱体位置，进而调整抓取姿态、运动轨迹和后续操作。整个过程涉及状态感知、结果判断、原因分析和动作重规划，任何一个环节缺失，都可能引发“蝴蝶效应”，进而导致任务失败。而想要实现自主纠错，首先要依赖更加丰富的感知反馈。意大利锡耶纳大学教授、WEAR-T公司联合创始人Domenico Praticchizzo在2026世界机器人大会上表示：“人形机器人要完成精细操作，不能只依靠视觉，必须获得力、触觉和温度等反馈。”IEEE机器人与自动化学会主席Nancy M.Amato也指出，当机器人进入复杂、长期和非结构化环境后，核心挑战将转向如何进行任务规划与协同，并将全局结构化规划与局部反应式协调结合起来。

此外，更高水平的自主性还要经受长时间运行的可靠性考验。刘少山认为，当前需要重点提高机器人长时间运行的可靠性等能力，并建立覆盖感知、决策、计算和执行全过程的安全机制。他指出，机器人能够以很快的速度完成百米比赛，并不意味着它已经可以在工厂里稳定工作8个小时。当运行时间不断延长，电量下降、关节和电机升温、机械部件磨损、传感器漂移以及通信和算力波动，都可能造成感知精度下降、动作变形甚至任务中断，任何局部故障都可能沿着连续任务链条向后传递。

去年，人形机器人还在赛场上跌跌撞撞；今年，机器人就已经在跑道上多次突破人类世界纪录。人形机器人的进化速度超乎想象。明年，又会是什么样的一番景象呢？