

# 人形机器人加速演进

特约撰稿 李佳师

然而，一切并不是看上去那样一帆风顺。与此同时，上述特斯拉与斯坦福大学的机器人视频被质疑造假：并非完全是机器人的自主行动而是有人遥控操作(Teleoperation)；而被ChatGPT加持的Figure 01也被要求再进行一些“整形手术”。为何“领头羊”机器人公司要在视频上“动手脚”？人形机器人是虚假繁荣还是“真元年”已至？究竟是什么在困扰人形机器人产业？

## “形”之惑：要不要人形？

在被网友质疑特斯拉擎天柱叠衣服视频造假之后，特斯拉首席执行官埃隆·马斯克在X上坦承了擎天柱的“不完美”：动作并非完全自主操作。斯坦福阿罗哈研究团队成员赵子豪也在社交媒体平台上展示了阿罗哈“笨手笨脚”的另一面：锅放不进橱柜里，炒完虾的碗没放对位置。南京邮电大学管理学院讲师刘德文认为，阿罗哈的很多高级技能属于遥控操作而非自主技能。而且，就算基于自主技能执行的任务，阿罗哈现有的平均成功执行率也不到90%，距离大规模商用还有很长一段路。

海银资本创始合伙人王熠全表示，擎天柱和阿罗哈都是有人遥控的，演示视频只是吸引了一波流量，人形机器人的这波热潮主要体现在“演示和造势”上。拥有明星机器人Digit的美国敏捷机器人公司(Agility Robotics)总裁Shelton表示，在他们的解决方案中，没有一种会采纳5根手指、27个自由度(人的肩部到手，大概有27个自由度)的方案。“我们的一些竞争对手曾公开声明，他们使用5根手指基本上是为了品牌推广。”Shelton说道。

阿罗哈研究团队成员符梓鹏在接受媒体采访时坦言：“我们的公开视频和研究资料是希望能够推动机器人领域的发展，吸引更多人才进入这个行业，解决很多科学和工程难题。”

“目前，所有的人形机器人都面临一个悖论：越是非人形机器人不可的工作，使用范围越窄；越是常见的应用场景，越不需要人形机器人。所以打开机器人市场的关键因素不是长得像人，而是找到常见场景里的高频需求。”王熠全认为，“擎天柱叠衣服，只需要机械臂就可以做到，很少有人需要一边走动一边叠衣服。同样是做家务，斯坦福的阿罗哈就不是人形的，只是轮式底盘加上机械臂，一样可以做各种各样的家务活。”

“人形”是不是机器人走向“通用”的必要选择？王熠全的想法代表业界一部分观点。特斯拉机器人团队的工程师杨硕也认为，通用机器人其实是一个伪命题。“以人类生活中最普遍、研究最深入的机电系统——汽车为例，并没有一款汽车能够完全代替其他的汽车。跑车、轿车、卡车等车型适用于不同的应用场景，不同的使用场景。未来，我们也需要应用于各种场景的各种专用人形机器人。”

2024年刚开春，人形机器人明星款就接踵而至。

不到一个月的时间，特斯拉三次发布人形机器人擎天柱(Optimus)性能更新的迭代视频；谷歌DeepMind与斯坦福大学联合推出的阿罗哈(Mobile Aloha)也两次迭代更新；沉寂许久的波士顿动力放出一段人形机器人Atlas拾起并搬运汽车零部件的视频；Figure将其人形机器人送进了宝马工厂。

随后，英伟达、OpenAI相继宣布开发机器人基础大模型。2月23日，Figure宣布完成6.75亿美元的融资，同时达成与OpenAI共同开发下一代人形机器人模型的合作。一天后，英伟达宣布成立通用具身智能研究实验室GEAR，GEAR负责人范麟熙表示，未来每一台移动的机器人都将是自主的机器人与模拟智能体，将像手机一样无处不在。3月13日，Figure发布OpenAI大模型加持的机器人Figure 01演示视频，该机器人能听人话且可同时执行两项任务。



波士顿人形机器人擎天柱二代已经能做90度的深蹲动作

## 通用机器人：人形是最佳形态？

与此同时，业界专家与企业代表都更倾向于将更通用、更高阶的机器人造得“像人”。图灵奖获得者、中国科学院院士、清华大学交叉信息研究院院长姚期智认为，具身智能体应该具备三个方面：身体、小脑和大脑。身体的具身必须首先有足够的硬件，具有传感器和执行器，其次是小脑主导视觉、触觉各种感知来控制身体，完成复杂的任务，最后是大脑部分主导上层的逻辑推理、决策、长时间的规划，以实现用自然语言和其他的智能体、环境进行交流。

“关于身体部分，具身AGI(通用人工智能)最理想身体的形式，我们认为应该就是人形机器人。因为人类的社会环境主要是为人类而定制的，比如楼梯的结构、门把手的高度、被子的形状等，都是为了人类的形状而定制的。所以如果我们能够打造一个泛应用的通用机器人，人形是最好、最适合的形态，人形机器人能够适应人类的各种环境。”姚期智说道。

除了无须为人形机器人设定特别的环境外，刘德文认为，人形机器人拥有和人类

相似的外表和行为，更有利于被人类接受；且人形机器人可以使用人类的工具，使得人类不用专门为机器人配备工具，这也方便机器人根据人类的演示进行模仿学习，进而扩展到更多任务中。业内人士普遍认为，人形机器人是通用具身智能的最佳载体。

基于这些判断，不难理解，为何特斯拉执着于让机器人更加贴近人形。《埃隆·马斯克传》作者沃尔特·艾萨克森透露，特斯拉正在探索将人形机器人擎天柱与超级计算机Dojo整合在一起，以创造“通用人工智能”。擎天柱最早在特斯拉工厂完成零部件搬运等枯燥、危险系数高的工作。按照王熠全的说法，这样的能力只需机械臂加移动轮盘就能实现，但要让机器人像人一样自主干活儿，仍有很多挑战。仅以行走为例，擎天柱三次更新的核心是行走，目前步速为0.6米/秒，与去年12月相比提升了30%，但也只是健康成年人步速的一半。

另外一个执着于将机器人造得贴近人形的是机器人创业公司Figure，其创始人Brett Adcock认为：“Figure的人形机器人是AGI的最终部署载体。”在宣布与OpenAI达成合作三周后，Figure公司放出了大模型加持下其最

新更新的机器人Figure 01演示视频，它已经可以听从人类的命令，递给人类苹果，将黑色塑料袋收拾进柜子里，将杯子和盘子归置放在沥水架上。该机器人能够听懂人类的命令，自动将复杂任务进行分解，然后行动，而且行动速度接近于人类，该公司称其行为为自主操作。因机器人能力的突飞猛进，Figure被科技巨头们竞相投资，这家2022年成立的公司，现在估值已突破26亿美元。

## “脑肌”结合：从形似到神似？

不仅仅是“形”似，更需要类脑智能的“神”似。中国科学院院士、中国科学院自动化所研究员乔红几年前就曾表示，现有机器人系统的性能仍然受到传感器精度、系统重复精度和辅助机械性能的限制。相比之下，人类在感官反馈绝对精度低、行为控制和分散计算能力有限的情况下，可以实现比每个身体单元独立完成能力更加出色的整体性能。这主要得益于人类能够整合多个脑区的信息，并且能充分利用身体结构的优势。所以应该通过模仿人类的生理结构、行为特征、智能原理和控制机制，开展类脑智能机

## 一个更懂制造的“百事通”

(上接第1版)比如，滚筒洗衣机内外桶配重块就无法自动装配，大部分流程仍依靠人工完成，因此往往会产生工艺响应不及时、工人无法快速掌握新工艺、质量不可靠等问题。

王超表示，如今，卡奥斯工业大模型COSMO-GPT的出现，让家电制造向“智”造更进化了一步。

工业大模型的推理和决策能力可以自动生成精确的机器运行控制指令。从利用3D扫描技术精确定位零部件，到根据零部件的型号自动匹配装配工艺，再到完成设备的柔性组装，工业大模型就像柔性生产背后那双“隐形之手”，使整个生产流程变得井井有条。

“在工业大模型赋能下，工艺设计环节效率可至少提高30%，换产调试环节效率可至少提高50%。”王超懂憬道：“在未来工业生产中，人的参与度会越来越低，在大模型驱动下逐步向工业智能体演变，各种工业要素之间高度协同、高度自治，工业将进入数据与智能双轮驱动的新阶段。”

## 一名能PK老师傅的“新员工”

在卡奥斯模具(青岛)工厂的注塑车间里，当技术人员给大模型输入指令后，一台白色“铁房子”内便开始闪烁橙色的光。在长达数十秒的轰鸣声过后，灯光暂停，一件塑料洗衣机模具产品做好了，产品误差不过头发丝粗细。

“COSMO-GPT能够自动识别指令，然后从庞大的数据库中调取数百个符合要求的工艺参数，并立即指挥注塑机开始工作。这个过程如果靠工艺师来做至少需要一个多小时，还未必能够保证准确性。”拿着刚做好的模具，卡奥斯模具(青岛)工厂首席注塑工艺工程师郑彦博再次向大模型提出了新指令，而这次的工艺参数却有不同。

当把两件模具放在一起时，记者发现，第二次生产出来的模具缺了一块，是个残次品。郑彦博指着有瑕疵的产品对记者说：“‘注塑’是家电生产的重要工序，其生产过程看起来不过是模具开合，但背后却牵扯着温度、压力、成型周期、模具健康、能耗等复杂的工艺和参数。能够熟练掌握合膜诀窍技术的基本上都是身经百战的老师傅。如果让新手完成合膜任务，他们至少要经历数十次，甚至上百次的失败，每失败一次不仅代表多产生了一套废件，还意味着转产时间延长。因此，注塑机设备便成了只能依靠人工经验调试的‘黑箱’。”

而将“新手”培养成经验丰富的“老师傅”至少需要7~10年。为了尽快缩短“新老”差距，就必须打开这只“黑箱”，将老师傅们的技术经验转化为可量化的数据和指标，让主观经验升华为科学易用的工业智能。

为此，卡奥斯工业大模型COSMO-GPT结合模具工厂共同打造了第一款注塑领域的“专业工业模型”。

尽管这名“新员工”上岗时间不过数月，但其表现已经让人刮目相看。

“过去，在调试生产的过程中产生的不合格产品都是作为废件处理的，现在的调试效率提升了约50%，大大减少了浪费。”郑彦博表示，不仅如此，应用工业大模型后，产线生产效率提升5%，节能降耗35%，量产调试降费50%。大模型还能对设备的健康状况进行提前预警，减少设备故障。

## 一位随时解惑的“行业专家”

工业制造领域“一米宽，百米深”，所有的行业知识深埋在各个细分领域。例如，家电生产就要历经铸造、焊接、冲压、锻造、切削、热处理等一系列工艺，且不同工艺流程各有一套自己的生产“诀窍”。

当工业大模型成为家电智能制造的超级“大脑”，为保证专业性和实践性，还需要不断补充“营养”，让自己变得越来越懂工业，越来越懂家电制造的细节。

海尔数十年积累的海量工艺数据为卡奥斯工业大模型COSMO-GPT提供了基础养分，而这些工艺数据来自一代又一代人的经验积累。

以前，无数经验都沉淀在老师傅的大脑里，技艺的传承只能靠车间老师傅“传帮带”，操作人员要想熟练掌握这些技艺必须经过长期的实践与积累。如今，卡奥斯的研发人员把大模型终端和工作场景相连，抓取工人

输入的参数，分析比对操作数据和操作效果的关联度，再经过工程知识“淬火”，老师傅的经验就变成了成熟可用、能够量化的数据和信息，然后再将这些“数字化”经验沉淀成训练卡奥斯工业大模型COSMO-GPT的语料，帮助大模型攻克家电制造中的难题。这样，工业大模型就可以通过知识问答的方式向年轻的注塑专业工艺员进行“经验传授”，让他们轻松获取老师傅的经验。

现在，海尔的每一员工身边都有属于自己的“行业专家”，可随时答疑解惑。“哪些设备参数对注塑质量有影响？”“注塑产品出现毛刺、飞边的原因有哪些？”“停机时的模具应如何保护？”……当记者尝试着向卡奥斯工业大模型提出问题后，便迅速得到了专业回复。卡奥斯工业智能研究院大模型产品架构师李锡源骄傲地说道：“现在，它的答复已经和一位在注塑领域工作十几年的老师傅不相上下了。”

按卡奥斯的畅想，随着工业智能技术的发展，以及数据、语料的积累逐渐丰富，人工智能工业大模型将最终走向自主学习、自主工作、自主纠错，帮助家电工业制造实现真正的无人化生产。

“这是一个全新的工业知识扩散体系。”青岛市人工智能产业协会会长陈维强表示，以GPT为代表的大模型为工业生产提供了新的思考方式和工具，正在打开一扇新的发展之门。

## 荣耀发布人工智能战略架构

本报讯 近日，荣耀正式在国内市场发布全新人工智能(AI)使能的全场景战略，展示荣耀全新的AI战略架构。

荣耀终端有限公司CEO赵明表示，荣耀的AI战略将智能终端的AI应用分为四个层级。

第一层用AI使能跨系统融合，让不同的操作系统通过AI来作决策，实现手机、平板电脑、PC等设备之间的无缝连接和数据共享。

## 苹果公开自研多模态大模型进展

本报讯 据相关媒体报道，近日，苹果公司研发团队发布了一篇论文，首次公布了多模态大模型MM1，展示了该公司在人工智能方面的进展与实力。

据悉，MM1最高参数量为300亿，该模型支持增强的上下文学习和多图像推理，在一些多模态基准测试中有着较好表现。MM1的独特之处在于其庞大的规模和架构创新，包括密集的模型和混合专家模型。这些模型证明了研究人员方法的有效性，将大规模预训练与战略性数据

第二层用AI重构操作系统，让手机“越用越懂你”、越用越好用，基于意图识别的人机交互新方式就是一个典型成果。

第三层是AI在端侧的应用，比如抠图、拍照渲染、总结文档。

第四层则是AI的端云协同，将网络AI大模型应用呈现。

据介绍，截至目前，荣耀AI研发费用累积已达100亿元，AI专利成果超2000项。

(文 编)

选择相结合，以增强模型的学习能力。

该研发团队称，MM1不仅在预训练指标中是“最先进的”，并且在一系列已建立的多模态基准测试中，经过微调后也保持了“有竞争力的”性能。

业内人士分析认为，目前MM1的整体性能还没有完全超过谷歌的Gemini和OpenAI MM1的独特之处在于其庞大的规模和架构创新，包括密集的模型和混合专家模型。这些模型证明了研究人员方法的有效性，将大规模预训练与战略性数据

(晓 文)