

# 人形机器人开启“灵魂”竞逐

## 机器人基础大模型：“机器人变局”关键？

黄仁勋将GTC大会主题演讲的最后主角留给了机器人，宣布启动通用机器人003计划：“这是AI领域最令人兴奋的课题。”这预示着AI与机器人深度融合将成为英伟达最大的目标之一，具身智能的时代来了。

人形机器人技术在去年取得了巨大的进步，Diffusion Policy、ALOHA/ACT就是比较有代表性的成果。这得益于模仿学习的突破，让学界有了收集高质量数据的方法，并发现生成建模技术在机器人领域能运用得很好。同时在机器人运动控制领域sim2real(仿真到现实)也有了非常大的进展。斯坦福Mobile ALOHA机器人团队成员赵子豪表示：“总的来说，我觉得AI的进步启发了机器人的进步。”但他同时认为，目前人形机器人技术仍处于发展早期，距离商业化仍有很长的路要走。

那么当下人形机器人产业发展面临哪些挑战？

北京航空航天大学副教授、博士生导师陶永表示，硬件的挑战一方面来自核心零部件，包括高精度伺服电机、伺服驱动器、减速器、控制器，也包括各种传感器，比如双目视觉传感器、激光雷达、末端的力传感器、触觉传感器等；另一方面则是如何将这些硬件进行集成。不过陶永表示，更大的挑战是软件层面的，底层架构需要人形机器人实时操作系统，上层需要环境感知能力，不仅需要各种传感器(力觉、触觉、视觉)、灵巧手，还需要高精度算法进行实际的感知。同时，机器人还需要精密的运动控制和运动规划能力，这需要精准感知、精准控制，还涉及与人、与外界的交互，全身协同控制等。

南京邮电大学管理学院讲师刘德文表示：“目前发展人形机器人的最大问题还是AI的技术问题，人形机器人工作场景中需要处理速度更快、鲁棒性、可扩展性和适应性更强的通用型算法。”

而最为核心的挑战是缺乏“机器人通用基础模型”。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，目前最大的挑战是机器人不能像大语言模型一样，通过基础大模型直接“一步到位”，做到最底层的控制。此外，机器人的算力、多维度感知融合、数据也都是挑战。黄仁勋不久前在接受媒体采访时同样表示，机器人基础模型是机器人产业大变局的关键。“大模型能生成文字、生成图像，那么能生成动作吗？答案可能是肯定的。如果模型能生成动作，就可以通过理解意图，生成相应可用的关节。状态空间模型(SSM)是AI的下一个重点，如果SSM实现突破，人形机器人的技术突破就指日可待了。”黄仁勋说道。

陶永认为，机器人基础大模型将加速人形机器人的进化。比如原来只能在平地上行走的四足机器人，通过具身智能大模型深度学习，通过反复学习和测试，能加速其泛化和智能控制的能力，慢慢地可以迁移到雪地里、沙地上行走。英伟达推出的图形化仿真环境，就可以将人形机器人的步态、行走等，先在仿真环境中进行训练，生成一些控制方法，再迁移到实际的人形机器人上，从而加速人形机器人的研发。

“在大模型出来之前，我是不看好人形机

特约撰稿 李佳师

在3月19日凌晨举行的英伟达GTC大会上，创始人兼CEO黄仁勋以人形机器人压轴，推出人形机器人通用基础模型Project GR00T和一款基于NVIDIA Thor系统级芯片(SoC)的新型人形机器人开发套件Jetson Thor。就在几天前，机器人AI公司Covariant宣布推出80亿参数的机器人基础模型RFM-1，其联合创始人Pieter Abbeel表示，RFM-1模型可以让机器人更流畅地转向新任务，为全球数十亿机器人提供动力。

英伟达的GR00T、Covariant的RFM-1，加上此前OpenAI与figureAI宣布联合研发的新一代通用机器人模型，全球机器人通用基础模型的竞争正拉开序幕。



图为Figure公司的机器人Figure 01

人的。因为就算你训练它会干一百万件事，但还会有第10001件事它干不了，它对这个世界观察、模仿、理解基本上是没有能力的。”360集团创始人周鸿祎表示，但是大模型出来后，大模型成人形机器人的灵魂，成人形机器人的核心，所以人形机器人这个产业在未来两年会获得突破性的增长。

## 英伟达通用基础模型：秀的是“硬”实力

机器人基础模型是人形机器人的核心和灵魂，不难理解英伟达为何如此重视人形机器人通用基础模型。

今年2月23日，Figure宣布与OpenAI共同开发下一代人形机器人模型，一天后，英伟达宣布成立通用具身智能研究实验室GEAR。GEAR可能是全球最有钱的具身智能实验室。GEAR实验室负责人范麒麟透露，GEAR团队有足够的资金一次性解决机器人基础模型、游戏基础模型和生成式模拟仿真。他认为，2024年将是机器人之年、游戏AI之年、模拟仿真之年。据介绍，GEAR研发团队专注于四大关键领域——多模态基础模型、通用型机器人研究、虚拟世界中的基础智能体和模拟与合成数据技术。

英伟达有机器人处理器、机器人计算平台、机器人开发套件和模拟仿真平台，还有历经五年建立起的机器人产业生态伙伴，再加上这个最有钱的GEAR，因此，英伟达的人形机器人通用基础模型之路被业界看好。

3月19日，英伟达在GTC上宣布推出的GR00T英文全称是Generalist Robot 00 Technology(通用机器人技术00)。按照英伟达的官方介绍，这是一个通用基础模型，能够通过理解自然语言文本、语音、视频和现场演示，然后模仿人类运动或是其他技能。为了配合Project GR00T，英伟达还推出了Jetson Thor计算平台和新版的Isaac平台。前者基于英伟达的Thor SoC，由高性能CPU集群和Blackwell架构的GPU组成，搭载了Transformer引擎，算力达到了800 TeraFlop，GPU性能比上一代Jetson Orin高8倍。而Isaac是一个端到端的平台，用于开发、模拟和部署人工智能机器人，能够通过基于Isaac Sim的Isaac实验室，利用GPU加速功能在虚拟环境中模拟测试、训练模型。

## OpenAI 机器人基础大模型：强强联手软硬互补

已经和英伟达在“机器人基础大模型”展开比拼的是OpenAI。其方式是联手Figure。是的，OpenAI“最懂模型”，而Figure“懂机器人”。Figure被誉为科技圈大佬钦点的“机器人领域的OpenAI”。该公司成立不到两年，三轮融资投资人集齐了硅谷科技圈有头有脸的大人物，公司成立一年即拿出了人形机器人Figure 01。周鸿祎曾这样评价此二者的合作：Figure AI、波士顿动力等机器人公司主要用舵机

来控制和操作机器人关节，有一套小脑逻辑把指令分解成动作序列，来控制机器人如何往前走，如何能够去拿起物体，但它是基于一定预设规则的，如果在这些规则之外，它就不会了。有了OpenAI的GPT加持，机器人不仅有了大脑，而且能够理解人的意图，两家公司的合作还能够在大脑和小脑之间建立连接，让机器人在理解和控制执行，以及泛化能力上得到极大的提升。

事实上，OpenAI在成立早期就试图建立机器人团队，但做硬件与软件毕竟是两条不同的路。OpenAI产品与合作伙伴副总裁Peter Welinder说：“我们一直计划回归机器人技术路线，通过Figure我们看到了探索人形机器人在高性能、多模式模型支持下可以实现目标的道路。”

“机器人基础大模型与语言大模型、视觉大模型，以及文生图、文成视频还是有很大差别的。”陶永表示，因为要生成动作、生成控制能力，不仅是模拟仿真环境实现就能完成，而且从模拟环境到现实环境还有很多新问题，包括与硬件的结合，类似手抓、身体平衡、胳膊的控制等，需要进行调试，涉及精密电机控制、传感器控制等多学科的问题。

基于陶永的分析，我们就能理解为何OpenAI自己不能做机器人基础大模型而必须选择一家在机器人领域有经验和积累的公司，Figure公司的核心技术来自其CTO杰瑞·普拉特(Jerry Pratt)，他从1998年开始研究机器人，所以能够在公司成立

一年就拿出惊艳业界的Figure 01。

## “新进入者”最被看好：一切充满变量

看重机器人基础大模型这一人形机器人“灵魂”的还有谷歌DeepMind和特斯拉。去年7月，谷歌DeepMind推出了全球第一个控制机器人的视觉—语言—动作(VLA)模型的RT-2，只需要像对话一样下达命令，它就能在一堆图片中辨认出泰勒·斯威夫特，并送给她一罐可乐。日前，谷歌公布了最新的具身智能进展RT-H，最新版RT-H能通过将复杂任务分解成简单的语言指令，再将这些指令转化为机器人行动，来提高任务执行的准确性和学习效率。目前，其在一系列机器人任务中的表现都优于RT-2。

近日，AI机器人创业公司Covariant备受关注。该公司成立于2017年，两位联合创始人来自OpenAI，之所以离开OpenAI，是“希望将AI从比特世界落地到真实世界，打造像人类一样通过推理来适应现实场景的机器人产品”。3月12日，其推出80亿参数的机器人基础模型RFM-1，可以将其视为一个视频生成器(就像Sora)。只要输入拾取某些物体的命令，系统就会使用其训练数据(形状、颜色、大小等)来识别它面前与该描述最匹配的物体，然后生成视频，预测自己试图拿起这个物体时可能发生什么，并确定最佳行动方案。从目前看，RFM-1最大的亮点就是泛化能力，它可以接受更多数据的训练来进一步扩展能力边界，其扩展的目标是“任何任务”。

中国同样有机会在这场人形机器人基础大模型的角逐中跑出优异成绩。中国科学院院士乔红带领的团队最近推出了Q系列人形机器人，突破了高爆发一体化关节、AI赋能设计、机器人基础模型、类人柔顺控制等核心技术。“神经网络之父”Geoffrey Hinton最近在牛津大学的公开演讲中表示，智能的本质是符号和规则；生物方法(即神经网络方法)则大不相同。看来，生物方法战胜了逻辑方法。Hinton的判断，进一步佐证了乔红从生物性的角度去探索和发展类脑智能机器人的思路，有望带来更优的解题路径。

去年年底，由小米机器人、优必选科技、京城机电、中国科学院自动化所等联合组建的北京人形机器人创新中心有限公司成立，其核心目标是面向人形机器人核心器件、通用本体、通用大模型、运动控制系统、工具链和开源社区等人形机器人行业短板和痛点开展技术攻关。该创新中心透露，近期将发布第一代通用开放人形机器人本体。赛迪智库未来产业研究中心人工智能研究室主任钟新龙在接受媒体采访时表示，中国大模型领域积累的技术优势具备向人形机器人领域转化的潜能；百度、阿里等互联网龙头企业及优必选、智元机器人等机器人初创公司均已展开大模型赋能人形机器人探索；中国业界围绕物理层、智能层和电力层三方面协同发力，有望实现人形机器人关键核心技术加速突破。

是否还有其他公司和机构成为踢出人形机器人快速发展“临门一脚”的最佳射手？应该说一切皆有可能，AI和机器人的世界有太多的变量。

# 中国电子报

## 全媒体

权威性高 传播力强 覆盖面广 影响力大

### 融媒体服务



- 报纸出版
- 官方网站(电子信息产业网www.cena.com.cn)
- 官方微信(公众号cena1984)
- 官方微博(http://weibo.com/cena1984)
- 视频平台
- 视频服务(视频制作、在线直播、在线会议等)
- 平台推广
- 内参专报
- 行业报告
- 图书出版

### 会展服务



- 会议活动
- 专业大赛
- 展览展示
- 专业培训
- 政府服务
- 指数发布
- 编辑推荐
- 产品评测
- 企业定制
- 舆情监测
- 数据营销
- 招商引资

## 立足电子信息业 服务新型工业化

中国电子报社创建于1984年。目前拥有集报纸、网站、微信、微博、音视频、第三方平台等全媒体服务，集会议活动、展览展示、专业大赛、定制服务等会展服务于一体的立体化、多介质系列产品，是促进行业高质量发展的“喉舌”与“纽带”。

《中国电子报》是具有机关报职能的权威媒体。《中国电子报》全媒体面向工业和信息化领域，聚焦集成电路、新型显示、智能终端、信息通信、人工智能、物联网、工业互联网、移动互联网、大数据、云计算、区块链、应用服务等电子信息完整产业链。

《中国电子报》全媒体日均触达用户量超过200万。

国内统一连续出版物号：CN11-0005 邮发代号：1-29

地址：北京市海淀区紫竹院路66号赛迪大厦18层  
电话：010-88558808/8838/9779/8853  
传真：010-88558805



官方微信



官方网站

在这里让我们一起把握行业脉动  
www.cena.com.cn