QR 代码

AI 生成的内容可能不正确。**AI与具身智能圆桌会议研讨纪要**

**From**：手工川

**To**：手工川的朋友们

组织者：南川Mark

讨论参与者:

杨洪兵，灵生科技创始人，具身智能领域专家。

周永，灵心巧手创始人，专注于灵巧手技术。

线下：来自AI、机器人、自动驾驶等领域的从业者与研究者。

本报告基于圆桌讨论内容整理，不代表任何具体个人及机构观点立场，未引用外部文献。

报告撰写人：芊雨Qianyu

日期：2025.02.09

问题目录：

1. 中美具身智能发展对比
2. 具身智能的应用场景（如何嵌入家庭？）
3. 具身VS智驾：谁会更早落地？
4. 汽车跟人工智能行业间的联系
5. 未来具身机器人的发展规模
6. 具身机器人的泛化涌现能力
7. 视频数据集的地位和未来发展

八、 灵巧手的技术难点

九、 零样本小样本学习对于机器人能够用吗？

十、 建筑场景智能机器人的发展现状

十一、 智能机器人未来的迭代速度

**Ⅰ. 具身智能的现状与发展**

**一、中美具身智能发展对比**

现在宇树非常厉害，直接上央视然后走向全球了。在现在的中美环境下，中国机器人的整个生产环节是不是在整个全球的生产业里面已经奠定了自己的根基了？机器人的产业发展是否能大幅地进入一个快车道？

* 美国：主要集中在纯具身大模型领域，注重软件和算法模型的研发，如特斯拉的自动驾驶技术。
* 中国：更注重软硬一体化发展，从上游零部件到整机再到本体，全产业链齐头并进，发展速度较快。
* 中国在具身智能领域的进展速度超过美国，主要得益于其完整的制造业产业链和丰富的硬件资源。

**杨洪兵**:我是从2023年2月份开始去进入到具身智能这个赛道，那我本人之前最早是在百度和腾讯做了很长时间人工智能，后面是在国内一家做人形机器人公司工作过一段时间。最近这两年具身智能的变化其实真的可以用这个一日千里来形容。就是之前我们其实对于这个有预期，但是没想到现在突破整个这个速度会这么快，尤其是在2024年。2024年这一年的发展其实是超出了我们很多人的想象，包括去年我们跟宇树其实也是有一些合作，然后一起在客户侧去交付一些具身智能的一些产品。

我们从地缘的角度分析中美确实在进行机器人的竞争。从 AI 到机器人，机器人是下一个大家决胜的竞赛领域，但我们现在看有一个不同，在美国它的竞争更多的是在纯具身大模型领域。所以你看它很少有一家公司同时去做软件核心，当然特斯拉这种是例外，因为本身它之前有做自动驾驶的一些经历。在这个机器人领域，在我看来做纯具身大模型的企业占了它很大的一个比重；

而在中国目前大家都在做软硬一体化的事情，这里面可能有些做偏零部件，但做零部件的本身也可能去做本体和模型。做本体的也会做很多这种算法和模型侧的一些研究，而做模型本身可能也会做本体，因为在中国它其实更加全面，因为中国有自己的一个制造业的场景，又有非常多的这种硬件产业链，可以说全世界最全的一个制造业的产业链。所以可以看到在具身领域，中国现在整个的进展速度比美国更快，核心还是说中国本身它从上游的零部件到整机再到本体，它整个是综合全面、齐头并进的一个发展，那美国他现在做具身的人，之前也是做人工智能或互联网出身的，所以他更愿意去把纯软件或纯算法模型的一些经验带入到这个具身领域。那从我个人的角度我认为中国的这种发展路线相对于美国速度会更快。

把整个机器人拉回到十年以前，那波士顿基本上已经引领了世界上机器人的发展，包括日本的一些人形机器人也会比较先进，但是如果我们把视角再拉回到近两年，那中国其实是有非常多的这个领先，包括我们在今年春晚看到这个机器人非常灵活，从本体然后到算法具有优秀的稳定性和一致性，这其实是咱们现在目前一个非常重要的优势。

**二、具身智能的应用场景（如何嵌入家庭）**

对于具身智能整个发展链路来说，未来要让它完全走进家庭要解决哪些核心的问题？资源技术问题可能需要多久的预期才能完全解决？

* 工业与商业场景：当前具身智能主要应用于半结构化场景，如工厂自动化、物流等。
* 家庭场景：家庭场景对机器人的泛化能力要求极高，目前仍面临技术挑战，但AI陪伴类产品（如电子宠物）已开始进入市场。
* 具身智能的终局是实现通用机器人，能够完成多种复杂任务，但这一目标仍需较长时间的技术积累。

**杨洪兵**:这是目前具身智能圈讨论最多的一个问题，到底这个具身智能，大家目前看到有一个终局就是现在包括说像特斯拉，他去说未来三年这样一个十万台甚至更多以上，不断在拉高他自己的一个预期，因为他对这个事是很乐观的。那同时一些机构预测后面每个家庭如果说有两台，那这个其实目前是100万亿，到150万亿台，这整个市场规模从100万亿到150万亿。大家其实对于具身的终局，都认为说是非常笃定这个事情的，越来越笃定，包括说大家也认为具身是实现最终AI甚至 ASI 最终一个必经之路，因为它涉及到你和物理世界直接的互动和交互。但是对于具体怎么去落实？其实现在有几种，一种是说从工业场景去切入，一个是从商业去切入，另外是您提到的家庭场景。

那目前普遍认为从工业和商业场景切入其实有很多，现在大家也看到在北京和上海很多这种头部的企业就是这样，在过去两年里有的可能拿到几十亿规模甚至以上的资金，可以说粮草弹药是非常充足的，那他们为什么现在主打一些相对工业和商业的场景？虽然做通用人形机器人是要解决专用人形机器人解决不了的场景，但是从路径上讲我们还是首先去选择那些半结构化场景或者说不是完全非标，属于半非标的这种场景，因为如果我们完全切入到一个完全非标的一个场景，它对于机器人的泛化能力要求是极高的，那可能是我们的终局，我们目前从技术的角度上还有一定的 gap 在里面。这就是回到刚才您说的，如果是从家庭场景去切入那是属于泛化性要求最高的场景。我举个例子，就是当我们作为一个 C 端的客户去购买一个机器人的时候，是希望说他可以去帮你去炒菜，帮你去做饭，然后帮你去叠衣服，然后帮你照顾老人，然后帮你照顾小孩儿，所以你是希望他可以 do anything for you ，这种情况就意味着这个机器人的稳定性要求从硬件到软件是极高的，所以这是为什么是最终我们能达到一个愿景，这个确实是比较难。但是我们从去年2024年其实也观察到，也包括说我们圈里面也会有一种新的赛道，大家在新闻上也可以看到这个 AI 陪伴，属于 AI 硬件里面的一个细分，那这个里面我们认为也是有一些机会，这个机会主要在于说它是以玩具的方式去切入到这个家用市场，而这种市场有一个非常好的一个，我们叫取巧也好，或者说我们觉得是非常巧妙的一点是它并不是让机器人真正去完成一些功能性的东西，因为包括说我们灵生去做巨型大脑，我们其实是让这个本体公司通过巨型大脑去完成很多实际落地的这个这样一些任务，所以我们说得更加通俗一点，就是我们真的想让大模型去干活，所以这是功能性的。但是从去年下半年就2024年年底到今年包括 q1，大家也看到很多属于 AI 陪伴，这个包括在1月份在美国举行的 CES 里面，大家也看到很多一些小型的电子宠物，我认为它非常好的一点就是大家对他的预期没那么高，所以它反而也是去切入到家用市场的一种方式。

**Ⅱ. 具身智能与自动驾驶的技术对比**

**三、具身VS智驾：谁会更早落地？**

具身智能机器人和自动驾驶在技术上面来讲，哪一个能更快落地到具体场景中？自动驾驶虽然现在很多技术很成熟，但有一个很核心的安全性问题要求特别苛刻。而机器人的功能性要求可能会更丰富，但安全性要求可能没那么严格。那他会不会比自动驾驶更早去大规模地在社会上应用？

* 技术相似性

感知-决策-执行闭环：具身智能和自动驾驶都遵循感知、决策、执行的闭环流程，技术上具有一定的可迁移性。

机器人定义：自动驾驶汽车可视为一种专用机器人，而具身智能则致力于开发通用机器人。

* 技术差异性
* 任务复杂度：具身智能的任务复杂度更高，要求机器人能够完成多种日常任务，而自动驾驶主要解决从A点到B点的位移问题。
* 安全性要求：具身智能的安全性要求更高，尤其是在家庭场景中，机器人不能对人类造成任何伤害。
* 未来发展

具身智能的技术迭代周期可能比自动驾驶更长，但大模型的加持有望加速这一进程。

**杨洪兵**: 这个问题就是这体现了，其实这两个行业有很强的差异性，但是它这里面也有一些可以复用和迁移的，这也是为什么我们看到很多之前做自动驾驶的一些人其实也在2024年，包括我也可能还没有官宣，2025年也会有很多具身的相关项目。原来是属于自动驾驶这个赛道，然后会切换到具身智能这个赛道。其实为什么会发生这个情况？第一，就是我们从相同性上讲就是智驾和具身有可迁移性，其实可以理解汽车也是一种机器人，不过它是属于一种专用机器人。它所完成的是位移和代步，就是你去对一个汽车的预期就是要帮你去完成一个位移。但是我们对于一个通用机器人的预期就会更高。所以从这个角度讲，他们两个都是机器人，所以他们会有很多的相同性。从另外技术的角度讲，智驾和具身机器人它都是从感知，然后到这个决策规划，然后再到执行整个这个一体化的链路，其实它都是整个这样一个闭环。从这一点上讲，它也是有一定的迁移性的。但是其实我这里更想强调他们的一个差异性，就是因为很多人觉得说，这些智驾的公司是不是直接去做具身，它就完全可以迁入？这里面也有很多的 gap 在里面。

首先就是智驾，我刚才提到它是一种机器人，但它又是另外一种意义上的专用机器人，但是我们现在要实现的具身智能是一种更加广义的通用机器人。所以坦白说，它的这种要求其实是比智驾更高的。虽然说从安全上讲，具身也有安全性要求，但是它的要求可能并不是高速场景，它是属于低速场景，它这种安全在家庭里面可以达到非常高的，几乎是就是百分之百。我们经常讲四个九，其实它比四个九要求更高，就你不能对于人类产生任何的这种伤害，我们对于这种叫 corner base ，就是这种错误的这种错误率其实要求是非常苛刻的，就甚至具身比那个智驾更加苛刻，因为它的这种场景更多。

然后第二个就是说，从技术难度上讲，我们认为从某种程度上讲，其实具身要比智驾可能实现起来是更难的，因为最终这个智驾它要完成的还是一个从 A 点到 B 点，就是完成最终这个位移就可以了，但是我们让通用机器人可以帮你做日常中的几千几万种甚至更多这样任务他都要去完成，这其实是非常复杂和困难的。其实我们知道在13年特斯拉开始去做智驾，那最早的时候他买的是 MOBILEL 的一套技术方案。然后后来他也是去自研，其实到今天也是基本上有12个年头了，智驾到今天其实没有完全落实下去，所以我们认为具身肯定也有很长的，非常长的一个迭代的周期。但是这个周期是不是会比智驾更长，其实现在不太好说，因为有一定的争议，因为相当于是我这个智驾之前走过的一些弯路，其实现在也是可以被具身所去借鉴的。在难度上讲，具身可能会更难。但是有大模型的加持，也会进行加速。它其实有正反两方面都在去做催化和反应，所以我总体而言，认为这个具身，它所带来的这个变化可能会更加深刻，对。

**四、汽车跟人工智能行业间的联系**

很多家动漫厂把哪吒托举起来了，中国的制造业也有希望把具身智能赛道共同托举起来。很多OEM主机厂来到这个机器人的赛道，但是特斯拉是最早去实践的。这两年广汽，奇瑞甚至小鹏都开始杀到这个赛道里来，是否会成为具身这个领域里面一个重要的力量，其优势以及可能掣肘的地方会有哪些？Tier 1公司做汽车底盘控制的方向跟具身智能有哪些可交叉的研发？

* 汽车主机厂优势：技术积累迁移、产业链资源、资金与市场渠道。
* 劣势：具身智能更高的复杂度、激烈的市场竞争、可能需要长期投入。
* Tier 1交叉研发：底盘控制、预控制器、传感器融合等技术可复用于机器人领域。
* 未来趋势：产业分工、技术融合、市场拓展将成为主要发展方向。

**杨洪兵**: 这两个问题我觉得都挺专业的，然后因为汽车领域，我可能没有您懂，然后我就试着回答一下吧，我觉得第一个是我的答案是 yes，因为我刚才也提到了，其实智驾和具身本身就是说汽车和机器人，它其实是有非常多的这种可迁移性可复用性的，所以我认为目前如果把具身划分成几种力量，就跟这个金庸小说里划分帮派一样，但是这个帮派不是说我们要对抗，它是说根据之前的这种来源和它对于不同的路径分类。

第一种大家也看到，有一些是属于教授创业，那我认为他们也是有之前自己从国外带回来的一些专业知识，第二个就是属于大厂。我本人其实属于大厂，就是大家也看到，包括腾讯，华为，小米，我认为是大厂。第三类，刚才您说的包括之前做智驾的，就是之前我认为它本身也是机器人，那它从智驾转到具身，其实也是一条路径。所以我认为第一个肯定是 yes ，毫无疑问， 这肯定是很重要的力量，而且目前这个力量也在不断更加壮大和充实。

然后第二的话就是我觉得从从技术的可复用上讲，您说您是 Tire 1，我理解您是那我先问一下您是就是纯软件还是软件一体？

**线下**: 转一体，我们主要做监控底盘包括转向制动和悬架，再加上它的控制算法。

**杨洪兵**: 就我理解是预控制器？

**线下**: 预控整个底盘预控加上机械部件。

**杨洪兵**: 这些部件okok。我觉得肯定在机器人领域也会有类似的这个产品，但是现在为什么还没有专门的这种 TY，其实刚才我也提到了，就是因为中美之间有个差异，美国现在是纯软件的这种，包括只做具身大模型的，像 physical intelligence 其实是非常重要的力量，但中国我们可以看到，其实基本上我们都在做纵向整合，就是基本上你看到一个公司，它从一些零部件，然后再到它的本体，然后再到它的大模型算法，它都在去做，这是因为我们现在还是属于一种比较混沌和非结构化的一个状态，大家其实都在拓荒，但是到后面我认为也会出现一定会出现像汽车那样一个产业的一个分工，就是就刚才您提到的哪吒就是一个非常细化的分工，大家一起把一个很优秀的产品最后去生产出来，其实我觉得机器人之后也是一样，就最后大家可能每个人也会要去做自己非常擅长和自己能够长期持续去投入去做长期主义积累的这样一些领域，所以最后也觉得一定会借鉴汽车的一些发展的路线。

**线下**: 就类似拓扑和 tesla 的关系？明白！

**Ⅲ. 具身智能的技术挑战**

**五、未来具身机器人的发展规模**

周总提到了未来机器人可能会改变世界，马斯克也说未来他的机器人可能是人类数量的十倍到100倍这个规模，大家对这个规模理解会有什么歧义或者看法吗？

* 数量规模：机器人数量可能是人类的十倍到一百倍，甚至更多。
* 生活方式：机器人将改变工作模式，并能够充当人类情感陪伴的角色。
* 社会结构：机器人将提升生产力，重塑社会结构。
* 人机融合：人类与机器人将实现灵智共生，而非对立威胁，共同创造未来。

**周永**: 我觉得从我的看法来说，机器人行业对人类社会的改变，大家的想象力都不够，就是说大家都不知道它带来的冲击力会多大。在我的想象中，他真正到来时他看现在的生活，简直就像我们看一群猴子一样的生活。就从我的角度来看变化是太大的。未来有可能第一常见的是大家都说认为他不用工作了，但它是一个渐进式的过程，可能后来先变成一周四天，最后就完全不用工作了。有可能未来大家都不会恋爱了，因为你会觉得机器人给你提供情绪价值更多。他就真的是生活中的伴侣。然后还有一些机器人他表现能力简直就会像神一样，就机器人跟人会相互融合的。我觉得它的数量不会少，只会多，未来机器人的数量有可能比马斯克说的还多，我觉得机器人行业是人类有史以来最大的变化，就是怎么说他的这种冲击力和变化都不足为过。

**杨洪兵**: 我基本同意周总的观点，我也用一下我们公司的一个命名来补充一下，我们为什么取名字叫灵生？其实咱们中国有句古话叫万物有灵，生生不息，所以我们相当于是取了中间两个字“灵生”。为什么叫万物有灵，生生不息？有两层含义，第一，就是后面我们周围所有的你能看到的物体它可能全部都是机器人，它并不一定局限于说一定要有双臂，一定要有双腿，或者说我们上春晚的那种才叫机器人，就像刚才周总所说的它给我们带来的一种变革绝对是非常革命性的，就相当于后面我们周围遇到的所有的，万事万物全部可以把它理解成一种智能机器人的一种形态。第二个，我认为就是为什么我们叫灵，我们也希望说这个灵生是灵智共生，是我们人类和机器人最后能够实现一种灵智共生，我们其实不是想那种比如说威胁论，或者说谁来消灭谁，就是说我们和机器人能够达成一种互相融合。

**六、具身机器人的泛化涌现能力**

大模型尤其是像现在的语言模型跟小模型有一个很大的差异就是它足够聪明之后，训练材料里他没学到的东西他能涌现出来，能解决。那机器人这边现在大一点的模型的响应周期实在不行，一个机器臂挥来挥去要算半天才做出一个动作。不知道现在具身智能这里是不是像这种大语言模型一样“悟出”什么东西了，还是一定要见到过的才能动？

* 泛化能力：当前具身智能模型已具备一定的泛化能力，能够在未见过的环境中完成任务，但尚未达到自主思考的水平。
* 涌现现象：与自然语言大模型不同，具身智能模型尚未出现明显的涌现现象，仍需进一步研究。
* 三年内可能出现具身智能的“ChatGPT时刻”，推动技术突破。

**杨洪兵**: 对您说的，我们现在叫具身大模型，其实是有别于现在火爆出圈的 deepseek，这个它核心还是属于自然语言模型，或者说再加多模态。为什么我们这个叫具身大模型？其实我们是说在空间理解上和空间操作执行这两个层面上。这里面包括感知，也包括最终真正的控制执行。这个是希望它能发生像自然语言大模型一样，也就是当你投喂给机器人的大模型更多的数据的时候，它会在他没有见过的任务，没有见过的环境和没有见过的物体上发生涌现。那就目前来看，这个涌现能力还没有，但这是我个人的观点，也有可能有些学者它会有更先进的观点。但目前看还没有特别多出现涌现，我指的是在机器人大模型上。但是为什么大家都在往这个方面方向去努力，包括学界和产业，就是因为虽然说还没有发现特别明显的涌现，但是现在已经有了一定的泛化，就包括说像国内外的，包括我们灵生做的这个具身的大模型，就是在他没有见过的这种物体和环境中 应用。我打个比方，你可能给了他一瓶矿泉水，这个时候他可以进行一个抓取，那后面我可能给了他一瓶可乐，或者给了他一瓶盒状的一个牛奶或者酸奶它同样可以完成抓取。整个这个过程并不依赖于说我对于它的预编程，或者说依赖于做了很多这种预先的人为干涉。所以这里面完全是通过这种数据的投喂和模型的训练带来了它的一个泛化能力。为什么我们没提涌现，就是我们认为这个泛化还不足以说到达说它可以完成一些自主的思考，更加惊艳的效果目前还没有看到。其实包括这个周总他们做灵巧手那个也给了一个补充，我认为灵巧手其实也是非常重要的，说可以去做很多这个非常复杂的操作的一个非常关键的部件，我们灵生现在也在去用一些。

**线下**: 我也补充一下，就是你的泛化是不是可以这样理解，就是可乐瓶子，矿泉水瓶子，还是只要是一个装水的容器这就叫泛化。

**杨洪兵**: 当然它有几个维度，我大概明白你意思就是说它从一个矿泉水瓶，然后再到一个容器。那再到说，比如说我要推门出去，我要去抓门把手。那这个其实它是一层一层去泛化的，那最终就变成只要是这个我们人类可以去做到的这些相应的动作。

**杨洪兵**: 那我们可以举一反三，就是人类为什么说可以从一种动作可以去迁移到很多不同的这种动作，他就是因为人类这个思维能力，包括大脑和小脑，是有举一反三的能力的。

**线下**: 那你觉得这一天大概什么时候会发生，就这种思考能力能够真正的工程化地跑到真正的机器人身上了？

**杨洪兵**: 其实具身大模型，就目前这个预测大家普遍是觉得说三年以内可能会出现这样的，在机器人领域出现 ChatGPT 时刻。但是具体的时间确实不太好直接给出来，但目前我们相对还是比较乐观的，因为国内外其实每天都会有很多的领域的 paper。有这样的一些进展，而且还在有很多其他领域的人也会进到这个领域里面，其实现在这个力量是一种加速度的。

**七、视频数据集的地位和未来发展**

作为主要是做多模态数据集服务的公司，我们现在主要的客户是做生成式 AI 像视频模型。一方面，我们现在已经开始跟世界物理模型做训练，包括3d 的公司在谈一些合作把我们的数据引入到他们的模型里；同时我们也在跟一些人型机器人的公司也在交流，包括训练智能的涌现，他们的机器人感知世界时需要有泛化的能力。我们做数据比较关心的，从训练本身的一个技术路径来讲，是在视频的领域。机器人去理解这个世界要学习泛化，像我们人一样。而人类小孩理解这个事情可以看大量的视频，不用去世界上走就可以学到很多东西。英伟达他们用了2000多万个小时做了那个 cosmos 那个模型训练。

未来这个视频数据会以什么样的形式呈现出来？未来在训练过程中帮助机器人训练泛化能力的视频数据重要吗？

* 视频数据集的重要性：视频数据集为生成式AI模型提供了丰富的时空信息，使模型能够学习物体运动、场景变化和事件发展等动态特征。字节跳动和灵生科技在内的多家公司已经在利用视频数据进行模型训练，这表明视频数据在具身智能领域已经具有重要的应用价值。
* 能助力提升机器人泛化能力：视频数据能够帮助机器人通过类似人类的学习方式，如“身教”来提升其对世界的理解能力，从而更好地完成具身任务。而未来“言传”，即通过自然语言与机器人建立沟通并进行训练的方式也会越来越重要。
* 有企业和学者用视频数据与一些非具身数据去训练世界模型生成器，从而生成更多的数据再喂给具身机器人。

**杨洪兵**: 是非常重要，就是您说的，当然也能看到您这个领域是非常专业的，其实并不是说未来只会更重要，事实上是现在很多具身的模型已经在利用了这个视频的这些数据，然后再去进行这个训练。但是比较典型的像字节的那个巨型的大模型，其实包括我们灵生自己的模型，我们现在也在去使用一些视频的数据，具体大家也可以关注一下我们灵生科技的这个我们的视频号和公众号，我们也有一些介绍。

刚才你提的很好，人类的就包括我自己的小孩，我们经常讲言传身教，首先言传是指，比如说你在告诉他一件事情是怎么去做，或者他做错的时候，你给他去纠错，这是通过自然语言。这也是我们后面认为和机器人真正去沟通一个最重要的一个方式，不是通过触屏，也不是通过键盘和鼠标是通过自然语言，所以我们叫言传。第二个维度，我们叫身教，所谓身教，就是您说的视频。因为我们人的眼睛看到的所有东西你都可以理解成实时的一个视频。也就是为什么说人类完全可以通过在视频里面学习到大量的技能和任务，其实就是身教，就是以人做示范。这也是为什么说我们认为要从这种人类的这种视频里面去提取这样一些数据，把它映射到机器人本体上，然后让这个机器人的本体训练最后完成最终这个具身的一个突破。我们认为这真正开始是在2023年底左右，那去年一年其实有大量的成果出来，我认为这个是后面非常重要的一条技术路线，甚至是最重要的几条路线之一吧。

当然，您刚才还提到了那个就是 nvidia 那个世界模型实验模型，它跟这个不同，就是有两种，一种是说我直接从大量的这种人类视频中去学习这个操作的这个数据，然后映射给具身。第二种是什么？就是我从人类视频数据里面学习到这种经验，我用来去训练什么？我用来去训练一个世界模型的生成器。您说的包括英伟达，包括我们也知道就是李菲菲的那个实验模型，包括国内我们现在也知道有很多团队在做很多，我们所知道这个清华有些实验室也在做这件事情就是相当于是他用这些视频数据再结合着其他的一些非具身的数据用来去训练这个世界模型。那这个世界模型后面又可以去生成大量的那种场景数据和操作数据。这些数据可以进一步再去给具身使用，它也算是两条路线，但是其实属于殊途同归。

**八、灵巧手的技术难点**

灵巧手现在贴合有很大的问题，很多灵巧手其实在力学上是没有传感器的，后面产品要怎么考虑？第一个是贴合的问题，第二个怎么跟力学做结合？

* 目前很多灵巧手在力学上缺乏传感器，导致贴合能力不足。灵生科技的产品默认配备触觉传感器，认为数据的重要性高于成本，即使要免费提供触觉功能。
* 具身智能的资源困境

1. 资源不足：具身大模型的资源远少于语言大模型，许多公司甚至没有足够的计算资源（如100个GPU）。
2. 数据稀缺：具身智能需要大量多模态数据，尤其是力学数据，但目前数据极度匮乏。

* 技术路线选择

1. 强化学习与大模型的平衡：灵生科技更倾向于强化学习，而非完全依赖大模型。同时，公司也与大模型团队保持合作。
2. 数据与规则的权衡：具身智能需要平衡多模态数据和物理规则。一方面，数据量需要极大增加；另一方面，理解物理约束和规则同样重要。

* 未来发展的关键

1. 大规模数据的需求：具身智能的真正泛化可能需要百万级的人形机器人提供数据支持，目前数据量远远不足。
2. 多模态数据的重要性：除了视觉数据，力学数据（如张力、肌肉力量等）同样重要，需要更多维度的数据来支持灵巧手的训练和优化。

**周永**: 我们的产品默认都带了触觉了，因为我们认为数据很重要，我们是没有不带触觉的版本，我们认为我们发过去了，如果是互联网模式，你都免费给用户去用，可以产生有触觉的数据，所以我们都是把这个带过去了。如果问刚才那些问题的就可能不一定是这个行业内的。第二点就是我认为具身大模型的公司，他们的资源应该比真正做语言大模型的少很多。

第一就是我们不一定走大模型的路线，我们是会偏强化学习一些。当然我们也有合作的很多老师做大模型的。整体上来说就是我认为有一些做大模型的，他们的路线实际需要的数据量是现在的一万倍。我给大家说一下，就比如说大家觉得2001年会诞生一个语言大模型吗？那个时候互联网网页都很少，youtube 都没有诞生，这是一个。第二个，比如说特斯拉，它是有100万辆车在那里跑的。大家去年中国真正交付的人形机器人可能不会超过3000台。很有可能对于置身大模型，我认为真正泛化的具身模型有可能是1000，100万个人形都在那里跑了，有100万个人形支持！

目前数据还是缺的厉害，第一是缺多模态数据，比如力学，我到底有多少张力，我肌肉有多少，多少秒，我抬下手有多少，本质来讲它需要大量的数据维度，视觉本身是辅助看视频这件事情，底层是辅助。理解多模态是一方面，理解物理规则，物理约束是一方面，才能形成动力学完整的途径。它不像我们做大语言模型，本质上是一个概率分布的一个召回。只要你要达成大模型达成六次法则，黄金法则相对而言，你达到一定的临界值就可以了。但是具身智能本身来讲，这个事就比较麻烦了，你究竟是以规则为中心，还是以数据量为主，以多模态数据量为中心？你要用规则的话这就没完没了，但是本质上讲很多场景都是以规则为中心，比如生物学要把你是我们把所有的 DNA 数据全部录入，那是世界上的所有电脑装不下。但理解它本身的生存规则就好。就这个事是这么回事，所以在某种意义上我们还是在跳过这么一种情况和控制算法，来做一些强化。

**九、零样本小样本学习对于机器人能够用吗？**

对于市场端的需求之前了解到一些工厂会比较看重零样本小样本学习，还有这个算法跟这个端侧的算力也是有一定的联系的，比如说端侧算力它支持什么样的算法，然后工厂产线的话目前我们看下来还是主要以机械臂为主诞生的，最后才是这个纯人形的，想问一下小样本和零样本学习，目前这个技术可能解决哪些问题，是否有希望？还有3d算法目前的数据格式是不是以高速扩建点这样的为主？

* 工厂场景的可行性：零样本和小样本学习在工厂环境中是可以实现的，因为工厂的任务相对固定且数据需求较为集中。
* 技术路线：

灵生科技采用小数据结合强化学习的路线，能够有效解决工厂等细分领域的具体问题。

预训练与微调（FT）：通过大规模预训练模型（使用大量无监督数据），然后针对具体任务进行微调（Fine-Tune），实现小样本学习。预训练阶段对数据质量要求不高，但数据量需较大；微调阶段则依赖少量高质量数据。

* 核心难点：工厂环境中，零样本小样本学习的核心难点在于提升精度和高频控制能力，而非学习样本数量。

**周永**: 好的。我觉得工厂没有太多难度，因为我觉得有难度的就是你做一个保姆的。就工厂的，零样本小样本我觉得都是可以解决的。因为我们本身自己是用的小数据，再加上强化学习的路线。在最后真正做出来一个通用大模型，就是类似于现在语言大模型，我觉得这是很有必要的。但是如果你只是解决某一些细分领域的问题，你说的那些我个人认为就是只要技术水平高一点都是能解决的。但是整体上来说，我还是非常乐观，我认为具身大模型可能得三到五年后，有可能一到两年内机器人就会很普及了，包括一些偏机器人形态的智能硬件可能很快就进入大家生活中了。

**杨洪兵**: 对零样本小样本我理解是这样，因为我们也是会先把这个模型做预训练，就是阶段会投喂给他，数据质量要求没有那么高但是我的数量级会比较大，用这样的数据去做预训练，那有了这样的训练之后，我们在真正去做某一件具体的任务的时候，这个时候会去做 FineTune微调，所以在 FT 的时候，其实就是小样本的一个学习，也就是通过在于训练阶段大量的这种相对于无监督的数据的这种投喂，再加上做具体任务。包括您说在工厂里面，它的核心的难度其实并不在于通过学习整个流程，通过小样本学习一定是可以做到的，它核心的点刚才有位同学也提到了，跟我们目前的这个观察是一致的，就是它核心其实说在真正的工厂里面它的精度和高频的控制上，其实这里面提升的空间会更大，对小样本学习其实并不难。

**周永**: 对你说的我们到时候也可以对一下，因为我们其实现在就是好几个客户，也就是用我们的灵巧手来做，你说的那个组装厂就是。

**十、建筑场景智能机器人的发展现状**

建筑机械的工作场景比较危险但也比较单一。我们希望能从自动驾驶的视觉系统上获得一些突破。视觉系统软硬件训练方面现在最先进到达了什么程度？从机器人的视觉系统来看，如果要适应建筑工程工地的场景的话，现在大概有一些什么样的差距？

* 现状：建筑场景的研究相对较少，验证过程较慢。目前首要任务是解决半结构化和复杂环境中的问题。有友商在研究机器人用于建筑任务（如墙壁粉刷），类似于工业场景。
* 核心需求
* 精度和准确性：必须保证任务的高质量完成。
* 控制速度：提高工作效率。
* 泛化能力：避免过度依赖预设程序，适应多样化的建筑任务。
* 目前技术仍在发展中，尚未完全成熟，且需要考虑经济性。成本和投入产出比是关键问题。机器人技术需要在经济性上具备竞争力，以满足建筑行业的实际需求。

**杨洪兵**: 您说的建筑这个领域我们目前所接触到研究并不是特别多，因为建筑这个场景它不是特别能够很快去做验证的，核心还是刚才我提到我们现在首要的去解决的，还是更多的去解除半结构化和碳飞镖的一些问题。那在结构化场景里，我们倒是知道有一些友商和一些伙伴在做用机器人去做这个墙壁的粉刷，类似场景有友商在研究。但是我们觉得它跟工业有点类似，就是说通过学习，不管是说包括触觉的这样一些联合数据的学习，它其实是可以解决你让他去做的某些动作，某些建筑的一些流程和工艺，它其实可以解决，在这里面核心的点是它在同时要保证，第一，我要保证它的这个精度和准确性。第二，我要保证说它的控制的这个速度，你理解它可以更高效率地去工作。 第三，还要解决一定的泛化，因为我觉得您一定不是不希望说我们做具身通用机器人。它所做的事情如果是我通过一个机械臂，或者我通过一个完全专门去为建筑设计的一个机器人，所有东西都是预先设置好地去做那样的事情，其实一定不是您希望的，因为他就是那种是完全没有任何泛化。所以这里面核心的逻辑就是它要解决的是我怎么能够在精度和控制频率以及泛化三者之间都能达到一个比较好的效果，其实我认为大家目前还是在进展中，还没有说现在有一款机器可以把这三个同时解决的很好。所以这里面三个最终一定带来的是成本的问题，投入产出比，因为我们知道现在大家希望通过这种机器人去解决更高成本和更高这种效率地进行工作。您这个产业上一定是比我更加专业的，他最终一定是，直白点就是一定还要去算整个的经济账。

**十一、智能机器人未来的迭代速度**

（1）未来五到十年，或者说三到五年内，具身大模型做出来了，那么人形机器人可能是会灭掉一批的。因为据说硅谷没有人在做大模型，都在做各种小场景的落地。类似的东西可能百分之八九十在未来的五到七年就被淘汰了。我们国家的综合竞争力在哪里？

（2）大概多少年可以淘汰多少，被一个什么最终的形态去解决掉，这就是一个比较抽象的问题。

（3）如果有好的技术，它的产品形态走 C 端还是走 B 端，从什么地方生出来，美国还是中国？

（仅讨论第二个问题）

**杨洪兵**：这个问题，我觉得其实也没有量化的一个答案，因为其实我觉得现在可能只能定性的去讲。

因为我觉得我比较同意周总的一个观点，就是人形机器人或者说得更加宽泛一点，就是通用的智能机器人，它对于我们整个生活的影响一定是会非常的深刻的。我也给大家推荐一部小说，我自己也比较喜欢的一部小说，之前得过雨果奖，北京折叠。那他讲的就是人类的生产率会越来越高吗？但是它一定会有一种情况，就像您说的，假如说我们的这个机器人的生产效率会非常高，美国也有一个那个有人工智能学家讲过他非常担忧的一个点，他当时担忧是50年之后，会出现一个人类造出一个机器人比人类更聪明，那他的一个最终的结果是说这个机器人一定会造他的第二代，比他自己更聪明，这样一步步迭代下去，一百代迭代之后，第99代机器人造出的第100代机器人会比人类聪明很多，它会导致比人类的的生产效率和思考的效率，思维的效率会质量高很多，所以这种情况会可能会导致一个结果就是可能有大量的人类的效率一定会被这个机器人所远远超过。所以这其实可能是有一个非常大的比例。但是具体这一部分人是否从事更多偏艺术性的工作，包括说像北京折叠里面所描述的那样一些情况，其实这个就不是一个技术问题了，它可能是一个哲学，甚至社会的一个考虑，这个就可能是要更专业的人去回答了。

（线下: 时间大概多久，比例大概多少？）这些我感觉有可能会很快，比如说我们之前也没有预想到2022年会突然出现大模型这样的一个产品出来，然后尤其这两年大模型它的能力迭代会很快。

**线下**: 其实我也想插个话，你要知道这个答案其实意义不大，完全没有意义。还有你要相信人的适应能力，是人肯定能适应。这不适应的可能会被淘汰，人几亿年还活着，还留着种子，肯定是不一样的。最近不是很多公司内部的人员面临变革，不会都裁掉，剩下的都是会 AI 的。淘汰是有结构性的，那已经发生过变革的人类肯定能适应出来。

**周永**: 因为我之前做了很长一段时间投资，最重要的就是要预测未来，精确到每一年你才能获得最大化收益。我每次都是提前告诉别人说投资什么在两年内可以翻倍的。我从2019年之后准确率高达95%以上，也经常会思索，比如说一年后两年后，五年后整个行业的现状和机器人的现状。因为我本身会思考另外一个方面，我会思考十年后每个国家的 GDP 公司的市值什么的，因为如果没有这个判断，你基本上无法做投资。所以如果你要问我，我个人觉得第一个，机器人在未来五年内是最大的旋律，它可能每年都增长五到十倍。第二点，如果你要问我三年后机器人怎么样，我说三年后的格局会更偏车企的格局一样，我觉得它会更像电脑。现在华为都占有一半的智驾了，未来所有的制造业巨头都会进来的，也可能华为推出一个方案，我打个比方，因为现在各地都在建人形机器人，就比如说现在可能很多头部的机器人厂商他做的非常棒，这里我不好说国内的公司，比如说波士顿动力做的很棒，但是也有一种可能有3、5年后，如果华为做了一套很好的系统，而湖北人形机器人，成都人形有可能他搭上华为这套系统，它再配上关节，再配上我们的灵巧手，有可能它这个机器就跟一个波士顿动力机器人看起来没两样。我认为在整个国家上百分之百确定的基本上就是中国领先第二名很多倍。特斯拉大概在6000亿市值的时候我就给大家说，特斯拉一定可以到2万亿市值，它现在很快就上来了，因为在美国基本上也就是它还能够对抗一下了。本质上来说是马斯克一个人遮住了整个硅谷制造业的脸面。对我说机器人这个行业本质上是不需要思索的，有很大的概率，中国占70%。