

| 证券研究报告 |

机器人系列报告一：具身智能——决定 机器人泛化能力天花板的“大小脑”

2025.03.09

分析师：闻学臣
执业证书编号：S0740519090007

分析师：苏仪
执业证书编号：S0740520060001

分析师：刘一哲
执业证书编号：S0740525030001

摘要

- 具身智能是一种**基于物理身体进行感知和行动**的智能系统，其通过**智能体与环境的交互**获取信息理解问题、做出决策并实现行动，从而产生**智能行为和适应性**。具体而言，具身智能在机器人上的应用体现可以划分为三阶段：感知、推理、执行。
- 具身智能是机器人与人工智能发展的交汇点。机器人的通用性取决于泛化性的发展程度，自20世纪50年代以来，经过多轮发展，机器人开始从传统的自动化工业场景机器人，向通用泛化场景的机器人方向发展。而人工智能在几十年发展过后，在今天迎来了智能程度、通识程度更高的生成式大模型时代。在当前时间点，人工智能可以真正为机器人赋予“大脑”，机器人也可为人工智能提供“身体”，两者深度融合，而融合的交点——具身智能，则是人工智能与机器人相互促进发展而形成的必然的产业趋势。
- 目前具身大模型可以分为两大流派，一类是端到端大模型，一类是分层具身大模型。**1) 端到端大模型**：能够直接实现从人类指令到机械臂执行，即输入图像及文本指令，输出夹爪末端动作。**2) 分层具身大模型**：不同层次模型协作，上层大模型进行感知与决策，底层硬件层和中间响应快的小模型进行决策的分解与执行。目前，由于受数据制约难以达到性能要求，端到端大模型尚未成为主流选择，更多厂商还是选择以分层模型为机器人的具身智能。
- **训练具身大模型的痛点：数据**。机器人需要用海量数据进行训练，使其变得更加智能。但机器人是非常新的领域，严重缺乏训练数据的积累。对应而言，目前主要的数据收集方法有四种：**1) 远程操作**，即由实验人员操作机械手柄，远程控制机器人做出相同动作，以此来积累数据；**2) AR**，即通过AR环境对机器人进行训练，积累数据；**3) 仿真**，即通过海量算力进行模拟运算，计算得出海量机器人训练数据集；**4) 视频学习**，即通过多模态大模型，直接让机器人通过视频学习人类动作，从而积累训练数据。
- **人形机器人产业未来展望：从专用到通用，从ToB到ToC**。短期来看，任务相对聚焦，对泛化能力要求不高工业制造场景下的任务正在更快进入商业化阶段。在工业制造场景实现商业化落地之后，海量机器人的具身数据叠加算力技术的进步，机器人的能力将循序渐进逐步解锁，并向商用服务、家庭服务等更开放的场景进行延伸，届时市场有望达万亿级。

摘要

- **投资建议：**作为机器人的“大小脑”，在硬件已达到较高工艺水平的情况下，具身智能的技术能力或将成为机器人发展的天花板。当前多家领军机器人企业均已各自的方式在具身智能领域投入较多资源与精力，以提升自家机器人的泛化行为能力。我们建议投资人持续关注具身智能产业的技术变革突破、商业化落地路线、节奏与进展等。具体到标的上，建议关注**索辰科技**（物理AI、机器人子公司）、**能科科技**（特种行业AI能力领先）等。
- **风险提示：**机器人产业相关政策落地不及预期；机器人商业化落地推进不及预期；AI与具身智能产业技术发展不及预期；行业竞争加剧；研报信息更新不及时的风险等。

目录

CONTENTS

- ① 什么是具身智能
- ② 具身智能：机器人与AI的交汇
- ③ 具身大模型赋能人形机器人蜕变
- ④ 具身智能赋能下，未来机器人发展方向



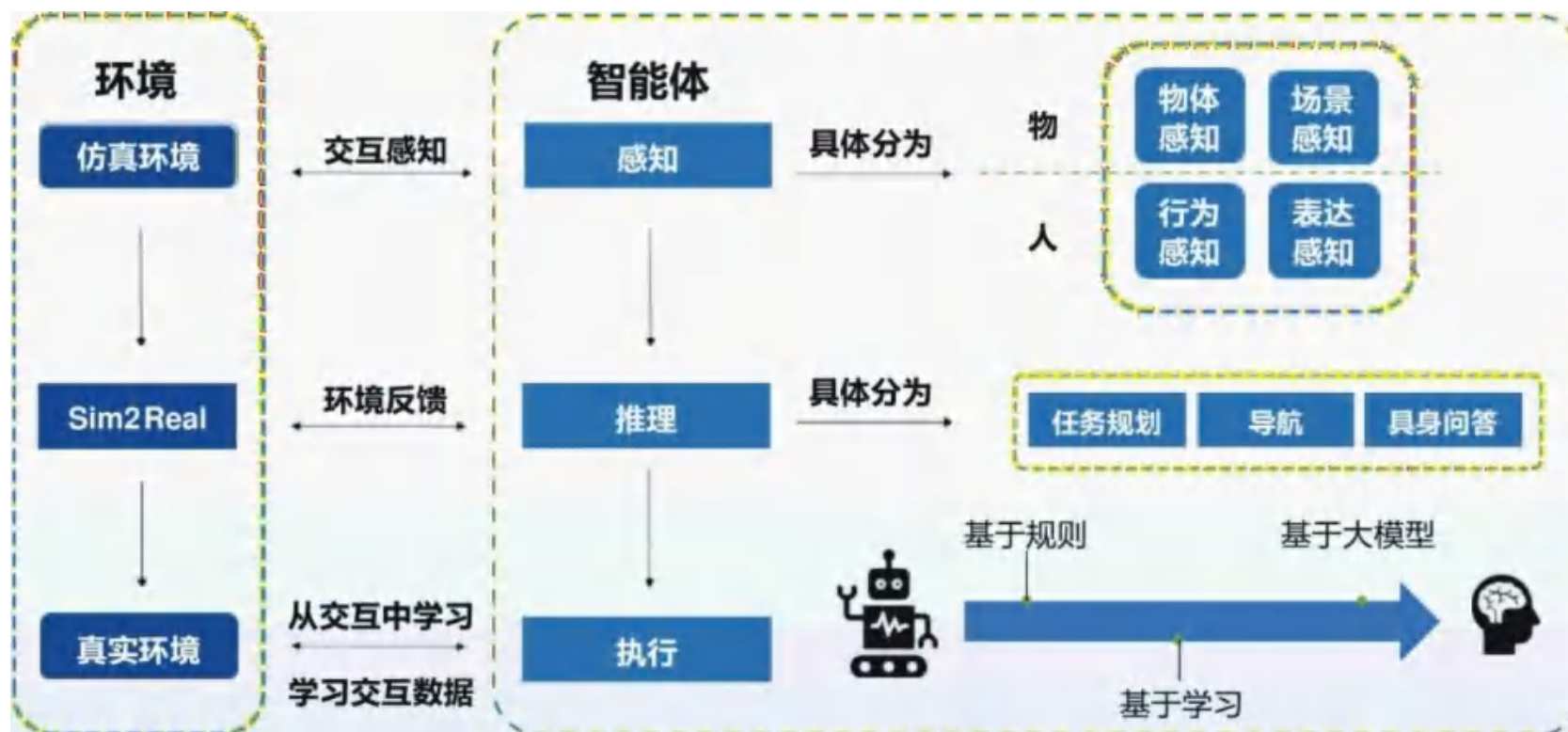
1

什么是具身智能

什么是具身智能

- 具身智能的定义：一种**基于物理身体进行感知和行动**的智能系统，其通过**智能体与环境的交互**获取信息理解问题、做出决策并实现行动，从而产生**智能行为和适应性**。
- 具身智能的实质：强调有物理身体的智能体通过与物理环境进行交互而获得智能的人工智能研究范式。
- 从机器人的角度出发，具身智能在机器人上的应用体现可以划分为三阶段：感知、推理、执行。

图表：具身智能的划分



1.1 具身感知

- 机器人需要具备环境感知能力，依据感知对象的不同，可以分为四类：物体感知、场景感知、行为感知、表达感知。
- 具身感知的过程主要包括以下几步：

图表：具身感知的过程分为以下几个步骤



资料来源：哈尔滨工业大学，中泰证券研究所

1.2 具身推理：任务规划

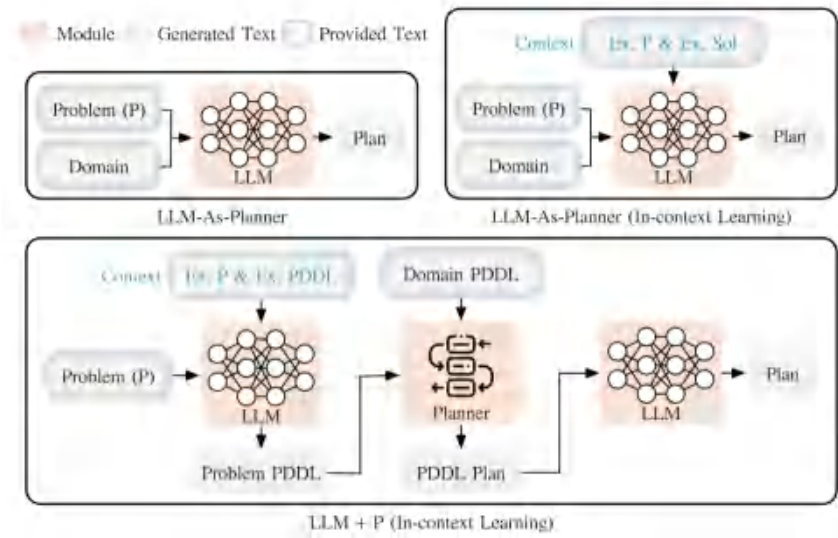
- 具身感知的过程主要包括以下几步：任务规划、导航、具身问答。
- **任务规划**：任务规划（Task Planning）是具身智能的核心任务之一，将抽象的非可执行人类指令转换为具体的可执行技能。这一步骤的完成需要两步：1) 将人类指令分解为机器人可执行的技能，2) 执行技能。
- 结合大模型的任务规划：大模型作为转换器、大模型作为分解器。
 - **作为转换器**：LLM+P，用LLM将状态信息描述成PDDL语言再进行规划，取代以往需要人工针对实际问题书写PDDL语言对任务进行建模；
 - **作为规划器**：可以zero-shot进行任务规划。

图表：任务规划将人类指令转换为可执行技能



资料来源：哈尔滨工业大学，中泰证券研究所

图表：大模型作为转换器的作用



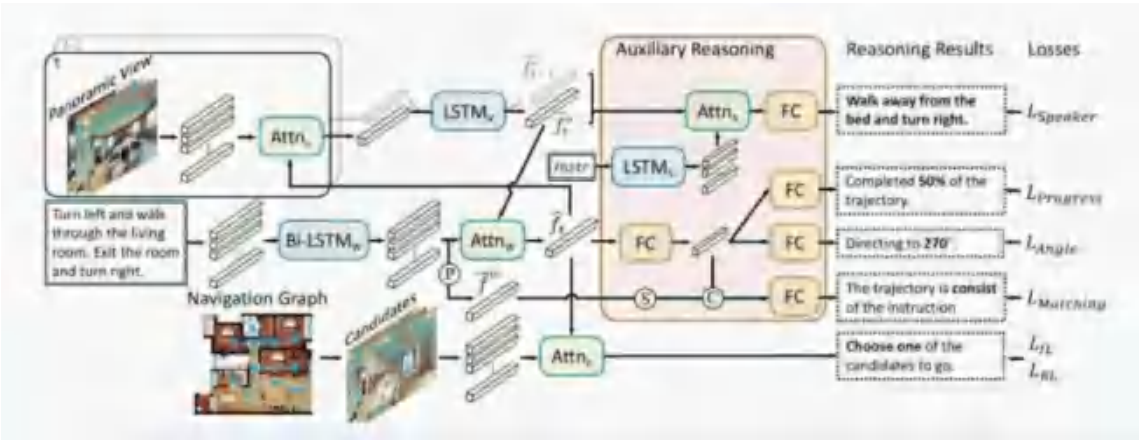
资料来源：哈尔滨工业大学，CSDN，中泰证券研究所

1.2 具身推理：导航

- 具身导航（Embodied Navigation）：智能体在3D环境中移动完成导航目标。
- 早期的具身导航，通过构建一系列基于规则的组件和算法，实现有效的环境感知、定位、路径规划和避障。而**基于学习的导航**则利用深度学习与强化学习技术，提高模型对复杂环境和新场景的泛化能力。不同于传统算法依赖预定义的规则和手工设计的特征，基于学习的导航算法从大量数据中学习环境特征和导航策略，实现强自适应性和高灵活性。
 - **视觉导航**：基于学习的导航的一个重要分支，它依靠计算机视觉来理解环境信息并做出导航决策。
 - **视觉语言导航**：通过自然语言指令和视觉图像进行导航的任务，其目标是开发一种能够与人类进行自然语言交流并在现实3D环境中导航的具身智能体。**1）大模型出现之前**：主要通过RNN，LSTM，Transformer等网络来提取命令中的语义信息；**2）大模型出现后**：利用大模型作为辅助来帮助规划器输出规划或者大模型直接作为规划器来输出规划。

图表：具身导航的发展阶段

图表：一种利用大模型的视觉语言导航方法机制



资料来源：哈尔滨工业大学，中泰证券研究所

资料来源：哈尔滨工业大学，中泰证券研究所

1.2 具身推理：具身问答

- **具身问答：**机器人需要主动探索环境，定位目标物体或位置获取环境中的信息，然后基于获取的信息回答问题。该任务可视为导航、VQA任务的结合。
- 相比于VQA等已有问答任务，具身问答的特点在于机器人具有主动行动能力。

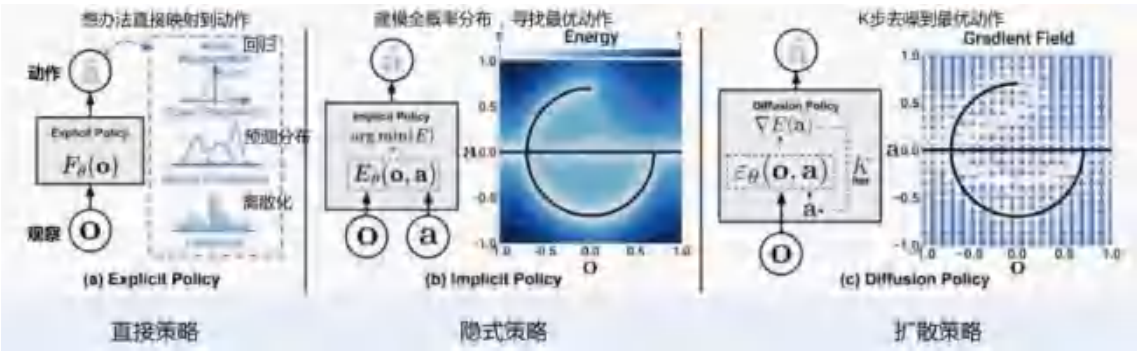
图表：具身问答示意



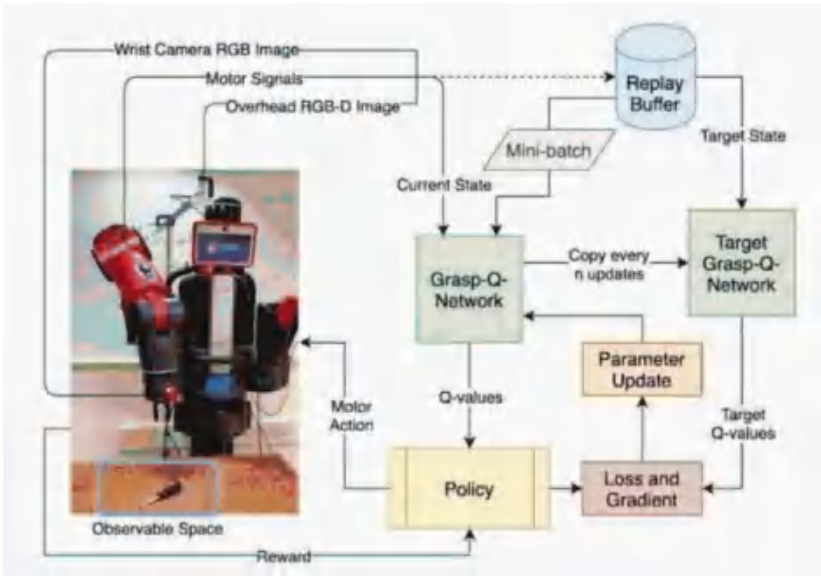
1.3 具身执行

- 具身执行，就是把任务具体执行出来，其中最重要的环节就是技能学习。
- 技能学习：以技能描述、环境观察为输入，输出完成技能所需的7Dof轨迹（主要指人手腕或者机械臂末端执行器的位置、朝向、末端状态）。技能学习主要有两类方法：
 - 模仿学习：收集专家演示数据，用神经网络拟合。
 - 强化学习：设计奖励函数，机器人通过交互学习行为策略。
- 两者的差别在于：模仿学习是从样例中学习，机器人学习过程中不与环境进行交互；而强化学习从交互中学习，机器人学习过程中与环境进行交互。

图表：模仿学习的三种策略

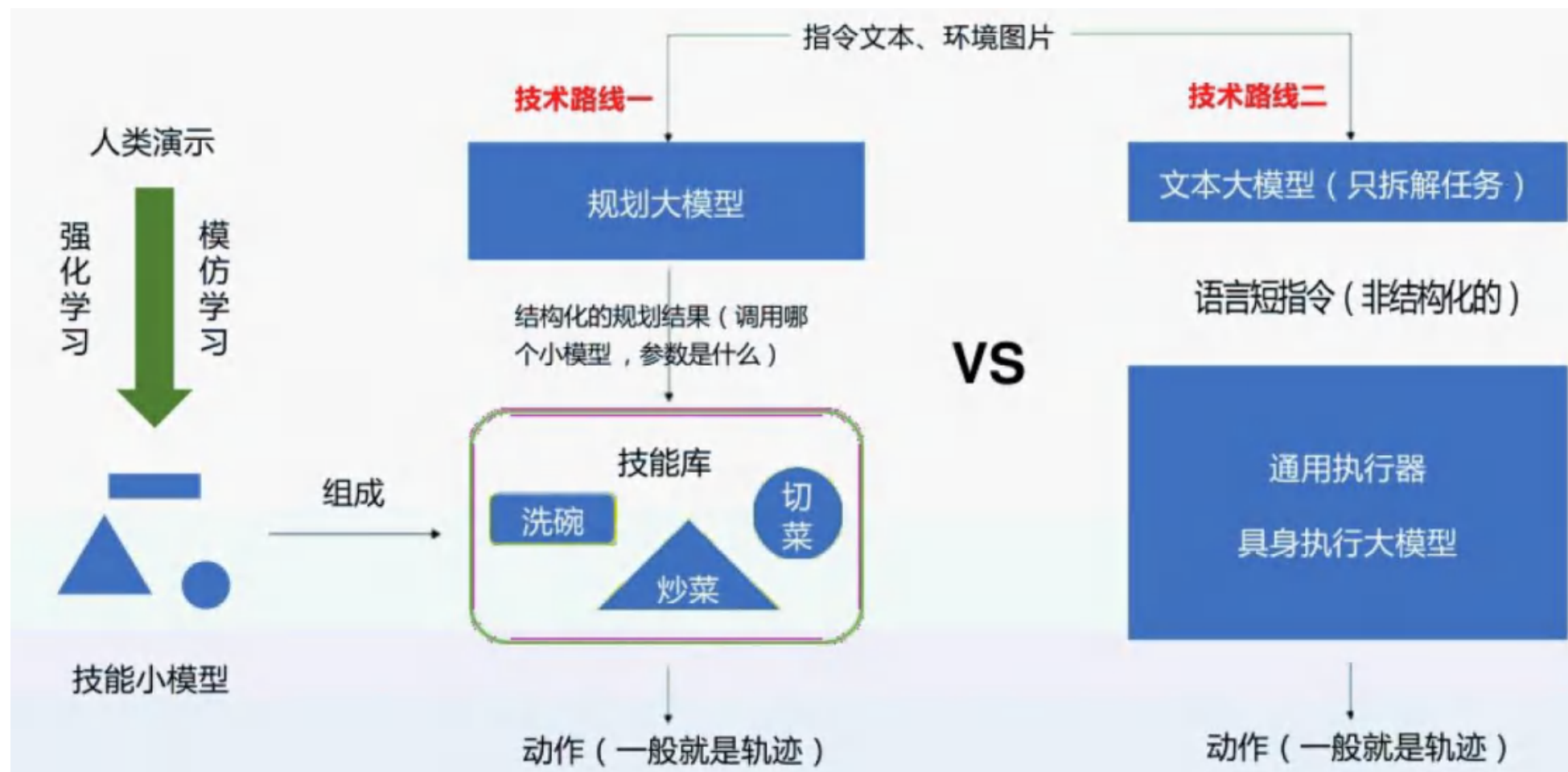


图表：一种强化学习的方法：Grasp-Q-Network Model-free



1.3 技能学习+大模型：通用执行模型结构

图表：技能学习+大模型：通用执行模型的两种结构





2

具身智能：机器人与AI的交汇

2.1 机器人的通用性取决于泛化性的发展程度

- 在上个世纪六十年代，人类已经制造出机器人并且可以控制其执行给定的动作。但是对于复杂抽象的人类指令，早期需要专家人工设计动作轨迹，然后再由机器人执行。一个理想中的智能机器人应该是“通用”的，可以适用于各种场景，进行推理决策并执行相关技能，完成各种任务。
- 从人工设计专门的程序到通用型智能机器人，存在着巨大的技术发展空间，其中最关键的问题之一即为泛化性，包括人类指令、环境配置、物体形状位置、机器人类别上的泛化性。泛化性描述了机器人因为学习场景和应用场景的任务设置不一致导致的性能变化情况，这衡量了机器人在特定维度上的通用性。
- 从泛化性的角度来看，智能机器人技术可以划分为以下几个阶段，其对应泛化性逐渐变强。

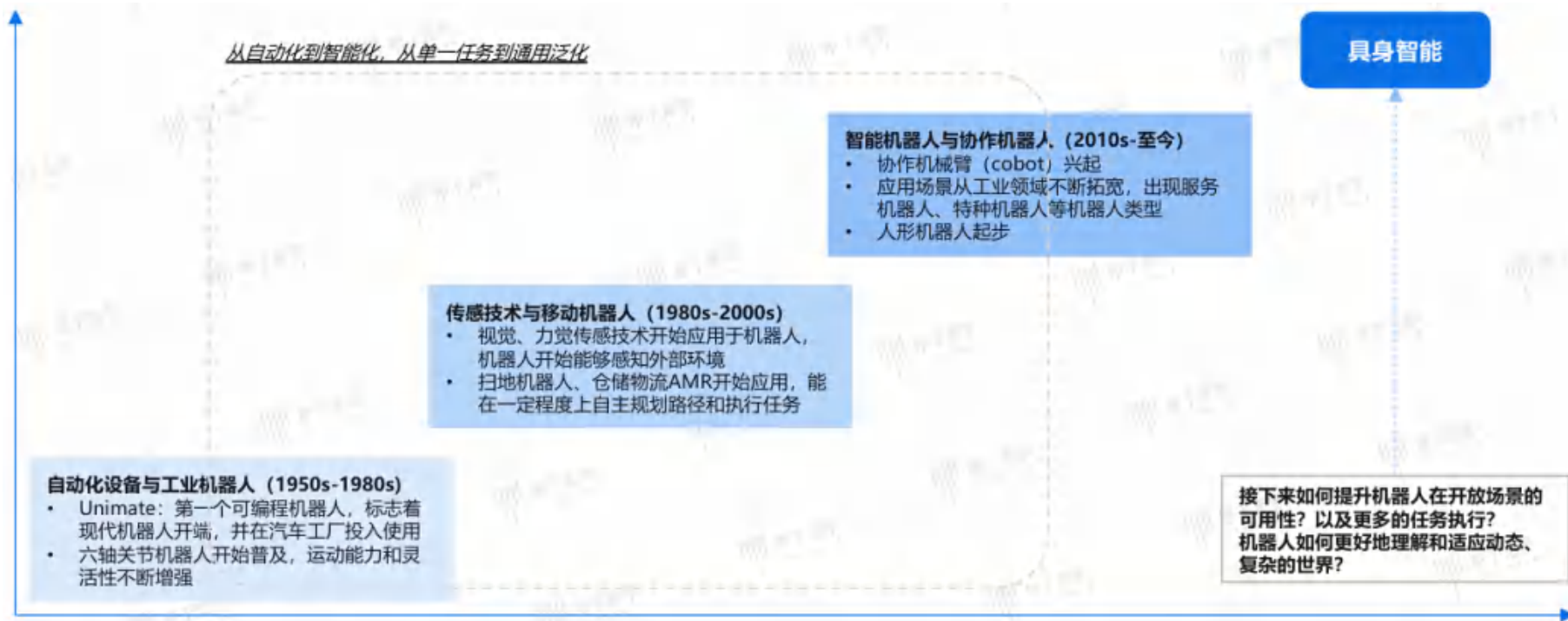
图表：泛化性角度下智能机器人技术的发展阶段



2.2 机器人发展线：从自动化工具到直接影响物理世界的智能体

- 从历史的角度来看，工业机器人的硬件进化不断发展，从自动化设备到移动机器人，从协作机器人到人形机器人，每一步都代表着技术的进步和创新。当下，已经进入了AI大模型与机器人在底层技术创新和应用融合的交叉时代。
- 技术的每一次进步都会带来机器人应用场景的突破，且一般会在精准、重复的工业场景作业流程中率先得到应用。未来，如何提升智能机器人在开放场景的可用性，是具身智能正在解决的问题。

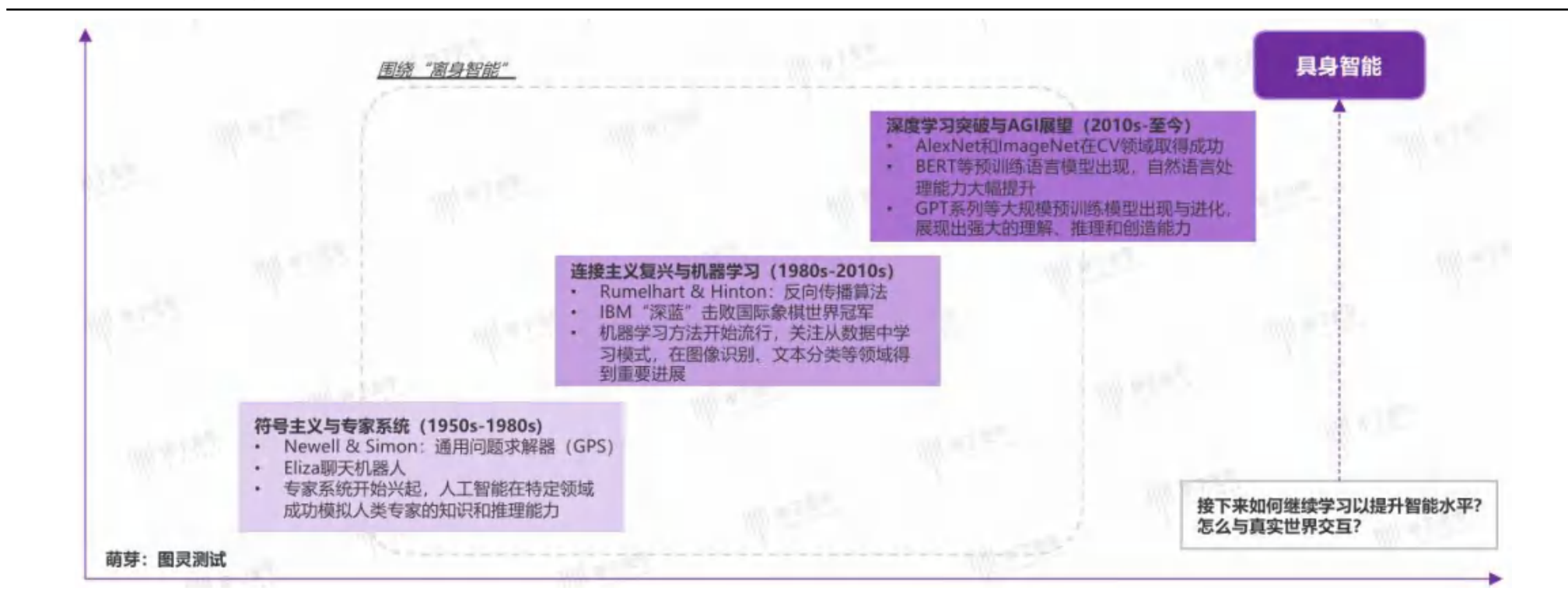
图表：机器人技术发展线



2.2 人工智能发展线：从理性主义到深度学习

- 在机器人领域的应用上，生成式AI技术正在兑现提升机器人关键能力的潜力，在环境感知、自主决策、学习与适应等多个方面均有表现。
- 基于网络数据训练的大模型似乎正在触及AI认知的边界，ScalingLaw下的预训练大模型性能似乎快要达到极限，参数量和数据量的扩充已经难以给模型带来质的突破。AI想要进一步发展，必须与物理世界建立更加紧密的联系，具身智能将是AI继续向AGI进步的关键途径。

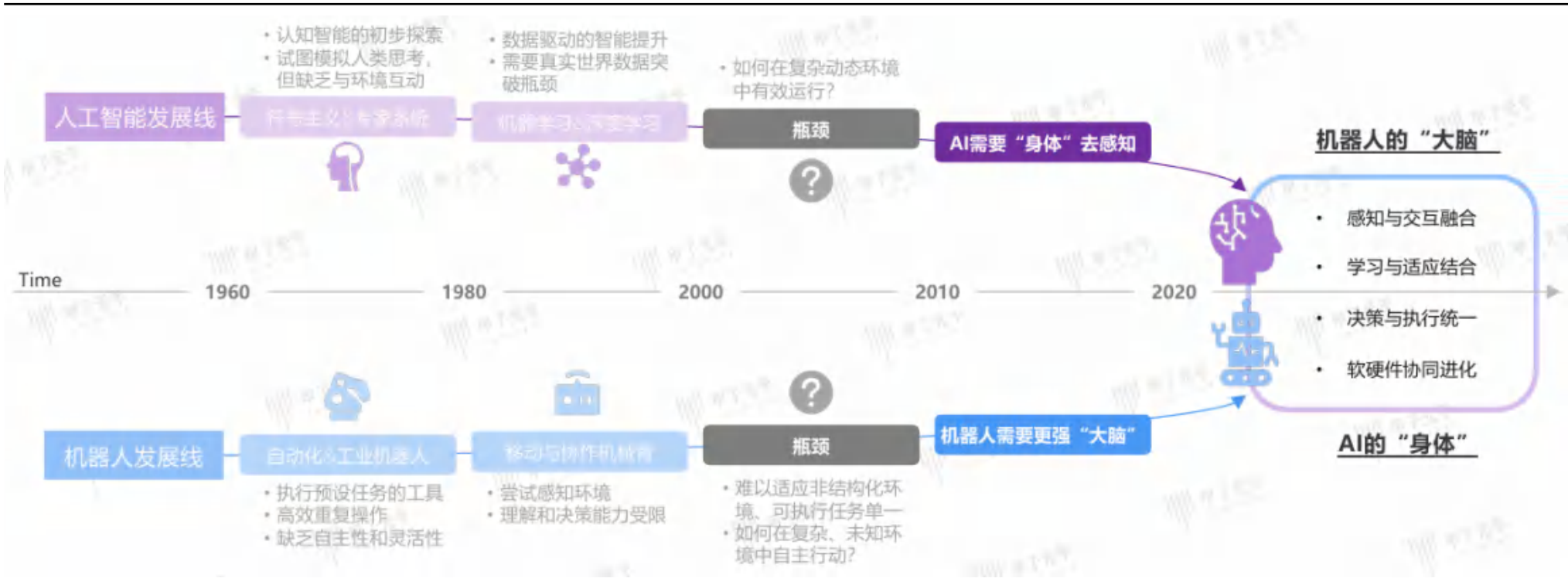
图表：人工智能技术发展线



2.2 具身智能：两条发展线的汇聚点

- 人工智能和机器人技术的发展并非孤立进行，而是相互促进、共同演进的。人工智能为机器人赋予了“大脑”，使其具备感知、思考和决策能力；而机器人则为人工智能提供了“身体”，使其能够与真实世界进行交互，获取经验和知识。
- 具身智能的兴起是人工智能和机器人技术各自发展到一定阶段的必然结果，是两者深度融合的体现。人工智能的持续进步需要与物理世界的交互能力，而机器人的未来发展也离不开更高级别的智能化水平。

图表：具身智能是AI与机器人两条线的交汇



2.3 2024年迎来具身智能投资热

- 如果说OpenAI的ChatGPT引爆了2023年对大语言模型的投资热潮，那么Tesla入局人形机器人和黄仁勋的“AI的下一个浪潮是具身智能”，则彻底带火了具身智能与人形机器人领域的投资热潮，成为2024年科技产业投资的最大热点。
- 从投资角度看，头部人形机器人的本体研发集成厂商已经历多轮融资，估值较高，行业整体投资热点正从人形机器人本体向具身智能模型和其他上游零部件迁移。

图表：2020-2024年国内人形机器人投资事件数量与金额变化



资料来源：IT桔子，甲子光年，中泰证券研究所

图表：从本体制造到具身大模型，国内外投资热点有所迁移

	公司名称	融资时间	融资金额	投后估值
国外	1X Technologies	2024 年1月	1亿美元	未提及
	Figure AI	2024 年2月	6.75亿美元	26亿美元
	Skild AI	2024 年7月	3亿美元	15亿美元
	Physical Intelligence	2024 年11月	4亿美元	24亿美元
	公司名称	融资时间	融资金额	
国内	千寻智能	2024年8月	近2亿元	
	穹彻智能	2024年9月	累计数亿元	
	星云智慧	2024年10月	2000万元	
	星海图	2024年11月	超2亿元	
	自变量机器人	2024年11月	亿元级	

资料来源：IT桔子，甲子光年，中泰证券研究所

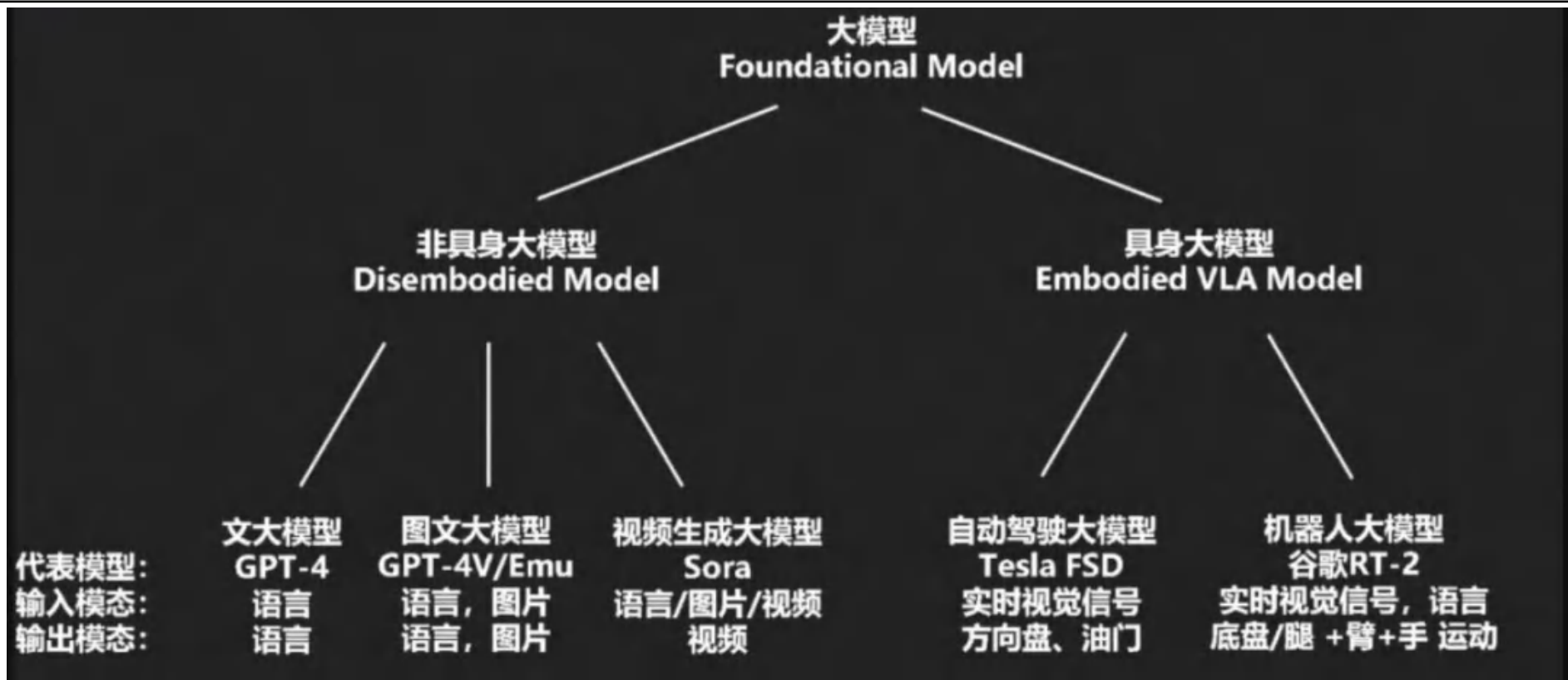
3

具身大模型赋能人形机器人 人蜕变

3.1 具身大模型&非具身大模型

- 从物理空间的角度来划分，大模型可以分为非具身大模型（Disembodied Model）、具身智能大模型（又被叫做机器人大模型）（Embodied VLA Model）。它们的区别是能否生成运动姿态（例如夹爪的末端位姿等），即是否能跟物理机器人联系起来。

图表：具身大模型与非具身大模型

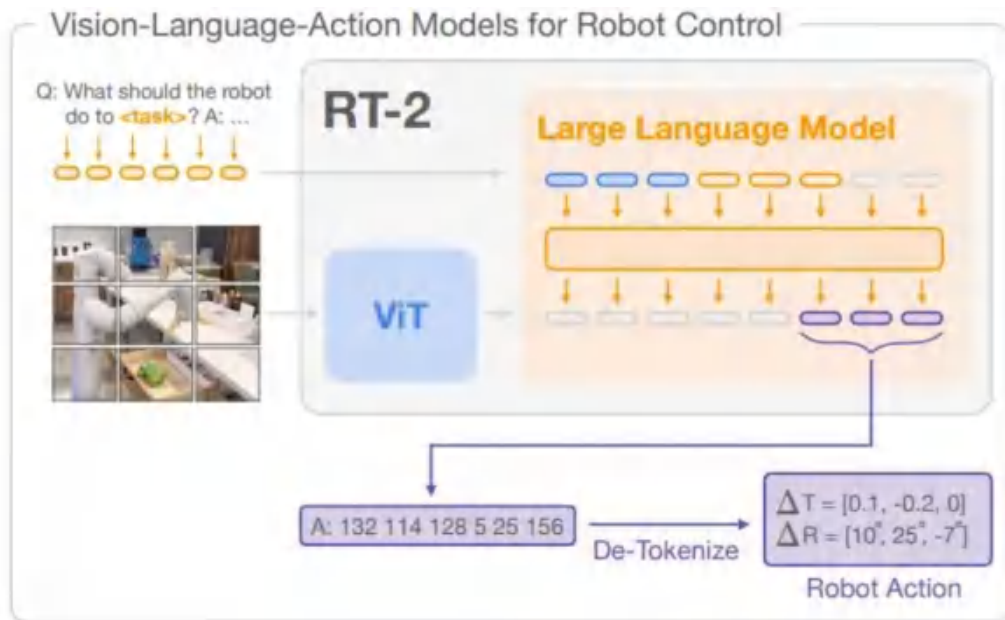


资料来源：CSDN，中泰证券研究所

3.1 具身大模型：分层模型or一体化端到端

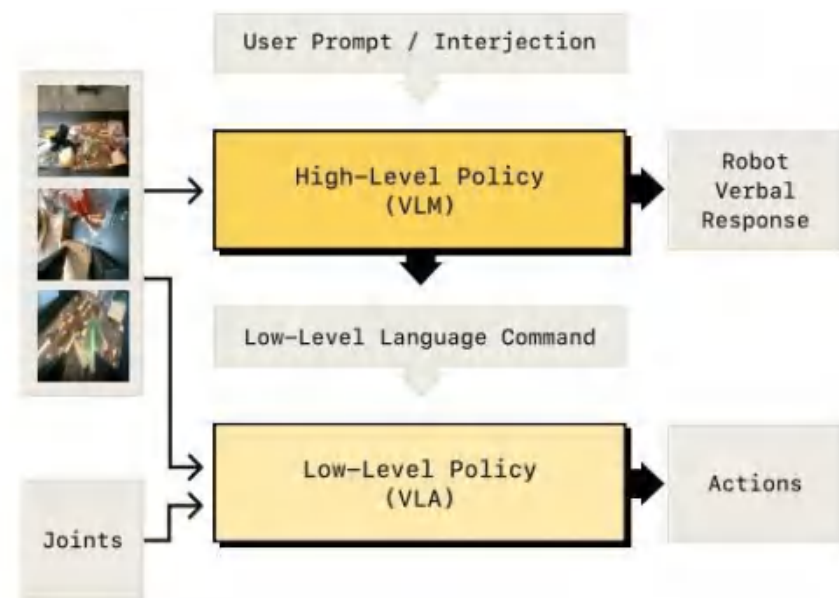
- 目前，具身大模型可以分为两大流派。一类是**端到端大模型**，一类是**分层具身大模型**。
- 从人体“大脑-小脑-肢体”的架构来看，分层大模型通过不同层次模型协作，利用底层硬件层和中间响应快的小模型弥补上层大语言模型的不足，推动机器人在实际场景的应用，并越来越多地采用基于学习的控制方法。
- VLA等端到端模型能够直接实现从人类指令到机械臂执行，即输入图像及文本指令，输出夹爪末端动作。
- 分层式路线一度是主流选择，因为端到端路线受数据制约难以达到性能要求；机器人数据正在逐步积累，端到端路线在未来可能成为主流。

图表：谷歌RT-2端到端大模型



资料来源：CSDN，中泰证券研究所

图表：Hi Robot的分层视觉-语言-动作模型



资料来源：Hi Robot: Open-Ended Instruction Following with Hierarchical Vision-Language-Action Model，中泰证券研究所

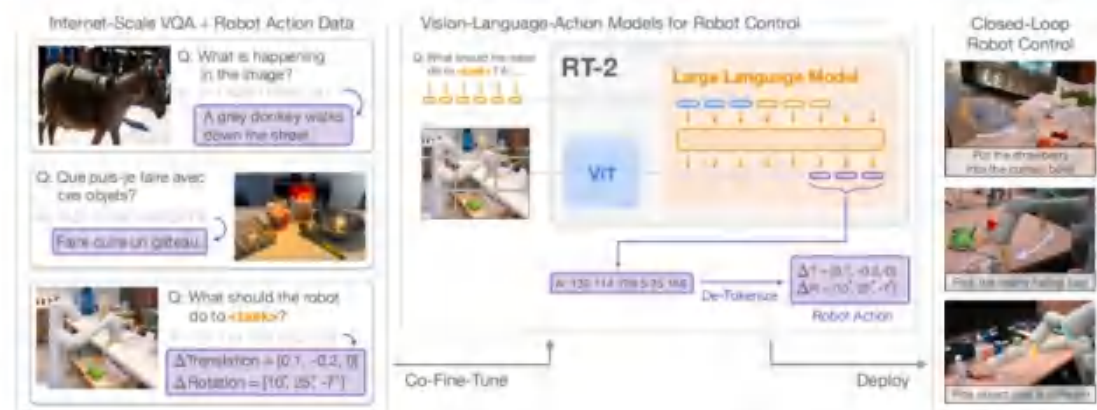
3.2 端到端大模型

- 在机器人中，端到端大模型可实现直接端到端地从人类指令到机械臂执行，即输入是图像及文本指令，输出是夹爪末端动作。以谷歌的RT-1到RT-2的进化迭代为例：
- **谷歌RT-1——关注泛化能力：**2022年，谷歌机器人研究团队历时17个月基于13个机器人得到了130k episodes 以及超过700个任务的数据，在该数据集的基础之上，基于模仿学习中行为克隆学习范式，把 Transformer 应用机器人的操纵任务上，提出了 RT-1模型。RT-1 的输入由图片序列、自然语言指令构成，输出由机械臂运动的目标位姿(Toll, pitch, yaw, gripper status)、基座的运动、模式转换指令构成。
- **谷歌RT-2——获得涌现能力：**2023年，Google 的 DeepMind 团队基于互联网上数据训练视觉-语言模型(VLM)，使其能够学习到更多关于视觉和语言之间映射关系的知识后，在机器人操纵任务上微调，提出了 RT-2。RT-2 的目标是训练一个学习机器人观测到动作的端到端模型，且能够利用大规模预训练视觉语言模型的益处，经测评，这类学习方法产生的Vision-Language-Action(VLA) 模型获得了涌现能力。

图表：谷歌RT-1端到端大模型



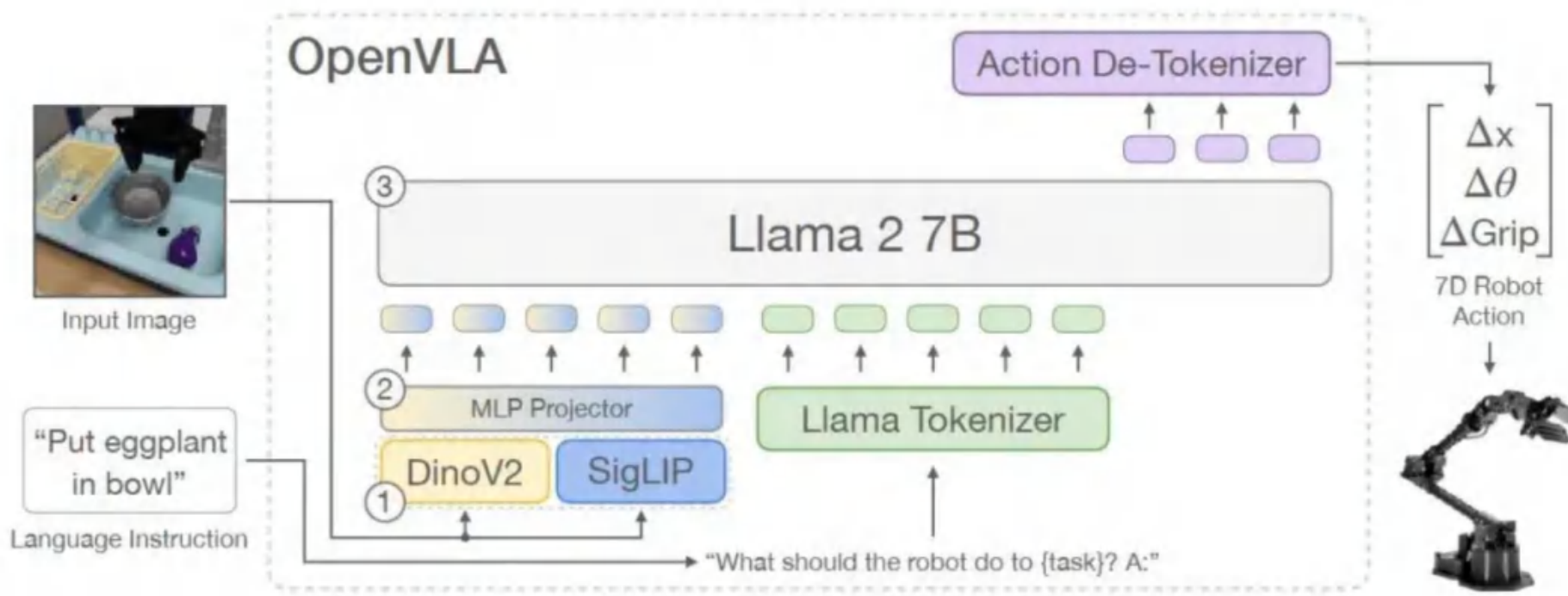
图表：谷歌RT-2端到端大模型



3.2 端到端大模型2.0阶段：VLA

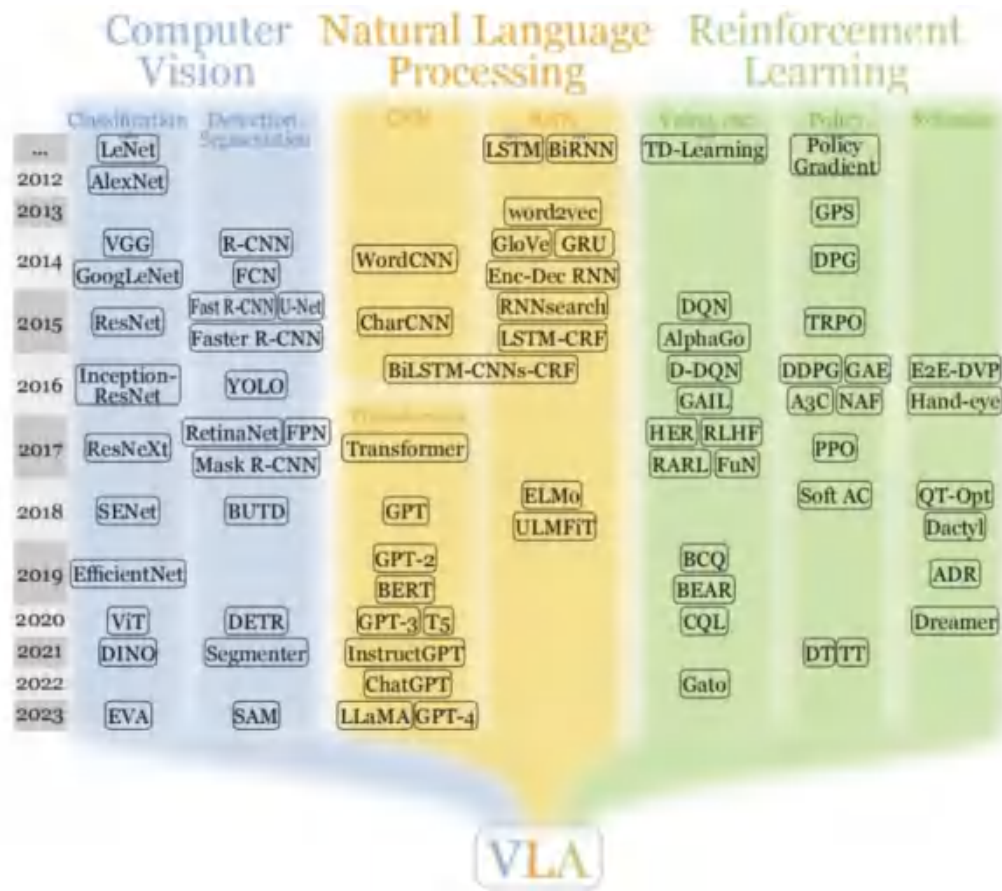
- 以谷歌RT-2发布为标志，VLA模型正式进入人们的视野中并快速发展。
- VLA模型是在视觉语言模型（VLM）的基础上发展而来的。在VLM基础上，它还利用了机器人或汽车运动轨迹的数据，进一步训练这些现有的VLM，以输出可用于机器人或汽车控制的动作序列。通过这种方式，VLA可以解释复杂的指令并在物理世界中执行相应的动作。

图表：VLA大模型工作原理简示



3.2 VLA大模型结合了视觉、语言和动作三种能力

图表：VLA大模型结合了视觉、语言和动作三种能力



资料来源：CSDN，中泰证券研究所

3.2 端到端大模型的主要优缺点

■ 优点（以VLA为例）：

- **端到端架构：**VLA是一个端到端的大模型，这意味着它可以简化传统上需要多个独立模块才能完成的任务流程。这不仅可以提高系统的效率，还能增强其灵活性和适应性。
- **泛化能力：**VLA具有强大的泛化能力。以谷歌DeepMind推出的RT-2为例，该模型可以在新的物体、背景和环境表现出显著改善的性能。它可以理解并响应那些在训练数据集中未曾出现过的命令，并基于底层语言模型提供的思路链进行推理，从而做出合理的决策。
- **通用性：**VLA具备高度的通用性。由于它是建立在一个通用的大规模预训练基础上，因此理论上几乎所有的“智能机器设备”都可以使用这套算法。无论是汽车、飞行器还是其他类型的机器人，只需要经过适当的微调就能满足特定应用场景的需求。

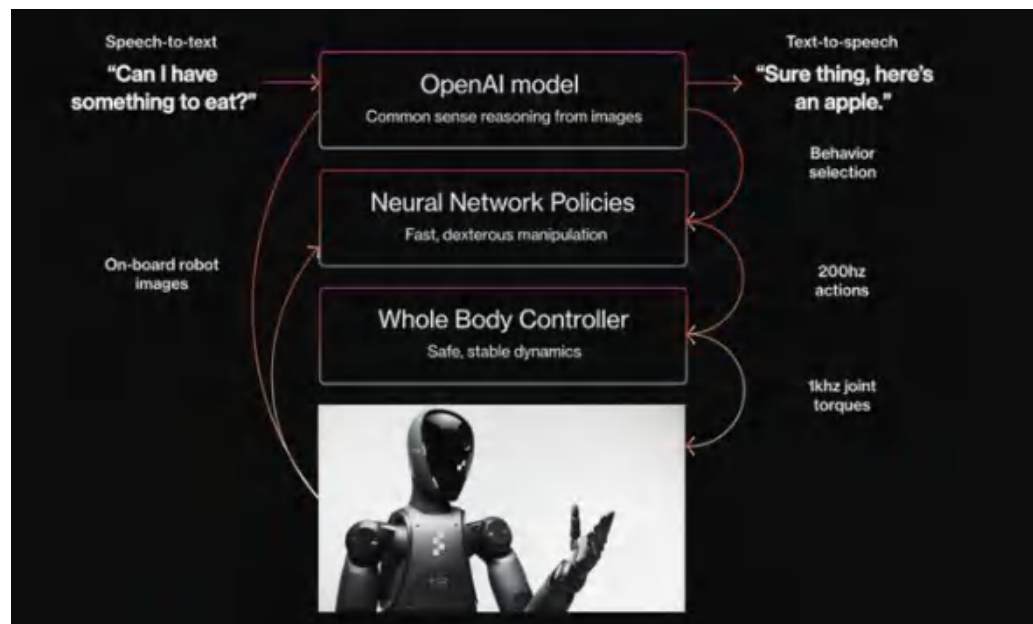
■ 缺点：

- **数据来源非常有限：**谷歌在Mountain Village（美国加州）办公室的厨房里采集了17个月，得到13万条数据，使得其机器人在谷歌的厨房里表现可以非常好。但一旦出了这个厨房，需要考察其环境泛化性，它的成功率就从97%骤降到30%左右。并且这种泛化是有选择的泛化，不是将其直接放到施工工地、非常嘈杂的后厨等场景中。
- **推理、响应速度仍有待提升：**RT-2大模型中包含了谷歌上一代大模型PaLM-E，它的速度只能达到1~3Hz，对应机器人的反射弧长达0.3秒甚至1秒。

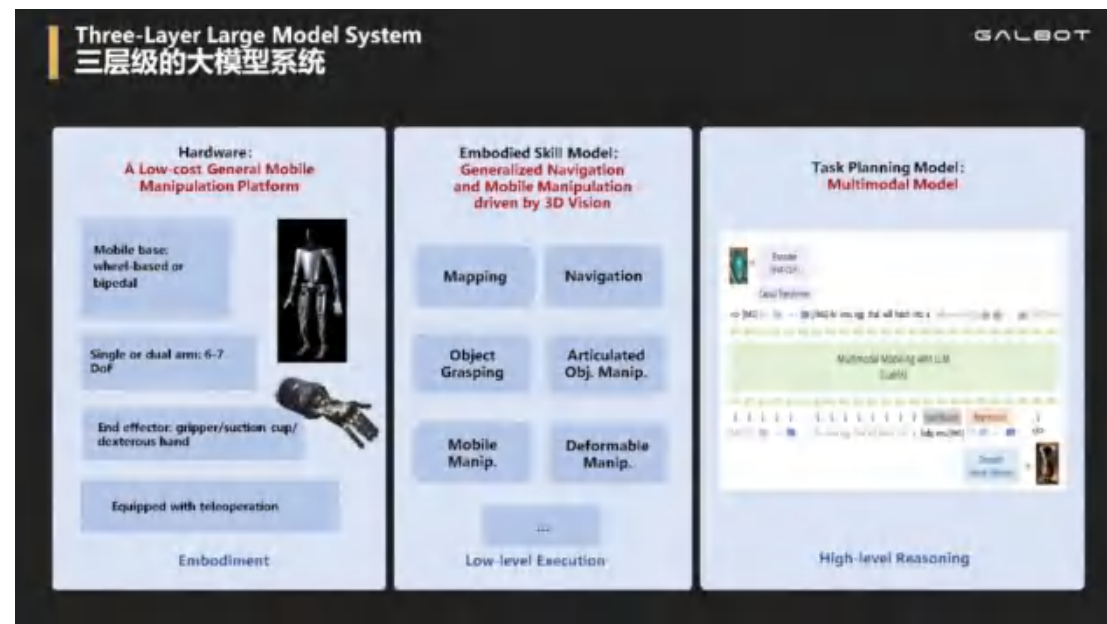
3.3 分层端到端大模型

- 分层决策模型通过将感知、规划决策、控制和执行各模块分解为多个层级，分别突破“大脑”和“小脑”，利用不同的神经网络进行训练，最终再整合起来。
- 一个典型案例如图Figure 02。Figure 02 采用三层级方案：1) 顶层集成了 OpenAI 的大模型，负责视觉推理和语言理解；2) 中间层是神经网络策略（NNP），负责快速、灵巧的操作，将视觉信息直接转换为动作指令，并输出这些指令；3) 底层是全身控制器，负责提供稳定的基础控制，在接收 NNP 的动作指令后，输出各关节的扭矩指令。
- 另一典例如银河通用的三层级大模型方案。1) 底层是硬件层，2) 中间层是通过仿真合成数据不用任何真实世界数据训练的泛化的技能，3) 最上层是大模型，可以实现完整的从任务的感知、规划到执行的全流程。

图表：Figure 02 的三层级方案



图表：银河通用的三层级大模型方案图示



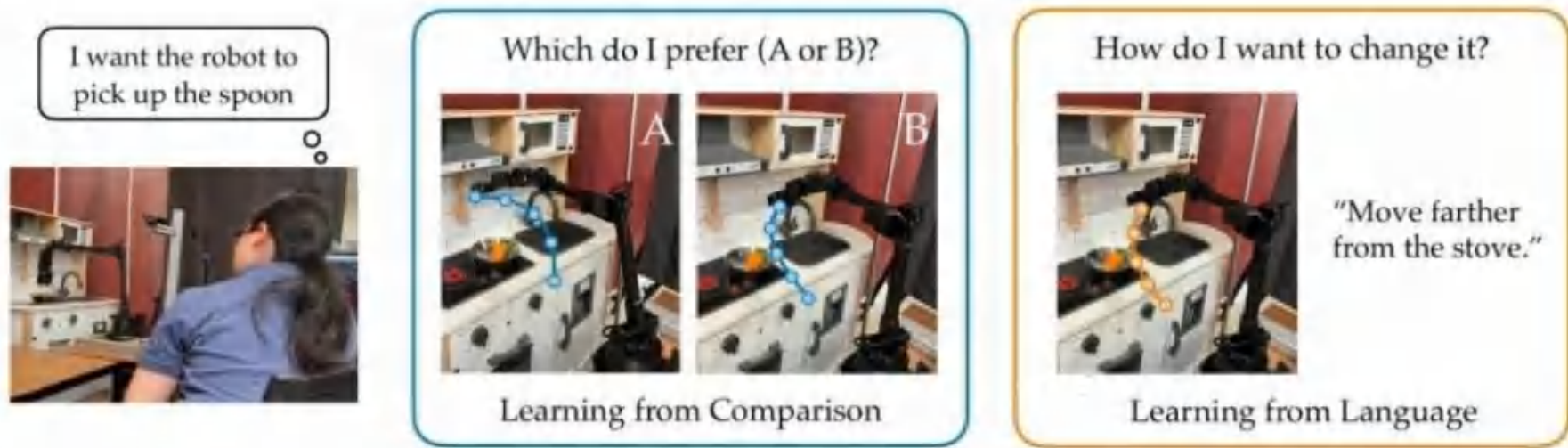
3.3 分层模型中，大模型为“大脑”

- 在分层端到端模型方案中，机器人的四大板块（感知、规划决策、控制和执行），逐渐被类人化地划分为负责解决高层次认知或决策问题（high level）的“大脑”，以及负责基础层面的功能性问题（low level）的“小脑”。两大系统各司其职又互相合作，“大脑”负责解析任务需求，整合来自传感器的信息，进行任务的细化和策略规划；“小脑”则专注于精细的运动控制，确保在“大脑”制定的策略指导下，机器人能够准确无误地执行动作并进行必要的调整。目前较为普遍的方向是，**机器人的“大脑”由大模型来担任，“小脑”则依然为底层控制算法。**
- **大脑：**负责 high level 的感知和规划决策系统，是多模态大模型。基于多模态大模型，机器人能学习、理解、融合和对齐各传感器采集而来的跨模态信息，实现对复杂环境的鲁棒建模与更精准、通用的感知，并根据环境情况，自主拆解任务、规划动作。
- 目前基于大模型的“大脑”技术路线正处在并行探索阶段，并逐渐向端到端的大模型演进。现阶段主要是4条技术路线：
 - **LLM（大语言模型）+VFM（视觉基础模型）：**实现人机语言交互、任务理解、推理和规划，目前最为成熟。主要代表是谷歌的SayCan 模型。
 - **VLM（视觉-语言模型）：**弥合语言与视觉理解间的差距，实现更准确的任务规划和决策，主要代表是清华大学的CoPa模型，利用嵌入在基础模型（比如视觉语言模型的代表GPT-4V）中的常识知识为开放世界机器人操控生成一系列的自由度末端执行器姿势，生成的操控任务分为任务导向抓取和感知运动规划。
 - **VLA（视觉-语言-动作模型）：**在VLM基础上增加运动控制，解决机器人运动轨迹决策问题。主要代表是谷歌RT-H模型，学习语言和运动，并使用视觉上下文，通过利用语言-视觉-动作结合的多任务数据集学习更强大和灵活的动作策略
 - **多模态大模型：**实现对物理世界环境的全面感知，是未来的主要研究方向。主要代表是麻省理工、IBM等共同研究的MultiPLY模型，将视觉、触觉、语音等3D环境的各类特征作为输入，以形成场景外观的初步印象，并通过多视图关联将印象中的输出融合到3D，最终得到以对象为中心的場景特征。

3.3 “小脑”：加载人工智能技术，运动方式更加拟人

- “小脑”的运动规划与控制是人形机器人实现自然和流畅动作的关键。目前，“小脑”运动控制包括基于模型的控制方法和基于学习的控制方式两个大类。
- **传统的基于模型的控制方法：**通过建立机器人的运动学和动力学模型，进行运动轨迹规划和平衡控制。这一类方法的特点是身体控制稳健，但步频较慢，且整体开发较为复杂，成本高，不利于产品快速迭代。基于模型的控制方法具体包括ZMP判据及预观控制、混杂零动态规划方法、虚拟模型解耦控制、模型预测控制+全身控制（MPC+WBC）等。
- **基于学习的控制方法：**使用端到端的人工智能技术，代替复杂的运动学模型，大幅度降低了“小脑”开发难度、提升了迭代速度，一般通过人类示教或自主学习建立运动执行策略，其中通过人类示教的方式也称为**模仿学习**，效果依赖高质量示范数据；通过自主学习的方式也称为**强化学习（Reinforcement Learning, RL）**，指通过精心设计学习目标，机器人不断在环境中探索逐渐发现最大化奖励的方式学习到最优的执行策略，效果依赖于仿真环境。

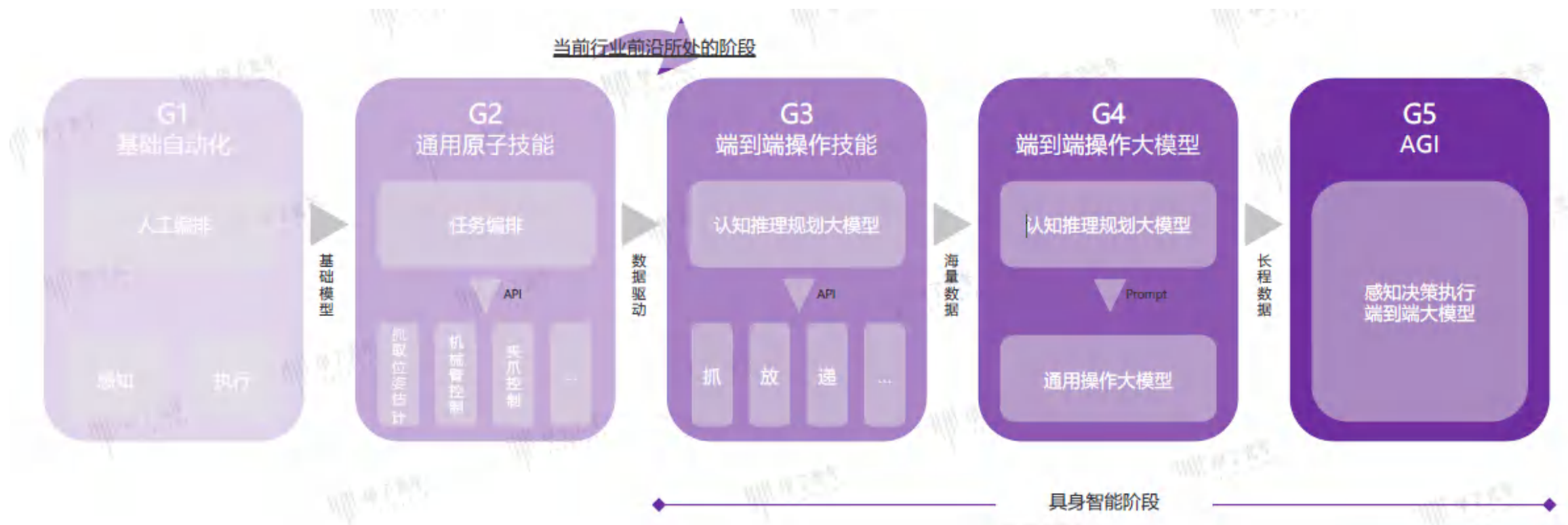
图表：一种利用比较性语言反馈学习人类偏好的机器人学习框架，更好地进行强化学习



3.3 通往One-Model端到端大模型是个循序渐进的过程

- 根据智元机器人定义的路线图，当前具身智能大脑已经具备认知、推理、规划的能力，且而小脑层面机器人技能任务的训练也是通过深度学习以驱动的。随着场景、数据的增多，多个特定小模型会逐渐泛化为通用操作大模型，并最终实现与上层模型的融合。

图表：具身智能的发展阶段，从模块化走向One Model端到端



3.4 人形机器人的核心壁垒与挑战——数据

- 一些对人类来说很简单的任务，对机器人来说可能并不容易，如灵巧性能力、空间感知能力、平衡恢复能力等，为了克服这些问题，需要用海量数据进行训练，使机器人变得更加智能。
- 但同时，机器人也是一个非常新的领域，严重缺乏训练数据的积累。对比不同模态下的最大数据集，文本模态约15T tokens，图片模态有6B图文配对数据，视频模态有2.6B视听特征数据。然而机器人模态只有240万个数据片段，相比其他模态而言，数据积累远远不够。
- 针对数据这一问题，业界形成了四种流派，以快速积累机器人训练数据，分别是远程操作、AR、仿真、视频学习。

图表：数据缺乏是当前机器人AI发展的主要瓶颈

Modality	Largest Dataset Size	Example AI Models
Text	15T Tokens	OpenAI GPT-4
Image	6B Image-Text Pairs	Midjourney
Video	2.6B Audio/Visual Features	runway
Robotics	2.4M Episodes	Not enough for general-purpose AI model

3.4.1 机器人数据积累方法——远程操作

- 远程操作，即由实验人员操作机械手柄，远程控制机器人做出相同动作，以此来积累数据。
- 近期的研究通过多个操作员协作收集了大规模多样化的数据集，结果表明，基于这些数据集训练的机器人表现出色，甚至能够在不同环境中泛化。
- 然而，数据采集是这一方法更广泛应用的主要瓶颈。
- 在单臂机器人任务中，数据采集通常需要多个操作员、机器人以及数月的人力投入。针对类人机器人的数据采集则更具挑战性，因为同时控制多只手臂和多指灵巧手的复杂性较高。
- 为了实现类人机器人的实时远程操作，研究人员开发了专门的远程操作界面，但这些界面的成本高昂，难以实现规模化。类人机器人的任务复杂性也更高，因此数据需求更大，但由于操作员需同时控制多臂和多指手，工作负担大幅增加，数据采集也更加困难。

图表：远程操作收集训练数据的操作示例



资料来源：Coatue，中泰证券研究所

3.4.2 机器人数据积累方法——仿真

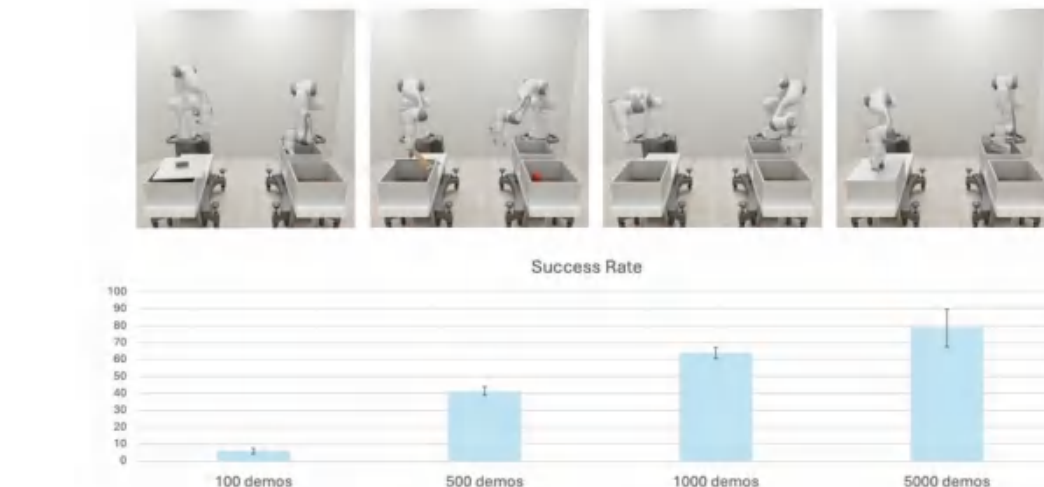
- 所谓仿真，即通过海量算力进行模拟运算，计算得出海量机器人训练数据集。仿真可能是目前最有可能做到规模化数据生成的路径，不过其背后需要巨大的算力支持。
- 2024年11月，来自英伟达、UT Austin和UCSD的研究人员推出了一种大规模自动化数据生成系统——DexMimicGen。它可基于少量人类演示合成类人机器人的灵巧手运动轨迹，帮助实现在更少人工干预下扩大类人数据的采集。
- DexMimicGen的核心思想是利用一小组人类演示，并通过在物理仿真中进行演示转换和重放，自动生成大量适用于双手灵巧操作场景中模仿学习的训练数据。例如，从仅5个源人类演示开始，DexMimicGen可以生成1000个双手灵巧任务的演示。研究团队从60个源人类演示中总共生成了21000个演示，涵盖了9种不同的任务。
- 研究团队发现，相较仅依赖源演示，DexMimicGen大幅提升了策略成功率，所有任务中基于DexMimicGen数据集训练的机器人表现明显优于仅基于少量源数据训练的机器人。且随着DexMimicGen数据量的增加，策略的性能也随之提升。

图表：DexMimicGen大幅提升了策略的成功率

图表：随着DexMimicGen数据量的增加，策略的性能也随之提升

Task	Source Demo DP	BC-RNN-GMM	BC-RNN	DP
Piece Assembly	3.3 ± 0.9	74.0 ± 2.8	74.7 ± 2.5	80.7 ± 0.9
Threading	1.3 ± 0.9	54.0 ± 4.3	55.3 ± 6.0	69.3 ± 1.9
Transport	52.7 ± 7.7	64.0 ± 3.3	60.0 ± 0.0	88.3 ± 0.9
Box Cleanup	62.0 ± 1.6	64.0 ± 10.0	96.7 ± 2.6	92.0 ± 4.3
Drawer Cleanup	0.7 ± 0.9	80.7 ± 5.0	50.7 ± 4.1	76.0 ± 0.0
Tray Lift	3.3 ± 0.9	66.0 ± 8.2	75.3 ± 7.5	88.7 ± 0.9
Pouring	0.7 ± 0.9	74.0 ± 8.6	86.0 ± 3.3	79.3 ± 0.9
Coffee	14.7 ± 0.9	12.0 ± 1.8	73.3 ± 12.3	77.3 ± 0.9
Can Sorting	0.7 ± 0.9	75.3 ± 1.0	96.0 ± 1.0	97.3 ± 0.9

TABLE 1: Success rates (3 seeds) of image-based policies trained with BC on the source demos and DexMimicGen datasets of 1000 trajectories.



3.4.3 机器人数据积累方法——AR

- 在一项名为《Explainable Human-Robot Training and Cooperation with Augmented Reality》的研究中，研究人员通过AR（增强现实）技术让人机交互过程具备更强的可解释性，从而进行数据积累。
- 2024年10月，斯坦福李飞飞团队发表论文《ARCap: Collecting High-quality Human Demonstrations for Robot Learning with Augmented Reality Feedback》，提出了ARCap系统。
- ARCap是一种便携式数据收集系统，它通过增强现实（AR）和触觉警告提供视觉反馈，以指导用户收集高质量的演示。ARCap使新手用户能够收集与机器人运动学相匹配的机器人可执行数据，并避免与场景发生碰撞。利用从ARCap收集的数据，机器人可以执行具有挑战性的任务，例如在杂乱的环境中进行操作和长视野跨具身操作。
- ARCap完全开源且易于标定；所有组件均由现成的产品构建。

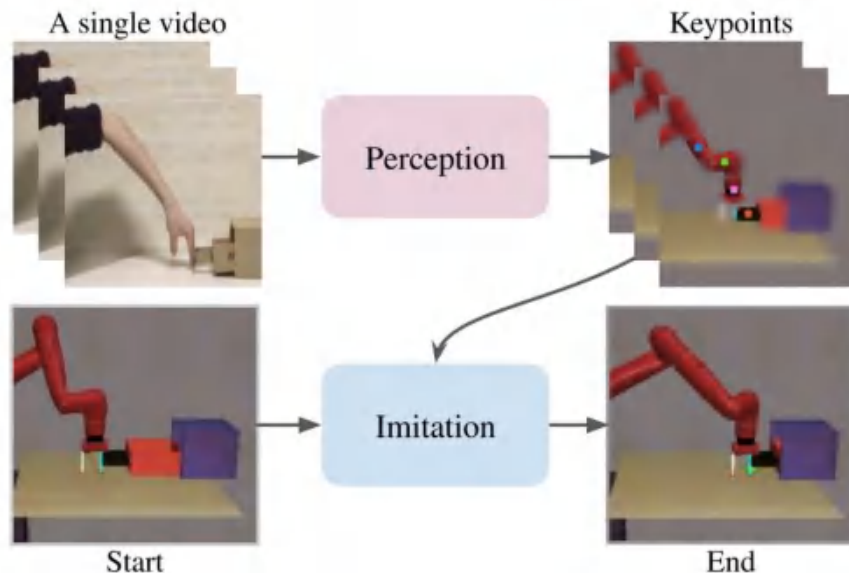
图表：研究人员通过AR指导机器人行为，机器人在AR场景中自行形成计划并行动，研究人员对结果给予反馈



3.4.4 机器人数据积累方法——视频学习

- 视频学习，即通过多模态大模型，直接让机器人通过视频学习人类动作，从而积累训练数据。
- 考虑到直接学习人类操作、人类远程操控等方式的示教成本较高，“从人类视频中学习”已成为机器人的主流训练方法之一，如直接通过YouTube视频进行策略学习。
- 但视频学习同样有自己难以解决的问题：1）控制上：不同于以往的示教方法（如遥操作示教及拖动示教），视频示教没有提供机器人本体的motion数据，agent需要仅根据演示视频生成action。2）感知方面，让机器人理解人类演示视频也具有很大挑战，机器人与人类的视角不同，并且人类演示视频中，每一帧是人手臂的运动图像，而机器人捕获的图像是机械臂的运动图像，这些原因导致了演示视频与机器人感知的图像之间存在gap。

图表：机器人从人类视频中实现物理模仿



3.5 主要厂商的做法：宇树机器人

- 2024年12月，宇树开源了一系列数据采集工具和数据集。
- 数据采集工具包括但不限于：1) 针对开源机器人系统R(obot)OS推出的模拟包，包含宇树系列机器人的质量、惯量、力矩等参数；2) 使用苹果Vision Pro对G1进行遥控操作控制的仓库，可以用于数据采集；3) RL GYM，用于机器人仿真训练，之前开源了英伟达Issac Gym平台上的训练代码，这次新增了对MuJoCo模拟仿真的支持。
- 数据集：包含五种操作，使用配备有三指灵巧手的宇树G1人形机器人收集，每张图分辨率为640×480，每个手臂和灵巧手的状态和动作维度为7，可使机器人完成拧瓶盖倒水、叠放积木等操作。
- 在数据采集环节，宇树运用了苹果Vision Pro进行G1的遥控操作控制，该数据集采用640x480分辨率的图像，并记录了机器人手臂和灵巧手的七维状态和动作数据。

图表：宇树开源的数据采集工具项目

Title	Description	Stars	Forks
unitree_robot	Robot 3D module for different environments. Robot's parameters can be found here.	1,234	567
unitree_rl_gym	RL Gym simulation package. It has unitree of all Unitree series robots, contains information such as mass, inertia, moment, limit and so on. Newly support G1.	890	123
unitree_sim	Develop Gazebo and ROS2 robots in the real environment. The interface types provided are consistent with unitree_sdk2.	456	789
unitree_simulator	Use Mucos as a simulator and file sim-to-real implementations, integrate with ROS2.	234	567
unitree_sdk2	An ROS2 simulation example for reinforcement learning, supports Gazebo, H1, G1.	567	890
unitree_sdk2_python	Use Apple Vision Pro to teleoperate Unitree G1, H1, G2.	123	456
unitree_sdk2_ros	Use Azure Kinect DK camera to teleoperate Unitree H1.	789	123
unitree_sdk2_ros2	Use G1 dual-arm robot for data collection, train and test, with modified Lefteris, an open-source training framework.	345	678
unitree_sdk2_ros2_ah	An ROS2 package used to develop Atinho, A1, G1 and H1 robots in real environments.	678	901
unitree_sdk2_ros2_g1	An ROS2 package used to develop G1, G2, H1, H2 robots in real environments.	901	234
unitree_sdk2_ros2_g1_python	Python interface for unitree_sdk2.	123	456

图表：宇树开源的数据集涵盖内容

Models 模型

None public yet 尚未公开

Datasets 数据集 5

14 Sort: Recently updated 排序: 最近更新

unitreerobotics/G1_BlockStacking_Dataset

unitreerobotics/G1_方块堆叠数据集

Preview 预览 · Updated 6 days ago 更新于 6 天前 · 1

unitreerobotics/G1_CameraPackaging_Dataset

unitreerobotics/G1_摄像头封装数据集

Preview 预览 · Updated 7 days ago 更新于 7 天前 · 1

unitreerobotics/G1_ObjectPlacement_Dataset

unitreerobotics/G1_对象放置数据集

Preview 预览 · Updated 7 days ago 更新于 7 天前 · 1

unitreerobotics/G1_Pouring_Dataset

unitreerobotics/G1_倒灌数据集

Preview 预览 · Updated 7 days ago 更新于 7 天前 · 1

unitreerobotics/UnitreeG1_DualArmGrasping

unitreerobotics/UnitreeG1_双臂抓取

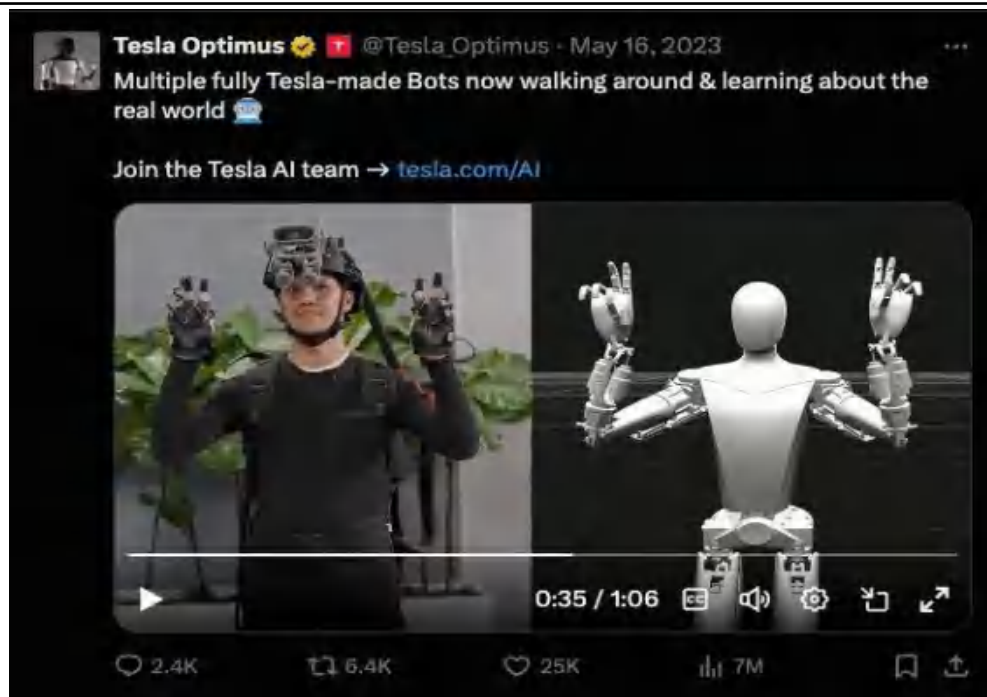
Preview 预览 · Updated 28 days ago 更新于 26 天前 · 107 · 123

公众号 · 量子位

3.5 主要厂商的做法：特斯拉

- 特斯拉在AI系统训练中采取的是以真实数据为驱动的策略。为训练Optimus机器人执行类人动作，特斯拉大量招募“数据收集操作员”，要求其穿戴动作捕捉服、VR头显等设备，模拟机器人需完成的动作（如行走、搬运物品）。Optimus的核心亮点之一在于其利用动作捕捉技术，以及VR和AR技术的融合应用，以实现更为真实的人类动作模拟。

图表：特斯拉的Optimus社交媒体账户发布的“数据收集操作员”的工作视频

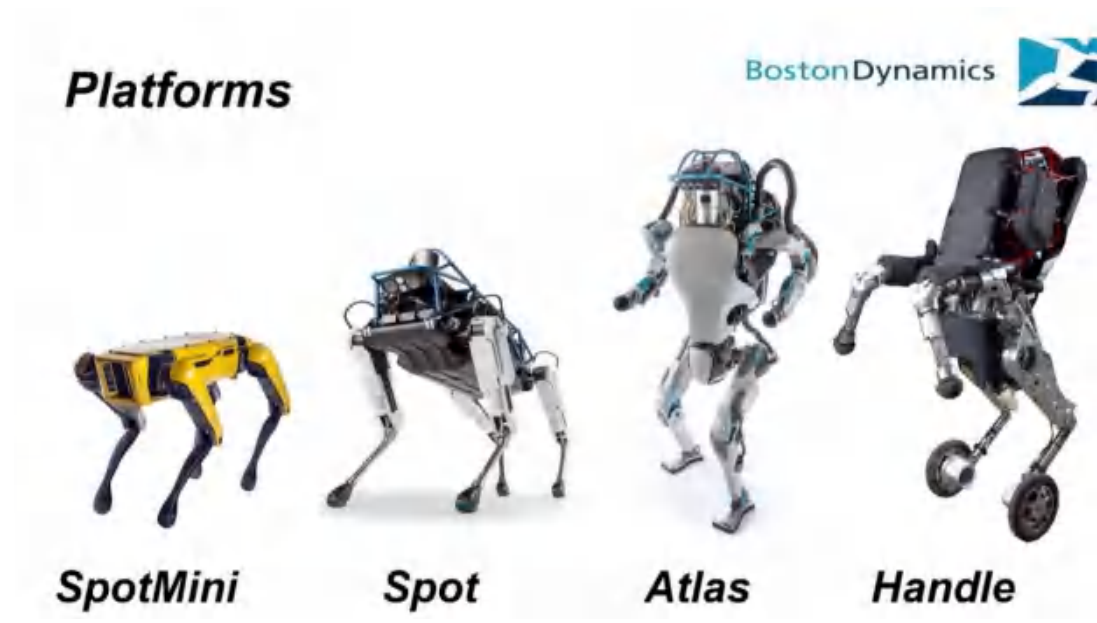


资料来源：新浪财经，中泰证券研究所

3.5 主要厂商的做法：波士顿动力

- 以波士顿动力的 Atlas 人形机器人为例，Atlas 依靠其视觉系统进行任务排序的推理，低层控制器进行行动规划。而不管是“大脑”视觉系统还是“小脑”控制系统，都离不开数据训练。
- Atlas 的训练过程融合了多个数据来源，包括真实工厂部署数据、实验室测试数据以及仿真环境中的模拟数据。通过对这些数据进行强化学习，不断训练、调优规控算法，Atlas 能够逐步掌握类似人类的精细操作技能。

图表：波士顿动力的几款主要机器人



资料来源：CSDN，中泰证券研究所

图表：波士顿动力机器人Atlas



资料来源：CSDN，中泰证券研究所

3.5 主要厂商的做法： Figure AI

- Figure AI创建于2022年，相较于国内外几家知名机器人公司成立时间较晚，侧重于研究具身人工智能发展。
- Figure AI基于端到端运动控制算法，结合OpenAI的GPT，先于特斯拉公开展示了人形机器人在无需预设程序的情况下的高度自主性和智能化交互，形成模型训练-云端设施-车厂落地的闭环。
- 近期，结束与Open AI的合作后，Figure AI实现将首个自研具身模型 Helix 装入机器人Figure。预告中展示了通过通用的“视觉-语言-感知”（VLA）模型Helix训练机器人，将感知、语言理解和学习控制统一起来，展示出较大的多机协作与强大的泛化功能。其原理包括双系统建构、端到端训练、解耦架构、优化推理部署。

图表：Figure目前达成合作的部分公司（截至2024年底）



资料来源：Figure AI官方社媒，51CTO.com，中泰证券研究所

图表：Figure最新发布的两机合作完成一系列家务的演示视频截图

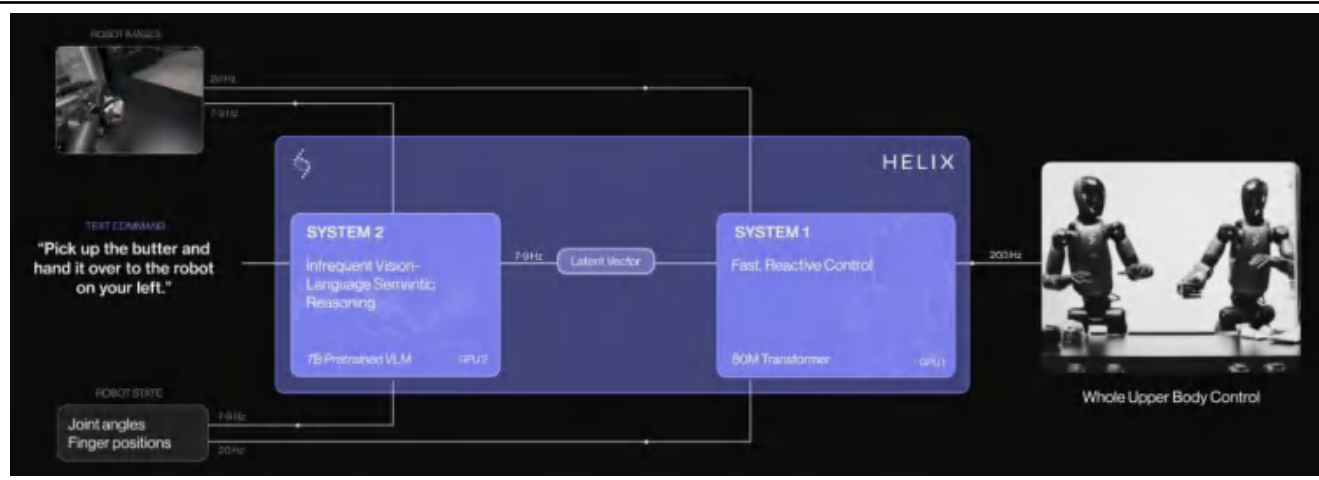


资料来源：Figure AI官方社媒，中泰证券研究所

3.5 Figure AI的Helix

- Helix是一个通用的“视觉-语言-感知”（VLA）模型，将感知、语言理解和学习控制统一起来，可以通过自然语言直接控制人形机器人整个上半身。
- Helix是首创的“系统1、系统2”VLA模型，用于对整个仿人上半身进行高速、灵巧的控制。之前的方法面临着一个基本的权衡：VLM骨架通用但不快速，而机器人视觉运动策略快速但不通用。Helix通过两个互补系统解决了这一问题，这两个系统经过端对端训练，可以进行通信：
 - 系统2：经过互联网训练的板载VLM，工作频率为7-9 Hz，用于场景理解和语言理解，可广泛应用于各种物体和环境。
 - 系统1：快速反应视觉运动策略，将系统2生成的潜在语义表征转化为200 Hz的精确连续机器人动作。
- 这种解耦架构允许每个系统以其最佳时间尺度运行，系统2可以“慢思考”高层次目标，而系统1则可以“快思考”实时执行和调整动作。
- 在训练效率方面，Helix以极少的资源实现了强大的对象泛化能力。研究团队总共使用了约500小时的高质量监督数据对Helix进行训练，其规模仅为之前收集的VLA数据集的一小部分(<5%)，并且不依赖于多机器人-具身收集或多阶段训练。
- 同时，Helix只需一个统一的模型，就能在不同任务中实现强大的性能。Helix的系统2参数量为7B，系统1则为8000万。

图表：Helix：用于整个上半身控制的“系统1、系统2”VLA



资料来源：澎湃新闻，中泰证券研究所

3.5 主要厂商的做法：银河通用

- 银河通用机器人有限公司成立于2023年，专注于通用具身多模态大模型机器人研发，并与北京大学等多所高校成立了具身智能相关联合实验室。银河通用研发的端到端具身抓取基础大模型GraspVLA及基于仿真数据驱动的训练范式，能够显著提升机器人在复杂环境下的操作精度与适应性，尤其在处理多变材质、动态工况等挑战场景中展现出卓越优势。
- GraspVLA 的训练包含预训练和后训练两部分。其中**预训练完全基于合成大数据**，训练数据达到了十亿帧“视觉-语言-动作”对，掌握泛化闭环抓取能力、达成基础模型。而针对特别需求，后训练仅需小样本学习即可迁移基础能力到特定场景，维持高泛化性的同时形成符合产品需求的专业技能。
- 整个开发过程无需大规模真实数据、仅通过合成数据达到基础模型，开创了VLA领域发展的全新范式。

图表：银河通用发布的第一代大模型人形机器人盖博特



资料来源：银河通用，中泰证券研究所

图表：由银河通用首次给出的 VLA 达到基础模型需满足的七大泛化金标准



资料来源：腾讯新闻，中泰证券研究所

3.5 主要厂商的做法：智元机器人

- 2024年，智元机器人在上海建了一座数据采集工厂，该数据采集工厂投入使用两个多月，就采集了超百万量级真机数据集，采集任务超一千种。现在智元数据采集工厂投放了近百台机器人，日均采集3-5w条数据。
- 12月30日，智元机器人宣布开源 AgiBot World（智元世界）。智元AgiBot World数据集包含超过100万条真实机器人操作数据，覆盖家居（40%）、餐饮（20%）、工业（20%）、商超（10%）和办公（10%）五大场景，涵盖80余种日常技能（如抓取、折叠、熨烫）和3000多种真实物品。数据质量从实验室级上升到工业级标准。
- 2025年2月，智元机器人宣布推出自主研发的大型仿真框架AgiBot Digital World，为机器人操作提供灵活的仿真数据生成方案、预训练的大规模仿真数据和统一的模型评测标准，同步开源海量仿真数据。

图表：智元具身智能数据系统AIDEA



资料来源：焉知科技，中泰证券研究所

图表：从事不同操作的智元机器人



资料来源：智元机器人，第一财经，中泰证券研究所

3.5 主要厂商的做法：优必选

- 在优必选内部，人形机器人训练的数据的收集有这样一个“二八定律”：用于人形机器人训练的20%的数据，由真实的机器人在实际场景中进行遥操作收集而来，80%的数据则是在仿真环境下生成而来。因此，优必选搭建了人形机器人智能仿真系统UNDERS2，可以低成本地为人形机器人模拟和生成多样化的场景。
- 同时，优必选通过与比亚迪、吉利、东风柳汽等车企合作，将人形机器人（如Walker S系列）部署在总装车间实训，打造人形机器人与无人物流车等协同作业的工业场景解决方案。
- 目前优必选正在人形机器人应用场景中验证DeepSeek技术的有效性，如多模态人机交互、复杂环境中的指令理解、工业场景中的任务分解与规划等，有望利用推理大模型的深度思考能力解决复杂任务中的挑战。

图表：优必选人形机器人智能仿真系统UNDERS2

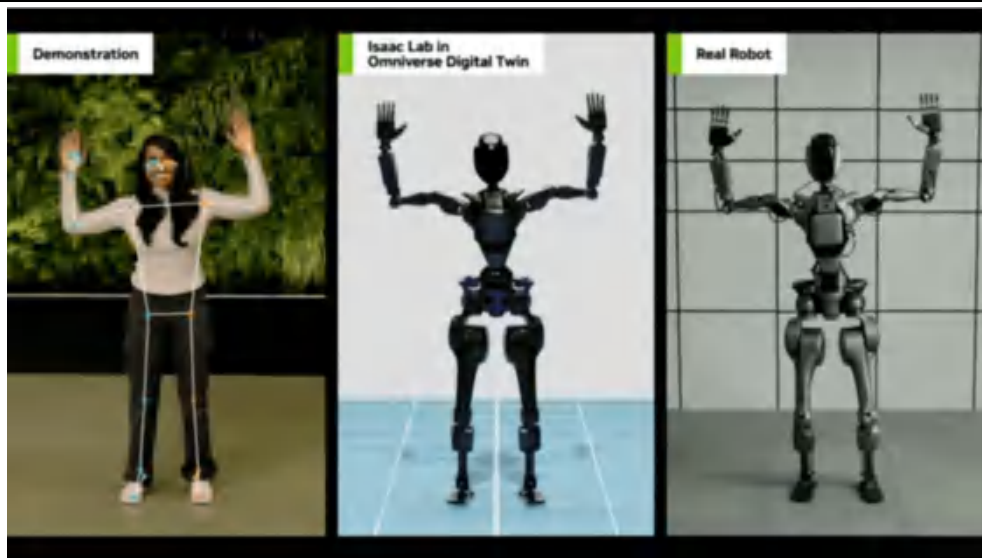
图表：极氪和优必选实现全球首例多台多场景多任务人形机器人协同实训



3.5 主要厂商的做法：傅利叶

- **GR-1: VR动作捕捉，实现Sim2Real。**傅利叶研发团队率先在安全可控、低成本的仿真环境中训练GR-1，再将训练好的模型迁移到现实世界（Sim2Real）。这种方式大幅提升了算法的迭代开发效率，降低了算法开发的总体成本。研发人员穿戴VR和动作捕捉设备，遥操作控制GR-1完成各种任务，记录动作数据，用于训练和改进仿真模型，优化机器人运动控制策略，增强在现实中的表现，像人类一样灵活决策，完成各种任务。
- **利用 NVIDIA Isaac Gym 开发人形机器人 GR-2：**傅利叶团队采用了 NVIDIA Isaac Gym（现已弃用）进行强化学习，利用 NVIDIA Isaac Gym 开发人形机器人 GR-2，目前正在将其 workflow 迁移到 NVIDIA Isaac Lab。同时利用 Isaac Gym 对抓取算法进行预训练，在实际部署前，对成功率进行仿真测算。这种方法显著减少了真实世界中的试错，节省了时间和资源。GR-2支持VR遥操作控制、拖拽示教、上位机末端控制等多种上肢示教方式。

图表：傅利叶机器人的方法——观察人类动作，进行模仿学习



资料来源：傅利叶官网，中泰证券研究所

图表：傅利叶GR-2



资料来源：机器人大学讲堂，中泰证券研究所



4

具身智能赋能下，未来机器人发展方向

4.1 人形机器人产业未来展望：从专用到通用，从ToB到ToC

- 从迭代路径看，To B市场是人形机器人发展的必经阶段，To C市场将成为远期重点方向。
- 从技术领域看，“大脑”“小脑”属于ICT领域，“肢体”属于工业装备领域，两个领域应协同发力，融合发展。但不同阶段的发力重点各有侧重。
- 在产业落地初期、应用规模较小时，人形机器人能否落地的重要因素就是其智能化水平；在大规模应用阶段，成本问题成为核心因素。
- 人形机器人落地应用节奏判断：
 - 1) 在工业制造场景，机器人的任务执行和流程和任务本身有高度规则性，机器人具备极高的劳动力替代性。2) 而服务场景更加开放，相关任务更加多样化，更加不可预测，需要机器人具备更强的自主决策能力与泛化适应性。

图表：具身智能机器人应用场景节奏预测

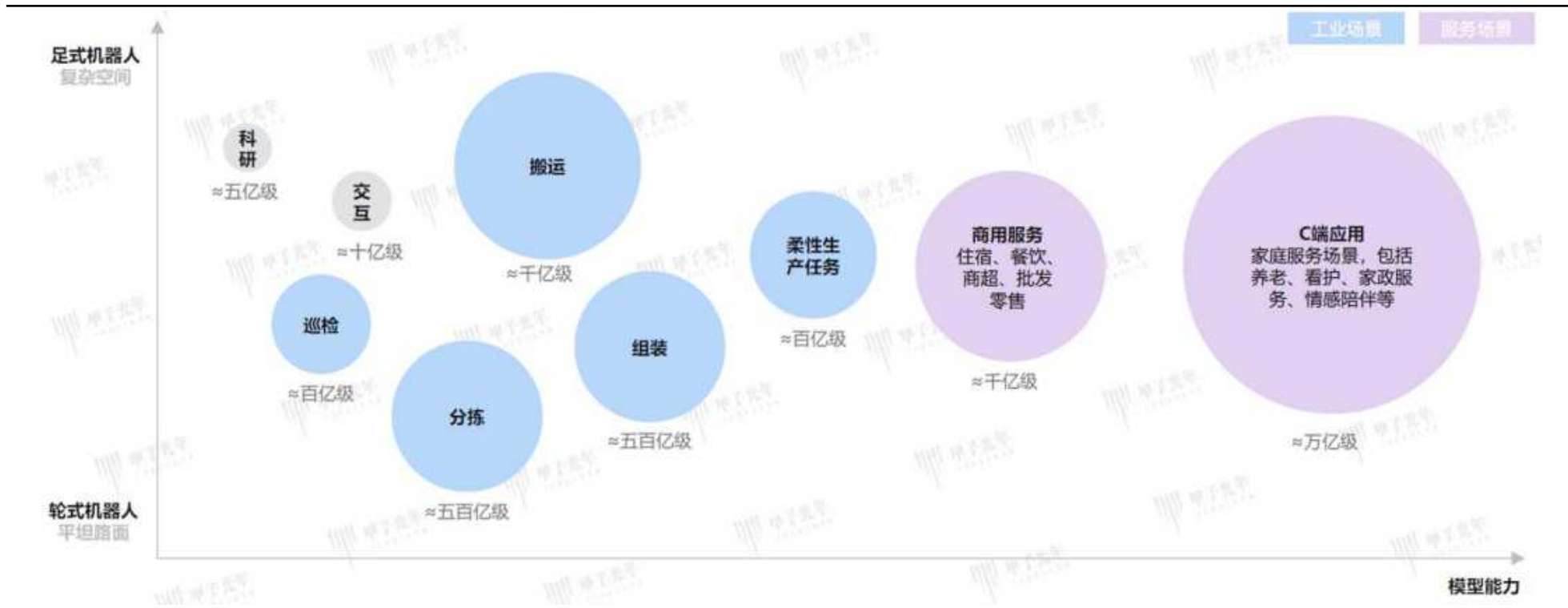


资料来源：甲子光年，中泰证券研究所

4.1 从工业制造的确定性走向家庭服务的灵活性

- 机器人在C端的应用最具想象力，但短期来看，任务相对聚焦，对泛化能力要求不高工业制造场景下的任务正在更快进入商业化阶段。
- 在工业制造场景实现商业化落地之后，海量机器人的具身数据叠加算力技术的进步，机器人的能力将循序渐进逐步解锁，并向商用服务、家庭服务等更开放的场景进行延伸，届时市场有望达万亿级。

图表：具身智能机器人在各类应用领域的市场空间分布



资料来源：甲子光年，中泰证券研究所

4.1 人形机器人商业化落地三阶段

■ 根据中国信通院的研究，人形机器人或将分三个阶段商业化落地。

- **第一阶段（2024—2030年）：**通过政策牵引，深入挖掘危险作业、极端环境等高价值场景，遴选一批可落地可推广的典型需求。政府搭桥，推动特种应用场景供需双方对接，定向开发一批产品并落地应用。
- **第二阶段（2030—2035年）：**加快探索工业制造、物流等制造业相关的大规模应用场景，从“替代相对简单且重复性的劳动”开始，成熟一代应用一代，在迭代中加速技术成熟、降低单体成本、提升整体性能，逐步提高对制造业场景的渗透率。
- **第三阶段（2035年后）：**加强人工智能与人形机器人的融合创新，实现更高水平的具身智能，并推动人形机器人进入医院、学校、商场、餐厅等服务业场景，最终走入千家万户

图表：人形机器人商业化落地三阶段



资料来源：中国信通院，中泰证券研究所

4.1 率先落地场景：工业制造

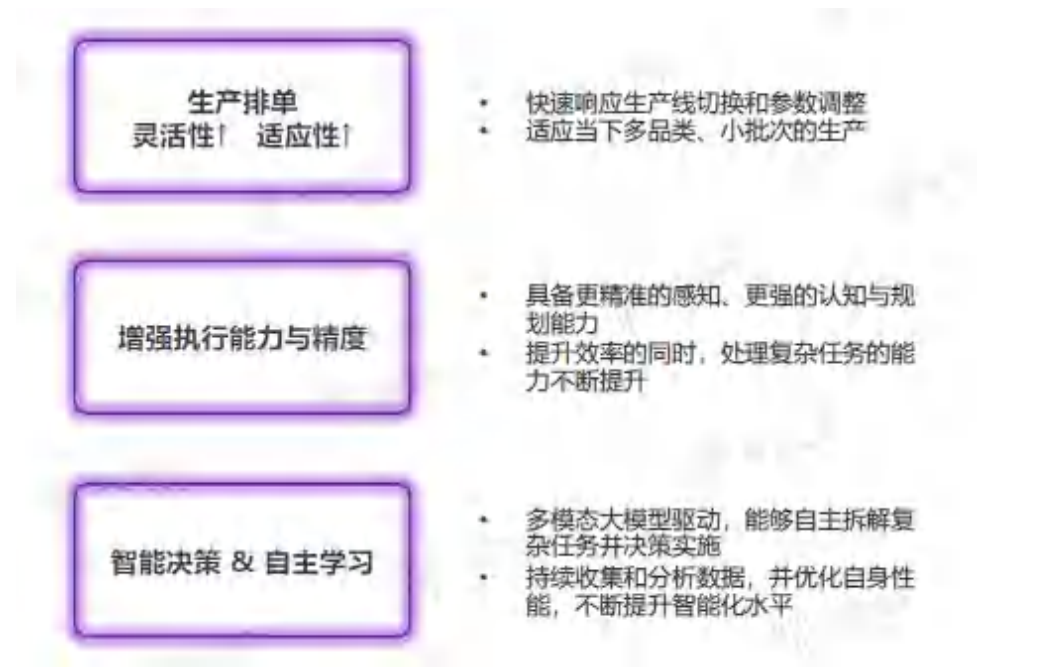
- 工业制造流程的特点使得该场景在具身智能技术落地应用方面占据先发优势，如柔性生产需求迫切、工作环境结构化程度高、成本效益优势突出等需求特点，刺激着工业制造场景客户对工业具身智能的应用更加期待。
- 工业具身智能机器人能够有效提升工厂生产任务的灵活性与适应性，并在作业过程中实现自主学习，不断增强其复杂任务执行能力与操作精度。
- 具身智能的训练需要数据，而质量更高的数据需要从真实场景中来，机器人落地最快的场景仍然是场景相对封闭的工业制造场景，越早实现规模化落地，就有望越早实现模型的能力提升。

图表：工业场景率先应用具身智能的原因



资料来源：甲子光年，中泰证券研究所

图表：具身智能机器人率先在工业制造场景创造价值



资料来源：甲子光年，中泰证券研究所

4.1 工业制造应用范例：配天机器人及其应用

- 配天机器人是一家专注于工业机器人、核心零部件及行业自动化解决方案的提供商，是京城机电旗下的国家级高新技术企业。公司始终致力于机器人技术的自主研发及高端装备技术瓶颈的突破，同时前瞻性布局打造以具身智能、机器人为大模型为核心的“AI+机器人”。
- 植根于工业制造场景，配天在“AI+机器人”已经成功研发免示教焊接软件模块，基于绎零机器人运动控制引擎，通过视觉检测和感知技术实现对焊接任务的快速识别和自主调整，无需人工示教即可投入使用，适应工厂灵活的排产需求。
- **应用案例：船厂BK板免示教焊接。**某船舶制造厂商专注于碳钢船体结构的生产，其生产线包含焊接工序。配天免示教焊接软件模块通过集成3D视觉系统，能够自主识别工件信息，无需预先进行人工标记。该模块可自动提取焊缝特征，并基于此选择合适的焊接路径规划和工艺参数，最终驱动机器人完成焊接作业，有效满足该产线的柔性化生产需求。

图表：配天机器人的工业自动化解决方案



图表：配天机器人AI赋能船厂BK板免示教焊接



4.1 工业制造应用范例：伟景智能及其应用

- 伟景智能创立于2016年，是一家专注于立体智能视觉系统和智能人形机器人研发、生产及应用的高科技人工智能公司。公司的核心子公司伟景机器人成立于2017年，专注于机器人和立体智能视觉系统研发与应用，拥有完整的智能机器人的生产线和3D立体智能相机整体解决方案，并形成了以自主核心算法、核心软件、核心硬件及行业系统解决方案为一体的全产业价值链。
- 以生产装配、危险作业为例：**1) 生产装配：**晓唯机器人可以在汽车制造、电子设备生产等工厂中，承担一些重复性、高精度的装配任务，例如在汽车生产线上安装零部件、拧紧螺丝等，助力产线自动化和智能化，提高生产效率和质量。**2) 危险作业：**在一些有毒、易爆、高空等危险作业环境下，晓唯机器人可以代替人类进行危险操作，减少或避免人员受伤甚至生命危险，可以最大程度地保障人员安全，提升生产环境的安全性和工作效率。

图表：伟景机器人的主要产品——双足版、采摘版、AGV版人形机器人



图表：晓唯机器人可用于生产装配工作，提升生产效率和质量



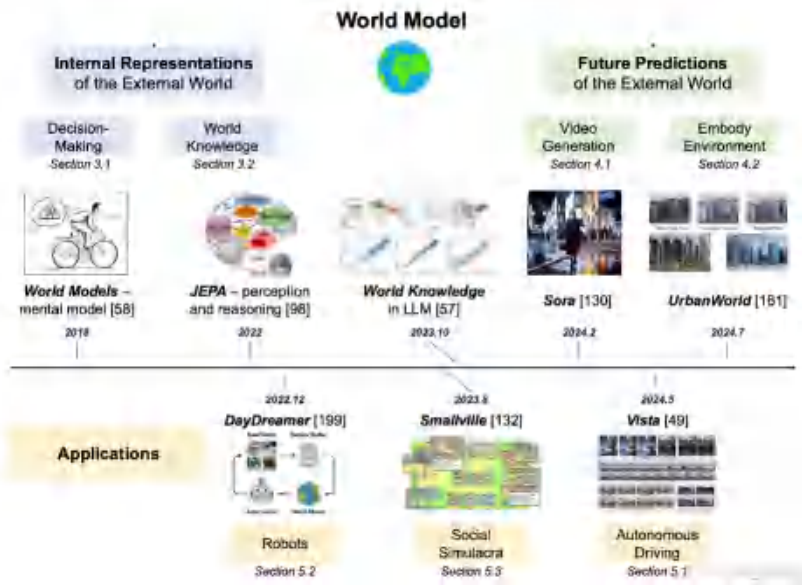
资料来源：伟景机器人，中泰证券研究所

资料来源：伟景机器人，中泰证券研究所

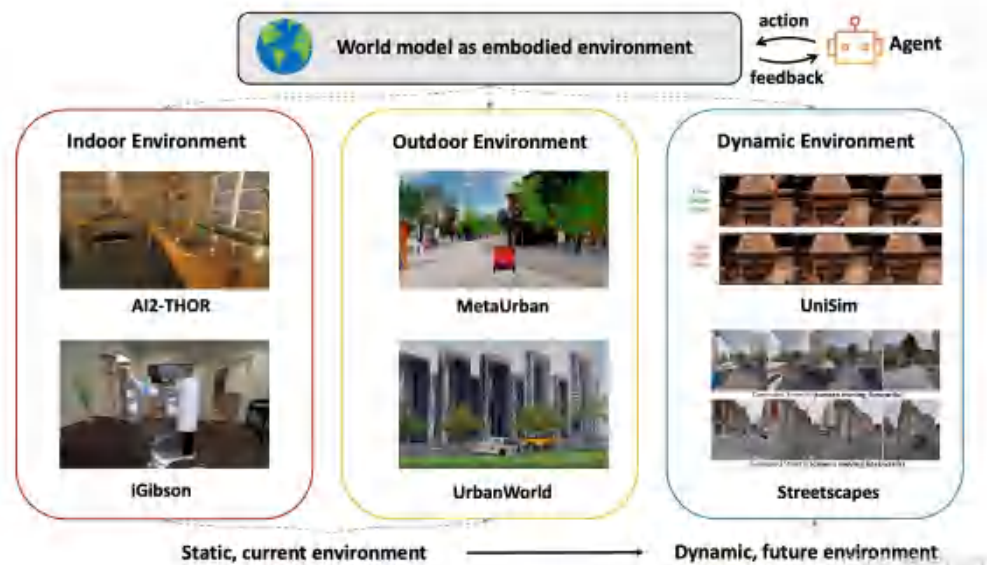
4.2 未来如何训练机器人：世界模型

- 世界模型：被认为是实现通用人工智能（AGI）的重要环节，终极目标是构建一个统一的模型能够复制基本的世界动态。
- 世界模型可大体分为两部分：内部表征和未来预测。1) 内部表征：用于理解世界机制；2) 未来预测：预测未来状态以模拟和指导决策。
- 作为具身环境的世界模型：具身环境世界模型的开发对于模拟和预测智能体如何与外部世界交互和适应外部世界至关重要。这些模型不仅代表世界的视觉元素，还结合更准确反映现实世界动态的空间和物理交互。通过整合空间表示并从基于视频的模拟过渡到沉浸式具身环境，世界模型现在可以为开发能够与复杂的现实世界环境交互的智能体提供更全面的平台。

图表：世界模型的分部及其研究发展历程



图表：作为具身智能的世界模型可分为三类：室内、室外和动态环境



4.2 世界模型正在打造具身智能技术进步的“演武场”与“弹药库”

- 对于具身智能，世界模型能够辅助具身大模型进行训练，提供的训练型的仿真：其生成的视频交给具身大模型，具身大模型通过它的规划执行接下来的动作，接下来的动作交互产生新的场景、新的视角，再通过世界模型继续生成新的数据，进行闭环仿真的测试，成为具身智能机器人的“演武场”。
- 以NVIDIA新发布的Cosmos世界基础模型为例，世界模型能够为自动驾驶、具身智能机器人等多种AI模型提供符合物理规律的逼真数据，大幅降低数据采集与标注的成本，充足的“弹药”有望在未来极大地加速具身智能的学习速度。

图表：几种主要世界模型的对比

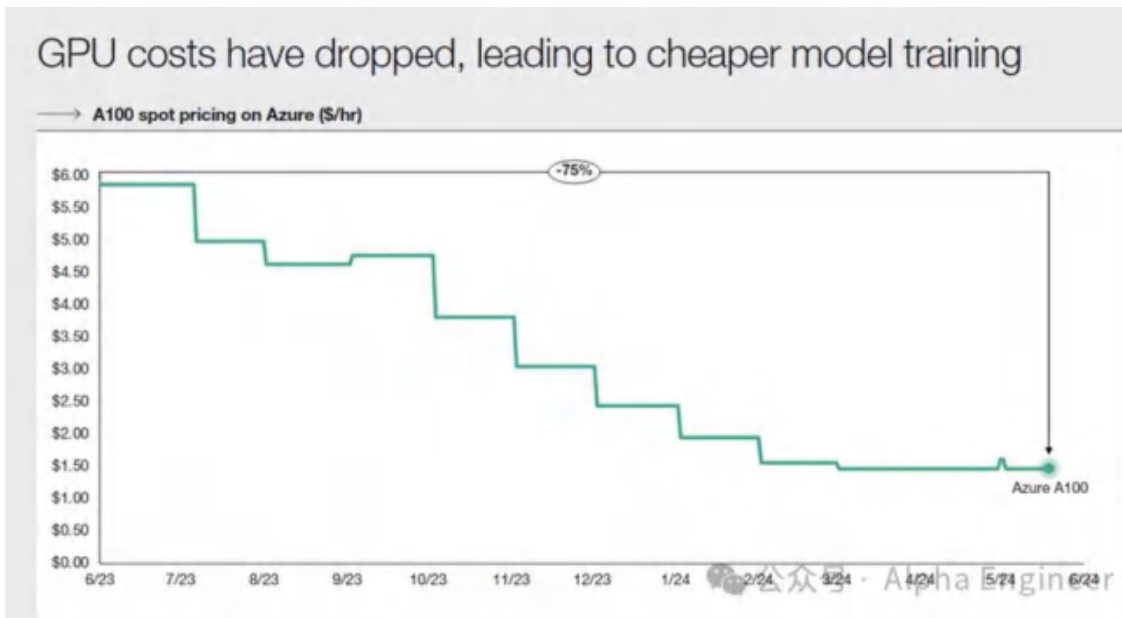
	World Labs	Deepmind-Genie 2	Nvidia-Cosmos
侧重点	从单张图像生成3D世界，强调空间智能	生成可交互、可玩的3D环境，用于训练智能体	构建物理感知AI模型，生成合成数据，强调物理准确性
技术方法	深度学习模型，3D重建	自回归潜在扩散模型，Transformer，CFG	世界基础模型平台，不同规格模型
核心能力	2D到3D的转换，遵循几何物理规则	生成可控的互动环境，模拟物理效果，长时间记忆	生成物理上准确的合成数据，理解物理因果关系
应用领域	3D内容生成、VR/AR、机器人导航	游戏开发、AI智能体训练、虚拟环境模拟	自动驾驶、机器人、物理模拟
未来应用场景	• 机器人导航：提供环境地图和物体识别； • 建筑设计、室内设计：从平面图生成3D模型； • 游戏开发：生成游戏场景、角色和道具； • 虚拟现实/增强现实：创建沉浸式体验； • 在线广告：生成商品3D模型，用于展示和虚拟试用； • 文化遗产保护：对古代建筑和文物进行3D重建。	• 机器人训练：在虚拟环境中训练机器人执行各种任务； • 游戏AI：训练更智能、更逼真的游戏角色； • 虚拟社交：创建虚拟社交环境，用于虚拟会议、虚拟活动等； • 教育和培训：创建交互式学习环境，例如虚拟实验室、虚拟手术室等； • 电影和动画制作：生成虚拟场景和特效。	• 合成数据生成：为各种AI模型提供训练数据，降低数据采集和标注的成本； • 自动驾驶：训练自动驾驶系统，提高其在复杂交通环境中的适应性和安全性； • 机器人：训练机器人在复杂环境中执行各种任务，例如工业自动化、物流配送等； • 物理模拟：用于科学研究和工程设计，例如模拟流体流动、材料形变等； • 灾难预测和模拟：模拟自然灾害或人为事故，用于应急预案制定和救援演练。

资料来源：甲子光年，中泰证券研究所

4.2 世界模型训练的可行性：机器人训练成本的大幅下降

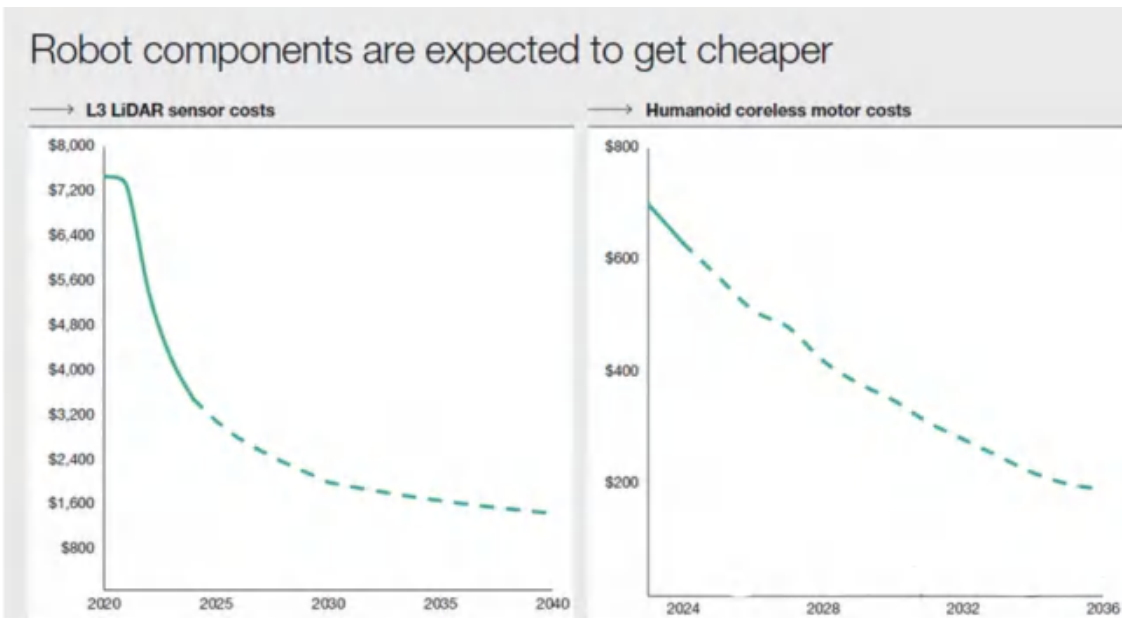
- **训练成本的降低：**随着GPU成本的下降，大模型训练的成本大幅降低。
 - 在截至2024年6月的过去一年中，Azure云平台上的A100显卡租赁价格从6美金/时下降到1.5美金/时，降幅达到75%。
- **硬件成本同样在快速降低。**
 - 2020年L3 LiDAR传感器的成本在7400美金左右，现在已经降至3200美金；类人无芯传感器的成本也出现了明显下降。

图表：GPU成本持续下降



资料来源：Coatue，中泰证券研究所

图表：硬件成本同样在快速降低



资料来源：Coatue，中泰证券研究所



5

投资建议&风险提示

投资建议

- 作为机器人的“大小脑”，在硬件已达到较高工艺水平的情况下，具身智能的技术能力或将成为机器人发展的天花板。当前多家领军机器人企业均已各自的方式在具身智能领域投入较多资源与精力，以提升自家机器人的泛化行为能力。
- 我们建议投资人持续关注具身智能产业的技术变革突破、商业化落地路线、节奏与进展等。具体到标的上，建议关注索辰科技（物理AI、机器人子公司）、能科科技（特种行业AI能力领先）等。

风险提示

- 机器人产业相关政策落地不及预期；
- 机器人商业化落地推进不及预期；
- AI与具身智能产业技术发展不及预期；
- 行业竞争加剧；
- 研报信息更新不及时的风险等。

重要声明

- 中泰证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会许可的证券投资咨询业务资格。本公司不会因接收人收到本报告而视其为客户。
- 本报告基于本公司及其研究人员认为可信的公开资料或实地调研资料，反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。本公司力求但不保证这些信息的准确性和完整性，且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，可能会随时调整。本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用，不构成任何投资、法律、会计或税务的最终操作建议，本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。
- 市场有风险，投资需谨慎。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。
- 投资者应注意，在法律允许的情况下，本公司及其本公司的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。本公司及其本公司的关联机构或个人可能在本报告公开发布之前已经使用或了解其中的信息。
- 本报告版权归“中泰证券股份有限公司”所有。事先未经本公司书面授权，任何机构和个人，不得对本报告进行任何形式的翻版、发布、复制、转载、刊登、篡改，且不得对本报告进行有悖原意的删节或修改。