



人身智能 初了解



具身智能目录

1. 什么是具身智能
2. 具身智能技术路线
3. 具身智能产业思考



2024

世界机器人大会



ZOMI

01 具身智能的定义



具身智能

- 通过在物理和数字世界中的学习和进化，实现理解世界、互动交互并完成任务的目标。由“本体”和“智能体”组成，并在真实的世界复杂环境中执行任务。

本体

实际的执行者，在物理或者虚拟世界中进行感知和任务执行

智能体

具身本体之上的智能核心，负责感知、理解、决策和控制

学习数据

泛化的关键，涉及机器人实际执行的数据稀缺且昂贵，目前分为仿真和实时采集

智能架构

智能体与物理世界的交互，适应新环境、学习新知识并强化解决问题的方法



具身智能

- **具身智能关键：**在于机器人能够通过视觉识别身边物理环境并做出决策和行动。
- **大模型关系：**多模态大模型 (MLM) 出现，使其成为具身智体大脑的较为有前途架构。

具身智能 = 视觉/触觉感知输入 + 算法决策（大模型） + 物理机械反馈执行



具身智体的典型架构

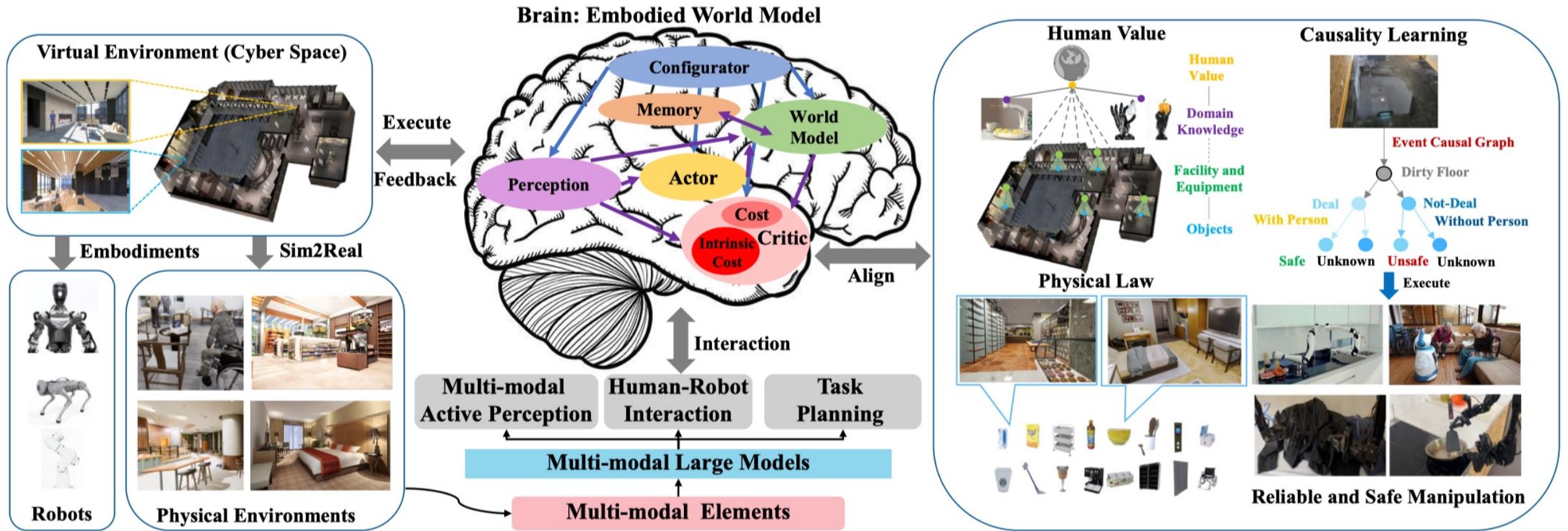
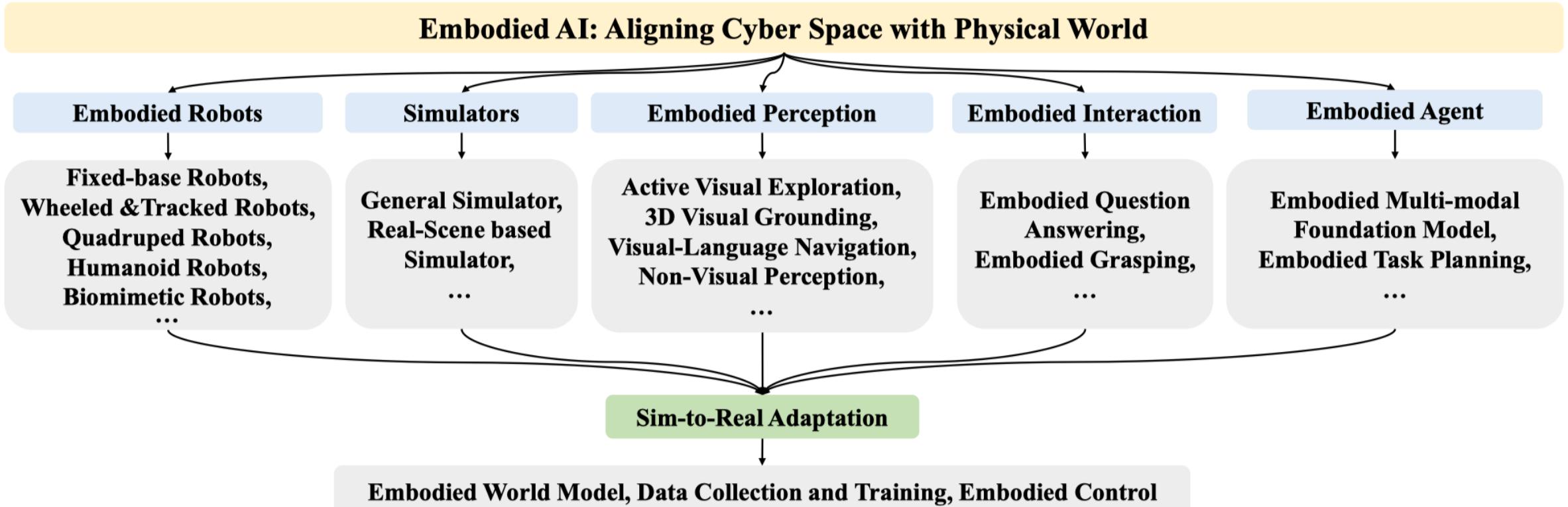


Fig. 2. The overall framework of the embodied agent based on MLMs and WMs. The embodied agent has a embodied world model as its “brain”. It has the capability to understand the virtual-physical environment and actively perceive multi-modal elements. It can fully understand human intention, align with human value and event causality, decompose complex tasks, and execute reliable actions, as well as interact with humans and utilize knowledge and tools.



具身智能涉及知识点

- 具身机器人；具身模拟器；具身感知；具身交互；具身智体；模拟到现实，包括具身WM、数据以及控制。



Applications Robotics Autonomous Driving Healthcare Domestic Assistance Industrial Automation Search and Rescue

Fig. 3. This survey focuses on comprehensive analysis of the latest advancements in embodied AI.



02 具身机器人



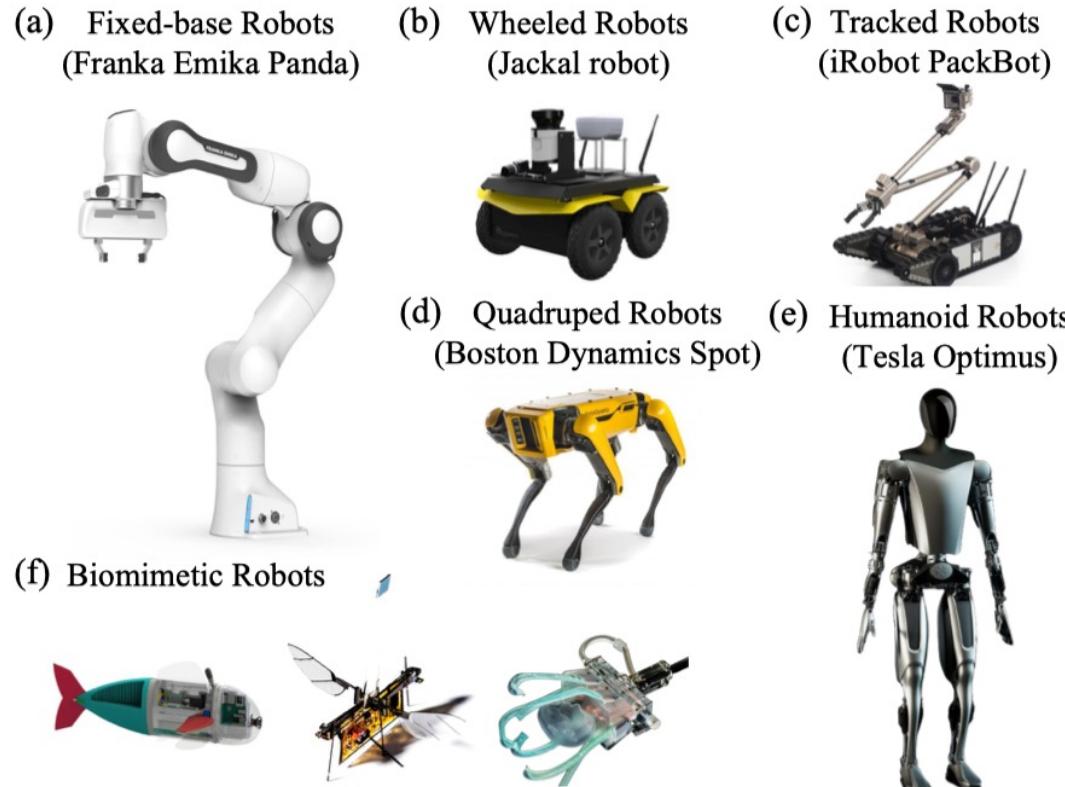
十八金刚 众创未来——人形机器人先锋阵列

Humanoid Robot Vanguard Matrix



具身的机器人形态

- 具身智体与物理环境互动，形态包括机器人、智能家电、智能眼镜和自动驾驶车辆等。其中，人型机器人作为最突出的具身形态之一，备受关注。



具身的机器人形态

- 具身机器人一般可分为：
 1. **固定基座型**，如机械臂，常应用在实验室自动化合成、教育、工业等领域中；
 2. **轮式机器人**，因高效的机动性而闻名，广泛应用于物流、仓储和安全检查；
 3. **履带机器人**，具有强大的越野能力和机动性，在农业、建筑和灾难场景的应对方面显示出潜力；
 4. **仿生机器人**，通过模拟自然生物的有效运动和功能，在复杂和动态的环境中执行任务。
 5. **四足机器人**，以其稳定性和适应性而闻名，非常适合复杂地形的探测、救援任务和军事应用。
 6. **人形机器人**，以其灵巧手为关键，在服务业、医疗保健和协作环境等领域广泛应用。



具身的机器人形态

- 具身机器人一般可分为：
 1. **固定基座型**，如机械臂，常应用在实验室自动化合成、教育、工业等领域中；
 2. **轮式机器人**，因高效的机动性而闻名，广泛应用于物流、仓储和安全检查；
 3. **履带机器人**，具有强大的越野能力和机动性，在农业、建筑和灾难场景的应对方面显示出潜力；
 4. **仿生机器人**，通过模拟自然生物的有效运动和功能，在复杂和动态的环境中执行任务。
 5. **四足机器人**，以其稳定性和适应性而闻名，非常适合复杂地形的探测、救援任务和军事应用。
 6. **人形机器人**，以其灵巧手为关键，在服务业、医疗保健和协作环境等领域广泛应用。

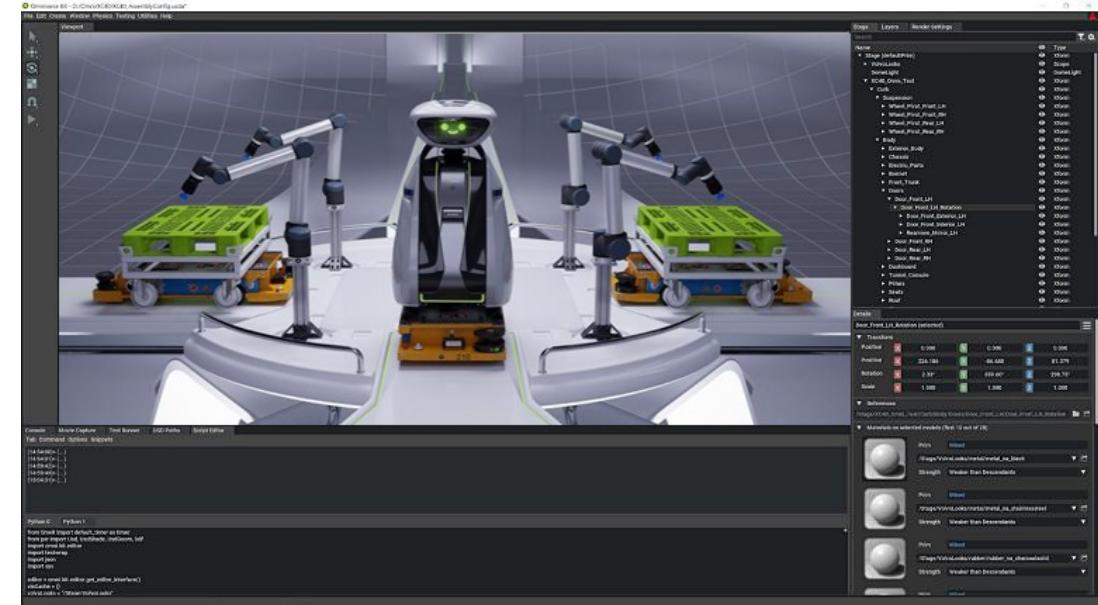
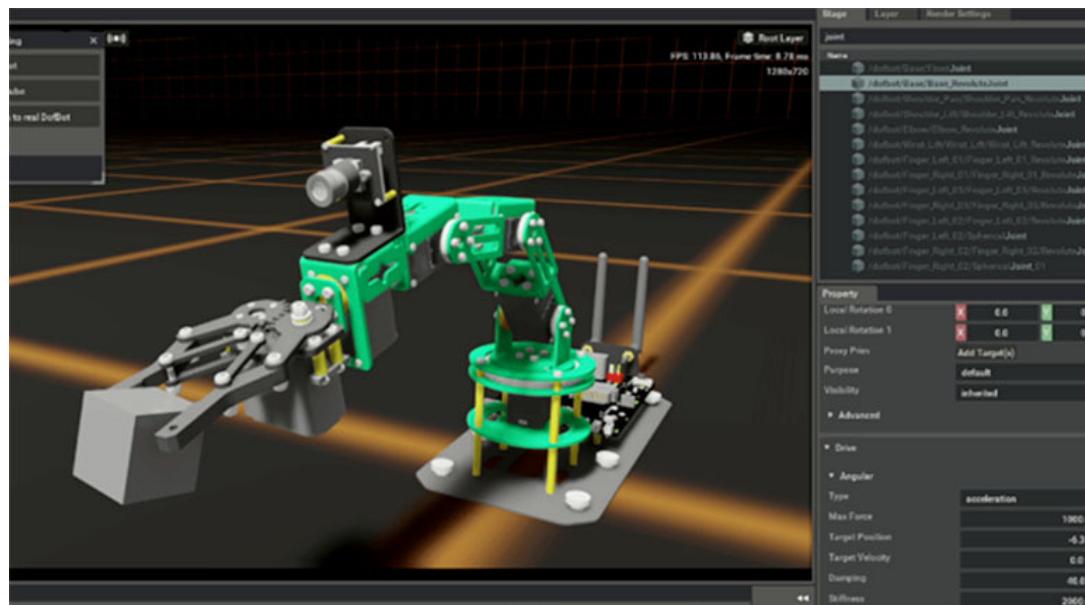


03 具身仿真



具身仿真

- 具有多样环境中进行测试的可扩展性，快速的原型设计能力，能够为广泛研究群体提供便利，提供用于精确研究的可控环境，生成用于训练和评估数据，并提供算法比较的标准化基准。
- 为了使智体能够与环境互动，必须构建一个物理的仿真模拟环境。需要考虑环境物理特性、目标属性及其相互作用的反馈结果。



[NVIDIA releases Isaac simulation on Omniverse](https://nvidia.github.io/isaac-sim/)





具身仿真

I. 基于底层仿真的通用平台

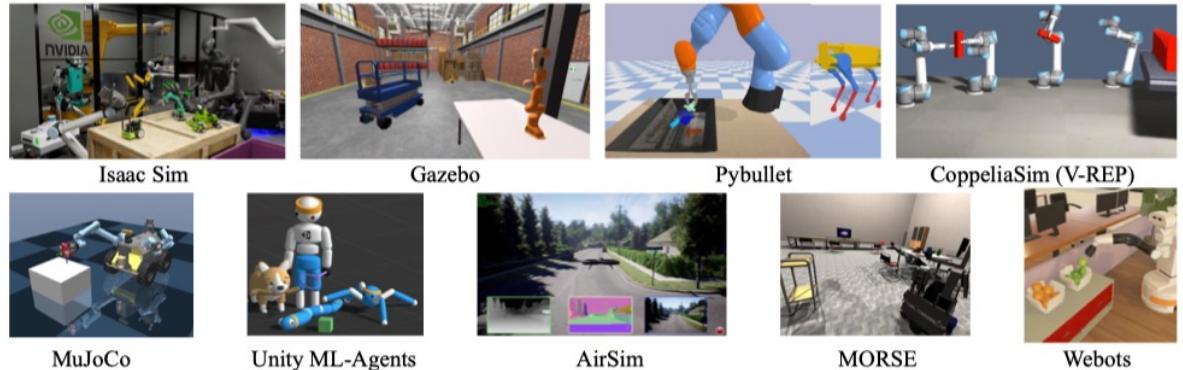


Fig. 5. Examples of General Simulators. The MuJoCo's figure is from [59].

2. 基于真实场景的仿真平台

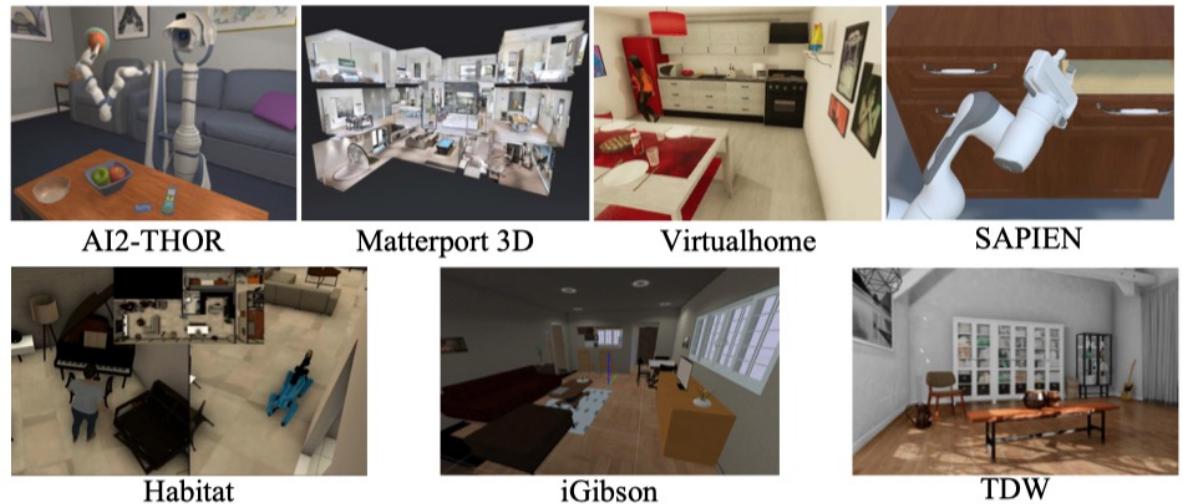


Fig. 6. Examples of Real-Scene Based Simulators.



1. 基于底层仿真的通用平台

- 通用平台，提供与物理世界高度相似的虚拟环境，用于算法开发和模型训练，具有显著的成本、时间和安全优势；

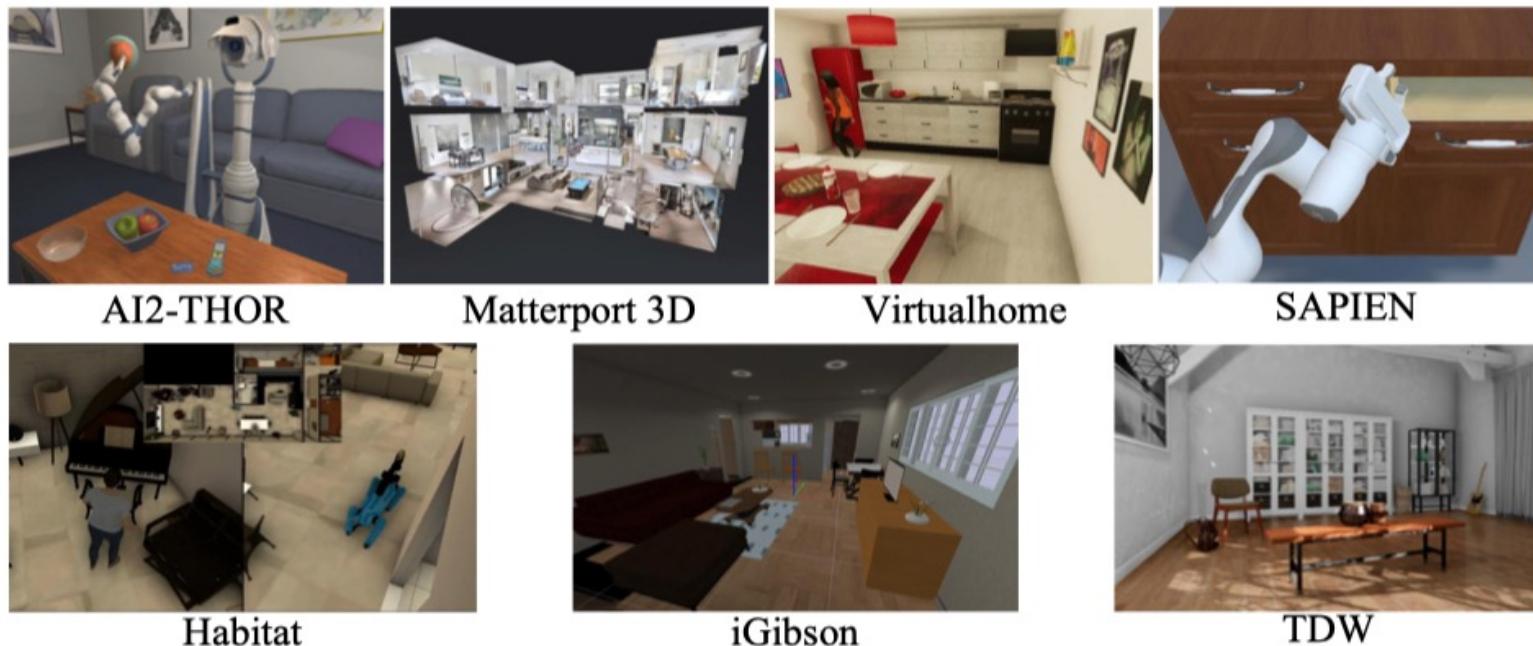


Fig. 6. Examples of Real-Scene Based Simulators.

2. 基于真实场景的仿真平台

- 基于现实世界环境的平台，从现实世界收集数据，创建逼真的 3D 资源，并使用 UE5 和 Unity 等 3D 游戏引擎构建场景。**具身智能研究首选。**

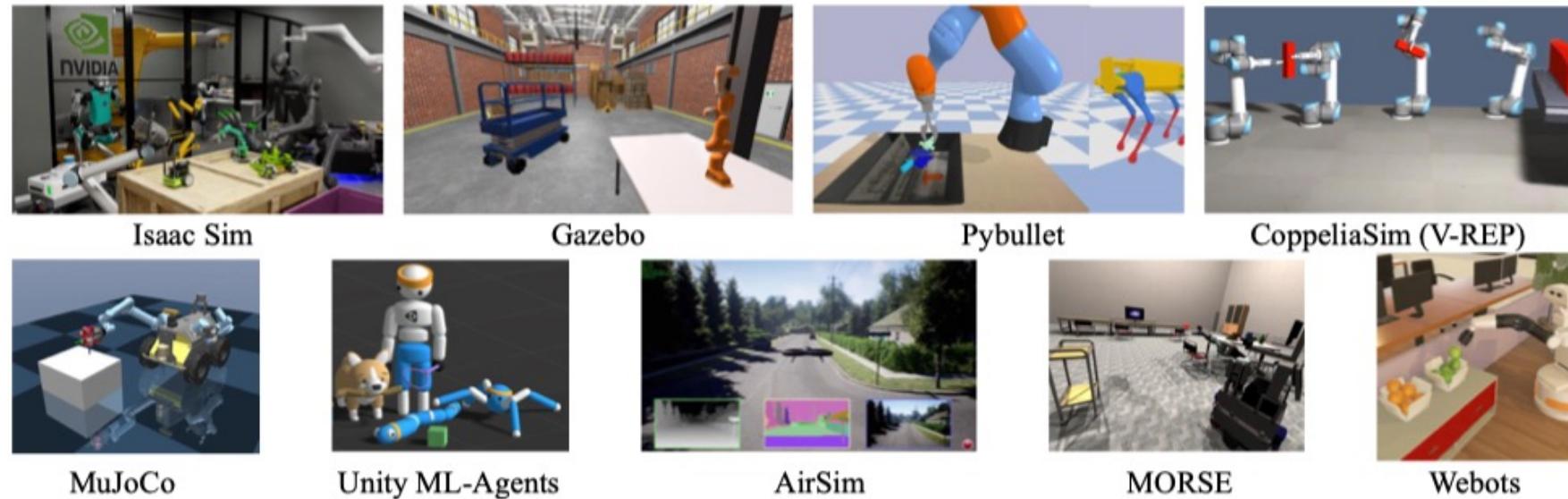


Fig. 5. Examples of General Simulators. The MuJoCo's figure is from [59].

04. 算法路线与 大模型关系



具身智能 Embodied Intelligence

- 具身智能（Embodied Intelligence）高级机器智能形式，它使机器人能够像人类一样感知和理解环境，并通过自主学习和适应性行为来完成任务。机器人的能力和实现过程抽象为：

感知 - 决策 - 执行



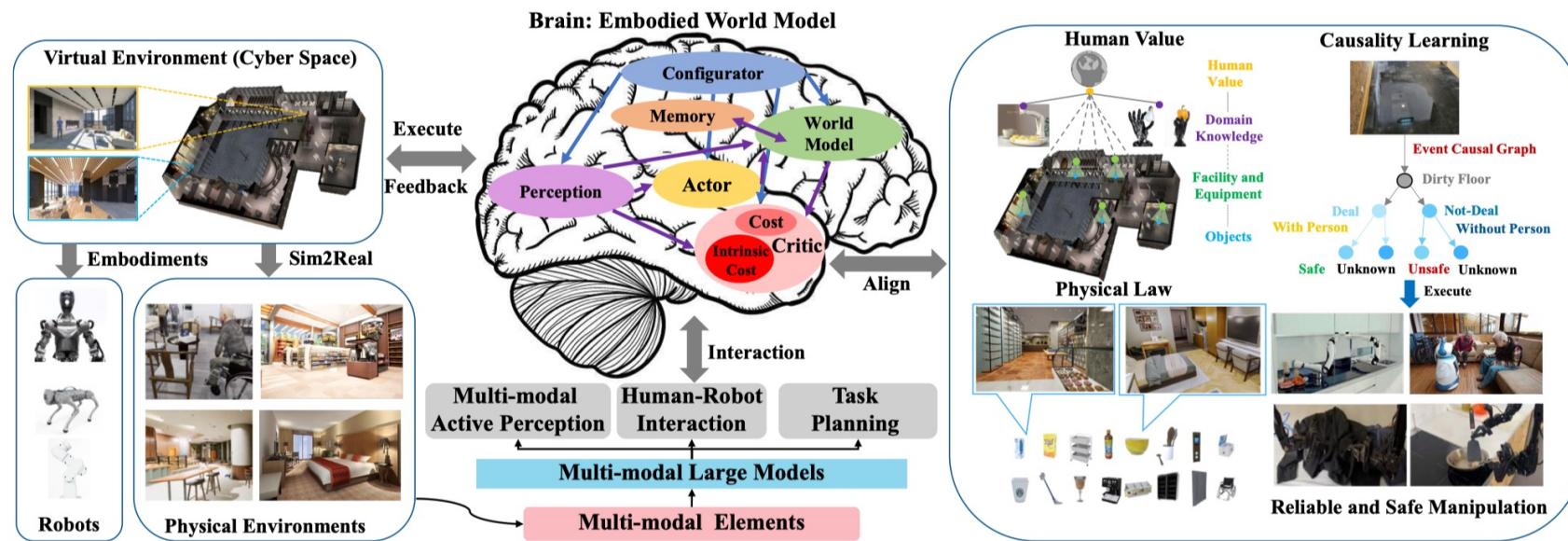
具身感知

- 负责获取和处理环境信息，为智能体的决策 Decision 和执行 Control/Action 提供依据。可以通过两种方式实现：
 1. 全感知，构建包含操作环境各种知识的大规模数据库或者环境；
 2. 具身交互感知，通过智能体与环境实时交互获取感知反馈，利用算法构建环境表征。



具身决策

- 决策是具身智能的核心，需要类似大脑的处理能力，以满足机器人在理解指令、分解任务、规划子任务、识别物体等需求。
- 大模型的突破为具身决策提供了新思路。大模型充当智体，利用强大语言理解和生成能力，与人类进行多维度交互，更好地理解指令和意图，生成恰当的响应和行为。



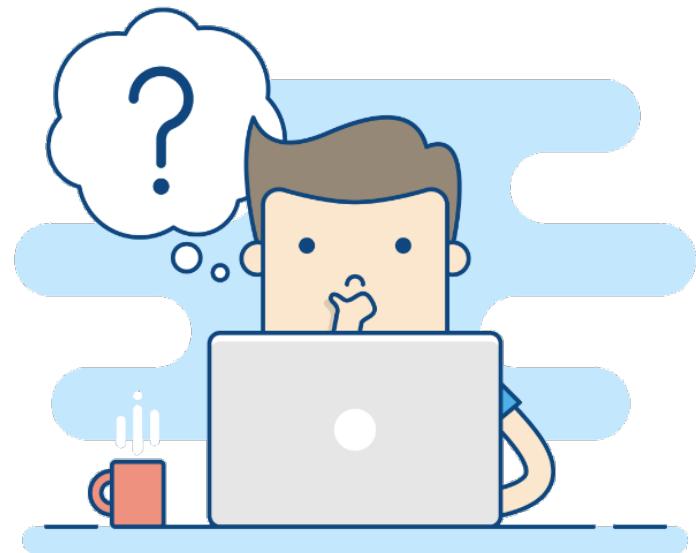
具身控制

- 控制根据感知信息和决策指令，协调机器人各部件的运动，实现智能化的行为控制。
- 多传感器融合通过综合利用多传感器信息，提高对环境感知和理解能力，实现更好的控制。



搞清楚一些概念

I. 具身感知、具身交互、具身大模型、具身控制之间的关系。



05 具身大模型



具身大模型

- 一方面商业上，初创公司并没有足够的科研、算法实力研究自己的大模型；
- 另一方面技术上，具身大模型不是真正的大模型，只是使用了大模型的 API 或者能力而已；



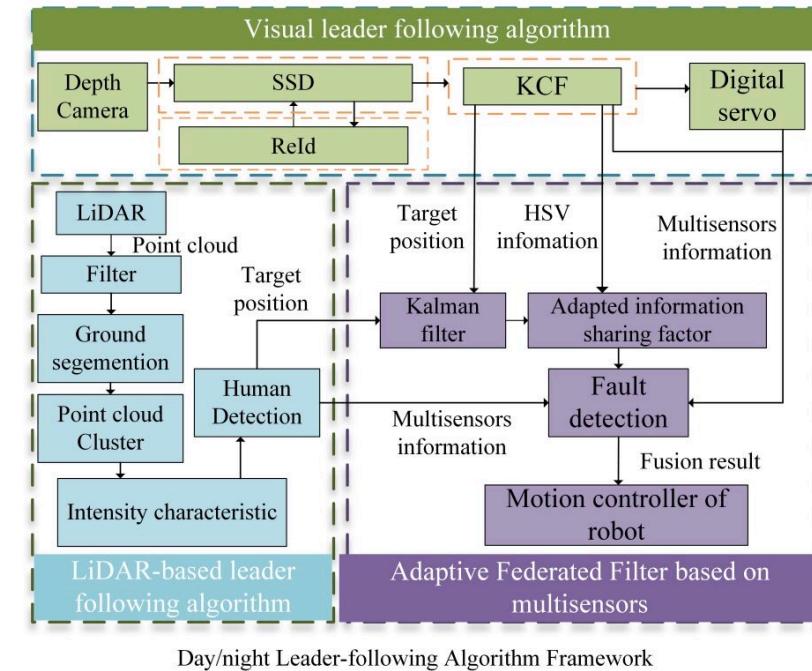
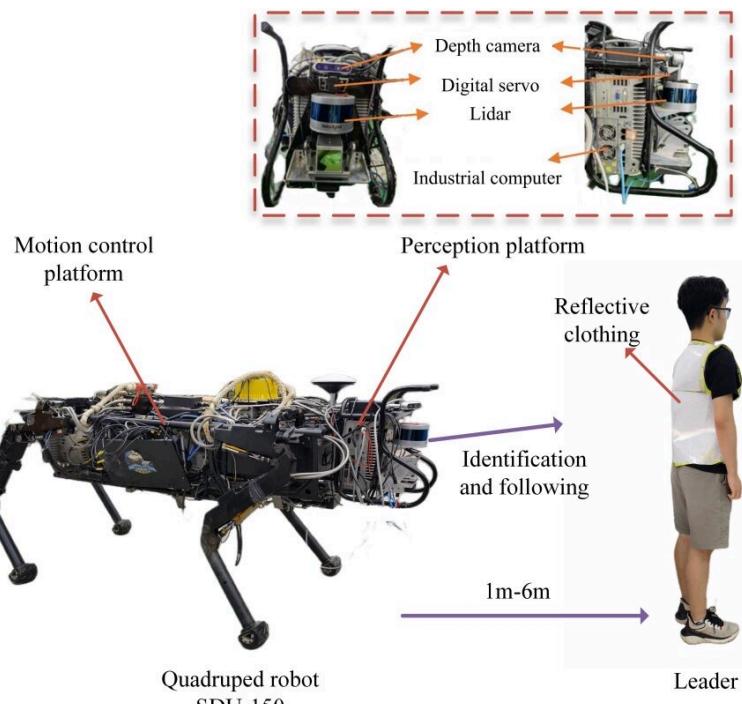
具身大模型

- 具身大模型的能力，体现在具身感知、具身决策、具身执行中使用了大模型。



大模型对具身智能的影响

- 业界具身智能大模型在实验室研发阶段，机器人通过视觉大模型对环境理解，语言大模型对任务进行分解，具身大模型生成机器人可执行的行为轨迹及代码。大模型让机器人在环境交互感知、自主规划、决策、行动方面有质的飞跃。

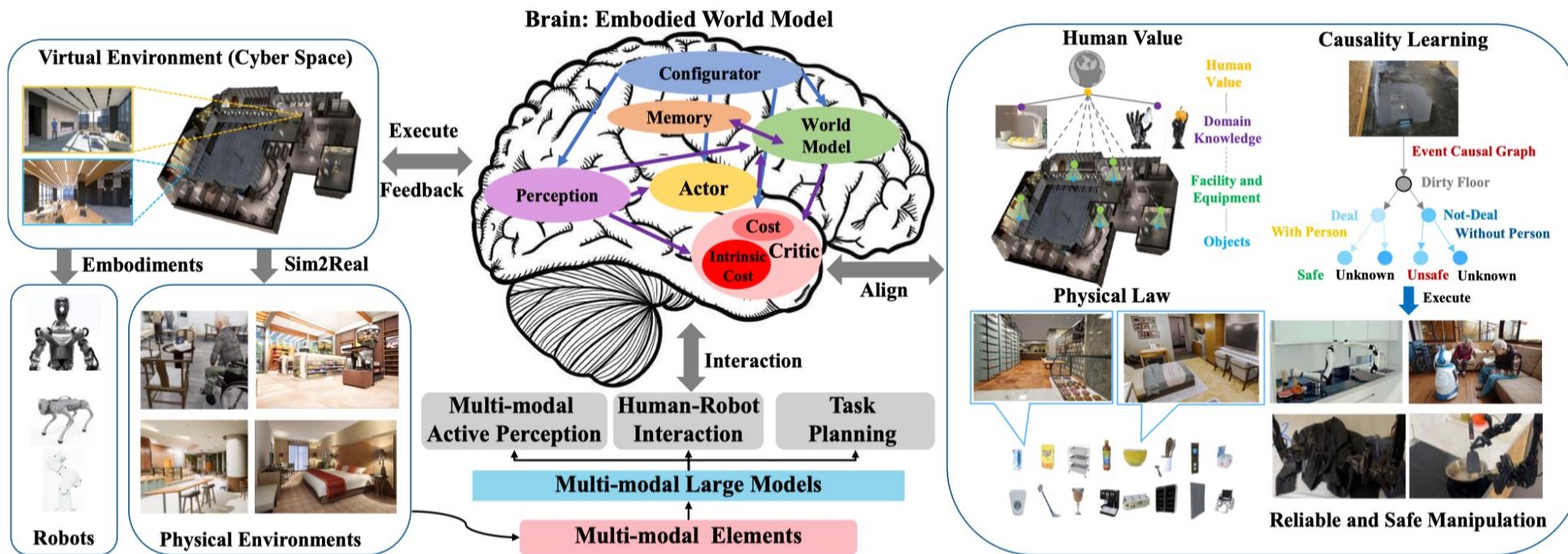


[A framework that allows four-legged robots to follow a leader in both daytime and nighttime conditions](#)



大模型对具身智能的影响

- 业界具身智能大模型在实验室研发阶段，机器人通过视觉大模型对环境理解，语言大模型对任务进行分解，具身大模型生成机器人可执行的行为轨迹及代码。大模型让机器人在环境交互感知、自主规划、决策、行动方面有质的飞跃。





Thank you

把AI系统带入每个开发者、每个家庭、
每个组织，构建万物互联的智能世界

Bring AI System to every person, home and
organization for a fully connected,
intelligent world.

Copyright © 2023 XXX Technologies Co., Ltd.
All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. XXX may change the information at any time without notice.



ZOMI

Course chenzomi12.github.io

GitHub github.com/chenzomi12/AIFoundation