

# 具身智能研讨会

具身智能正在成为人工智能的一个新方向

做让智能体感知、理解环境，同时让机器与环境交互

应用：学术、产业都飞速发展

会议目的：分享最新科研成果，以及应用落地的一些经验

碰撞出火花、提供新的研究思路或应用的方法

## 一. 王鹤 (北大助理教授、智源学者)

### —— 面向通用机器人的具身多模态大模型系统

#### ① 通用机器人

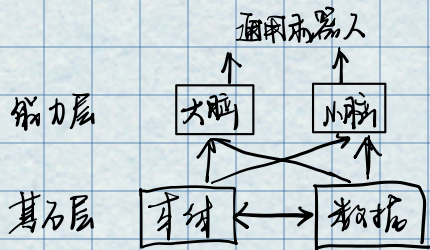
Generalist Robot

Open AI + Figure / NVIDIA Project GR00T / Tesla Optimus

• 任务通用：能化零为整，"举一反三"

• 环境通用：在各种地形环境中穿梭

图灵测试 → 测试到的元素去共同思考他们之间的关系。  
(通用机器人二难)



大脑：感知、决策问题

小脑：把感知决策转化成动作控制问题

身体：机器人的本体 { 自然语言描述 → 学习数据的 data.  
人：没有翅膀 → 无法学习飞行的 data

## 一. 身体

通用机器人

(通用机器人)

腿式人形机器人 最终形态，但不意味着所有机器人都要学习

思考：一般机器人和人一样

渐进式地去发展

上半身是首要的，没有上半身就什么都干不了

腿的身体是实现环境通用，辅助手实现任务通用 (手摸不到腿)

目前腿式人形机器人的腿部技能需要靠低姿态和跑，无法弯腰、蹲下捡、抓取高处

如果我们想让这个机器人像人一样在各种环境中穿梭，各种个路手到擒来，对腿的要求必须

多样化

人形和机器人学都会用到图灵测试，但不 → 今天没有解决

银河图灵 Galbot G1:



马路上，捡东西用右腿  
可抬高 2.4m.



希望未来人工智能公司

在 交互 的基础上, 把 交互 体系发展出来

## 二. 数据

通用机器入的能力层基于具身智能实现:

而这部分 data 目前必须依赖 选择件 的数据采集.



排列组合问题

非常复杂, 那为什么谷歌和特斯拉都要去做

: 特斯拉把自动驾驶做成了 (而成为行业)

而人工智能不行, : 这就要看人工智能本身  $1 \rightarrow$  每个数据材料好。

死锁  
循环

核心: 数据问题

$0 \rightarrow 1$ : 合成数据

learning on 3D synthetic data and then Sim2Real

零部件, 灵巧手, 深度传感器等仿真  $\rightarrow$  CV Sim2Real gap

为什么需要: 可以提供非常好的 三维数据

是知道具体 data 为多少, 对学习环境来说更有信息量

比 2D 有 low 信噪比 sample efficiency

能发现大规模量 CV

控制选择, 涉及自身能力

现有问题: 灵巧手  $\rightarrow$  目前已有的大模型 b 有缺陷

(还没有 5 自由度 2 指头抓 和 干啥)

$\rightarrow$  高自由度意味着更难学习, 需要更大 data  $\approx$  目前

## 三. 大模型

答: 大模型本身: RT-2: VLA 大模型, 排列组合通用, "算法不用"

Input: language, vision, other proprioceptive

Output: action

既能决定下啥干啥, 又能输出 action

我们提出: 三层级的大模型

多模态大模型

Sim2Real 模型

大模型耦合

$\rightarrow$  强人自由交流



应用: BAAI+GALBOT: 具身大模型机器人 G1 无人值守

货架自动 3DGS 建模识别 → 解析商品规格  
根据下单自动拣货