人工智能技术
Artificial Intelligence

——人工智能: 逻辑智能+计算智能+学习智能
AI: Logical Intelligence+Computational Intelligence+Learning Intelligence

理论课：王鸿鹏、王润花、韩明静

实验课：许丽、靖智博

南开大学人工智能学院

概论

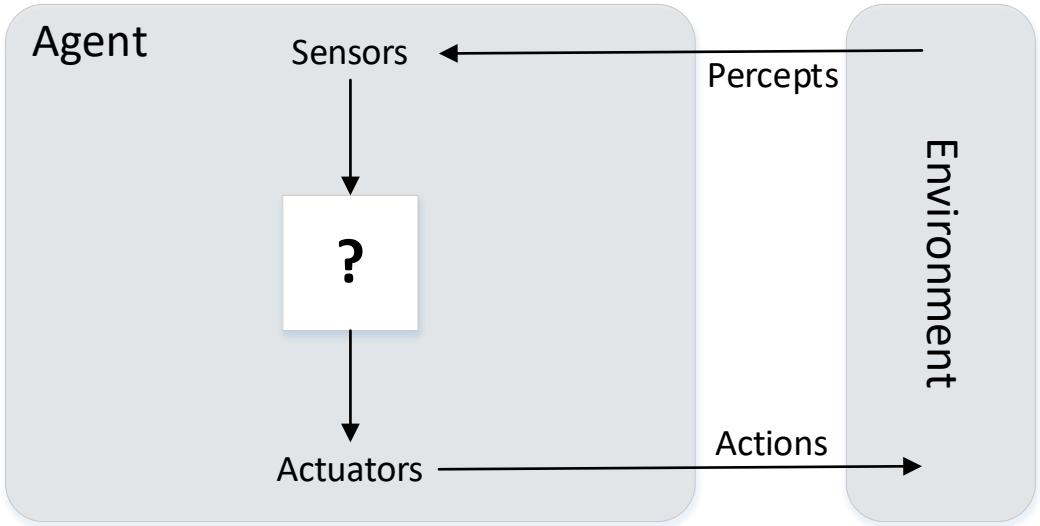
Introduction

——智能化智能体

Chapter 2-1: Intelligent Agent

Agents and Environments

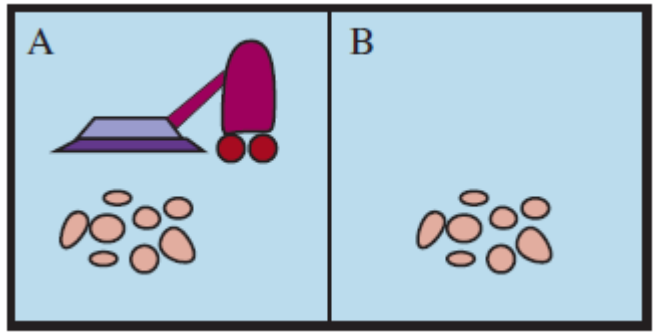
- Agent（智能体）：具有智能的实体，是驻留在某一环境下，能持续自主地发挥作用，具备驻留性、反应性、社会性、主动性等特征的计算实体。
- 智能体是人工智能领域中一个很重要的概念。任何独立的能够思想并可以同环境交互的实体都可以抽象为智能体。



图：Agent通过传感器和执行器与环境进行交互

Intelligent Agent 是这样一种智能体，给定它所感知到的和它拥有的先验知识，以一种被期望最大化其性能指标的方式运行（性能指标定义了智能体成功的标准）。此类智能体也成为理性智能体(Rational Agent)。智能体的合理性是通过其性能指标，其拥有的先验知识，它可以感知的环境及其可执行的操作来衡量的。这个概念是人工智能的核心。

Agents and Environments



一个只有两个位置的吸尘器世界。每个位置都可以是干净的或脏的，智能体可以向左或向右移动，并且可以清洁它所在的方格。不同版本的吸尘器世界允许不同的规则，例如智能体可以感知什么，它的行动是否总是成功，等等。

Percept sequence	Action
[A, Clean]	Right
[A, Dirty]	Suck
[B, Clean]	Left
[B, Dirty]	Suck
[A, Clean], [A, Clean]	Right
[A, Clean], [A, Dirty]	Suck
⋮	⋮
[A, Clean], [A, Clean], [A, Clean]	Right
[A, Clean], [A, Clean], [A, Dirty]	Suck
⋮	⋮

吸尘器世界中，一个简单智能体函数的部分列表。如果当前方格是脏的，智能体会清洁该方格，否则它会移动到另一个方格。请注意，除非对可能的感知序列的长度进行限制，否则该表的大小是无限的。

好的行为：理性的概念

- 自治性(Autonomy):智能体能根据外界环境的变化，而自动地对自己的行为和状态进行调整，而不是仅仅被动地接受外界的刺激，具有自我管理自我调节的能力。
- 反应性(Reactive):能对外界的刺激作出反应的能力。
- 主动性(Proactive):对于外界环境的改变，智能体能主动采取话动的能力。
- 社会性(Social):智能体具有与其它智能体或人进行合作的能力，不同的智能体可根据各自的意图与其它智能体进行交互，以达到解决问题的目的。
- 进化性:智能体能积累或学习经验和知识，并修改自己的行为以适应新环境。

Agent的特性

- 自治性(Autonomy):智能体能根据外界环境的变化，而自动地对自己的行为和状态进行调整，而不是仅仅被动地接受外界的刺激，具有自我管理自我调节的能力。
- 反应性(Reactive):能对外界的刺激作出反应的能力。
- 主动性(Proactive):对于外界环境的改变，智能体能主动采取话动的能力。
- 社会性(Social):智能体具有与其它智能体或人进行合作的能力，不同的智能体可根据各自的意图与其它智能体进行交互，以达到解决问题的目的。
- 进化性:智能体能积累或学习经验和知识，并修改自己的行为以适应新环境。

Agent概念的提出

- 现在IT界的智能体概念是由麻省理工学院的著名计算机学家和人工智能学科创始人之一的Minsky提出来的，他在“Society of Mind”一书中将社会与社会行为概念引入计算系统。
- 传统的计算系统是封闭的，要满足一致性的要求，然而社会机制是开放的，不能满足一致性条件，这 种机制下的部分个体在矛盾的情况下，需要通过某种协商机制达成一个可接受的解。Minsky将计算社会中的这种个体称为智能体。这些个体的有机组合则 构成计算社会——多智能体系统。
- Simon的有限性理论是多智能体系统形成的另一个重要的理论基础，Simon认为一个大 的结构把许多个体组织起来可以弥补个体工作能力的有限；每个个体负责一项专门的任务可以弥补个体学习新任务的能力的有限；社会机构间有组织的信息流动可以弥补个体知识的有限；精确的社会机构和明确的个体任务可以弥补个体处理信息和应用信息的能力的有限。

好的行为表现：理性的概念

- 理性智能体是做事正确的智能体：对的行动就是使得智能体更加成功的行动。
- 性能度量：

我们需要某种方法来度量一个智能体的成功与否。连同对智能体所处的环境、智能体的传感器和执行器的描述，这将为智能体面临的任务提供一个完整的规范说明。

性能度量是智能体成功程度标准的具体化。

- 理性：

对于每个可能的感知序列，根据已知的感知序列提供的证据和智能体内建的先验知识、理性智能体应该选择期望能使其性能度量最大化的行动。

环境的性质

- PEAS: Performance, Environment, Actuators & Sensors。

以无人自主驾驶汽车为例：

- 性能P：安全性、时间、合法驾驶、舒适性。
- 环境E：道路、其他汽车、行人、路标。
- 执行器A：转向、加速器、制动器、信号、喇叭。
- 传感器S：相机、声纳、GPS、速度计、里程计、加速度计、发动机传感器、键盘。

Agent Type	Performance Measure	Environment	Actuators	Sensors
Taxi driver	Safe, fast, legal, comfortable trip, maximize profits, minimize impact on other road users	Roads, other traffic, police, pedestrians, customers, weather	Steering, accelerator, brake, signal, horn, display, speech	Cameras, radar, speedometer, GPS, engine sensors, accelerometer, microphones, touchscreen

自动出租车司机任务环境的PEAS描述

环境的性质

- 完全可观察和部分可观察：如果是完全可观察的，智能体的传感器可以在每个时间点访问环境的完整状态，否则不能。例如，国际象棋是一个完全可观察的环境，而扑克则不是。
- 确定性和非确定性：环境的下一个状态完全由当前状态和由智能体接下来所执行的操作决定的。（如果环境是确定性的，而其他智能体的行为不确定，那么环境是随机性的）。随机环境在本质上是随机的，不能完全确定。例如，8数码难题（8-puzzle）这个在线拼图游戏有一个确定性的环境，但无人驾驶的汽车没有。
- 静态和动态：当智能体在进行协商（deliberate）时，静态环境没有任何变化。（环境是半动态的，环境本身并没有随着时间的流逝而变化，但智能体的性能得分则是会发生相应变化的）。另一方面，动态环境却改变了。西洋双陆棋具有静态环境，而扫地机器人roomba具有动态环境。
- 离散和连续：有限数量的明确定义的感知和行为，构成了一个离散的环境。例如，跳棋就是离散环境的一个范例，而自动驾驶汽车则需要在连续环境下运行。
- 单一智能体和多智能体：仅有自身操作的智能体本身就有一个单一智能体环境。但是如果还有其他智能体包含在内，那么它就是一个多智能体环境。自动驾驶汽车就具有多智能体环境。

环境的性质

- 完全可观察与部分可观察
- 单智能体与多智能体
- 确定性与非确定性
- 情景性与序列性
- 静态的与动态的
- 已知的与为止的

任务环境 Task Environment	可观察的 Observable	智能体 Agents	确定性 Deterministic	片段/延续式 Episodic	静态/动态 Static	离散/连续式 Discrete
填字游戏 Crossword puzzle 计时国际象棋 Chess with a clock	Fully Fully	Single Multi	Deterministic Deterministic	Sequential Sequential	Static Semi	Discrete Discrete
扑克 Poker 双陆棋 Backgammon	Partially Fully	Multi Multi	Stochastic Stochastic	Sequential Sequential	Static Static	Discrete Discrete
出租车 Taxi driving 医疗诊断 Medical diagnosis	Partially Partially	Multi Single	Stochastic Stochastic	Sequential Sequential	Dynamic Dynamic	Continuous Continuous
图像分析 Image analysis 分拣机器人 Part-picking robot	Fully Partially	Single Single	Deterministic Stochastic	Episodic Episodic	Semi Dynamic	Continuous Continuous
炼油厂控制器 Refinery controller 互动英语导师 Interactive English tutor	Partially Partially	Single Multi	Stochastic Stochastic	Sequential Sequential	Dynamic Dynamic	Continuous Discrete

任务环境及其特征的示例

智能体的结构

智能体的体系结构：智能体 = 体系结构 + 程序

```
function TABLE-DRIVEN-AGENT(percept) returns an action
  persistent: percepts, a sequence, initially empty
              table, a table of actions, indexed by percept sequences, initially fully specified

  append percept to the end of percepts
  action ← LOOKUP(percepts, table)
  return action
```

TABLE-DRIVEN-AGENT程序在每个新的感知出现时被调用，并每次返回一个动作。它将完整的感知序列保留在内存中

智能体程序

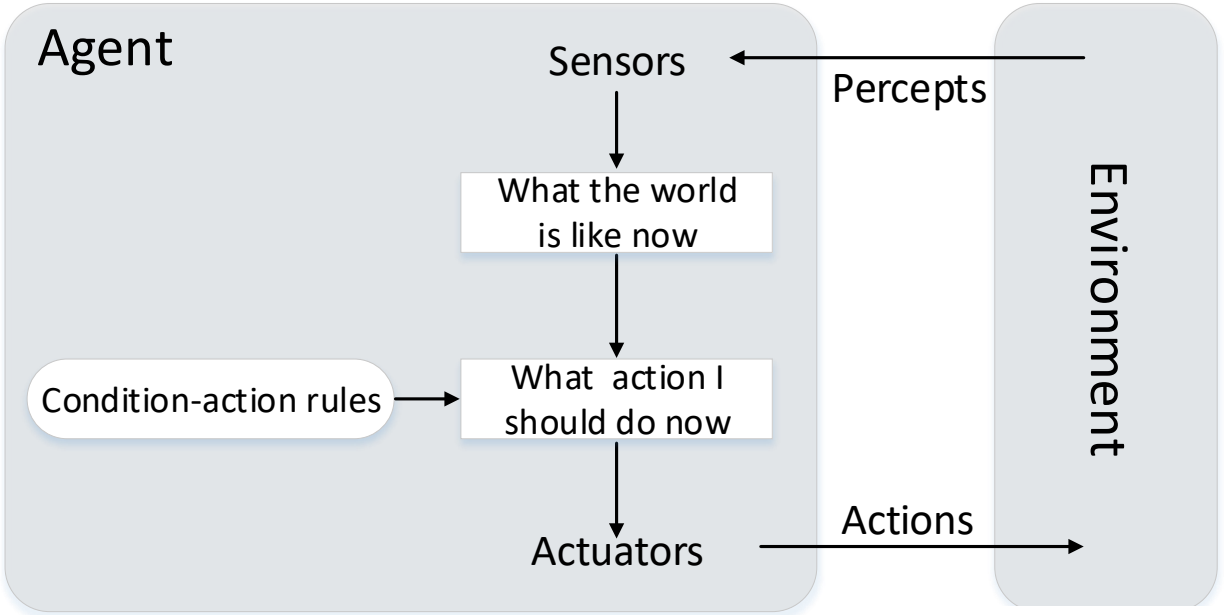
```
function REFLEX-VACUUM-AGENT([location,status]) returns an action

  if status = Dirty then return Suck
  else if location = A then return Right
  else if location = B then return Left
```

在两个位置的吸尘器环境中，一个简单反射智能体的智能体程序

智能体的类型-1

- 一般有4种类型的智能体，根据智能水平或其能够执行任务的复杂性不同而区分。
- 所有类型都可以随着时间的推移改进性能并产生更好的操作。这些可以概括为学习智能体（learning agents）。

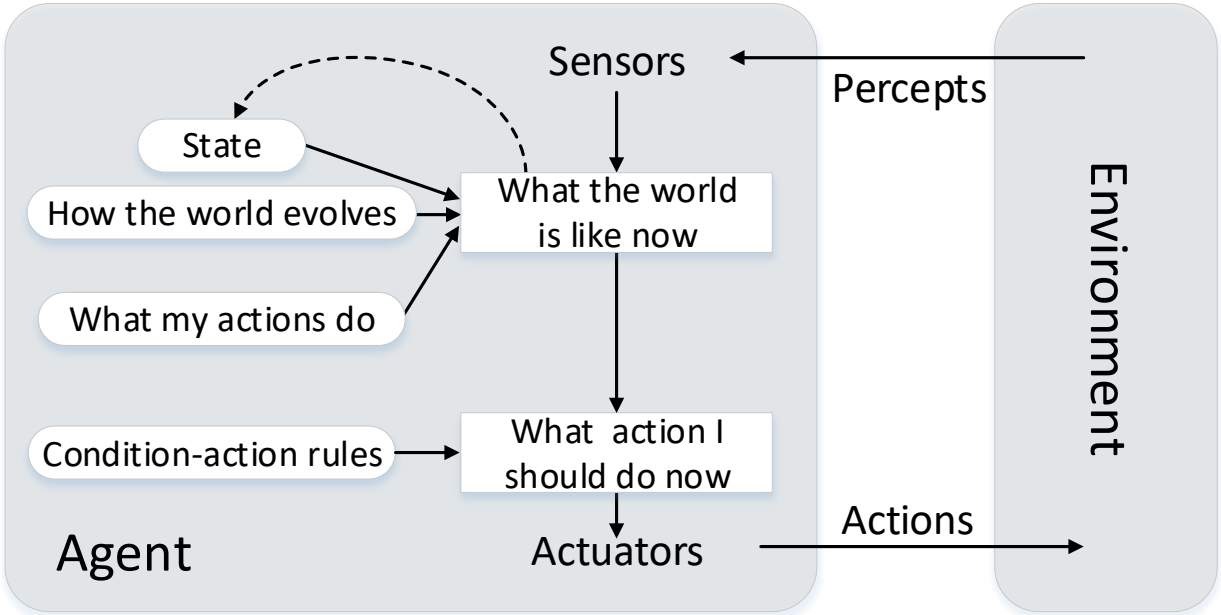
单反射性智能体（Simple reflex agents）



if *car-in-front-is-braking* **then** *initiate-braking*.

智能体的类型-2

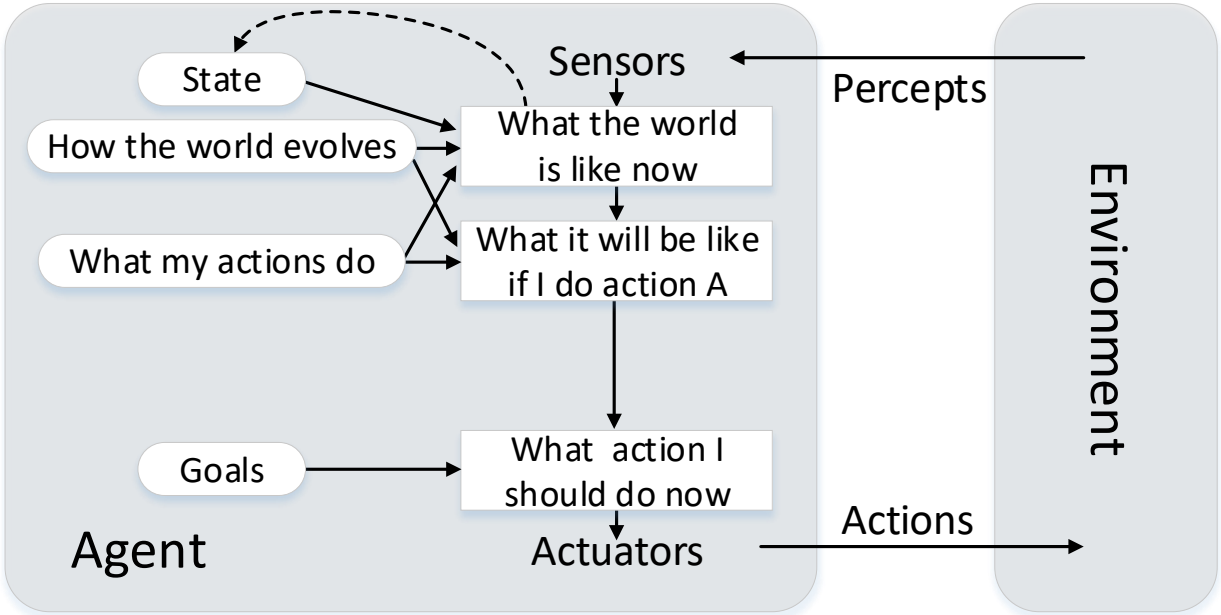
基于模型的反射性智能体(Model-based reflex agents)



智能体跟踪部分可观察的环境。这些内部状态取决于感知历史。环境/世界的建模是基于它如何从智能体中独立演化，以及智能体行为如何影响世界。

智能体的类型-3

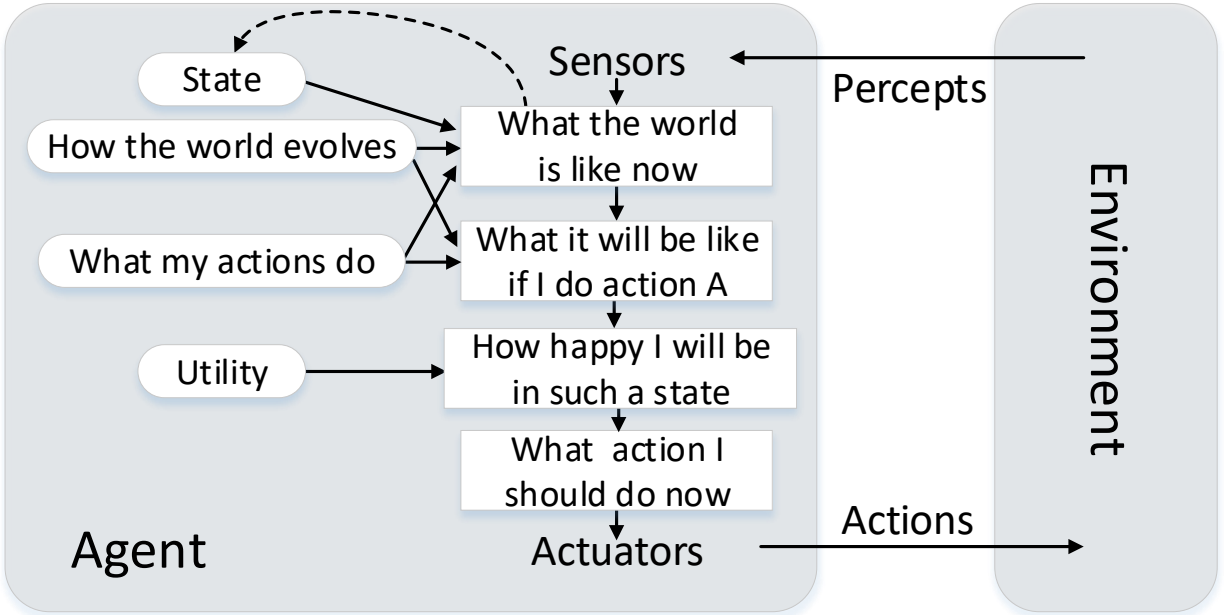
基于目标的智能体（Goal-based agents）



这是对基于模型的智能体的改进，并且在知道当前环境状态不足的情况下使用。智能体将提供的目标信息与环境模型相结合，选择实现该目标的行动。

智能体的类型-4

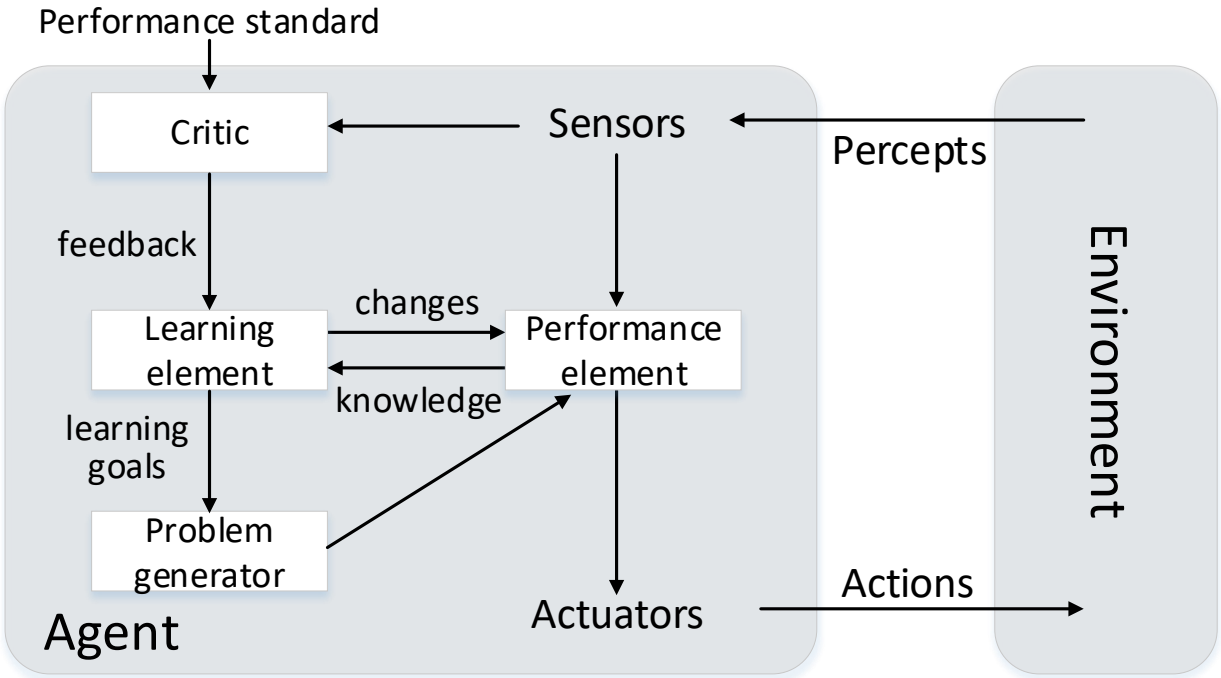
基于效用的智能体（Utility-based agents）



对基于目标的智能体进行改进，在实现预期目标方面有所帮助是不够的。我们可能需要考虑成本。例如，我们可能会寻找更快、更安全、更便宜的旅程到达目的地。这由一个效用函数标记。效用智能体将选择使期望效用最大化的操作。

智能体的类型-5

学习智能体（Learning Agents）



学习要素：负责改进

性能要素：负责选择外部行为，这是截至目前我们通常认为的智能体。

评论：关于确定的性能标准，智能体做得如何？

问题生成器：允许智能体探索。

多智能体(Multi-Agent System)

- 多智能体系统是多个智能体组成的集合，它的目标是将大而复杂的系统建设成小的、彼此互相通信和协调的，易于管理的系统。
- 多智能体系统是分布式人工智能（DAI，Distributed Artificial Intelligence）的一个重要分支，是20世纪末至21世纪初国际上人工智能的前沿学科。研究的目的在于解决大型、复杂的现实问题，而解决这类问题已超出了单个智能体的能力。
- 多智能体系统研究领域，主要包括：多智能体规划、学习、推理、协商、交互机制等等理论，及其实际应用。

多智能体(Multi-Agent System)

- MAS的研究涉及智能体的知识、目标、技能、规划以及如何使智能体采取协调行动解决问题等。研究者主要研究智能体之间的交互通信、协调合作、冲突消解等方面，强调多个智能体之间的紧密群体合作，而非个体能力的自治和发挥，主要说明如何分析、设计和集成多个智能体构成相互协作的系统。
- 同时，人们也意识到，人类智能的本质是一种社会性智能，人类绝大部分活动都涉及多个人构成的社会团体，大型复杂问题的求解需要多个专业人员或组织协调完成。要对社会性的智能进行研究，构成社会的基本构件物——人的对应物——智能体理所当然成为人工智能研究的基本对象，而社会的对应物——多智能体系统，也成为人工智能研究的基本对象，从而促进了对多智能体系统的行为理论、体系结构和通信语言的深入研究，这极大的繁荣了智能体技术的研究与开发。

多智能体(Multi-Agent System)的优势、特点

- 多智能体系统在表达实际系统时， 通过各智能体间的通讯、合作、互解、协调、调度、管理及控制来表达系统的结构、功能及行为特性。
- 多智能体系统具有自主性、分布性、协调性， 并具有自组织能力、学习能力和推理能力。采用多智能体系统解决实际应用问题， 具有很强的鲁棒性和可靠性， 并具有较高的问题求解效率。
- 在多智能体系统中，每个智能体具有独立性和自主性，能够解决给定的子问题，自主地推理和规划并选择适当的策略，并以特定的方式影响环境。
- 多智能体系统支持分布式应用，所以具有良好的模块性、易于扩展性和设计灵活简单，克服了建设一个庞大的系统所造成的管理和扩展的困难，能有效降低系统的总成本；
- 在多智能体系统的实现过程中，不追求单个庞大复杂的体系，而是按面向对象的方法构造多层次，多元化的智能体，其结果降低了系统的复杂性，也降低了各个智能体问题求解的复杂性；

多智能体(Multi-Agent System)的应用领域

- 在智能机器人中，信息集成和协调是一项关键性技术，它直接关系到机器人的性能和智能化程度。一个智能机器人应包括多种信息处理子系统，如二维或三维视觉处理、信息融合、规划决策以及自动驾驶等。各子系统是相互依赖、互为条件的，它们需要共享信息、相互协调，才能有效地完成总体任务，其目标是用来结合、协调、集成智能机器人系统的各种关键技术及功能子系统，使之成为一个整体以执行各种自主任务。利用多智能体系统，将每个机器人作为一个智能体，建立多智能体机器人协调系统，可实现多个机器人的相互协调与合作，完成复杂的并行作业任务。

概论

Introduction

——具身智能

Chapter 2-2: Embodied Artificial Intelligence

内容目录

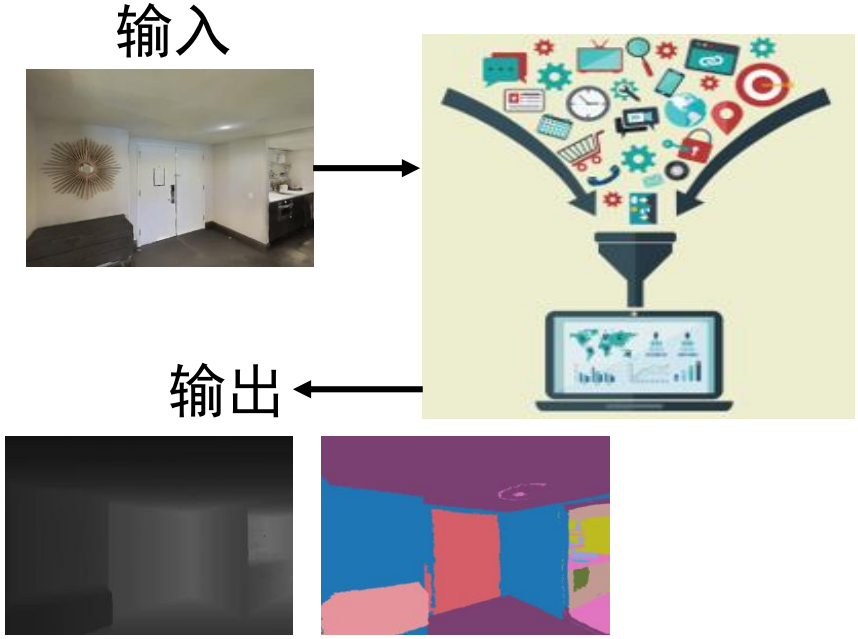
EAI：Embodied Artificial Intelligence 具身智能

- 具身智能（Embodied Artificial Intelligence，EAI）是指一种基于物理实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来获取信息、理解问题、作出决策并执行行动，从而展现出智能行为和适应性。
- 具身智能注重通过物理实体的感知、运动以及与外部环境的交互来实现认知，从而构成“感知 - 思考 - 行动”的闭环。
- 具身智能系统的实现形式并非局限于人形结构，而是可根据场景需求适配多样化智能实体。例如具备环境感知能力的智能扫地机器人、用于高空作业的无人机、已进入路测阶段的自动驾驶汽车等，都是具身智能的具体应用载体。

具身智能概述

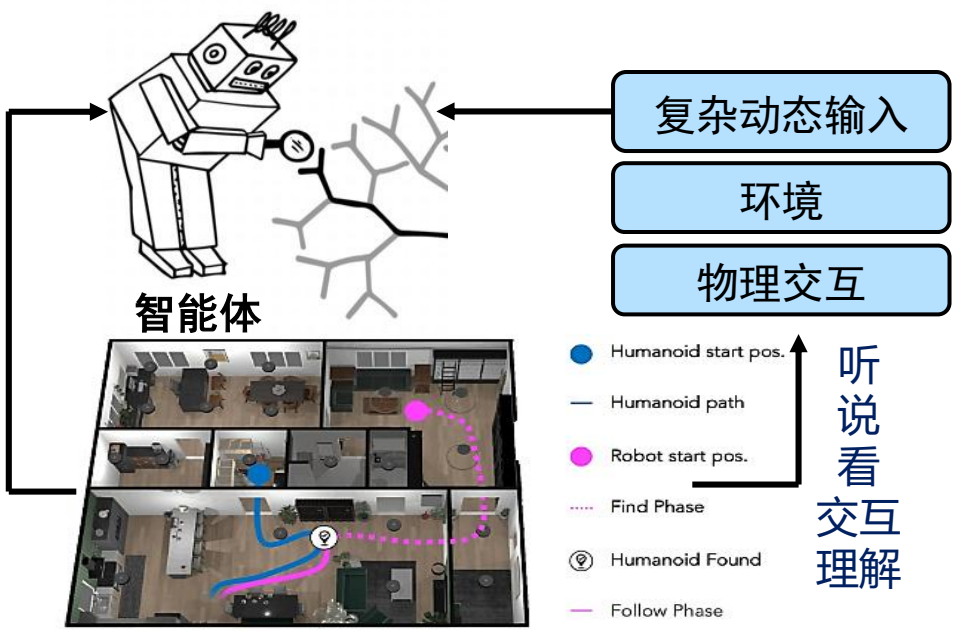
具身智能定义：以智能体作为本体支撑，不再局限于被动响应，而是能够像生物体一样，主动适应环境变化，应对噪声干扰，并适时调整自身行为。

离身智能



单一的符号智能往往与真实世界相脱节，认知与身体解耦

具身智能



智能是具身化和情景化的，具身智能可通过与真实世界的交互完成任务

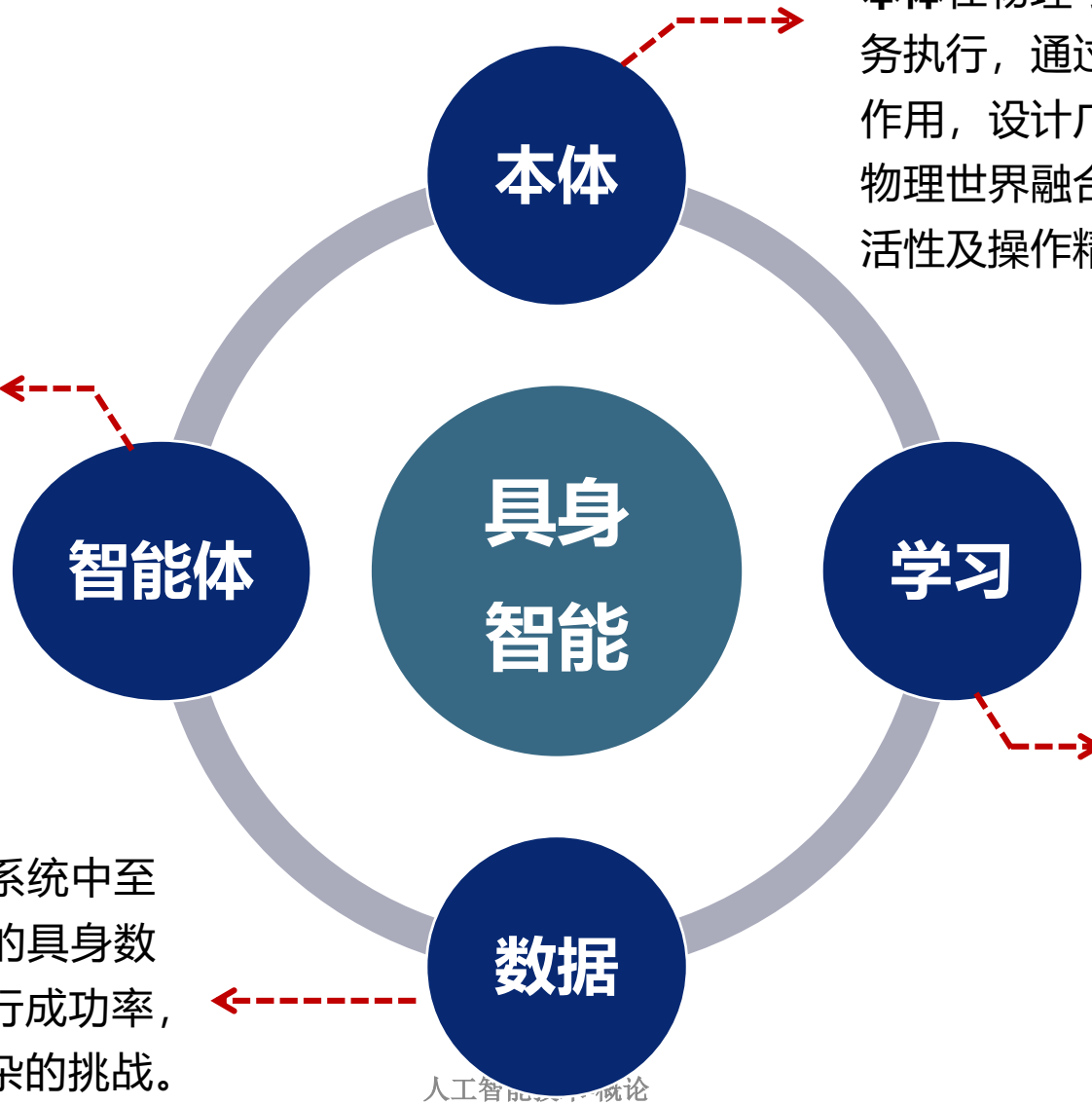
具身智能概述

- **具身（Embodiment）**：智能系统所依附的、能够支持丰富感官体验与灵活运动能力的物理实体，是智能体与环境互动的基础。
- **具身的（Embodied）**：具有身体的，可参与交互、感知的。
- **具身智能（Embodied AI）**：特指那些拥有物理形态，并能直接参与物理世界交互的智能系统，如服务型机器人、智能无人驾驶车辆等。它们通过“身体力行”的方式，展现出高度的环境适应性与任务执行能力。
- **具身任务**：像人类一样通过观察、移动、对话以及与世界互动从而完成的一系列任务。
- **多模态**：一个模型或系统能够处理多种不同类型的输入数据并融合它们生成输出，这些类型包括文本、图像、音频和视频等。这种能力对于提升智能系统的环境感知与决策能力至关重要。
- **主动交互**：机器人或智能体与环境的实时交互过程，从而提高智能体的学习、交流与处理问题的能力，是具身智能实现高效任务执行的关键。

具身智能概述

具身智能的核心要素

智能体作为本体的智能核心，具备敏锐的感知能力和动态决策机制，能够解析复杂环境并高效执行任务；借助深度学习和多模态模型的发展，智能体实现了从单一任务向多功能通用应用的跃升，具备自我进化和持续优化的能力。



本体在物理与虚拟空间中承担环境感知和任务执行，通过多种形式的机器人展现其关键作用，设计广泛适应性的本体是实现数字与物理世界融合的基础。其感知能力、运动灵活性及操作精度共同决定了本体的多维性能。

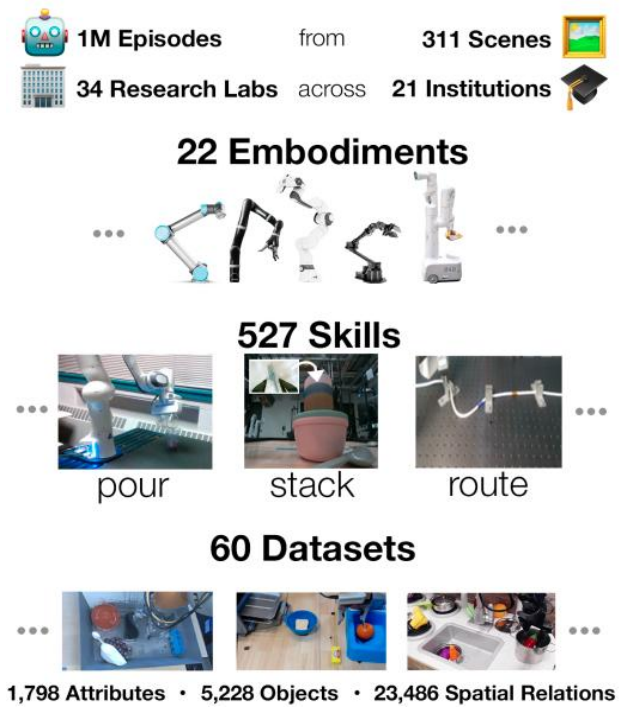
具身学习通过智能体与环境及人类的互动，构建“感知-决策-行动”闭环，利用人类-智能体交互数据强化多模态系统，推动智能体进化与性能提升。在交互过程中，系统提供多样化输出选项供用户反馈，以此优化未来性能，并通过人机协作纠正错误，增强系统的安全性和可靠性。

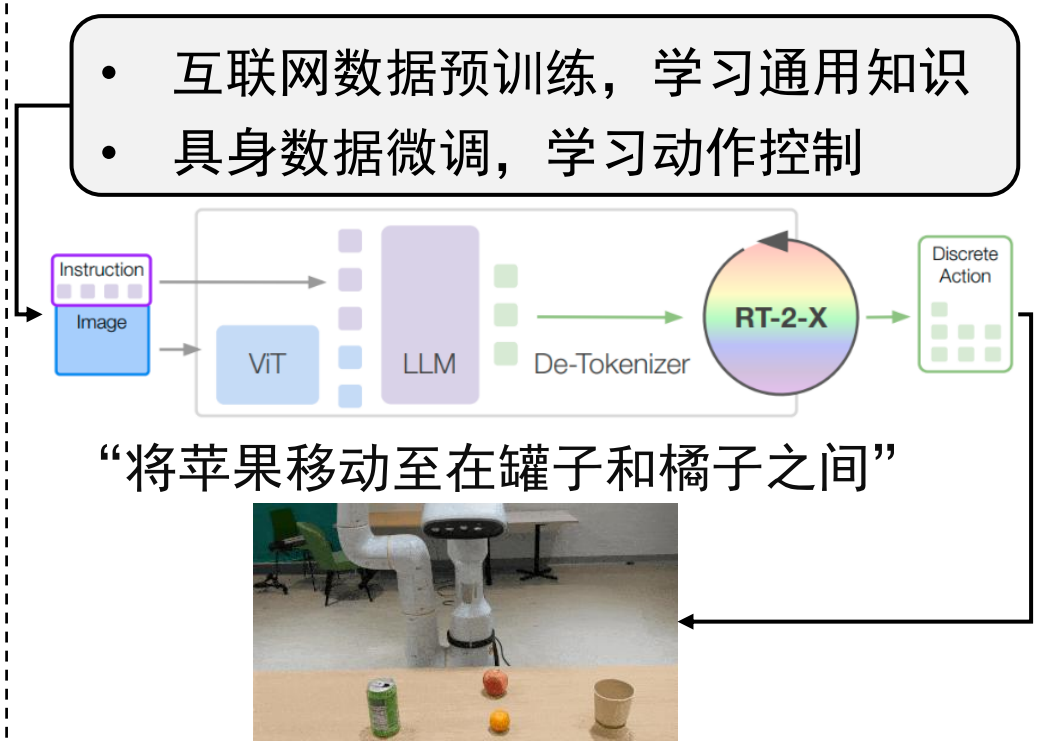
数据在机器学习与具身智能系统中至关重要，通过整合大量多样的具身数据，提升了智能体的任务执行成功率，但仍面临数据采集与结构复杂的挑战。

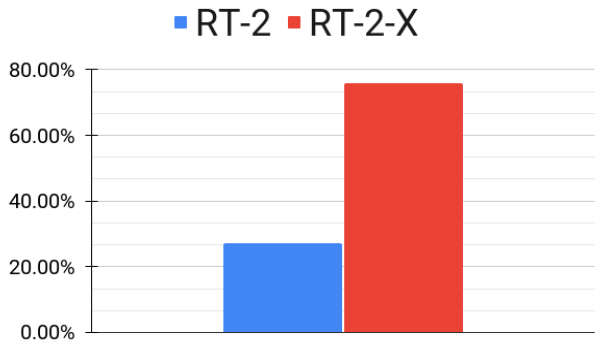
具身智能概述

■ 具身智能的核心要素

- 具身数据的获取方案：RT-X项目







基于Open X-Embodiment数据集训练，RT-2-X模型在众多具身任务上平均成功率提升近50%。

RT-X项目构建通用具身数据集，涵盖多种机器人类型、任务和场景，整合了来自34家研究实验室的60个数据集，数据集总量惊人地达到1,402,930条记录。

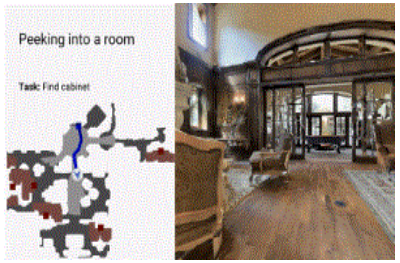
具身智能概述

■ 具身智能的核心要素

● 具身智能领域四种数据积累方法

虚拟式方法

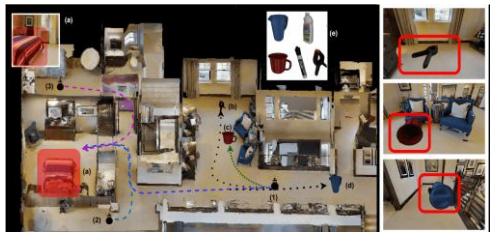
建立虚拟仿真环境，模拟现实环境在虚拟环境中训练智能体。

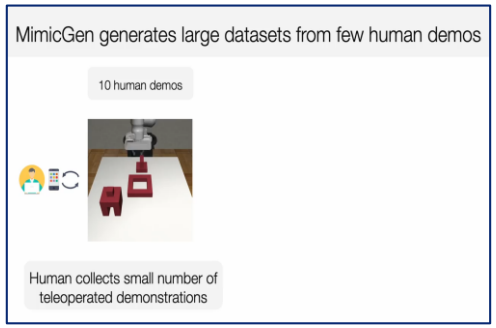


虚拟环境中收集人类演示数据。

生成式方法

训练生成模型，生成具身数据。

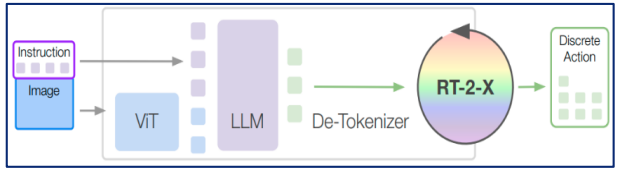




根据人类演示数据，生成更多训练数据。

网络式方法

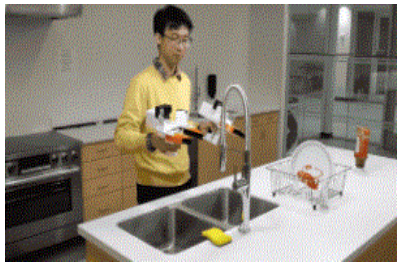
完成互联网数据的预训练，学习通用知识，并实现具身数据微调和动作控制。



机械臂抓取等实际场景。

表演式方法

少量的人类演示。



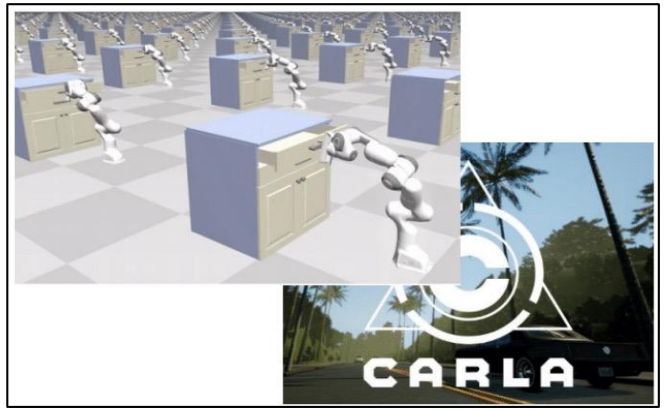
机器人从收集的数据中学习。

具身智能概述

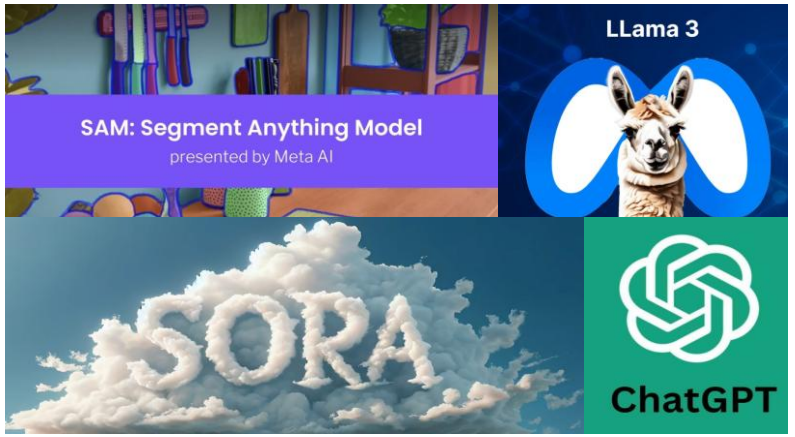
■ 具身智能的核心要素

● 具身智能系统中四种常见的策略泛化方法

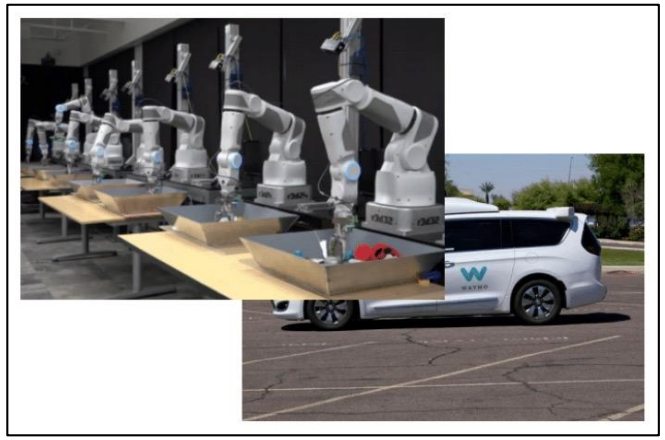
1. 多任务/多场景/多技能决策



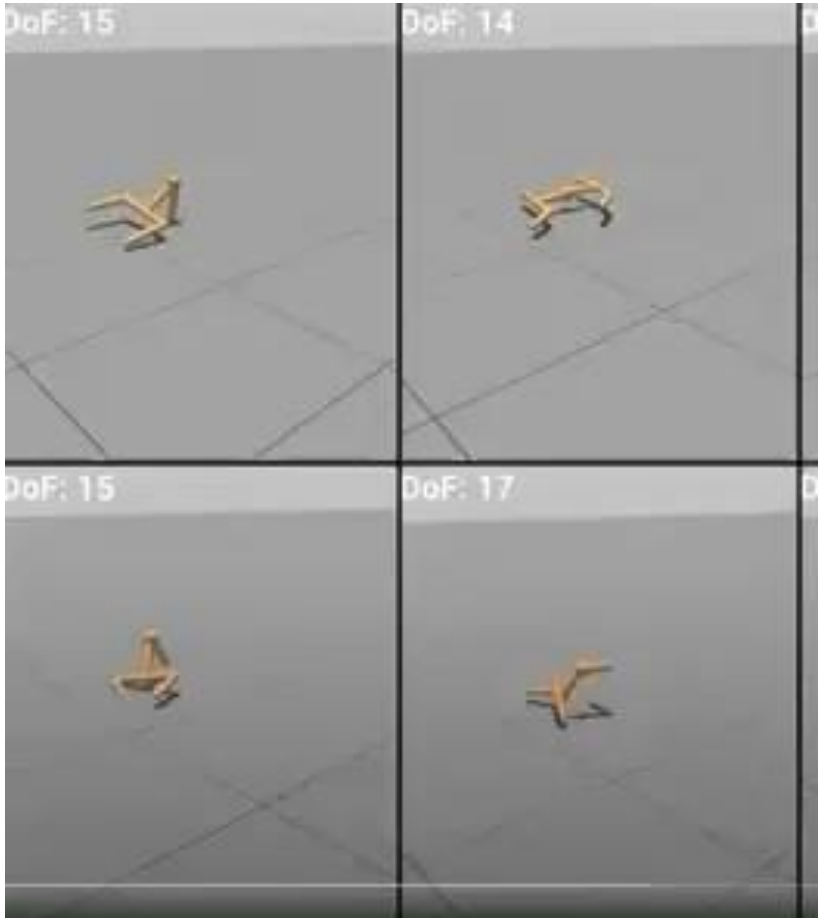
2. 仿真训练



3. 大模型技术



4. 真实训练



具身智能的进化学习示例

具身智能概述

■ 具身智能与人工智能

	传统人工智能	具身智能
概念定义	强调通过计算机技术模拟和实现人类智能，一般以软件形式存在。	强调智能系统与物理实体之间的交互。例如机器人系统。
实现路径	传统的算法和模型，例如机器学习、神经网络等。	不仅依赖于传统AI算法，还依赖于传感器、执行器和物理动力学的结合。
研究焦点	聚焦于抽象问题解决、符号知识表示与逻辑推理过程，以及在已知或可建模环境中提供决策支持，较少涉及实际物理环境中的动态交互。	强调智能体与物理环境之间的交互，关注感知与行动的结合、自适应学习，以及智能体如何基于自身物理特性在不同情境性作出反应。
应用领域	医疗数据分析、图像识别、语音识别以及自然语言处理等领域。	机器人、自动化制造、仓储物流等需要与物理环境交互的场景。

具身智能概述

■ 具身智能的意义与价值

● 核心理念

- 智能体与环境动态互动，超越静态数据处理方法。
- 强调嵌入物理环境，通过感知、理解和行动适应与改变环境。

● 研究进展

- 整合视觉、语言处理及决策制定。
- 在虚拟仿真环境中展示应对复杂挑战的能力。
- AI模拟器作为理论与实践的桥梁，如AI2-THOR，支持多任务广泛训练。

具身智能概述

■ 具身智能的意义与价值

● 应用前景

- 提升人机协同效率，实现情感交流与策略制定。
- 在环境保护、资源管理、教育公平、医疗普惠等领域发挥作用。
- 执行危险任务，减轻人类负担，精准调控资源利用。

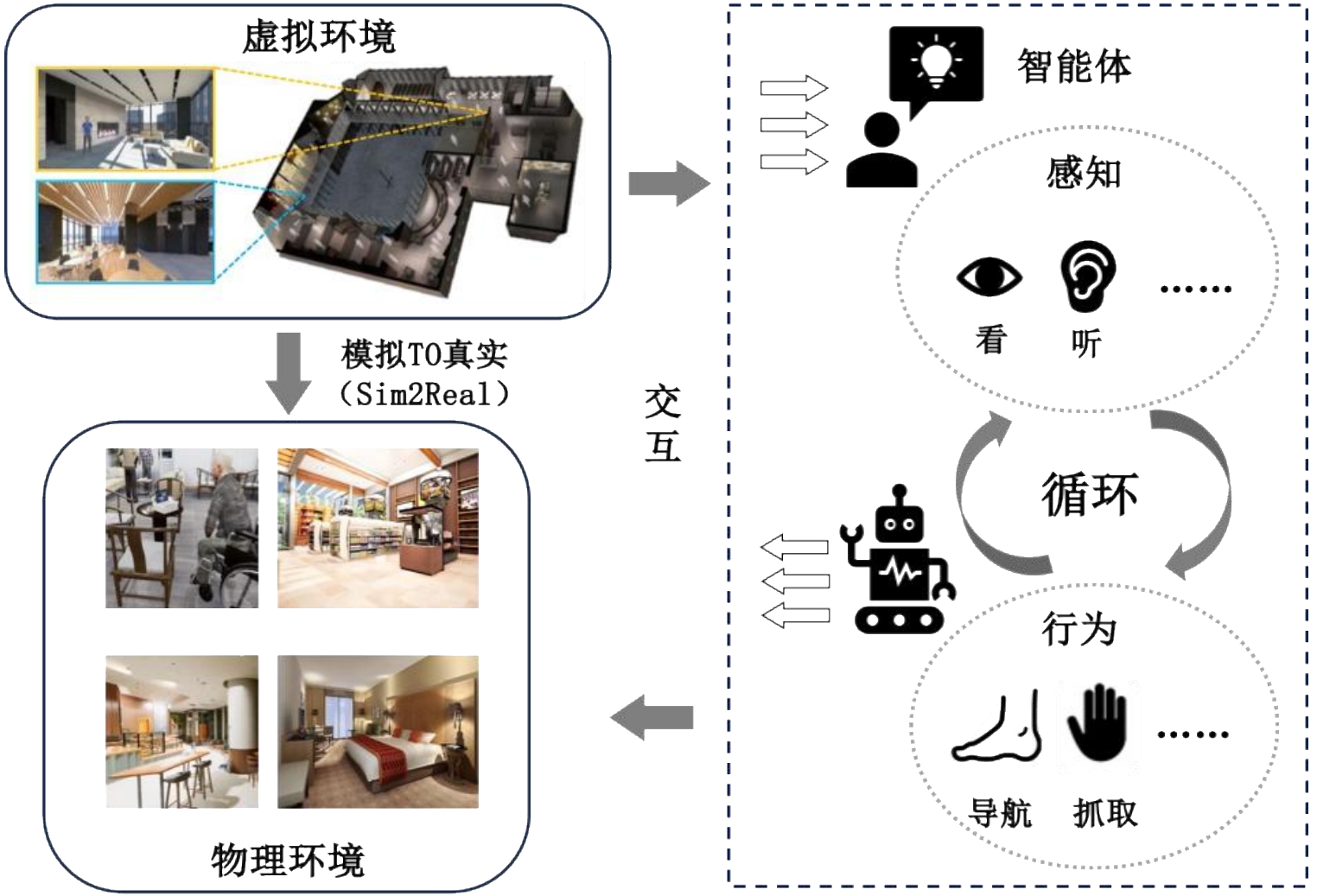
● 未来展望

- 代表人工智能的重大技术飞跃。
- 推动社会智能化、和谐化发展的关键驱动力。

内容目录

具身智能的核心技术

■ 具身智能的系统框架



核心技术：

- **具身感知：**深度融入物理世界的智慧触角
- **行为模块：**复杂任务达成的执行者
- **具身交互：**构建人机协作的新生态
- **强化学习与模仿学习**
- **仿真到真实的迁移**

具身智能的核心技术

■ 具身智能的核心技术：具身感知

● 主动视觉感知

智能体能够自主控制感知设备，如选择最佳视角和运用注意力机制。这种能力允许智能体主动探索环境，优化信息获取，从而提高任务执行效率。例如，通过调整摄像头的角度和焦距，智能体可以聚焦于最相关的视觉线索。

● 三维视觉定位与物体感知

智能体需具备在三维空间中定位自身及周围物体的能力，这对导航和物体操作至关重要。现代视觉编码器预训练技术增强了对物体类别、姿态和几何形状的精确估计，使智能体能在复杂动态环境中全面感知。这使得智能体能够准确理解其所在环境的三维布局，并据此作出决策。

● 多模态感知整合

除了视觉之外，触觉和听觉等感知模态同样重要，它们为智能体提供额外的环境信息。触觉传感帮助智能体感知物体的质地、重量和形状，支持精确的物体操作。整合多模态感知数据，能够显著提升智能体对环境的整体理解能力，使其在执行任务时更加灵活和高效。

具身智能的核心技术

■ 具身智能的核心技术：具身感知

● 具身感知模式的发展

感知大模型

- SAM:视觉分割大模型
- DINO:视觉分割大模型

静态环境识别精度与人类相当

被动感知

第三人称

具身主动感知

通过主动获取图像，相比现有大模型的目标检测性能有显著提升。

主动感知

第一人称

被门挡住了视线?

交互后，推开门看看是什么，能干什么

具身交互感知

第一人称

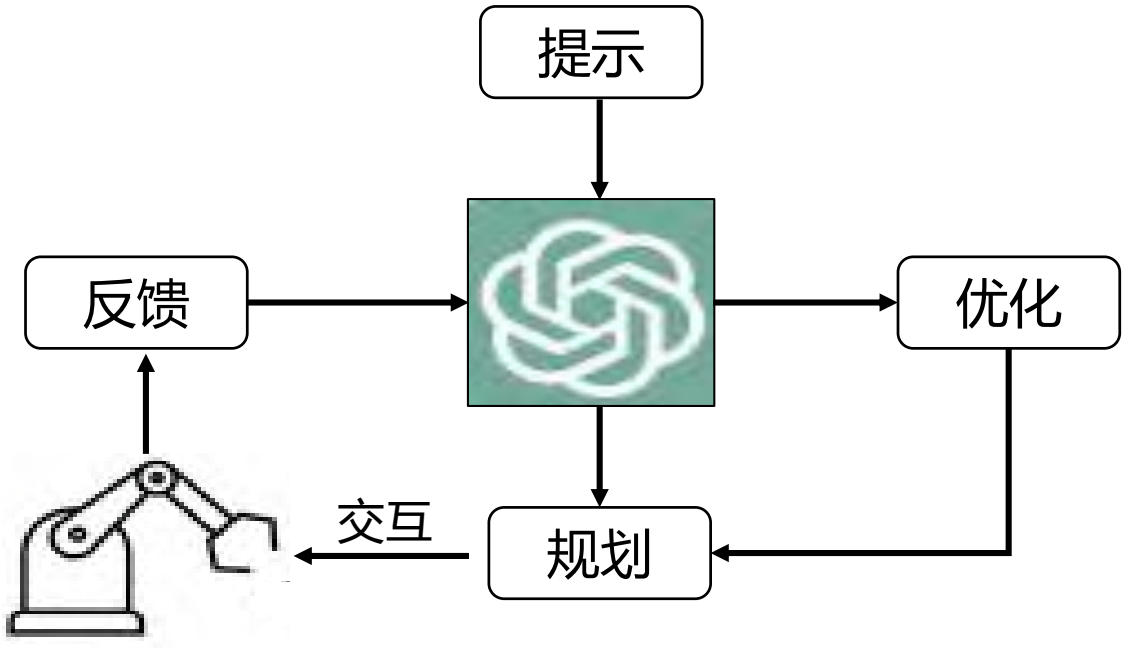
行为交互+感知

具身感知模式从被动到主动交互感知方向发展

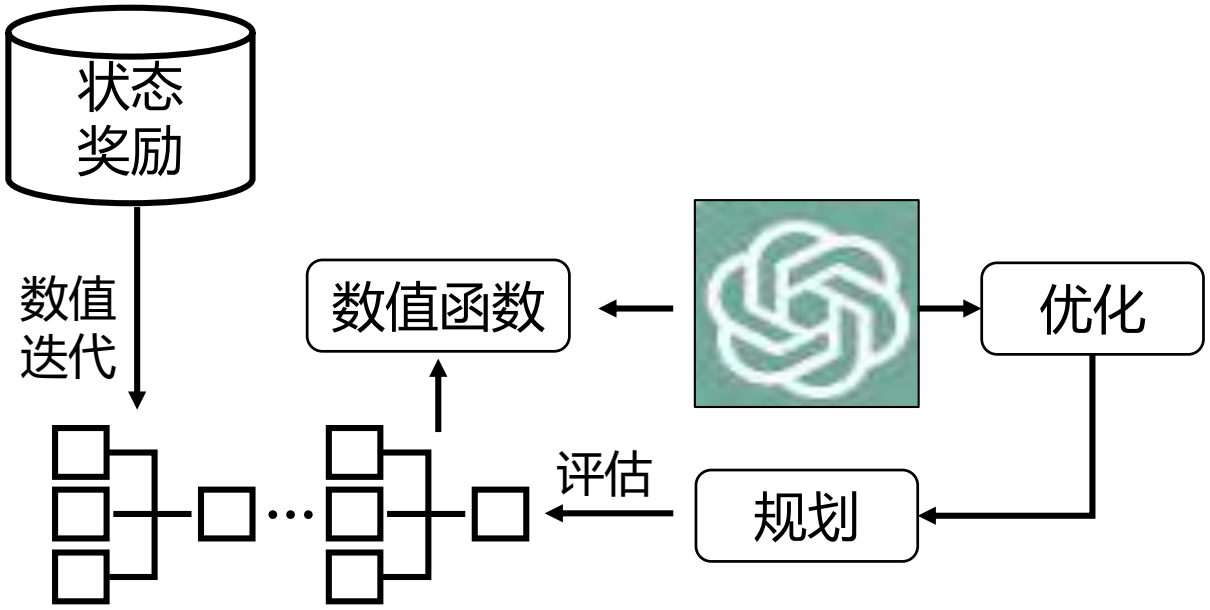
具身智能的核心技术

■ 具身智能的核心技术：行为模块

行为模块是连接感知与行动的纽带，它基于丰富的感知数据或人类指令，操纵智能体执行复杂的物体操作任务。这一过程融合了语义理解、场景感知、决策制定与稳健的控制规划。



基于物理反馈的规划



基于强化学习的规划

具身智能的核心技术

■ 具身智能的核心技术：具身交互

● 人类监督与反馈的重要性

人类在监督智能体行为轨迹的同时，确保其行动符合需求，并保障交互的安全、合法及道德边界。尤其在医学诊断等敏感领域，人类监督能有效弥补数据局限性与算法能力的不足。

● 从被动感知到主动交互的转变

智能体通过在线互动实现模型发展与进化，从第三人称的被动感知转向第一人称的主动交互感知。如智能体能够通过行为交互主动适应实际场景，如“被门挡住视线”的情况。

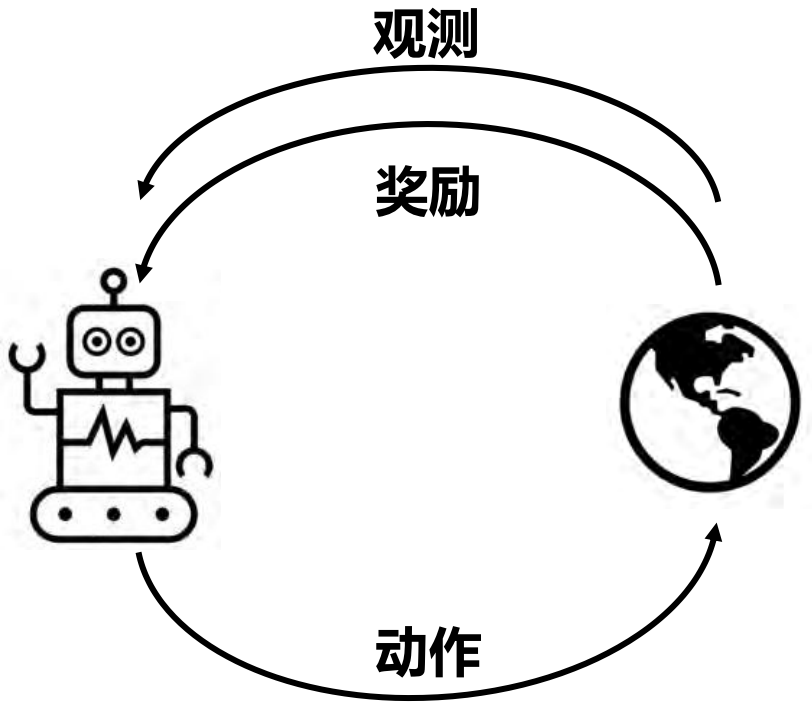
● 人类与智能体交互的两种范式

具身交互分为“不平等互动”模式，即“指导者-执行者”范式，人类发布指令，智能体辅助完成任务；以及“平等互动”模式，智能体与人类共同决策，预示更加协同的未来。

具身智能的核心技术

■ 具身智能的学习框架：强化学习

- 强化学习是一种通过智能体与环境交互来学习最优策略的方法。在具身智能中，智能体通过执行动作并接收环境反馈（奖励或惩罚）来优化行为，从而不断尝试新的动作组合以最大化累积奖励。



- 环境的下一时刻状态的概率分布将由当前状态s_t和智能体的动作a_t共同决定，可以表示为：

$$S_t \sim P(\cdot \mid S_t, a_t)$$

具身智能的核心技术

■ 具身智能的学习框架：强化学习

- 不同于有监督学习最小化预测误差思路，强化学习的最终优化目标是最大化智能体策略在动态环境交互过程中的价值。策略的价值可以等价转换为奖励函数在策略占用度量上的期望，即：

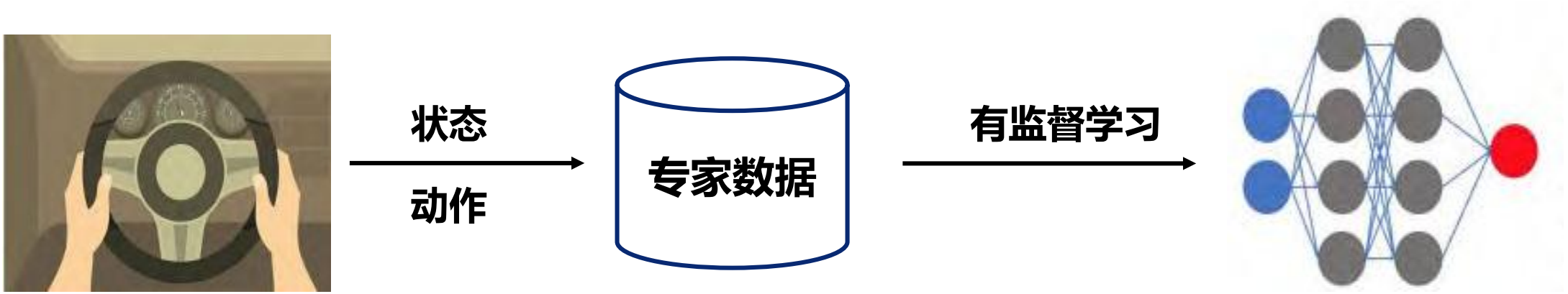
$$\text{最优策略} = argmax\mathbb{E}_{(\text{状态,动作})\sim\text{策略的占用度量}}[\text{奖励函数(状态,动作)}]$$

- 在具身智能的应用中，强化学习不仅能够帮助智能体学会执行基本任务（如行走、抓取等），还能够通过不断试错和自我优化，提高智能体在复杂环境中的适应性和鲁棒性。

具身智能的核心技术

■ 具身智能的学习框架：模仿学习

- 假设存在一个专家智能体，其策略可以看成一个理想的最优策略，那么具身智能体就可以通过模仿这个专家在环境中交互的状态动作数据来训练一个策略，并且不需要用到环境提供的奖励信号。这类方法我们称之为**模仿学习**。**与强化学习不同，它是一种通过观察专家演示来学习行为的方法**。



具身智能的核心技术

■ 具身智能的学习框架：模仿学习

- 在具身智能的上下文中，模仿学习通常涉及收集专家（如人类操作者）在执行特定任务时的行为数据（如动作序列、轨迹等），统称为**状态动作对**$\{(s_t, a_t)\}$，表示了专家在环境s_t下做出a_t的动作，而模仿者的任务则是利用这些数据在**无须奖励信号的条件下**训练一个智能体模型，使其能够复现专家的行为。

- **典型的模仿学习方法包括**：
  - **行为克隆（Behavior Cloning，BC）**
  - 逆强化学习（inverse RL）
  - **生成对抗模仿学习（Generative Adversarial Imitation Learning，GAIL）**

具身智能的核心技术

■ 具身智能的学习框架：行为克隆

- **行为克隆**采用直接的有监督学习框架，将专家数据对$\{(\boldsymbol{s_t}, \boldsymbol{a_t})\}$中的状态$\boldsymbol{s_t}$作为样本输入，将动作$\boldsymbol{a_t}$视为标签。因此，BC算法的学习目标可以表示为：

$$\theta^* = \arg\min_{\theta} E_{(s,a)\sim B}[L(\pi_\theta(s), a)]$$

其中，B属于专家数据集，$\mathcal{L}$为监督学习框架下的损失函数。如果动作a呈现出离散序列的形式，损失函数可以采用最大似然估计来优化；如果动作a是连续序列，则可以采用均方误差函数。

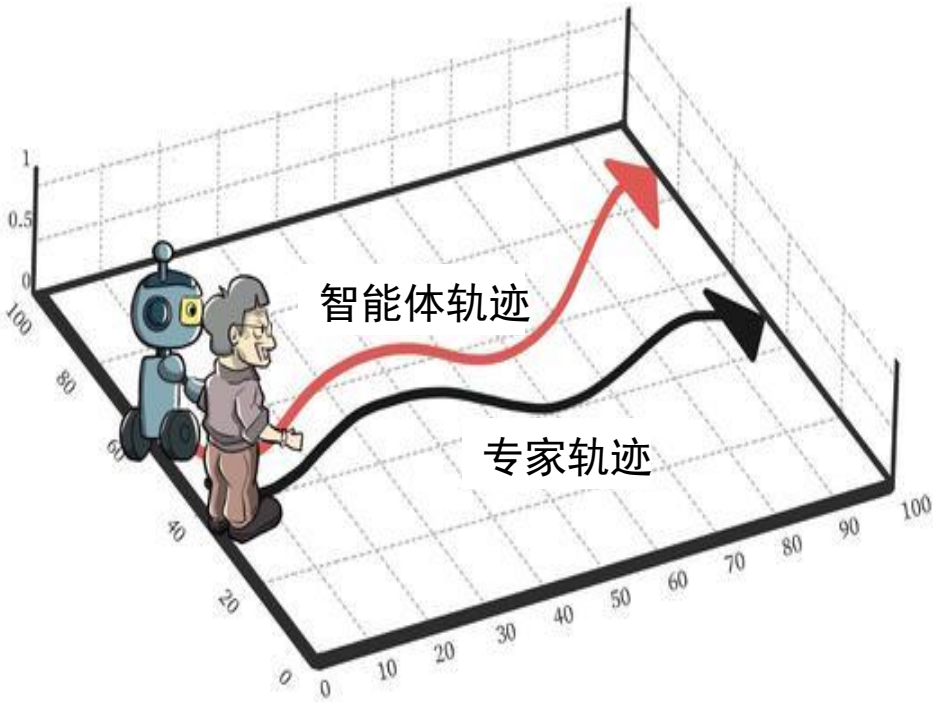
具身智能的核心技术

■ 具身智能的学习框架：行为克隆

● 行为克隆中的复合误差问题：

- 行为克隆算法仅仅基于一小部分专家数据进行训练，因此其策略仅能在这些专家数据的状态分布范围内做出准确预测。

- 然而，强化学习涉及的是序贯决策问题，这意味着通过行为克隆学习到的策略在与环境进行交互时无法完全达到最优。一旦策略出现偏差，所遇到的下一个状态可能从未在专家数据中出现过。 **分布偏移问题**



具身智能的核心技术

■ 具身智能的学习框架：生成对抗模仿学习

- **定义**：借鉴生成对抗网络思想，使学习得到的策略所产生的状态-动作对分布尽可能接近专家策略的分布，即智能体占用度量ρ_π尽量接近于专家的占用度量ρ_E 。
- **原理**：GAIL中的策略（类似于GAN中的生成器 ）需要与环境进行互动，通过执行动作并观察结果来逐步调整自身；而判别器D的作用则是评估状态-动作对(s,a)是否源自专家，输出一个介于0到1之间的值，用来估计状态-动作对(s,a)来自学习策略而非专家的概率。判别器的目标是最大程度地区分专家数据与学习策略生成数据。
- **对比**：行为克隆算法则无需此类环境交互即可直接从专家数据中学习策略。

具身智能的核心技术

■ 具身智能的学习框架：生成对抗模仿学习

- **判别器D**对应的目标函数定义为：

$$\mathbb{L}(\phi) = -E_{\rho_\pi}\left[\log D_\phi(s,a)\right] - E_{\rho_\pi}\left[\log\left(1 - D_\phi(s,a)\right)\right]$$

其中，判别器D的参数ϕ决定了其区分能力。

- **模仿者的优化目标**是生成能够欺骗判别器的轨迹，使其难以分辨这些轨迹是否出自专家。为此，判别器的输出作为奖励信号用于训练模仿者策略，即**当模仿者在状态s下执行动作a，对应的状态-动作对(s,a)被提交给判别器，其输出值作为奖励**。利用标准的强化学习算法，依据这些奖励优化模仿者策略，使其生成的数据分布逐渐逼近专家的真实数据分布，实现有效的模仿学习。

具身智能的核心技术

■ 具身智能的学习框架：总结

- **模仿学习的优点**：通过不断与环境交互，采样最新的状态-动作对，具备快速学习能力。由于直接利用专家的先验知识，这种方法可以显著减少智能体在试错过程中所需的时间和资源。此外，模仿学习还能避免强化学习中常见的奖励稀疏或奖励欺骗问题，提供更稳定的引导。

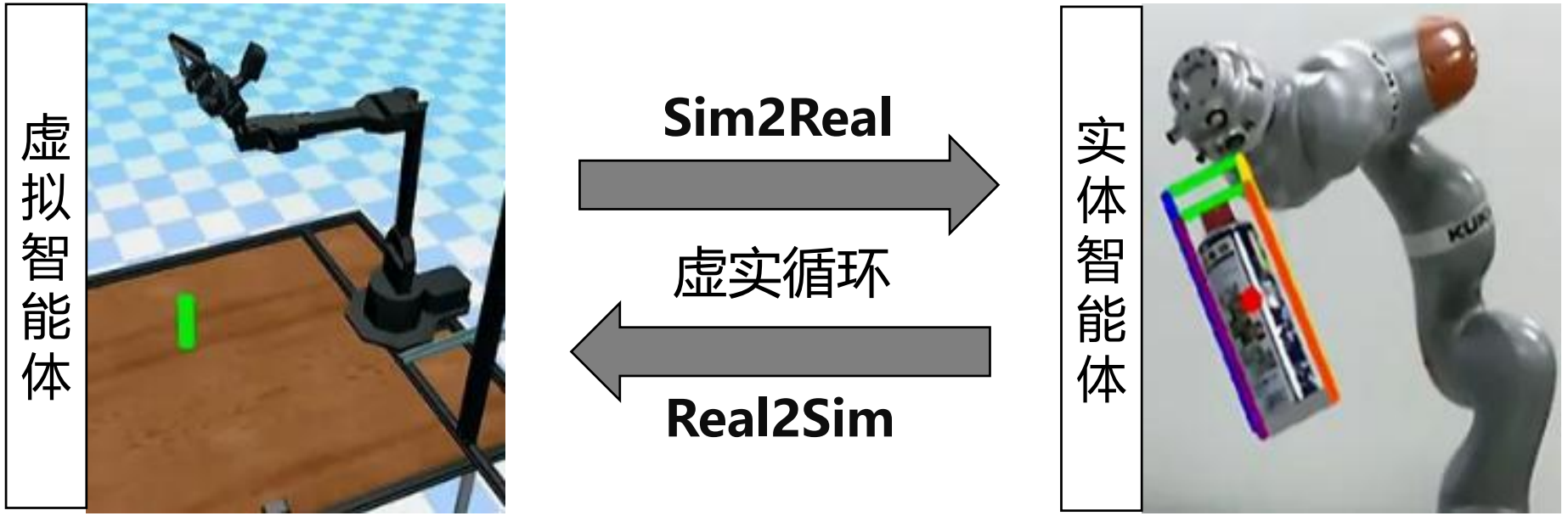
- **模仿学习的缺点**：仍面临专家演示数据局限性和偏见的问题，可能导致学习到的策略泛化能力不足。此外，模仿学习通常难以直接处理环境或任务变化带来的不确定性，需要与其他方法结合以提高适应性。

- **模仿学习与强化学习相融合**：首先通过模仿学习快速构建基本行为模型，再利用强化学习进行微调和优化，提高模型的泛化能力和鲁棒性。例如，在机器人抓取任务中，先通过模仿学习掌握基本动作，再通过强化学习优化以提高成功率。

具身智能的核心技术

■ 具身智能的核心技术：仿真到真实的迁移

- **仿真到真实的迁移（Simulation-to-Real，Sim2Real）**关注于将在仿真环境（如虚拟世界、模拟器等）中训练得到的模型、算法或策略成功地迁移到现实世界中的物理实体（如机器人、自动驾驶汽车等）上，并确保其在实际应用中表现出良好的性能和稳定性。



具身智能的核心技术

■ 具身智能的核心技术：仿真到真实的迁移

● Sim2Real的实现方法：

(1) 构建高精度、高逼真度的仿真环境

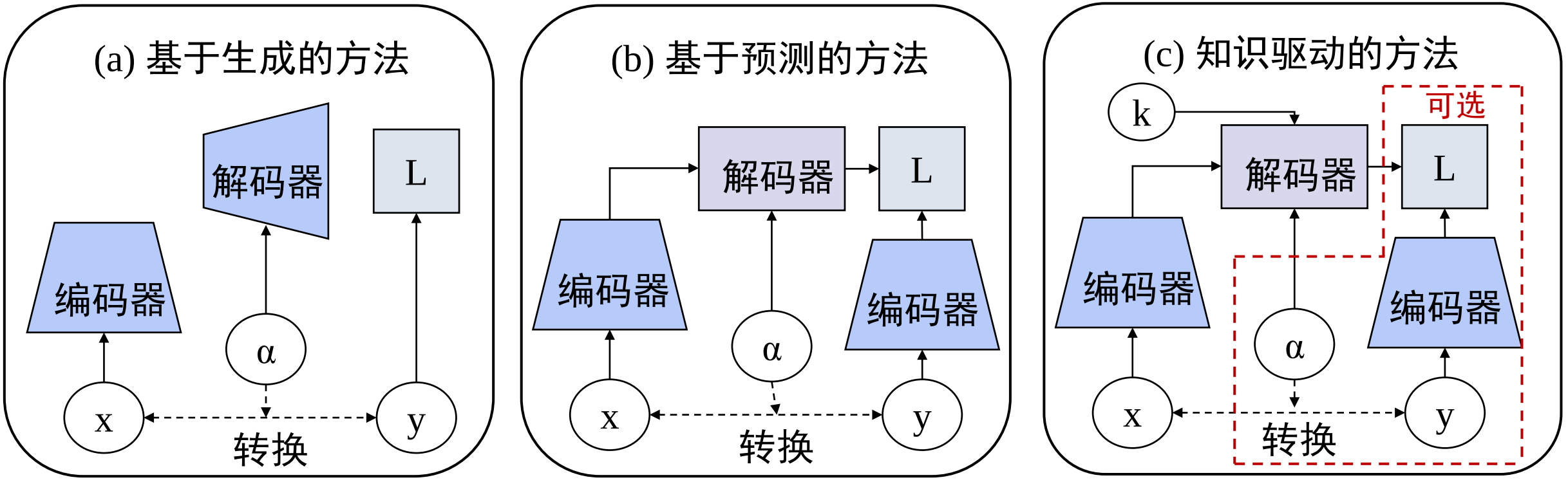
- **概述**：构建高精度仿真环境是实现Sim2Real的关键步骤，通过理解现实世界、选择合适工具、精细建模与校准等方法，可以创建接近现实的仿真环境。引入不确定性因素并进行验证与优化，有助于提高模型在现实世界中的性能和稳定性。
- **方法：世界模型**通过模拟环境状态变化和预测策略效果，为Sim2Real提供准确可靠的仿真环境。它处理感知信息和数据建模，实现对物体、场景、动作等要素的准确抽象和模拟。

具身智能的核心技术

■ 具身智能的核心技术：仿真到真实的迁移

- **Sim2Real的实现方法：**

(1) 构建高精度、高逼真度的仿真环境



具身世界模型的设计方法

具身智能的核心技术

■ 具身智能的核心技术：仿真到真实的迁移

● Sim2Real的实现方法：

(2) 数据驱动的方法

- **生成多样化模拟数据**：数据驱动方法首先强调从仿真环境中生成大量、多样化的数据，覆盖各种物理条件、环境变化和任务场景。这些数据为模型提供丰富的训练素材，确保其在模拟环境中表现出色并具备泛化能力。
- **收集与整合现实数据**：现实数据的收集与整合是数据驱动方法的关键，通过部署传感器和记录设备获取高质量现实数据，校准模拟环境并验证模型表现。这些数据帮助发现潜在偏差，提升模型在现实世界中的适用性。

具身智能的核心技术

■ 具身智能的核心技术：仿真到真实的迁移

● Sim2Real的实现方法：

(2) 数据驱动的方法

- **结合模拟与现实数据优化模型**：在模型训练阶段，结合模拟数据与现实数据来优化模型参数，通过预训练和微调策略，使模型在模拟环境中学习基本技能，并通过现实数据反馈进行优化。这种方式增强了模型适应现实环境的能力。
- **持续学习与迭代优化**：数据驱动方法强调模型的持续学习与优化，通过不断收集新的现实数据并将其用于模型再训练与更新，确保模型紧跟环境变迁与任务需求变化。这一过程提高了模型在现实世界中的稳定性和可靠性。

具身智能的核心技术

■ 具身智能的核心技术：仿真到真实的迁移

- **Sim2Real的实现方法：**

(2) 数据驱动的方法

机器人演示数据

打开机柜

将胡萝卜放在盘中

人体演示数据

切割甜椒

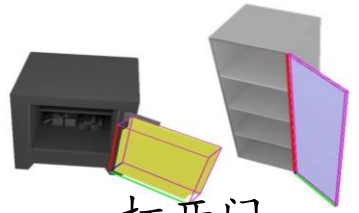
清洗玻璃

专家演示数据

将蓝色块放到绿色碗中

用微波炉加热汤

标注数据



打开门

只用两根手指握住刀柄

视频

点云

RGB+深度图

文本

声音

触觉

模拟

数据格式

具身智能的核心技术

■ 具身智能的核心技术：仿真到真实的迁移

● Sim2Real的实现方法：

(3)域适应与域随机化

- **域适应**：域适应使模型能够在不同但相关的数据分布上保持高性能，即从仿真环境迁移到现实环境中。通过识别仿真与现实环境的主要差异，并利用特征对齐等技术来缩小这些差异，帮助模型忽略特定噪声，关注有用特征。
- **特征对齐与生成对抗网络**：特征对齐通过学习共享特征空间使仿真与现实数据表示接近，而生成对抗网络生成具有现实特性的仿真数据，作为补充训练集帮助模型适应现实环境。无监督或自监督方法利用未标记的现实数据微调模型，设计自监督算法捕捉现实世界的本质特征。

具身智能的核心技术

■ 具身智能的核心技术：仿真到真实的迁移

- **Sim2Real的实现方法：**

(3)域适应与域随机化。

- **域随机化：**域随机化通过增加仿真环境的复杂性和多样性来提高模型训练的鲁棒性，通常在仿真训练阶段随机化环境参数。这种方法不需要现实数据，依赖模拟器和广泛的随机化策略，鼓励模型学习不依赖特定环境参数的特征，从而提高泛化能力，并可作为域适应的预处理步骤以加速适应过程。

内容目录

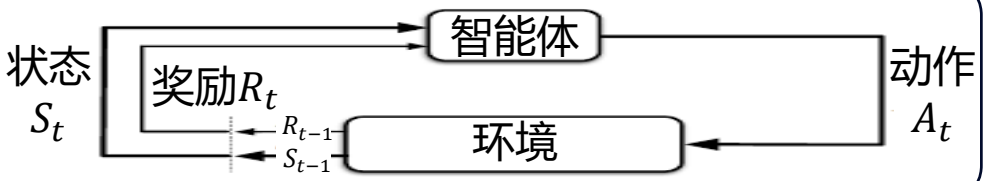
具身智能的典型案例

■ 具身智能任务的树状结构

- 从人工智能的发展范式出发，具身系统的研究焦点在于**如何更有效地适应未知环境，特别是在机器人规划与导航等复杂任务中**。

具身智能

通过与环境交互得到的反馈进行学习：反馈-动作-反馈-动作……



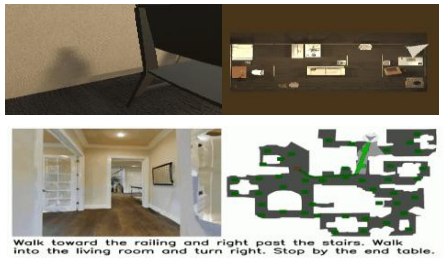
智能体

具身于本体之上的智能核心，负责感知、理解、决策、控制等核心工作。

智能机器人操作

机械臂通过与环境进行多模态交互，将物体移动到指定地点。

服务机器人导航



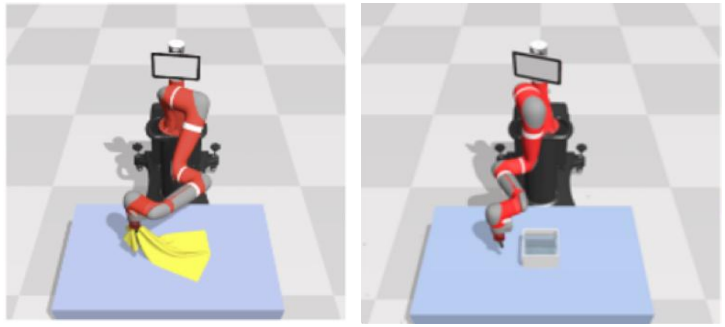
机器人通过与环境进行多模态交互，移动到指定地点。

具身智能的典型案例

■ 智能机器人操作任务

- 智能机器人操作是一个综合性的领域，它集成了视觉、语言等多模态输入处理能力，旨在输出精准的机器人动作以执行多样化的具身智能任务，如**物体抓取**任务。

(a) 刚性物体操作-ManiSkill



(b) 柔性物体操作-SoftGym

提示： 给定 <img> 我需要钉钉子，场景中有什么物件可能有用？
预测： 石头。
行动：1 129 138 122 132 135 106 127

(c) 视觉语言操作

具身智能的典型案例

■ 智能机器人操作任务

● 视觉-语言-动作（VLAs）的基本概念与操作策略

- **定义：**VLAs是一种结合了视觉、语言与动作执行的更高级别任务处理框架，旨在结合视觉与语言信息，指导机器人或智能系统完成复杂任务（如清理桌面、拿取物品）。其核心在于其强大的多模态处理能力。

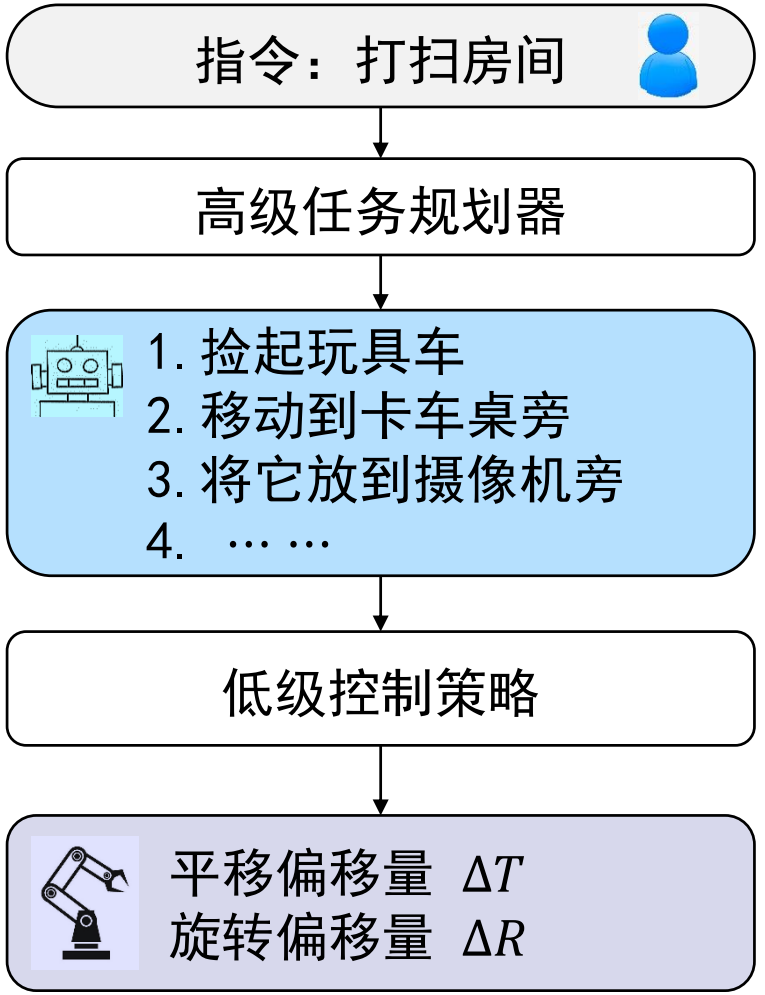
- **VLA模型组成：**视觉模块负责解析图像数据，语言模块则理解自然语言指令，动作模块据此生成动作指令并控制机器人执行相应的动作。三者之间通过深度协作与交互，使得模型不仅能理解复杂的场景与指令，还能灵活地执行任务，促进机器人综合能力的全面优化与提升。

具身智能的典型案例

■ 智能机器人操作任务

● 视觉-语言-动作（VLAs）的基本概念与操作策略

分层机器人操作策略

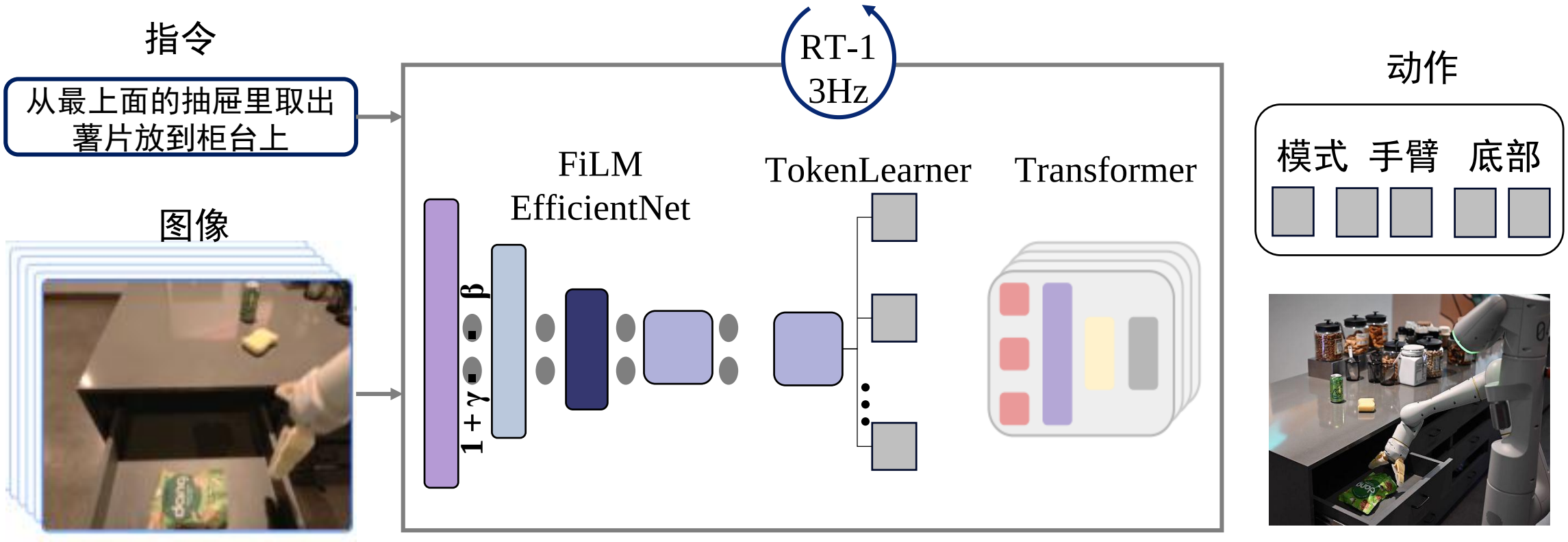


具身智能的典型案例

■ 智能机器人操作任务

● 视觉-语言-动作（VLAs）的具体实现——Robotics Transformer系列模型

- 2022年12月，谷歌推出了名为Robotics Transformer 1（RT-1）的具身智能模型，这是一种多任务处理模型，能够将机器人的输入和输出动作转换为Token形式，从而提升实时控制。

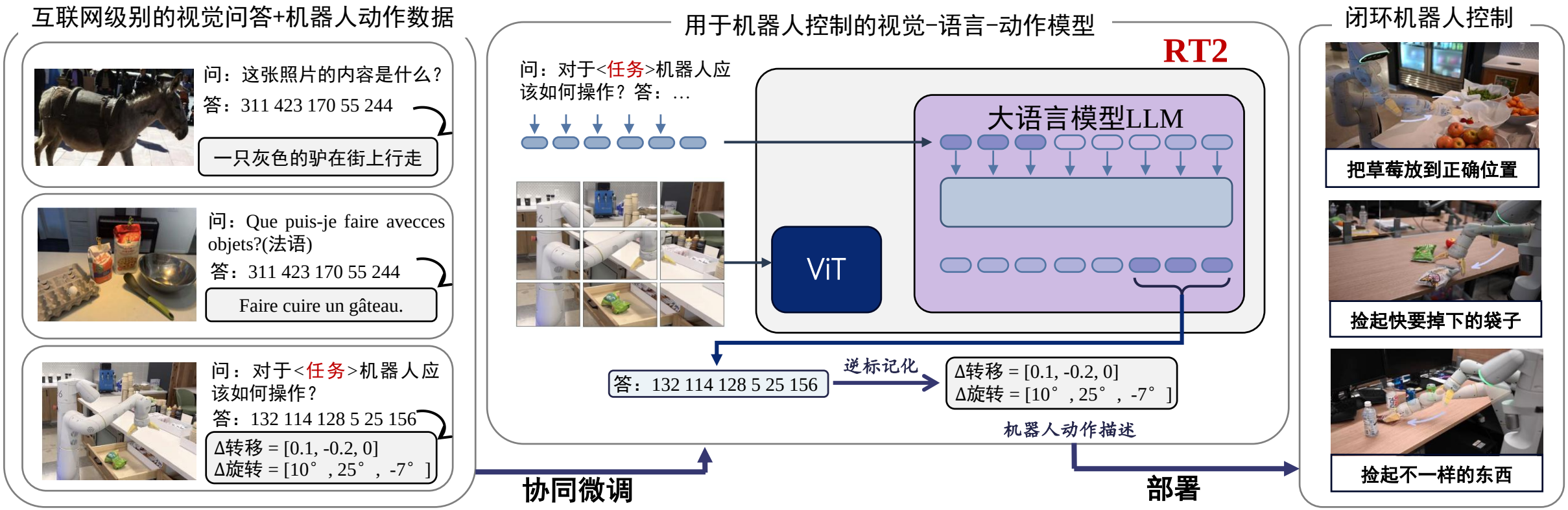


具身智能的典型案例

■ 智能机器人操作任务

● 视觉-语言-动作（VLAs）的具体实现——Robotics Transformer系列模型

- RT-2在模型设计上进行了重大创新，它将机器人的动作编码成一种独特的文本标记语言，这种创新性的表示方式使得RT-2能够利用互联网级别的庞大视觉-语言数据集进行训练。

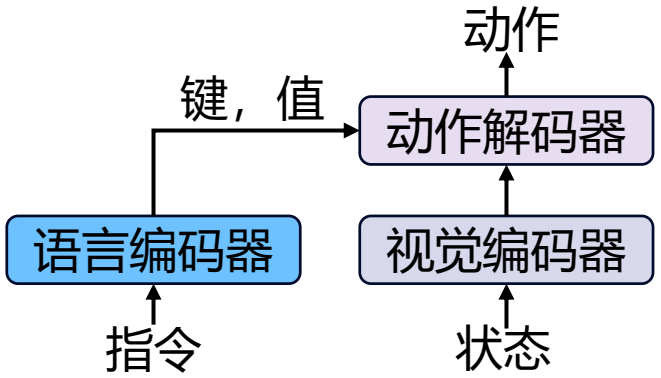


具身智能的典型案例

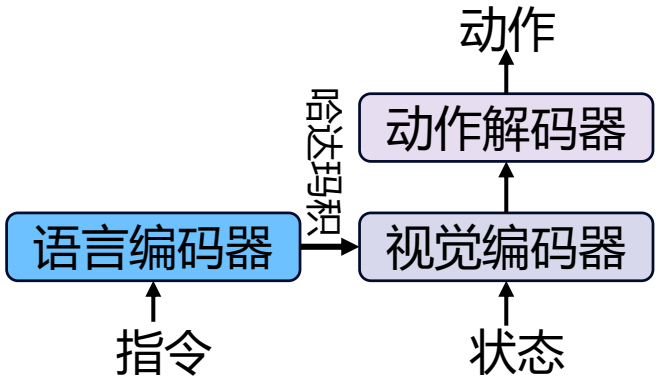
■ 智能机器人操作任务

● 其他VLA的典型技术

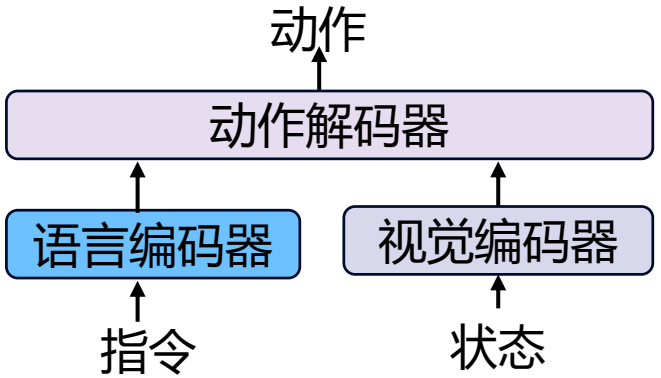
- **预训练视觉编码器**：通过大规模数据集的训练，获得能够捕捉复杂视觉特征并生成高质量视觉表示的模型。
- **环境动力学建模**：包括利用前向动力学方程来预测物体在给定力作用下的运动轨迹，以及利用逆向动力学方程来推断产生特定运动所需的力或力矩。**世界模型**
- **视觉-语言融合机制**：



(a) 交叉注意力



(b) FiLM



(c) 拼接

具身智能的典型案例

■ 服务机器人导航任务

- **服务机器人导航**要求机器人在未知且复杂的环境中，仅凭目标位置和多个视角的观测（主要是视觉信息），通过集成的感知硬件与先进算法进行深度分析，并在与环境的持续交互与反馈中，高效且准确地在限定步数内抵达指定位置。

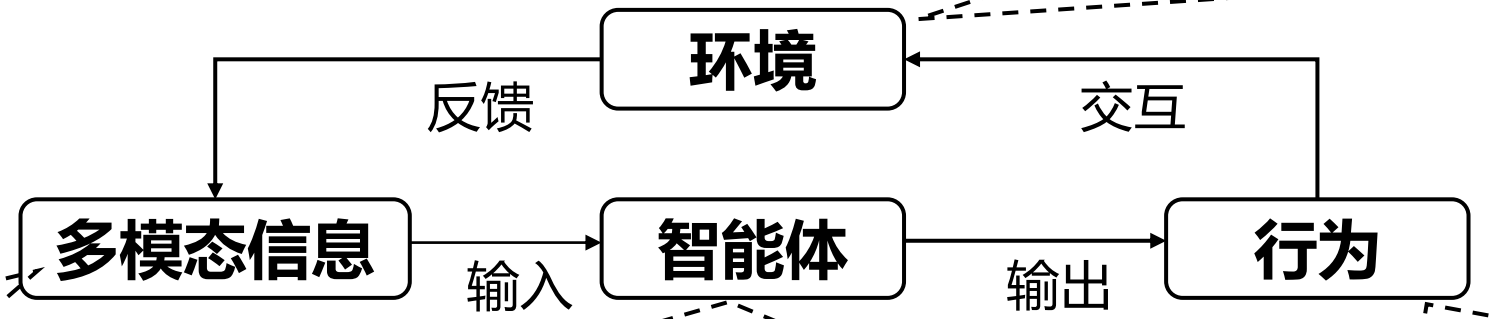
视觉目标导航任务示例

RGB视图　　第三视角俯视图

具身智能的典型案例

■ 服务机器人导航任务

服务机器人导航的整体流程



- **感知硬件**



RGB-D摄像头　激光雷达　臂端摄像头

- **感知算法**

目标检测、视觉分割、视觉预训练感知

- **硬件平台**

不同类型的服务机器人

- **导航规划算法**

语义地图、边界跟踪、强化导航算法

- **执行硬件**

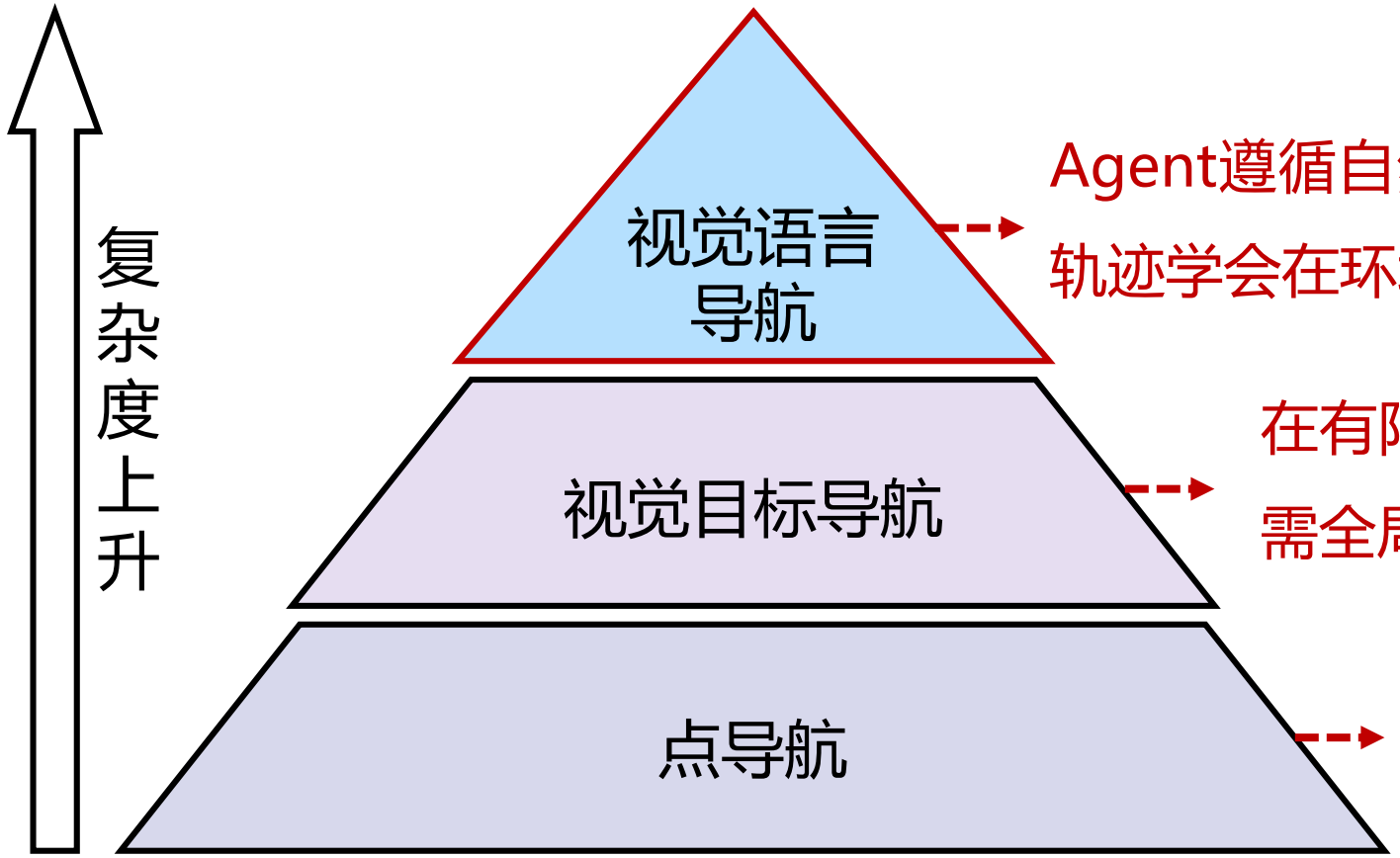
轮式　足式　复合式

- **行为预测算法**

轨迹预测、长短期目标预测、智能决策

具身智能的典型案例

■ 智能导航任务的金字塔结构



Agent遵循自然语言指令并结合视觉观察以及历史轨迹学会在环境中逐步导航。

在有限部分观测输入条件下，即缺少导航所需全局信息，输出动作对目标位置的推测。

Agent导航至距离某个特定点有一定距离的位置。

具身智能的典型案例

■ 智能导航任务——点导航

- **初始化与目标定位：**Agent通常在环境的原点（0,0,0）初始化，目标点通过相对原点的三维坐标指定。为了完成任务，Agent需要具备视觉感知、情景记忆构建、逻辑推理、路径规划及导航等能力。
- **导航硬件与位置感知：**Agent集成GPS和指南针等硬件来确定自身相对于目标的方向位置，目标坐标可以是静态的或动态的。然而，由于室内环境中定位的不准确性，目前视觉导航工作转向基于RGB-D的在线定位，不再依赖传统的GPS和指南针。
- **学习型导航方法：**基于学习的点导航方法探索端到端解决方案来处理未知环境中的导航，利用多种感官输入（如彩色图像、深度图及最近的观测动作），无需真实地图或精确姿态信息。

具身智能的典型案例

■ 智能导航任务——视觉目标导航

- **任务定义与数学描述：**

  - **任务定义：**Agent仅依据第一视图的RGB图像导航至目标物体t，该物体属于目标类别集合T，并且Agent事先并不了解环境。

  - **数学描述：**假设在每一次的导航过程中，Agent和目标物体的初始位置都是随机选取的，Agent可以通过策略学习网络 π 根据当前的RGB图像I和目标物体的特征向量w_t采样动作a，即$a \sim \pi_\theta(I, w_t)$。

    其中，θ是策略学习网络的参数权重，动作a属于集合A，其中前行距离和旋转角度可以根据实际任务需求来定义，最后“Done”表示Agent已经找到目标时的状态，从而结束这一导航过程。

具身智能的典型案例

■ 智能导航任务——视觉目标导航

- **示例：**

左图为成功示例，其中绿色轨迹表示成功的导航路径；白色三角形表示智能体的视角，蓝色方框标记的是目标物体。

(a) 目标导航路径

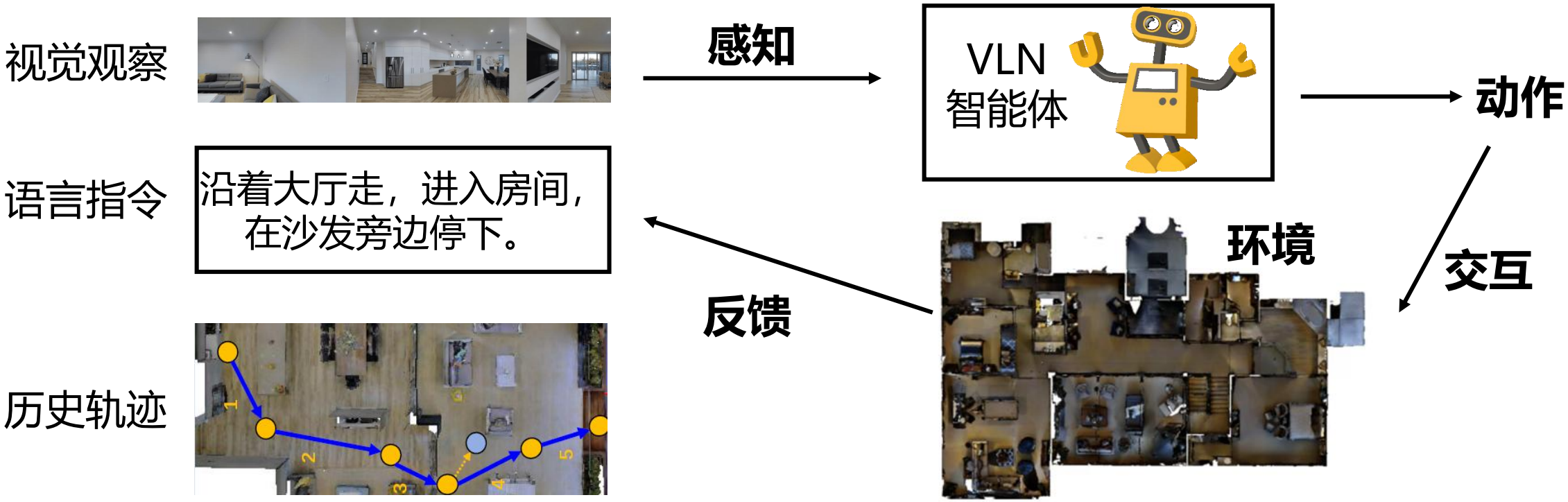
(b) 智能体观测视角与环境

具身智能的典型案例

■ 智能导航任务——视觉语言导航（VLN）

- **VLN的一般框架：**

视觉-语言-导航（Visual Language Navigation，VLN）任务旨在使得Agent遵循自然语言指令并结合视觉观察以及历史轨迹学会在环境中逐步导航。

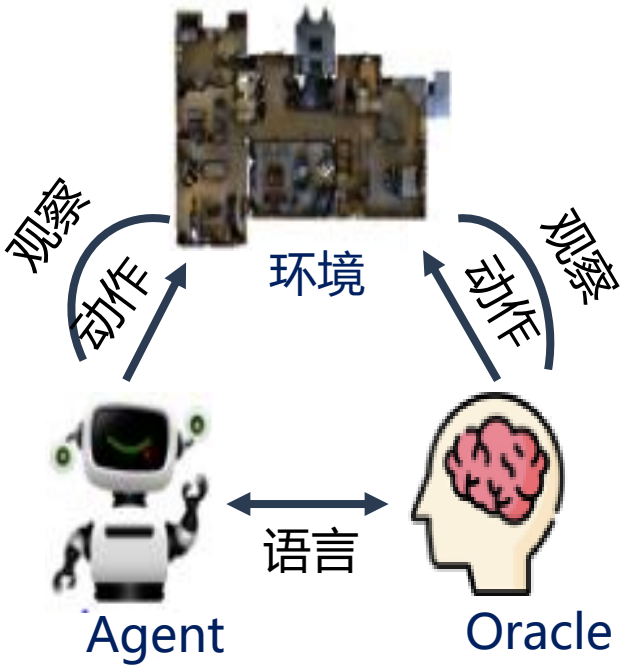


具身智能的典型案例

■ 智能导航任务——视觉语言导航（VLN）

- **基于VLN的室内导航方案：**

借助于视觉观察、环境交互以及奖励机制，构建强化学习框架；利用语言指令指导Agent完成语言理解、视觉与语言关联以及动作预测，使得智能体移动到指定位置。



走向围栏，随后向右经过楼梯。走进起居室后右转，在桌子前停下。

(1)理解语言

(2)关联视觉语言

(3)动作预测

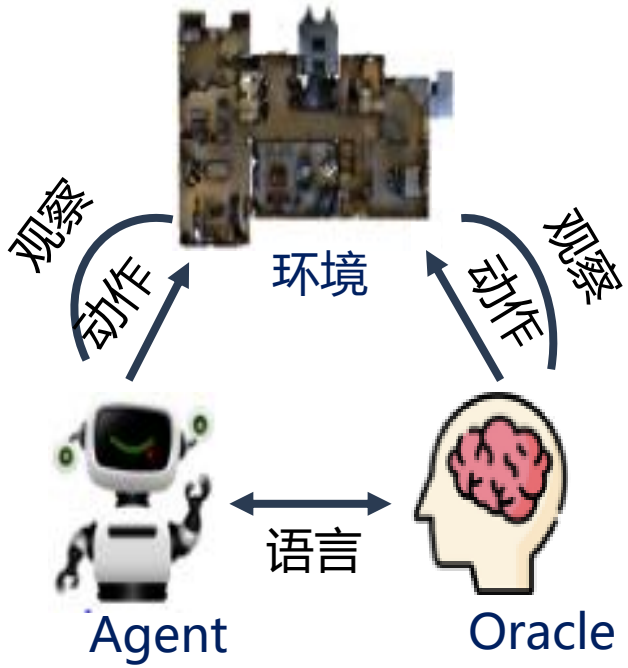
朝着栏杆行走，然后从楼梯右经过，走进客厅向右转，然后停在边桌旁边。

具身智能的典型案例

■ 智能导航任务——视觉语言导航（VLN）

- **基于VLN的室内导航方案：**

VLN框架的核心组成：自然语言理解与处理 视觉感知与识别 导航规划与控制 跨模态融合与协同



走向围栏，随后向右经过楼梯。走进起居室后右转，在桌子前停下。

(1)理解语言

(2)关联视觉语言

(3)动作预测

朝着栏杆行走，然后从楼梯右经过，走进客厅向右转，然后停在边桌旁边。

具身智能的典型案例

■ 智能导航任务——视觉语言导航（VLN）

● VLN的具体实现——Robo-VLN模型

- Robo-VLN（机器人视觉和语言导航）利用分层跨模态Agent，通过模块化训练与分层决策，将VLN定位为逼真模拟中的连续控制问题，从而完成长期跨模态任务。
- 智能体由一个**高级策略**和一个相应的**低级策略**组成。

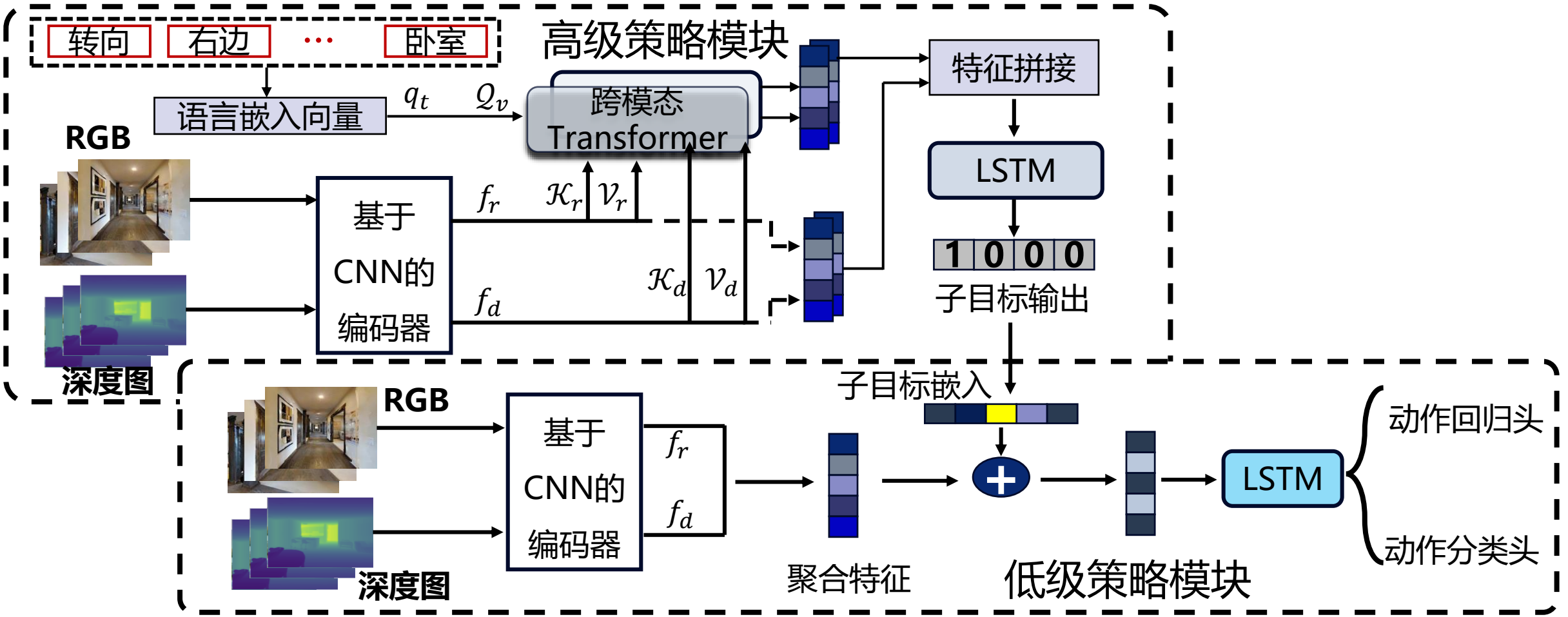
高级策略是由编码器-解码器架构组成，其任务是将相关指令 与观察到的视觉信息通过交叉注意力机制进行特征对齐，同时利用多模态注意力解码器获取跨时间信息。

低级策略利用模仿学习策略将子目标信息和观察到的视觉状态转换为线速度和角速度，然后计算低级动作运行与停止的分类概率。

具身智能的典型案例

■ 智能导航任务——视觉语言导航（VLN）

● VLN的具体实现——Robo-VLN模型



内容目录

具身智能的前沿与展望

■ 具身智能大模型

- **定义**：具身智能大模型是指那些能够赋予机器人或其它具身智能体以感知、理解和互动于物理世界能力的模型。这些模型通常包含了先进的机器学习技术和算法，尤其是深度学习和强化学习。
- **特点**：
  - **多模态感知**：具身智能大模型能够学习并处理来自多种传感器的不同类型数据，如视觉、听觉甚至是触觉，这使得机器人将能够实时获取环境信息，从而更全面地理解周围环境。
  - **决策与行动的智能化**：通过学习大量的决策案例和行动规则，为智能体提供决策支持。然后，将这些决策转化为实际的行动，从而实现对未知环境的有效干预和改变。
  - **学习与适应的持续进化**：在与环境的互动过程中，大模型可以通过持续学习和优化算法，不断提升自身的泛化能力和适应能力。通过微调或优化，推动智能体的不断进化和升级。

具身智能的前沿与展望

■ 具身智能的未来挑战

● 提升非结构化真实环境的快速适应能力：

- 需具备灵活的计算能力，以适应非结构化环境中信息稀缺与场景多变的挑战；当前研究虽在具身导航与抓取任务上取得进展，但仍需开发更灵活可扩展的智能体架构，实现感知、理解、规划与执行的深度整合与闭环循环。

● 提升复杂环境的准确认知与执行能力：

- 要求智能体精准感知并理解真实环境，当前研究虽利用大语言模型进行任务规划，但仍缺乏整体性认知，提升知识迁移与泛化能力是实现全面认知的关键；为执行长期复杂任务，如打扫厨房，需开发融合强大感知、丰富常识与高性能规划算法的新型规划器，以动态调整策略，适应复杂场景，确保高效稳定的任务执行。

具身智能的前沿与展望

■ 具身智能的未来挑战

● 发展多实体协作的群体智能：

- 具身智能需要发展类似生物群体中个体间协同作用的机制，以实现分工协作和动态任务分配，从而灵活应对多变情境；研究人员正探索开发支持多实体间高效协作的智能系统，以模拟自组织特性，提升智能协作水平。

● 应对数据安全与伦理挑战：

- 具身智能在与真实环境交互时需确保数据安全与隐私保护，特别是在家庭护理等场景中，防止隐私泄露风险，并在决策中遵循伦理准则；为应对这些挑战，研究人员探索数据加密与隐私保护技术，同时制定严格的伦理规范，以增强数据安全性和用户信任度。

Q&A

THANKS!