



南京大學
NANJING UNIVERSITY

人工智能导论

Summary

郭兰哲

南京大学 智能科学与技术学院

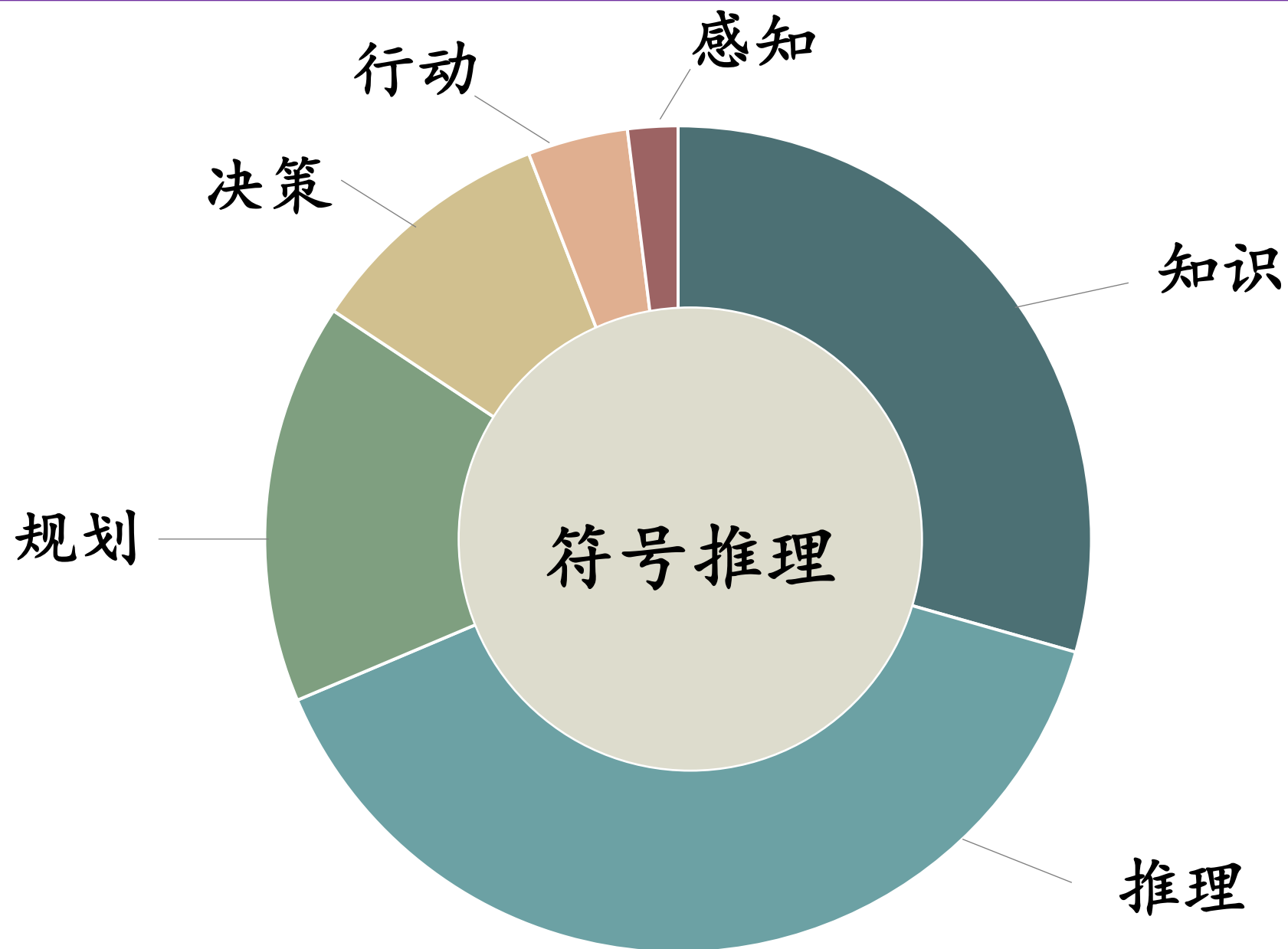
<https://www.lamda.nju.edu.cn/guolz/IntroAI/fall2025/index.html>

Email: guolz@nju.edu.cn

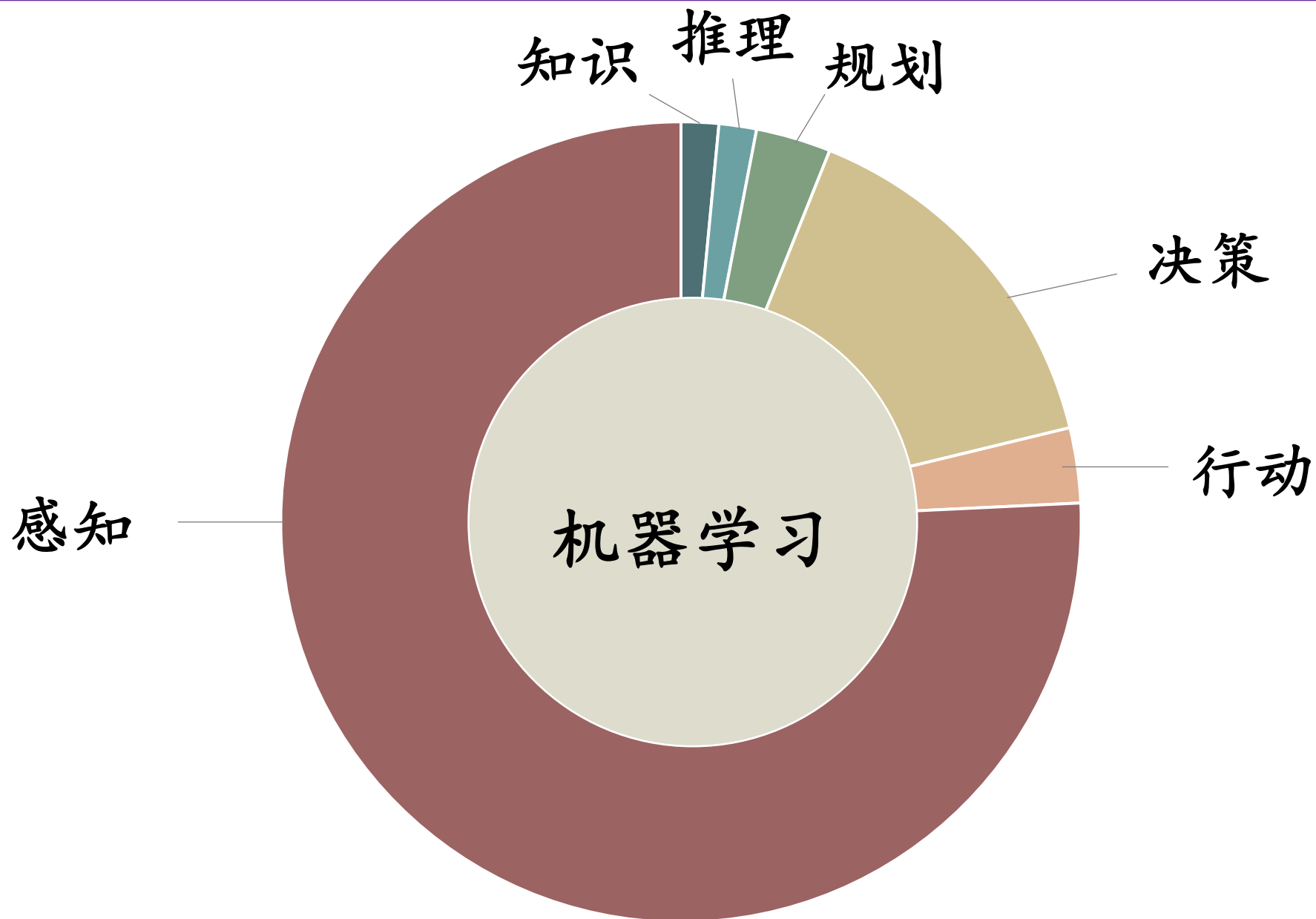
人工智能



第一代人工智能：符号主义



第二代人工智能：联结主义



当下，大模型时代

- 感知：多模态大模型，形成通用感知能力
- 知识：大模型成为知识载体(参数化知识)，也开始融合外部知识库
- 推理：基于LLM Next-Token Prediction的链式思考(COT)
- 规划：大模型初步具备规划能力，能够进行任务分解，反思，回溯
- 决策：基于LLM的智能体，初步实现工具调用、策略执行
- 行动：多模态Agent+机器人，初步实现感知-推理-行动的闭环控制

未来，人工智能有可能朝什么
方向发展

现存的挑战

- **技术(Technology)**
 - 仍以数据驱动的机器学习为主，难以充分利用大量知识
- **能力(Capability)**
 - 模型可信性、稳健性、可解释性不足
- **形态(Deployment)**
 - 主要以数字世界的智能系统为主，如何应用到真实物理世界仍然是挑战

一些可能的方向？

方向1：数据知识双驱动的人工智能

- 符号主义：数据利用效率低、知识难以提取、泛化性弱
- 连接主义：难解释、难可信，距离真正的人脑相差甚远，在概率近似正确的机器学习范式下看不到解决希望

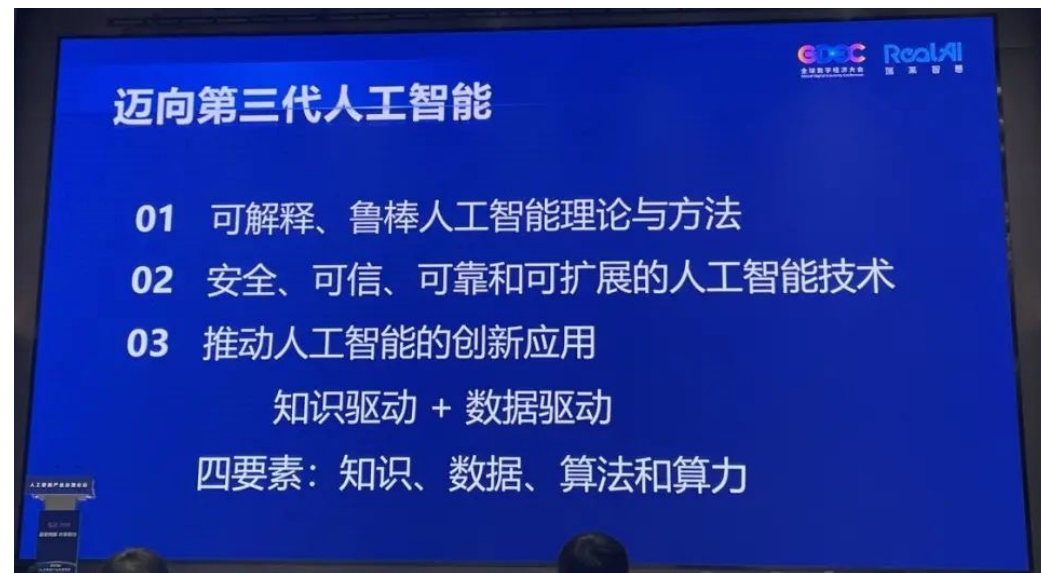
未来的人工智能应当是融合数据驱动与知识驱动的

四要素：知识、数据、算法、算力

方向1：数据知识双驱动的人工智能

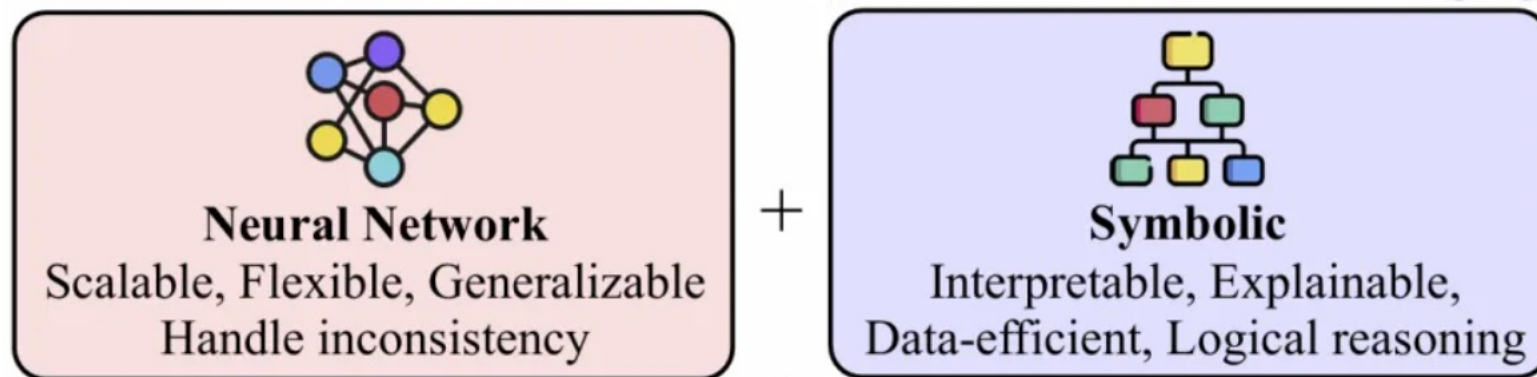


南京大学周志华院士
知识数据双驱动人工智能



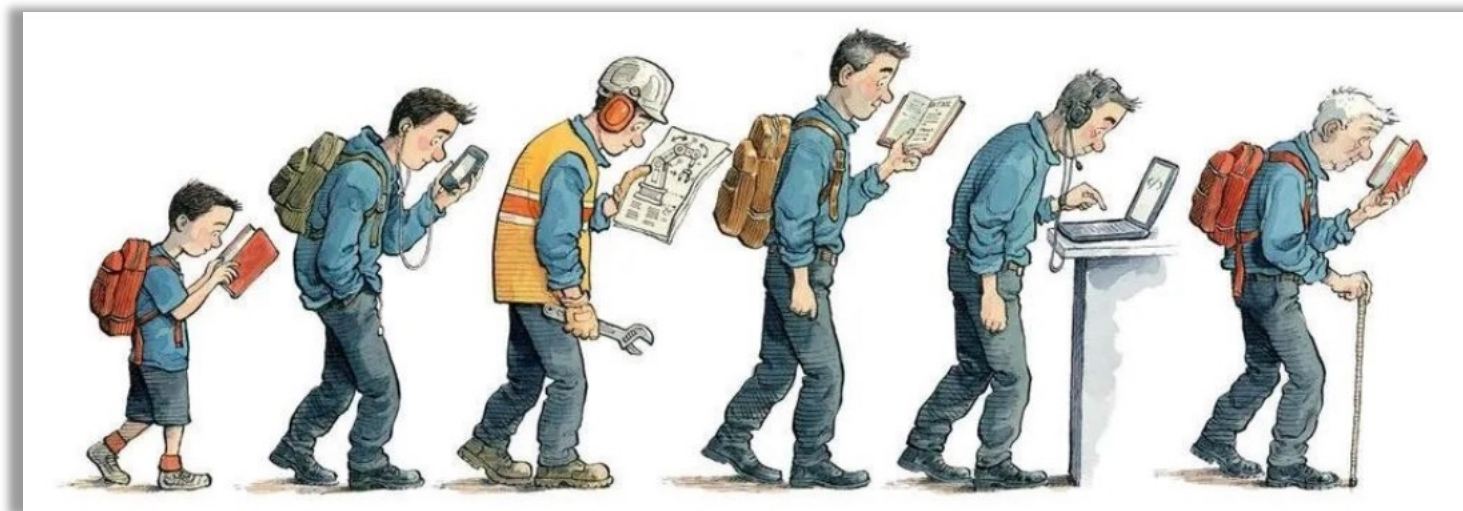
清华大学张钹院士
迈向第三代人工智能

方向1：数据知识双驱动的人工智能



- 数据知识融合的理论基石?
- 知识如何表示? 如何利用? 如何更新?
- 如何实现神经与符号的高效联合优化?
- 如何实现系统大规模、可扩展?
-

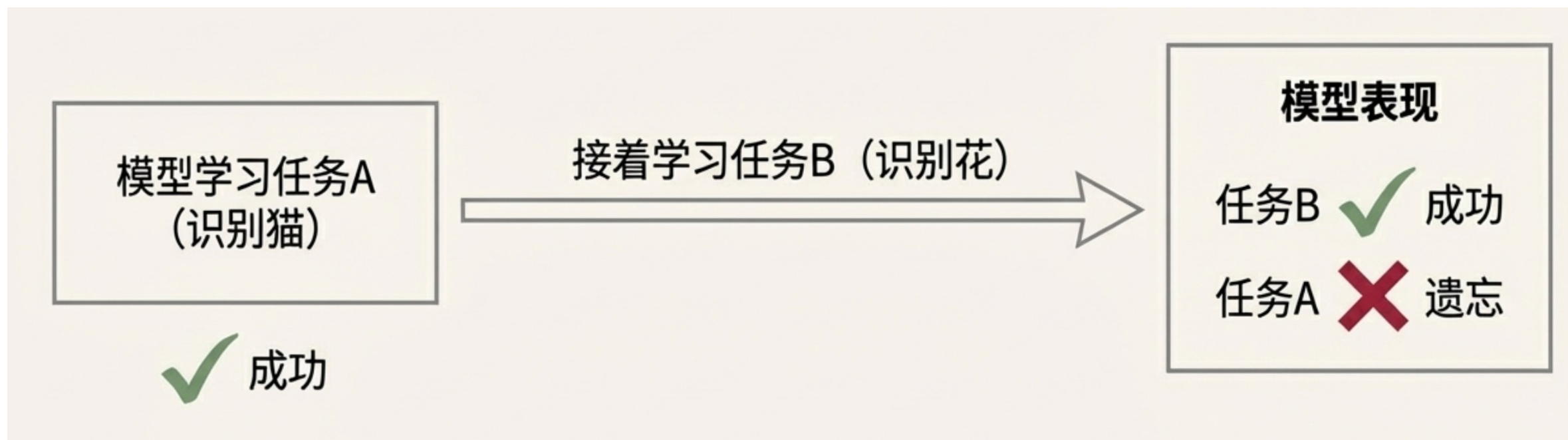
方向2: Life-Long, Continual Learning



智能体在整个生命周期中持续学习、积累知识、更新技能，
并能够在新任务下灵活应用过去的经验

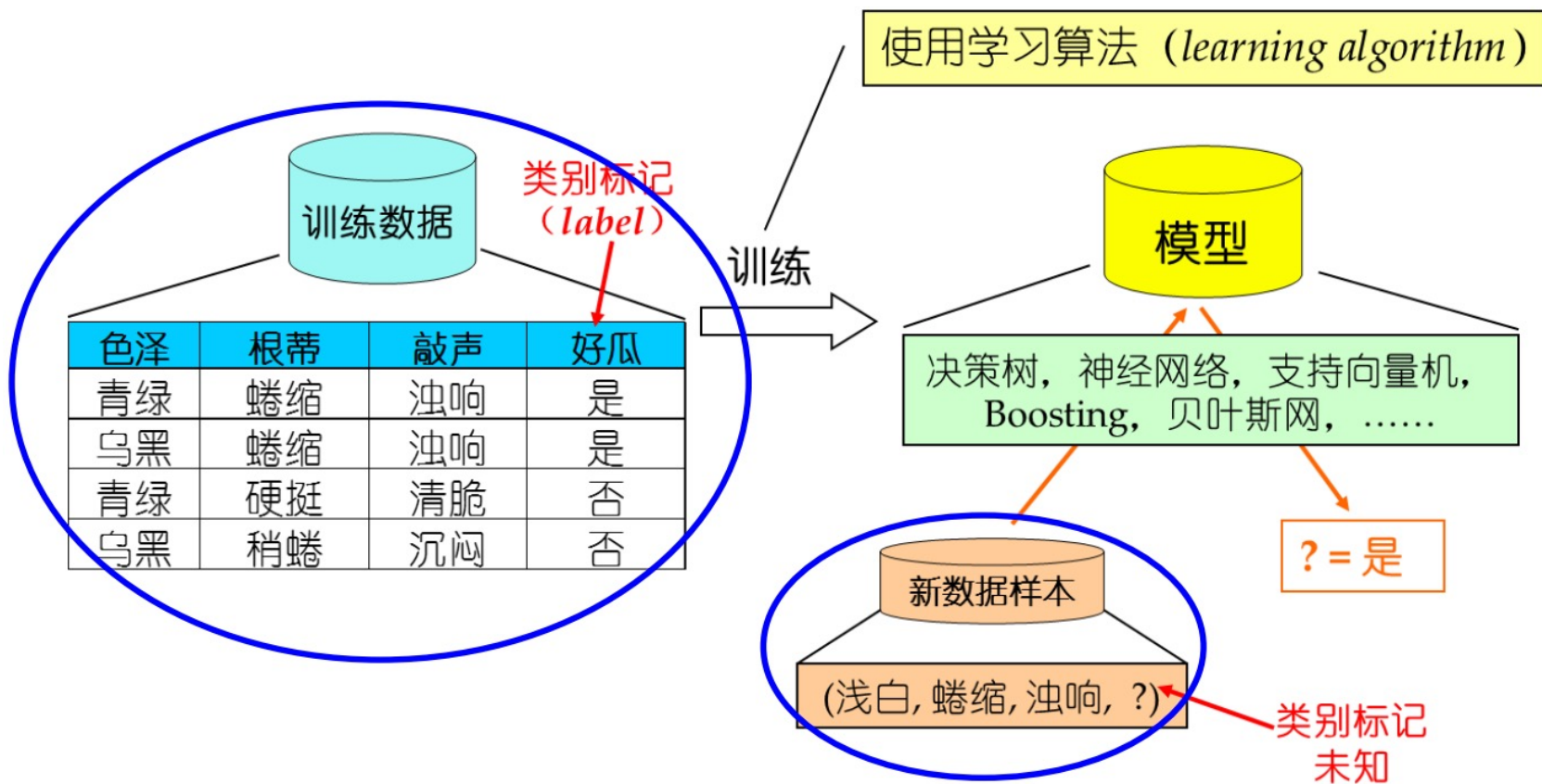
方向2: Life-Long, Continual Learning

主要挑战: **灾难性遗忘(Catastrophic Forgetting)**, 即学会新任务后, 旧任务的性能会严重下降



方向3：开放动态环境的机器学习

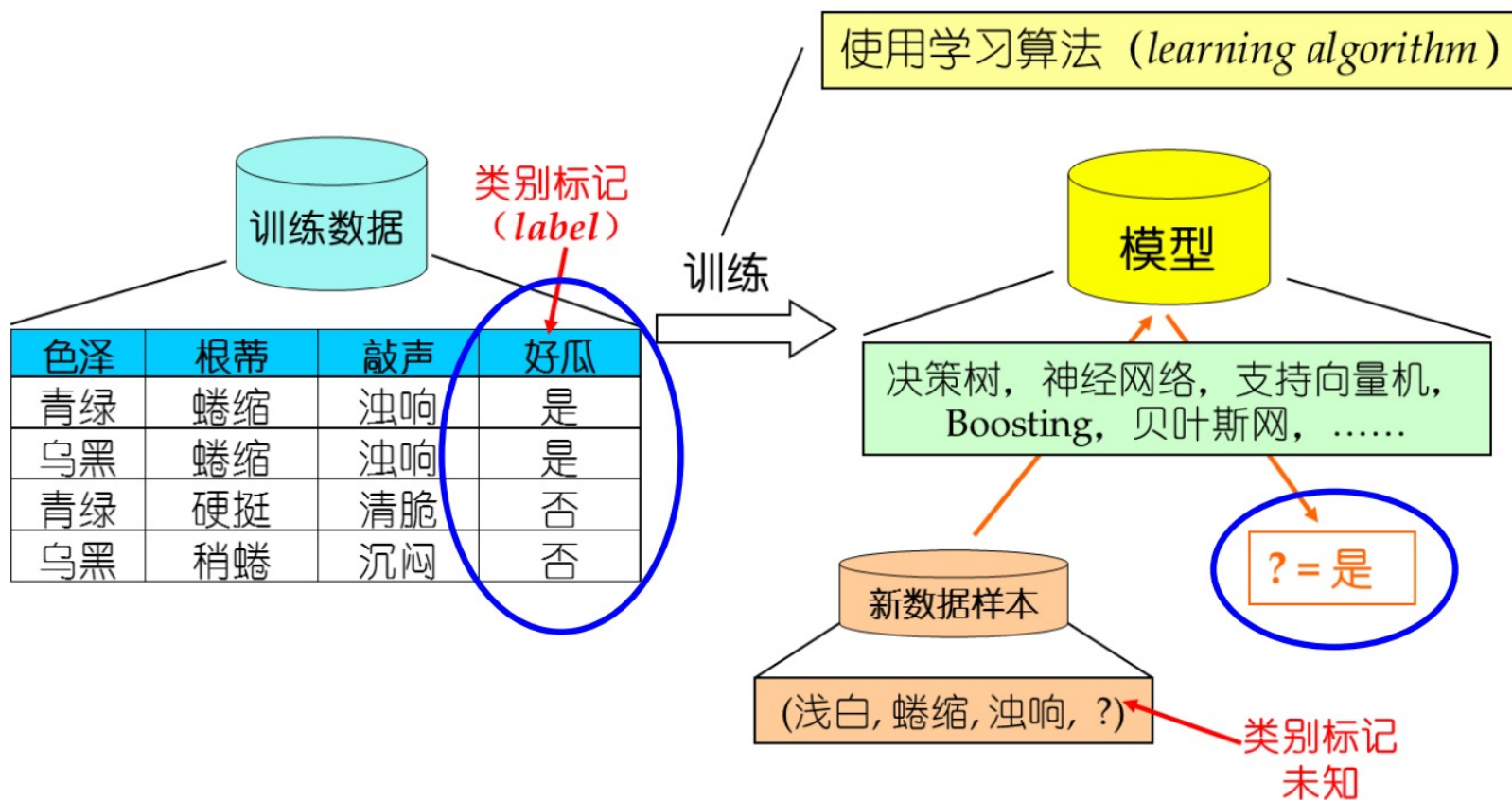
主要针对封闭静态环境（重要因素大多是“定”的）



数据分布恒定

方向3：开放动态环境的机器学习

主要针对封闭静态环境（重要因素大多是“定”的）

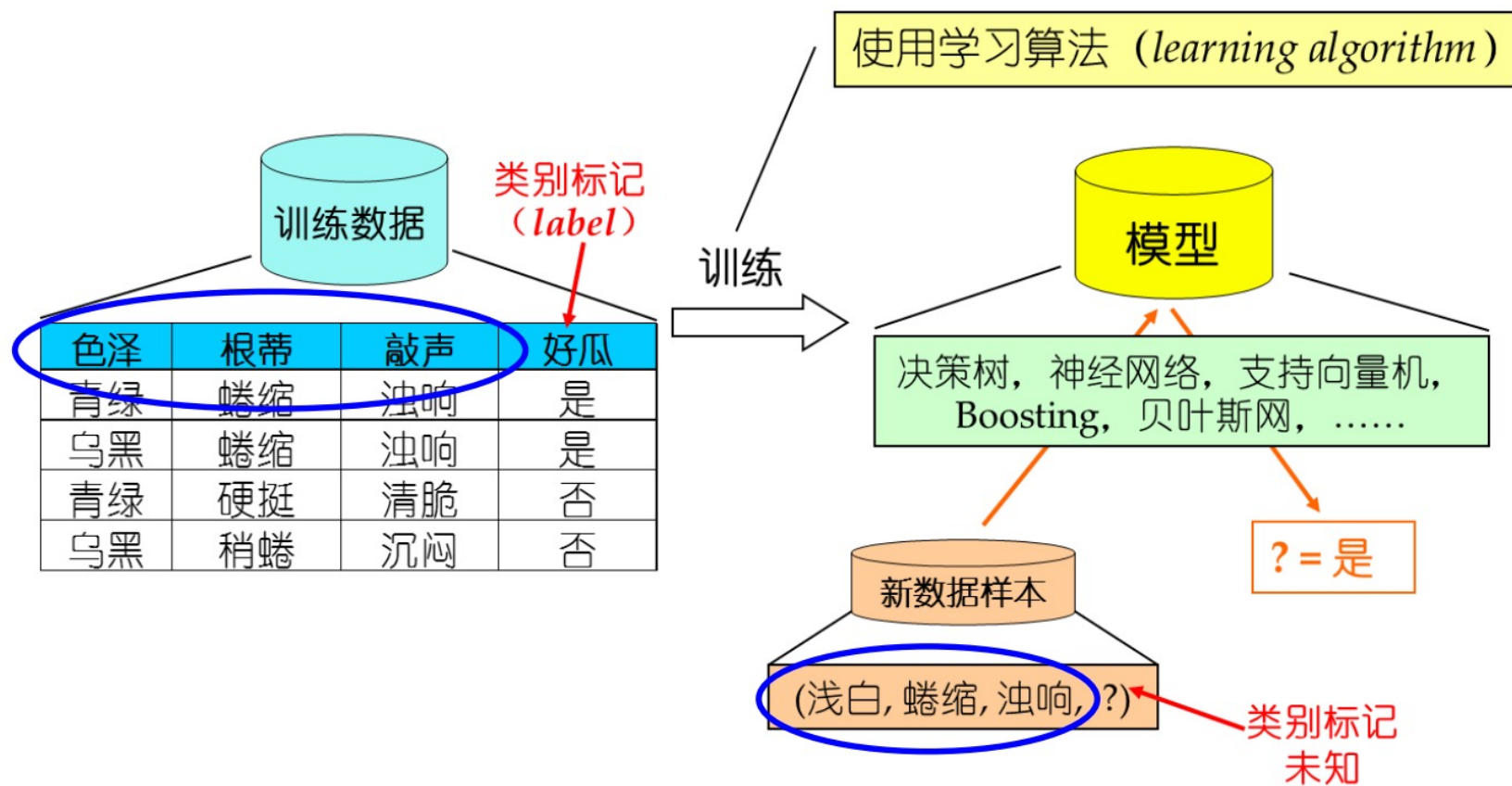


数据分布恒定

样本类别恒定

方向3：开放动态环境的机器学习

主要针对封闭静态环境（重要因素大多是“定”的）



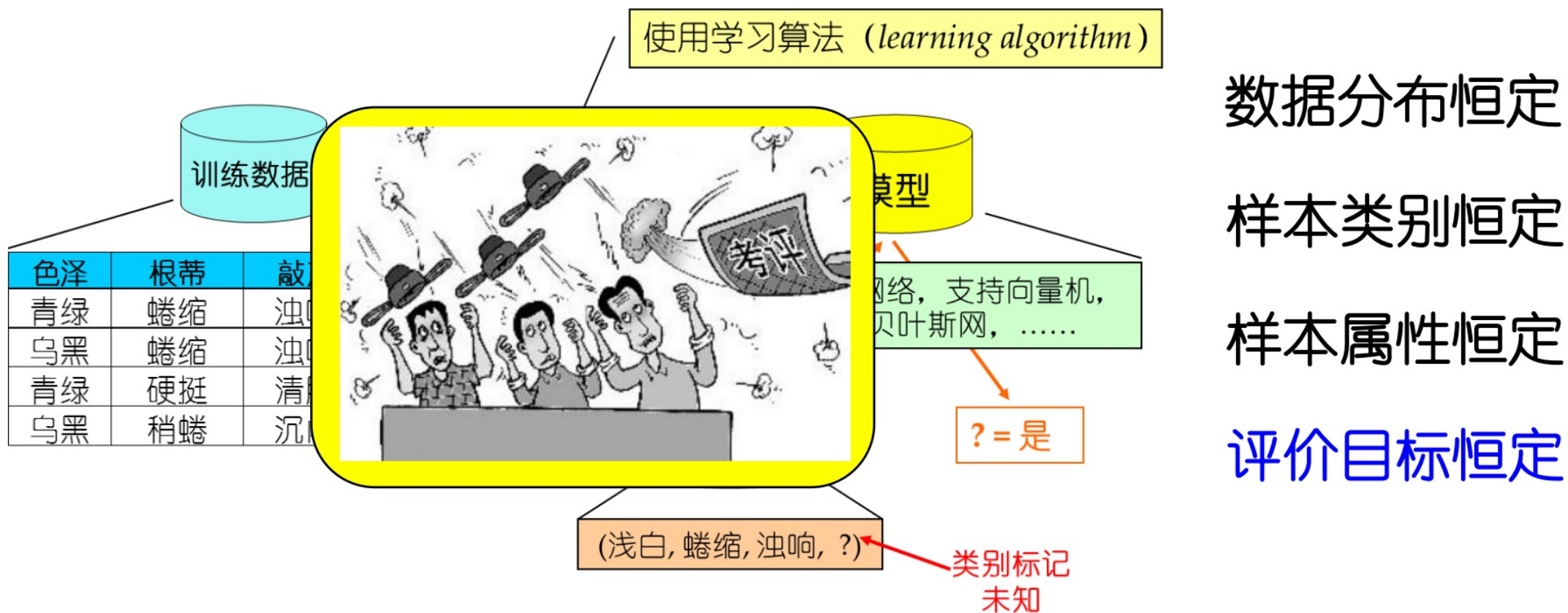
数据分布恒定

样本类别恒定

样本属性恒定

方向3：开放动态环境的机器学习

主要针对封闭静态环境（重要因素大多是“定”的）



方向3：开放动态环境的机器学习

主要针对封闭静态环境（重要因素大多是“定”的）

使用学习算法 (*learning algorithm*)

数据分布恒定

类别恒定

属性恒定

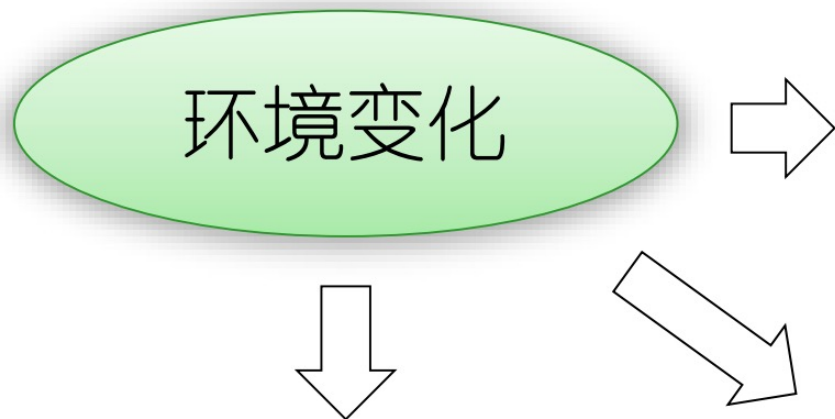
目标恒定

封闭静态环境 → 开放动态环境

一切都可能“变”！

色泽	根蒂
青绿	蜷缩
乌黑	蜷缩
青绿	硬挺
乌黑	稍蜷

方向3：开放动态环境的机器学习



环境监控



60%传感器失效，新增60%传感器

难点：属性变动

物种监测



背景随季节改变

难点：分布变化

自动驾驶



汽车厂商设想的场景

上路后遇到的场景



难点：类别增加

方向3：AI安全、对齐、可靠性

AAAI “主席报告”

(“Presidential
Address”)

2016.02.14



Tom Dietterich

AAAI/AAAS/ACM Fellow

AAAI 现任主席

国际机器学习学会创始主席 (2001-2008)

STEPS TOWARD ROBUST
ARTIFICIAL INTELLIGENCE

走向鲁棒的人工智能

Tom Dietterich
President, Association for the Advancement of Artificial
Intelligence

方向3：AI安全、对齐、可靠性

T. Dietterich强调：随着人工智能技术的发展，越来越多地面临“高风险应用”

因此，必须要有“鲁棒的AI”

- 对人类用户错误鲁棒
- 对网络攻击鲁棒
- 对错误目标鲁棒
- 对不正确模型鲁棒
- 对未建模现象鲁棒



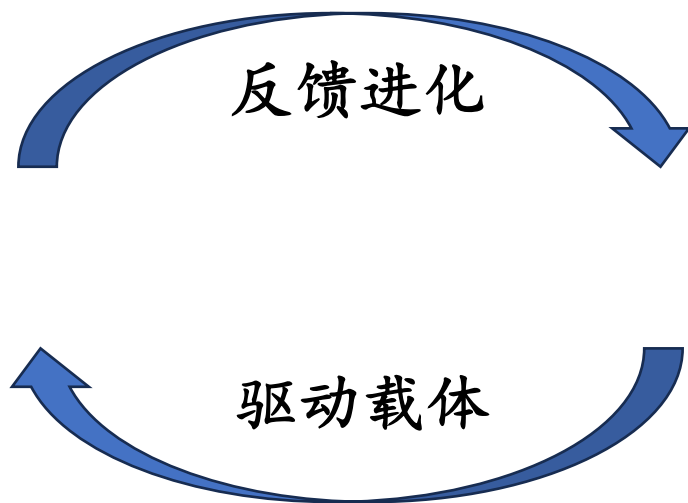
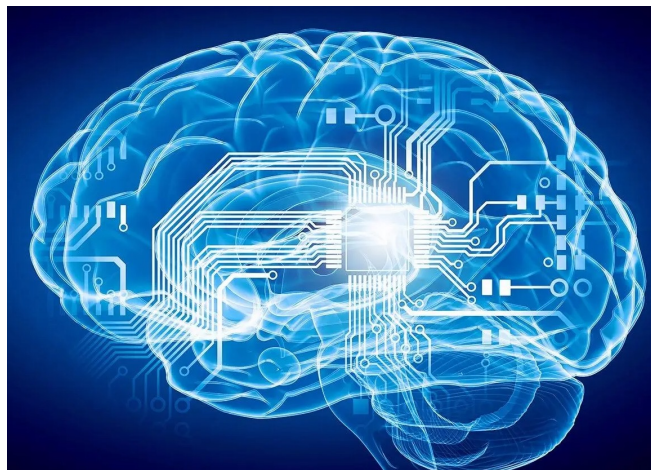
方向5：具身智能 (Embodied AI)

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

机器人本体



具身智能



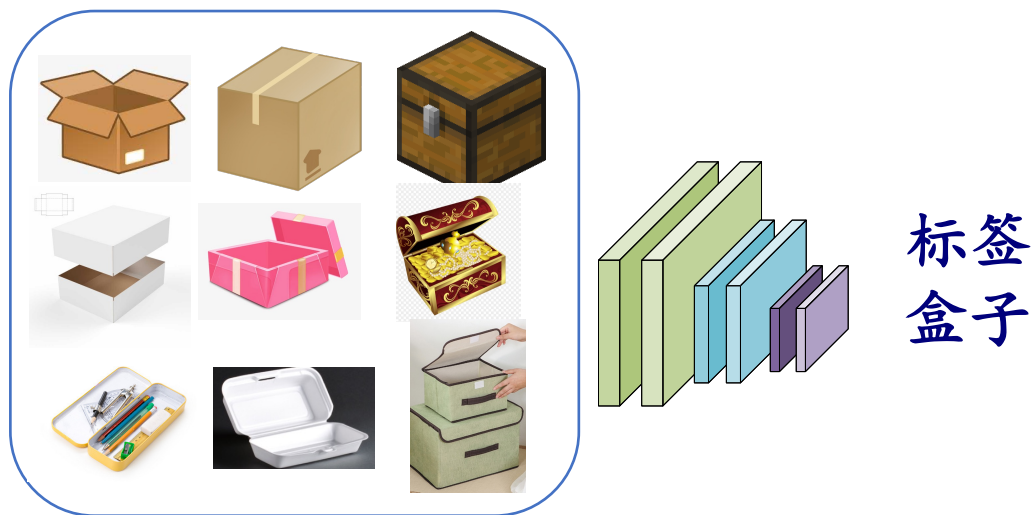
基于身体，理解物理世界、并与物理世界互动达成人类需求的智能系统

方向5：具身智能 (Embodied AI)



方向5：具身智能 (Embodied AI)

离身智能 (当前主流AI学习方式)

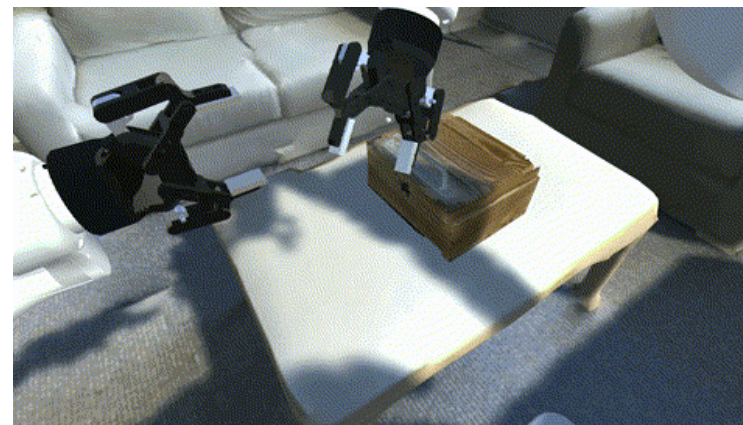


人类告诉我长这样子是盒子
(“标签”学习)

什么是“盒子”

具身智能：具有身体体验智能

具身智能



亲身体会盒子是什么
可以打开，可以装东西

方向5：具身智能 (Embodied AI)



硬件方面：

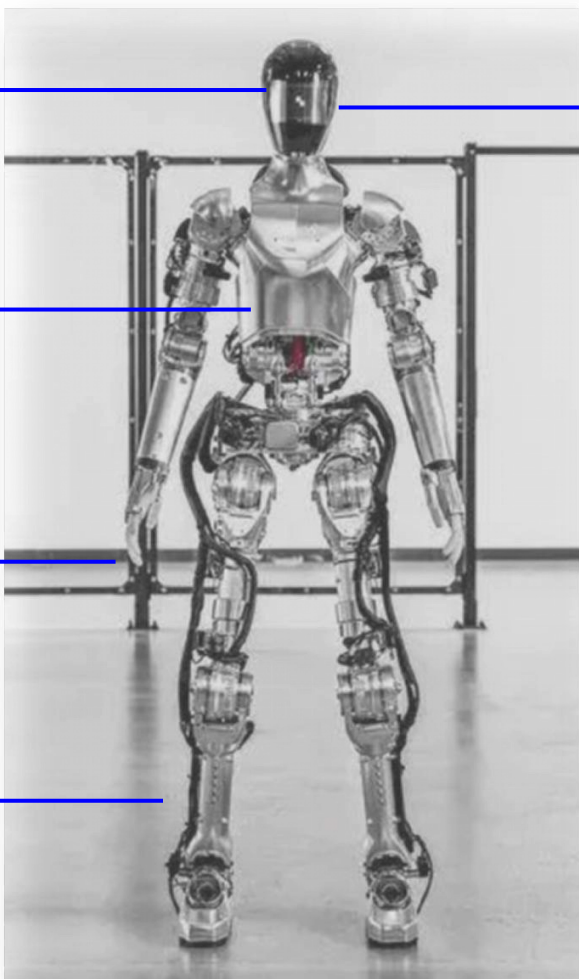
2D视觉信号或
3D点云信号

语音信号

机器人躯体的
硬件结构

触觉信号或力
反馈信号

位姿信号



软件及算法方面：

大
脑

- 收集所有传感器采集的环境信息和自身状态并综合分析当前所有状态 **(具身感知)**
- 根据当前状态，对自身下一步的运动做出决策和规划 **(具身推理)**

小
脑

- 向下位机发送运动指令 **(具身执行)**
- 下位机通过运控技术执行指令

方向5：具身智能 (Embodied AI)

机器人视觉传感器信号



- 收集所有传感器采集的环境信息和自身状态，并综合分析当前状态（具身感知）

机器人采集视觉信息，分析出应该对咖啡进行清理

- 根据当前状态，对自身下一步运动做出决策和规划（具身推理）

- 向下位机发送运动指令（具身执行）

生成机器人的运动轨迹，包括手臂如何运动、手掌如何运动、腿部如何运动等

清理咖啡需要如下几步：

- 1.扶正杯子并拿起杯盖
- 2.找到抹布
- 3.用抹布擦拭地面
- 4.将抹布放回
- 5.将杯子和杯盖扔掉

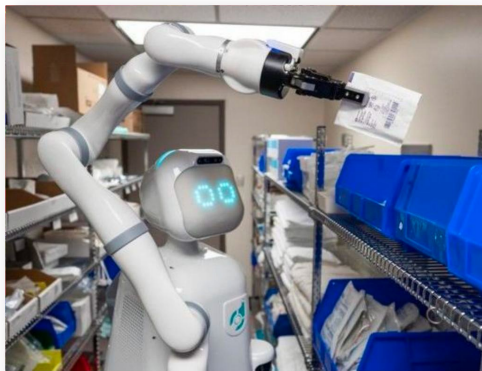
- 下位机通过运控技术执行指令

机器人执行

方向5：具身智能 (Embodied AI)

- 硬件方面：能造出具备基本性能的机器人硬件和高精度传感器
- 软件及算法层面：运控技术相对比较成熟，**如何实现更好的具身感知和具身推理还存在诸多问题**
- 当前最先进的大模型+最先进的机器人能否实现具身智能？**不能！**

原因：大模型本身推理规划能力有限，稳健可信性不足



通用机器人是对
人工智能的终级幻想

**革命尚未成功，
同志仍需努力！**

这门课会结束，但我们的思考与实践，永无止境

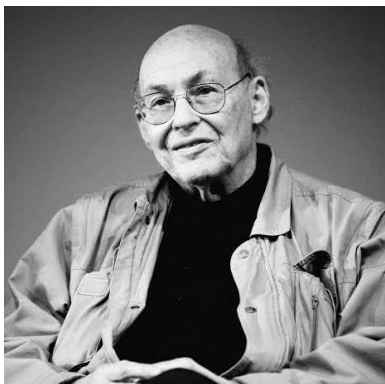
去思考根本的问题

去做真正值得做的研究

去做AI能够赋能的产业

保留对根本问题的执着，对好奇心的忠诚

人工智能历史中的图灵奖得主



Marvin Minsky
1969年图灵奖



John McCarthy
1971年图灵奖



Allen Newell
1975年图灵奖



Herbert Simon
1975年图灵奖



Ed Feigenbaum
1994年图灵奖



Judea Pearl
2011年图灵奖



Yoshua Bengio
2018年图灵奖



Geoffrey Hinton
2018年图灵奖

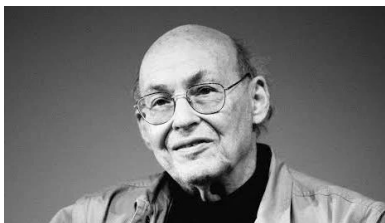


Yann Lecun
2018年图灵奖



Richard Sutton
2024年图灵奖

人工智能历史中的图灵奖得主



俱往矣，数风流人物，还看今朝。

——毛泽东



Judea Pearl
2011年图灵奖



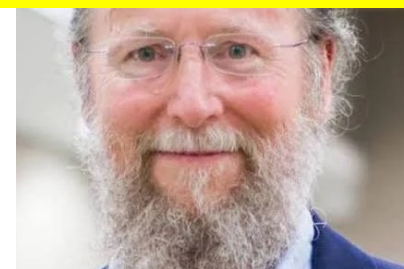
Yoshua Bengio
2018年图灵奖



Geoffrey Hinton
2018年图灵奖



Yann Lecun
2018年图灵奖



Richard Sutton
2024年图灵奖

祝大家在人工智能的道路上
一帆风顺、学有所成！