通识选修课 · 2026秋季学期

具身智能

从看见到行动

第1章 · 从科幻到现实：什么是具身智能

本章学习目标

- 理解具身智能的基本概念
- 建立PCA闭环思维
- 区分离身智能与具身智能



两个改变世界的瞬间

VS



AlphaGo vs 李世石

4:1 大胜，但第78手暴露了深层局限

李世石"神之一手"让AlphaGo陷入混乱

拥有**1920个CPU+280个GPU**的超级大脑

却永远无法离开棋盘

思考：AlphaGo能去厨房倒杯水吗？



波士顿动力 Atlas

从后空翻到**连续空翻**

身体知道答案

拥有**56个自由度**的全电动躯体

在物理世界中摸爬滚打、学习成长

思考：Atlas能下围棋吗？

AlphaGo的"局限"：棋盘上无敌，棋盘外无知



2016年3月 AlphaGo vs 李世石 对弈现场

本质

1920个CPU + 280个GPU的超级模式匹配系统

第78手"神之一手"的深层原因

深度学习学到的是统计知识，不是因果逻辑

根本问题

没有身体 = 没有感知 = 没有与世界的物理交互

核心概念 · 离身智能

离身智能（Disembodied Intelligence）： 智能被锁死在特定符号空间里

类比：你读了无数本骑自行车的书，但第一次上车还是会摔倒 —— 知识 ≠ 技能

Atlas的后空翻：身体知道答案

物理构造

56个自由度关节 · 360度全向摄像头 · 全电动驱动

"失稳"的教育意义

CES展会落地不稳、保护罩甩飞

物理世界的复杂性永远无法完全建模

学习闭环

感知→认知→行动→新感知→新认知→新行动



1个月 从踉跄到连续空翻

关节记住了"落地时膝盖弯曲15度最能吸收冲击力"——这些知识不是写在代码里的，而是从身体与物理世界的交互中"生长"出来的

智能的"完整形态"是什么？

AlphaGo

超级大脑 + 无身体 = 离身智能巅峰

- 1920个CPU + 280个GPU
- 看过三千万盘棋谱
- 离开棋盘 = 一无所知
- 智能被锁死在符号空间里

Atlas

灵活身体 + 有限大脑 = 具身智能雏形

- 56个自由度全电动驱动
- 在物理世界中摸爬滚打
- 从体验中"生长"知识
- 不会下围棋，不会写诗

智能是否必须依赖身体？

哲学脉络

柏拉图
身体是囚禁灵魂的"牢笼"

笛卡尔
"我思故我在"——思维与身体分离

梅洛-庞蒂
身体不是心智的容器

离身智能："缸中之脑"的幻想

思想实验：缸中之脑

如果你的所有感官体验都是电信号模拟的，你怎么知道自己是真实的？—— 普特南，1981

传统AI的假设

智能 = 大脑的符号运算

身体只是无关紧要的"硬件载体"

就像一台电脑：

CPU才是重要的，机箱是什么形状、什么颜色，根本不影响计算能力

骑自行车的案例

知识：平衡原理、力学公式、重心计算

技能：身体记住的平衡感

身体"知道"的东西，大脑"解释"不清

你不需要"计算"怎么走路

根本局限

离身智能只能处理**符号**，无法获得**技能**；无法真正"理解"它处理的内容

AlphaGo可以下出完美的围棋，但它不知道"围棋"是什么

具身认知：第二代认知科学的革命

> *"身体不是心智居住的容器，而是我们'拥有'一个世界、生成所有意义的原初视角"*
>
> —— 梅洛-庞蒂《知觉现象学》，1945

三大经典实验

咬笔实验

嘴里横着咬一支铅笔

（面部肌肉呈"微笑"状态）

然后看卡通图片

结果：咬着铅笔的人认为图片更好笑

身体倾斜实验

身体向前倾斜

→ 对积极词语反应更快

身体向后倾斜

→ 对消极词语反应更快

姿态影响情绪判断

热咖啡实验

拿到热咖啡的人

→ 认为对方更"温暖"

拿到冰咖啡的人

→ 认为对方更"冷淡"

温度影响人格判断

> 核心发现：你的身体状态——哪怕只是手心的温度——会**无意识地渗透**到认知判断中

具身认知三大核心命题

01

身体结构规定认知能力的范围

蝙蝠用超声波感知世界

鹰的视觉清晰度是人类的8倍

不同的身体结构

= 完全不同的"世界版本"

02

身体结构决定认知产生的机理

"伸手"这个动作本身

就在塑造空间认知

盲人通过触摸

来"看见"物体的形状

动作即认知

03

身体、认知和环境组成动态统一体

你不是"在你房间里"

你和房间是不可分割的系统

你坐在椅子上

椅子支撑着你

你也同时"感知"着椅子

核心结论：你的大脑不是认知的"司令部"，你的身体也不是认知的"执行者"。**认知是整个身体-环境系统的涌现属性。**

从哲学到工程：具身智能的诞生

> *"具身智能通过构建具有本体感知与行动能力的智能体，利用多模态传感器实时捕获环境状态，利用执行机构施加物理作用，并在连续时空维度中形成'感知-认知-决策-行动'的闭环学习系统。"*
>
> —— 郑南宁院士

具身智能公式

具身智能 = 智能体 + 传感器 + 执行器 + 算法大脑 + 闭环学习

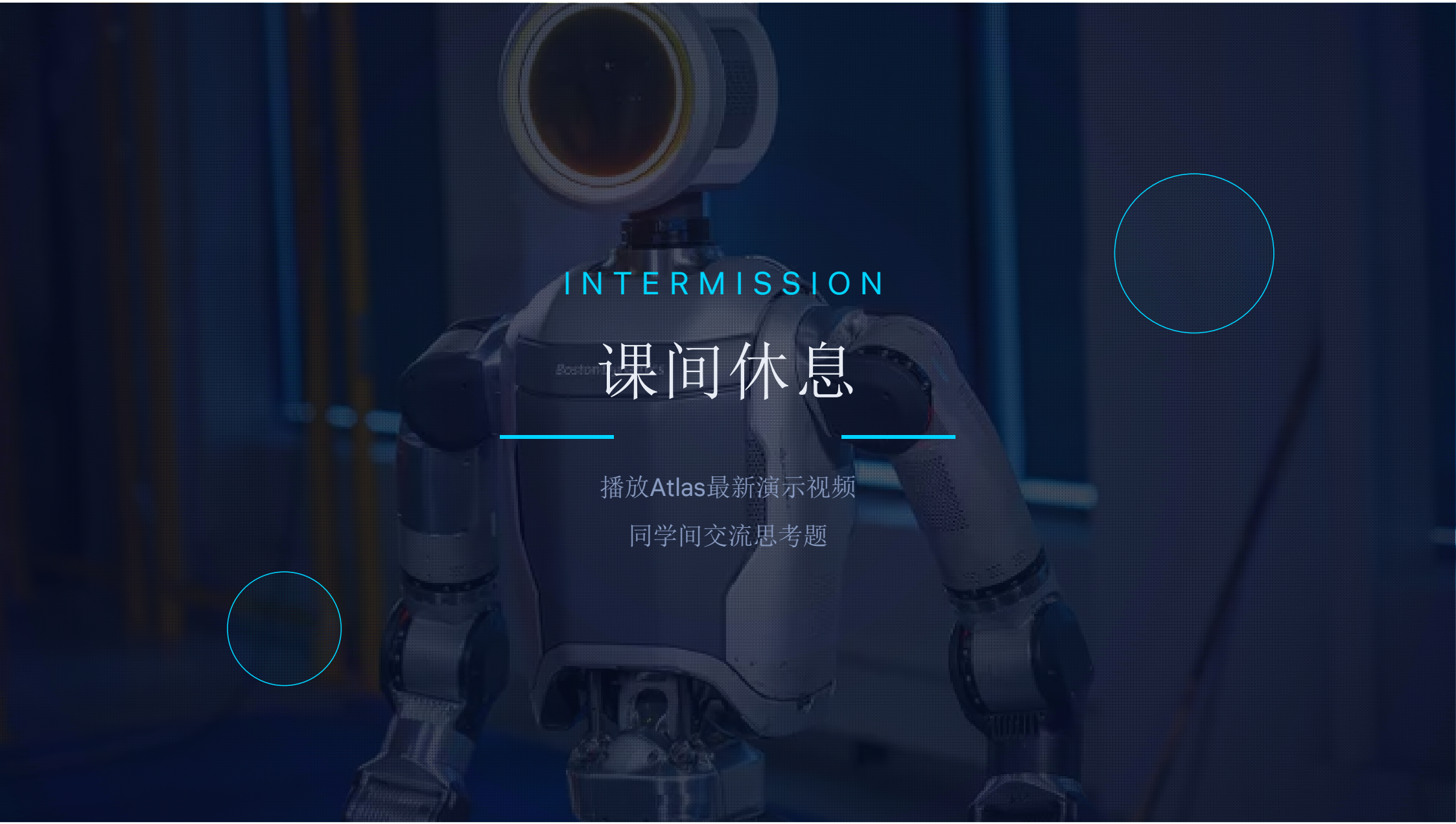
关键澄清

不是"有身体"，而是**"用身体与世界交互并从中学习"**

回顾对比

AlphaGo（离身）：没有传感器，没有执行器，没有与物理世界的任何交互

Atlas（具身）：有摄像头，有电机关节，在每一次交互中学习、调整、进化



INTERMISSION

课间休息

播放Atlas最新演示视频

同学间交流思考题

感知（Perception）：世界的"入口"

从"拍照片"到"看懂"——六类"人造感官"

摄像头 像素眼

最常用的"眼睛"

单目拍平面，双目算深度

模仿人眼的双眼视差

激光雷达 蝙蝠技能

发射激光束测距

输出"点云"——三维数字骨架

误差仅几厘米，不受光照影响

毫米波雷达 穿山甲

穿透雨雾、雾霾

直接测量相对速度

分辨率较低但全天候工作

IMU 内耳平衡感

加速度计 + 陀螺仪

感知加速度、角速度和姿态

手机横竖屏切换就靠它

超声波 海豚技能

成本极低、结构简单

有效距离几米

倒车雷达"滴滴"声来源

GNSS 方向感

GPS + 北斗卫星导航

精度可达米级甚至厘米级

告诉机器"我在哪里"

多传感器融合

当一种传感器失效时，其他传感器可以"补位"

就像人类五感各有分工——蒙上眼睛还能靠听觉和触觉

认知（Cognition）：大脑的"加工"

从"原材料"到"有意义的信息"——认知的四个层次

01 目标检测

在传感器数据中"找到"物体

"那里有辆车"

"那边有个行人"

→

02 语义分割

不仅要"找到"，还要"理解"场景

"这是可行驶区域"

"那是人行道"

→

03 预测建模

基于当前状态预测未来

"行人要过马路"

"那辆车要变道"

→

04 因果推理

理解"为什么"

"信号灯红了→车停了"

"路面有积水→减速"

关键洞察

具身认知 vs 离身认知：认知是"嵌入的"，不是"真空的"——同一个大脑，不同环境，不同"认知风格"

生活案例：你开车时看到前车刹车灯亮了，你的脚自动移到了刹车上——你"思考"了吗？

行动（Action）：身体的"输出"

接一杯水的惊人复杂性

距离估算

杯子约30厘米远

与肩膀齐平高度

→

轨迹规划

不能太快（会打翻）

不能太慢（水会凉）

→

力矩调节

太松了会滑落

太紧了会捏坏纸杯

→

反馈修正

没抓住？调整策略

手指实时感知温度

核心挑战

精确性 vs 鲁棒性

动作要足够精确（抓取鸡蛋）

又要足够鲁棒（应对地面不平）

"戴着镣铐跳舞"

电机有最大转速和扭矩

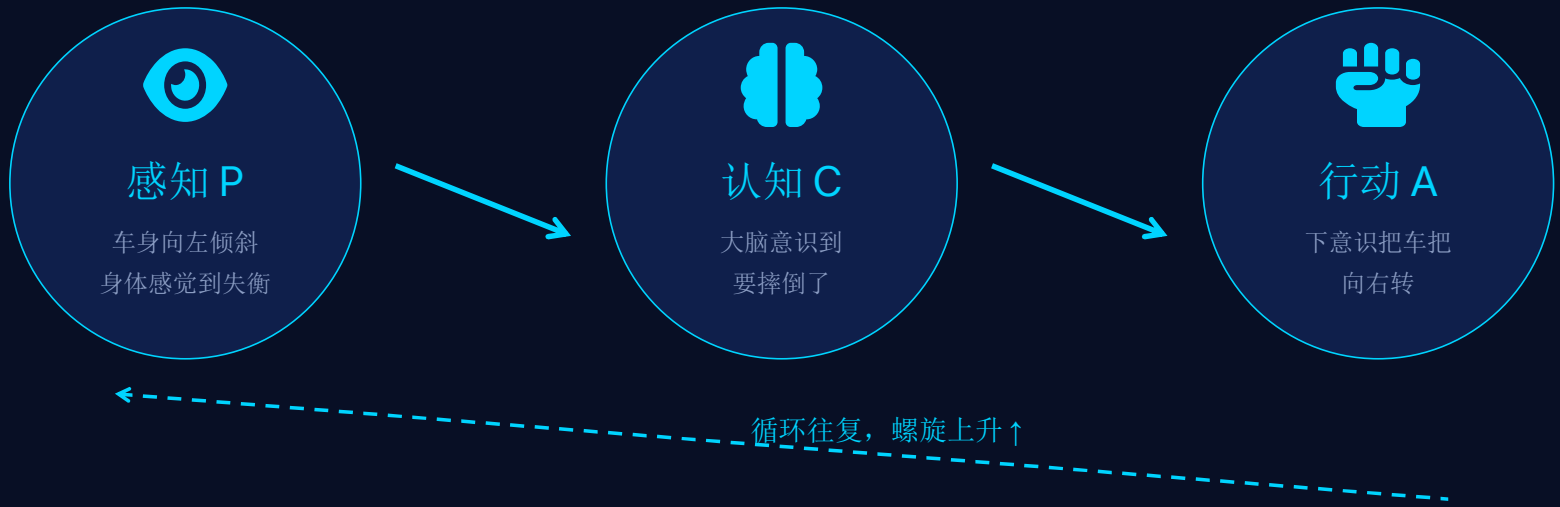
电池有最大输出功率

好的系统学会在约束中找到最优解

思考：你伸手拿桌上的笔时，能感觉到自己在"计算"轨迹吗？这些计算已经变成了"自动化"的技能

PCA闭环：感知-认知-行动的"共舞"

不是流水线，是循环往复、相互塑造的动态系统



三大特征

实时性

毫秒级响应

持续不断运转

不是"先感知完再行动"

耦合性

感知、认知、行动

紧密耦合在一起

你中有我，我中有你

学习性

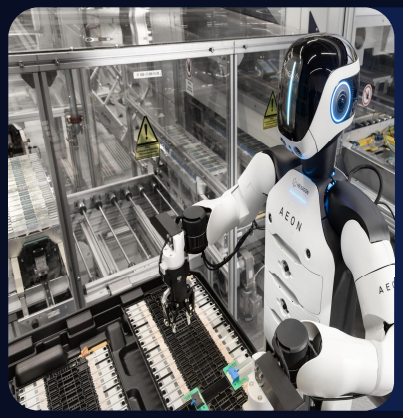
每次运转产生结果

成功或失败

用来更新内部模型

具身智能的"物种图谱"

不同身体 = 不同"世界版本" = 不同智能形态



人形机器人

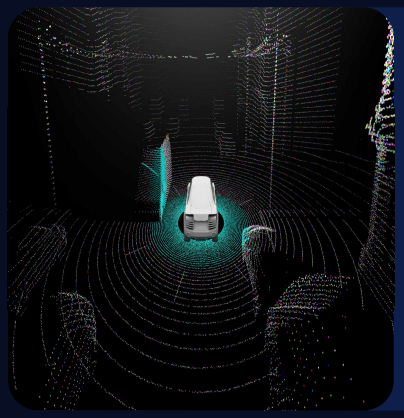
Atlas、Figure 02、宇树H1

特斯拉Optimus

为什么是人形？

人类世界的基础设施

是为人类身体设计的



自动驾驶汽车

"轮式机器人"

商业化程度最高

特斯拉FSD数十亿英里

L3级已获法规批准

PCA闭环每天运转数十亿次



四足/多足机器人

波士顿动力Spot

宇树Go2

超越人类的身体

崎岖地形行走

废墟搜救、管道巡检

软体与微型机器人

柔性材料模拟生物运动

微观尺度工作

另类的"身体"

血管药物输送

精准授粉、精密检测

第1章核心总结

一个核心思想

身体决定心智，交互产生智能

2 两个关键对比

离身智能：AlphaGo

超级大脑，无身体

具身智能：Atlas

灵活身体，从体验中学习

3 三大哲学命题

身体结构规定认知范围

身体结构决定认知机理

身体-认知-环境统一体

1 一个系统框架

PCA闭环

感知 → 认知 → 行动

循环往复、相互塑造



一个物种图谱

人形 · 轮式 · 四足 · 软体 —— 不同身体，不同智能形态

你不需要懂数学公式，不需要会写代码

只需要记住：**身体决定心智，交互产生智能**

课堂讨论：智能是否必须依赖身体？

分组辩论

正方 **智能必须依赖身体**

- 没有身体就没有与物理世界的交互
- 没有交互就没有真正的理解和技能
- AlphaGo不能回答"今天天气怎么样"
- ChatGPT不能骑自行车

身体 = 世界入口

反方 **智能可以不依赖身体**

- AlphaGo的围棋智能是真实的
- ChatGPT理解和生成语言的能力是真实的
- 智能的本质是信息处理
- 身体只是可选的载体

大脑 = 智能核心



综合思考

未来的"终极智能"应该更像AlphaGo还是更像Atlas？还是两者结合？

下章预告：我们将拆开机器的"感官"——从人眼到像素眼，从激光雷达到多传感器融合



谢谢聆听

THANK YOU

智能不是被"编程"出来的

而是从身体与世界的交互中"生长"出来的

课后任务

阅读专栏1-1、1-2、1-3 | 观看《西部世界》《流浪地球》

预习第2章：机器的"感官"