

生成式人工智能应用实战（慕课版）

第1章：AI 与 AIGC 概述

北京航空航天大学 29系人文与社会科学高等研究院

清华大学 新闻与传播学院、人工智能学院 双聘教授 @新媒沈阳 团队

何静（副教授）

（一篇科普报告每多一个公式读者数量就减半。该报告仅用于科普，面向对象为小白。）

AI是什么？

AI：人工智能（Artificial Intelligence），指让机器具备像人一样思考、学习和决策的能力。通俗来讲，它就像给机器装上了一颗大脑，让他们不再是简单的工具，而是能理解我们，陪伴我们，和我们一起学习的伙伴。

感知

人类用眼睛看、耳朵听、手触摸等等来感受外部世界
AI可以用摄像头“看”、麦克风“听”、传感器“触摸”

学习

人类通过模仿、练习、总结来获得经验和知识
AI通过分析海量数据来学习，“消化”的数据越多，就会变得越“聪明”

- 人学认猫：看几只猫，老师讲一讲，就记住了
- AI学认猫：看10万张猫的图片，从中总结“猫的特征”

这4步很像人类的大脑
“从看到 → 到思考 → 再行动”的过程

决策

人类面对选择会“判断利弊”然后行动
AI也能基于计算结果“做出最优选择”

- 比如导航系统根据路况决定推荐哪条路线
- 推荐系统根据你行为，决定给你推什么短视频

推理

人类能分析信息、做出判断，比如“这动物是不是猫？”
AI会根据已有的数据和模型，计算出最可能的答案

- 比如 AI识图时，判断这张图更像“猫”还是“狗”

日常中的案例

➤ 人们用眼睛看，AI用摄像头“看”

手机能认出照片里的人是不是你

➤ 人们用耳朵听，AI用麦克风“听”

你说“嘿Siri”手机就会回答你

➤ 人们学习知识，AI从数据中“学习”

让AI看10万张猫的照片，它就能学会认猫

➤ 人们思考决策，AI计算后“做决定”

地图APP帮你选出不堵车的最快路线

AI 是如何学会的？

AI 并不是程序员写死一堆规则，而是通过“学习”来获得“智能”



机器学习靠人“讲重点”，深度学习自己“抓重点”



人工智能 vs 机器学习 vs 深度学习

- 1 **人工智能** 是总目标：让机器变得“像人一样聪明”
- 2 **机器学习** 是达成这个目标的关键方法之一，通过大量数据自己“总结规律”
- 3 **深度学习** 是机器学习里目前最强的技术路线，模仿人脑结构，自动“分层”提取信息

对比点	机器学习（ML）	深度学习（DL）
特征提取	人工设计	AI 自动学出来
适用任务	简单结构化任务（分类、推荐）	复杂任务（图像、自然语言等）
类比	老师教学生	学生自学

举例（识别猫）



喂它猫的图，人还要提前告诉它“猫耳朵尖、有胡须”等特征

举例（识别猫）



喂它猫的图，但不用人告诉特征，它能自动从图像中“学”出这些特征（耳朵、胡须、眼神等）

AIGC是什么？

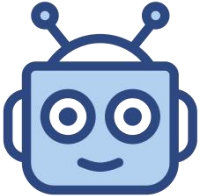
AIGC:AI Generated Content,又称生成式AI,是“AI生成内容”的意思。
它是指人工智能不仅能“识别”、“判断”,还可以主动“创作”出新的内容,比如文字、图像、音乐,以及视频。

深度学习的两种模式

判别式

生成式

喻为：分类大师
任务：区分 & 判断
典型应用：人脸识别 → 张三 or 李四？
情感分析 → 正面 or 负面？

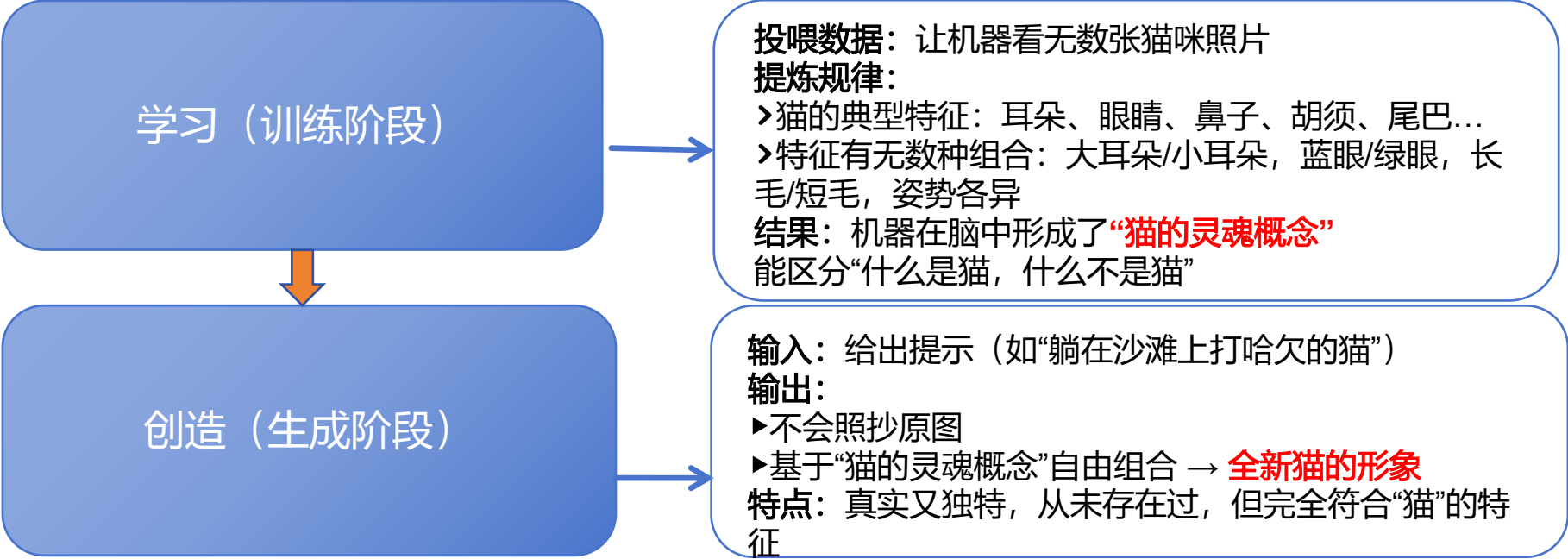


喻为：创造大师
任务：理解 & 创造
典型应用：广告文案生成 → 创意营销内容
医疗辅助诊断 → 给出诊断建议



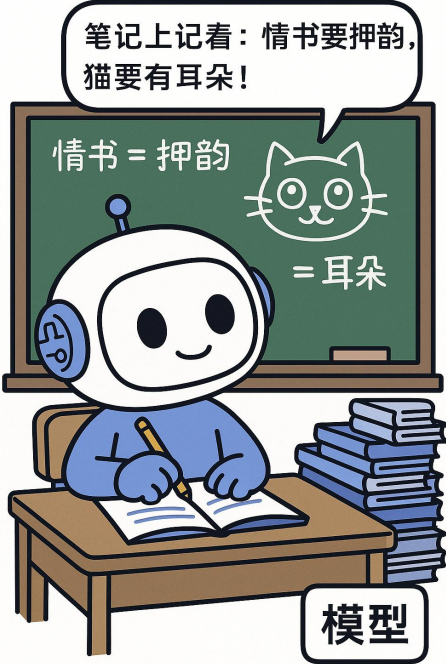
生成内容类型	示例	代表性工具
文字	写作文、写脚本、写代码	DeepSeek、文心一言、ChatGPT
图像	画头像、插画、设计草图	即梦、豆包、Midjourney
音频	AI 配音、朗读文章、变声效果	海绵音乐、网易天香、suno
视频	数字人解说、AI换脸、视频生成	可灵、海螺、sora

AIGC 是如何生成内容的？



AIGC 能生成内容，靠的是“三件法宝”——数据、模型、算法

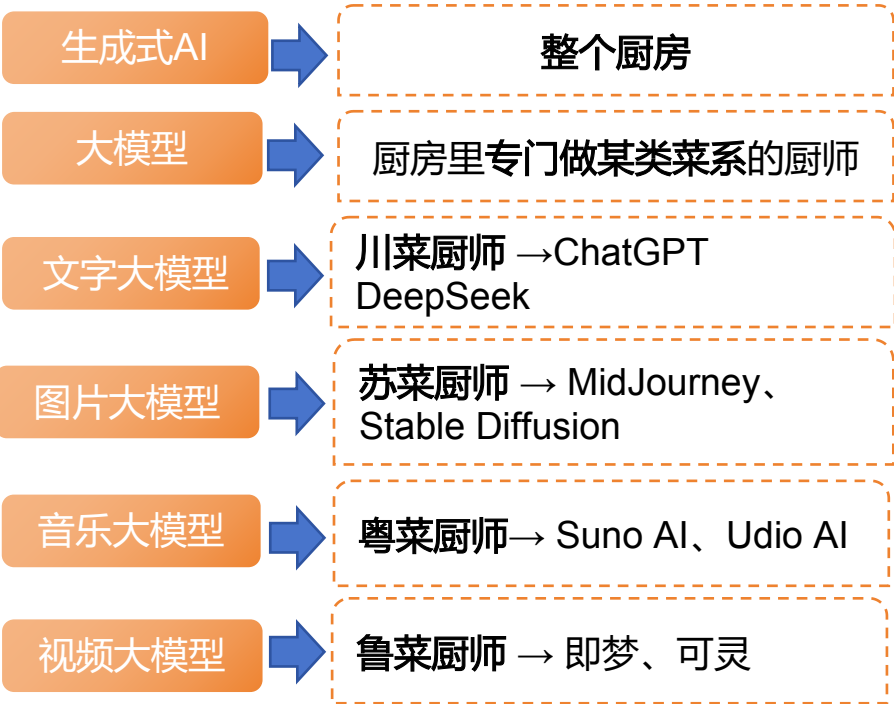
关键要素	作用	解释
训练数据	学习素材	AI 看了海量文本/图像，学会了“怎么说话、怎么画图”
大模型	理解能力	像“看过几百万本书”的AI大脑，能理解你的要求
生成算法	表达能力	像“动笔写作/动手画图”的过程，把想法输出成作品



大模型是什么？

大模型： Large Model。是指在海量数据和算力支持下训练出的、参数规模庞大、具备跨领域泛化和生成能力的人工智能模型。

厨房类比



常见大模型对话产品

类型	代表产品	模型类型说明
文生文	ChatGPT、DeepSeek	大语言模型
文生图	MidJourney、Stable Diffusion	图像生成模型
文生音乐	Suno AI、Udio AI	音乐生成模型
文生视频	即梦、可灵	视频生成模型

注意：对话产品和基座大模型实际上是两个东西

大模型的原理是什么？

大模型类似于一个“超级学霸”。这个学霸的学习和思考过程主要分三步：

第一步

学霸读遍互联网上的书、文章、网页，目的不是“背书”，而是 **总结语言规律**

海量读书
(训练)

- 它会判断每个词跟其它词的关系有多重要
例子：“中国的首都是__” → 北京；“猫喜欢吃__” → 鱼 / 老鼠
- 结果：形成一个复杂的 概率网络（神经网络）

第二步

用户提问 → “为什么天空是蓝色的？”学霸**将问题切分成 Token**（词或字碎片）

接到问题
(输入)

操作步骤	类比说明
把句子拆成 Token	把一句话像拼图一样分成一个个“词块”
遮掉一个词	模型猜这个词应该是什么
猜得准就得分，错就调整	模型通过反复猜 → 更新参数 → 越猜越聪明

Token（词元）是
大模型“看懂语言”
的最小拼图单位。

第三步

像玩“疯狂填词”一样，**逐字逐词预测**

预测下一个字
(生成)

根据问题，选出最合理的开头词→例子：“为什么天空是蓝色的？”→最可能的开头是“因为”
逐字逐词预测下一个最合理的词→例子：“因为”→“太阳”→“光”→“进入”→“大气层”→...
不断调整句子结构，保证通顺和逻辑→最终生成完整答案“因为太阳光进入大气层后，发生了瑞利散射...”

大模型就是一个“根据输入，**预测下一个最可能出现的词**”的超级机器。但，它有时候会犯错（猜错词），或者编造信息（因为它的目标是让句子看起来合理，而不是100%正确），就像学霸有时候也会凭感觉答错题一样。

智能体是什么？

智能体： 英文是Agent，是一种具有**感知、决策和行动能力**的实体，它生活在电脑、手机、互联网里，能理解任务 → 自主分析 → 自动执行

对比维度	一般 AI 模型	智能体（AI Agent）
角色定位	被动工具（像搜索引擎、问答机器人）	主动助手（像私人助理、执行机器人）
工作方式	一问一答、输出结果，完成一小步就结束	拆解任务、制定计划、执行多步行动，直到完成目标
主动性	没有主动性，只能等用户发指令	有主动性，会根据目标自己思考和行动
示例行为	回答问题、翻译句子、生成一段文字	自动写报告、搜索信息、调用日历、发送邮件，一站式完成任务
指令依赖	高度依赖用户提示：你不说，它不做	可以接受一个简单目标，自主规划怎么完成，甚至自己补全信息

案例

用户

请帮我订两张明天晚上7点左右的《奥本海默》电影票，要IMAX厅的，最好靠中间位置。再在开场前1小时提醒我。

今天 14:30

电影助手

已为您预订：

影片：《奥本海默》

时间：明天 19:10 (8月12日)

影院：万达影城(中心店) IMAX厅

座位：8排5座、8排6座 (最佳观影区)

取票码：X5J9K2

支付成功：156元

已设置提醒：明天18:10发送提醒到您的手机

祝您观影愉快！

...

普通程序：需要你手动一步步操作
(查 → 选 → 比 → 买)

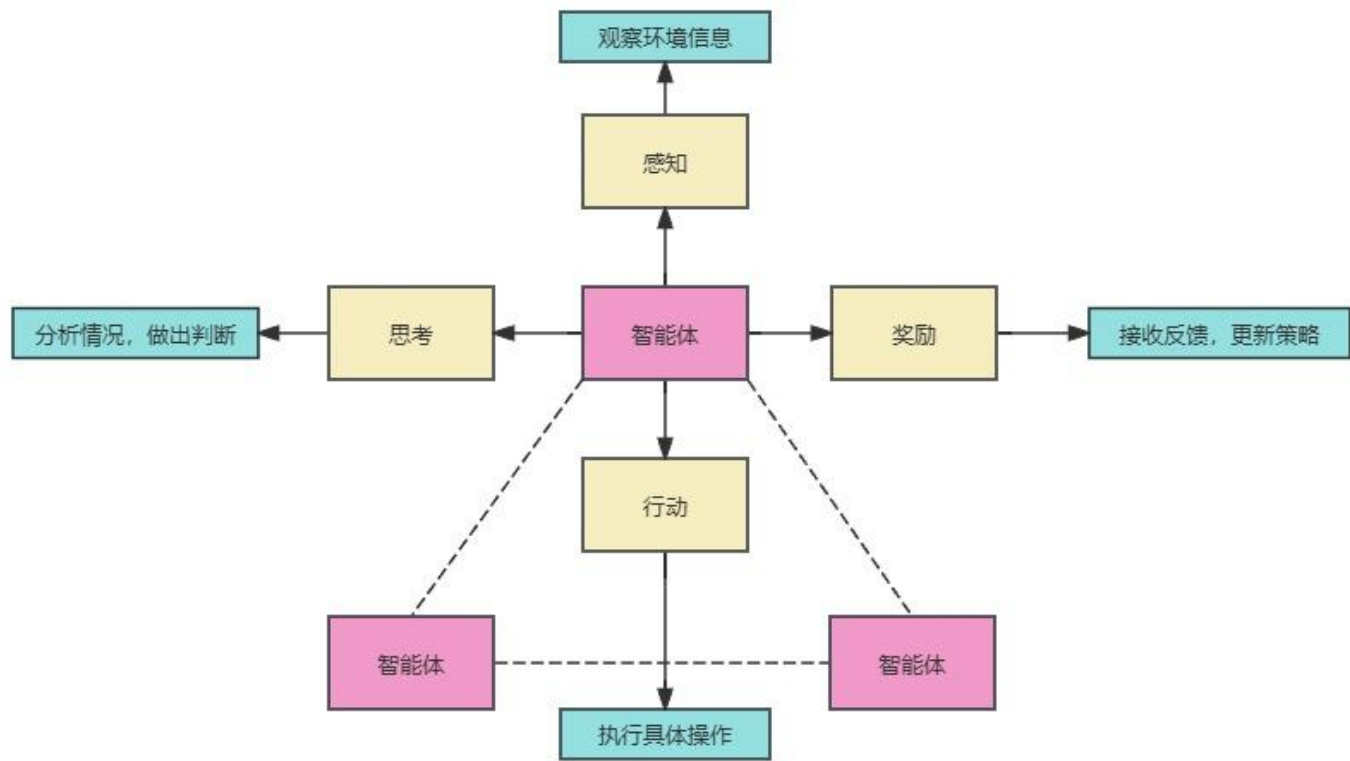
智能体：只需目标指令，**它会全流程代劳**

➢ 能理解模糊需求

➢ 会主动做出判断

➢ 真正做到自动化、智能化

智能体是如何完成任务的？



以“让智能体玩俄罗斯方块”为例

感知

看到屏幕当前堆积的方块形状和下一个即将落下的方块（是长条形的还是正方形的）

思考

“如果把下一个长条形的方块横过来放在这个缺口，就能消除整整四行，得到高分！”

行动

控制手柄，将方块移动到目标位置并旋转。

奖励

成功消除四行，游戏给它加了很高的分数（奖励信号），它就知道这个决策棒极了。

总结

智能体的核心循环：感知 → 思考 → 行动 → 奖励

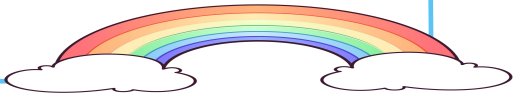
通过不断尝试和反馈，智能体逐渐学会最优策略

从玩游戏到订票、开车、理财，本质都是这个闭环，只是任务复杂度不同

具身智能是什么？

具身智能： Embodied Intelligence，是指拥有“身体”的人工智能，它不仅能思考，还能看世界、听声音、动手动脚、和环境互动。

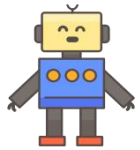
让具身智能玩俄罗斯方块



核心理念



聪明的大脑



聪明的身体

人类理解世界靠的不仅是大脑，还有身体

关键挑战

从复杂环境中提取有效信息（视觉感知）
协调身体动作，精准执行指令（运动控制）
处理物理世界的不确定性与延迟（真实交互）

例如

“眼睛” = 摄像头 → 从屏幕中获取方块信息
“手指” = 机械手指 → 操作手柄完成移动、旋转

它的知识和智能，来自于“眼睛”、“手指”与真实物理设备（游戏机、电视）的交互，而不仅仅是处理抽象的数据。

具身智能 = AI 脑袋 + 感知五官 + 行动四肢



真正走出屏幕、进入现实世界的人工智能！

总结

传统 AI = 博览群书但从不动的“学霸”



具身智能 = 喜欢动手实践、在运动中学习的“高手”

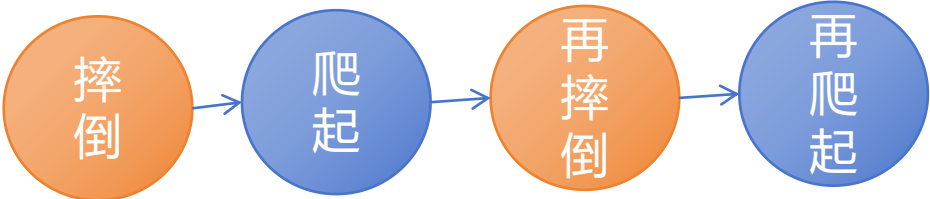
具身智能是如何动起来的？

具身智能的运动规律 = 感知 → 思考 → 行动的循环

通过不断试错与反馈，在真实环境中逐渐学会更聪明、更灵活的动作

类比：小孩学走路，摔跤 → 调整 → 走稳

小孩儿是怎么学会走路的呢？

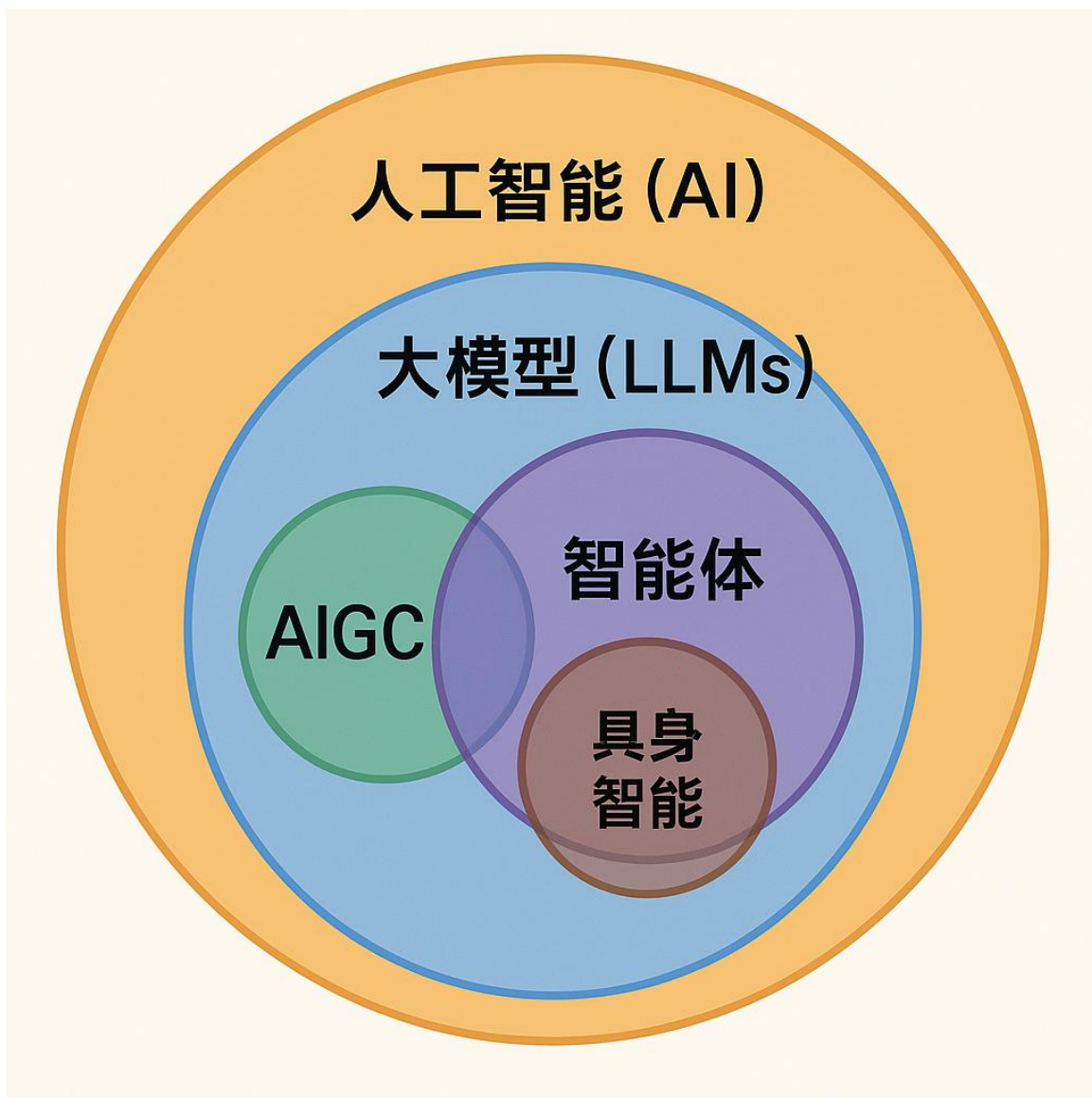


案例：
六足机器人走沙滩



核心技术	作用	详细描述	示例应用
传感器	感知环境信息	通过触觉、视觉、听觉等传感器收集物理世界的的数据，为机器人提供实时反馈	六足机器人感知地面硬度调整步态
强化学习	通过试错提升决策能力	机器人通过与环境互动，基于奖励机制优化行为策略，逐步学会复杂任务	Optimus 在搬运中学习最佳抓取力度
仿生设计	模仿生物结构和行为	借鉴自然界生物的形态和运动方式，设计更灵活、高效的机器人身体和动作	Cheetah 机器人模仿猎豹奔跑姿势
模拟环境	提供安全高效的训练空间	在虚拟环境中模拟现实场景，让机器人预先训练和测试，降低物理实验成本和风险	Octobot 在虚拟管道中优化蠕动路径
深度学习	处理复杂数据并提取模式	利用神经网络分析传感器数据，识别环境特征并预测行动结果，提升智能水平	机器人通过图像识别区分障碍物类别

它们之间是什么关系呢？



关系总结

- **AI 是最大集合**，包含了所有技术与子领域。
- **大模型**是当前 AI 的主力技术范式，它既支撑 **AIGC**，也支撑智能体的发展。
- **AIGC** 是基于大模型的内容生成，是一种应用形式。
- **智能体 (Agent)** 是基于大模型的行动系统，可以调用工具、与环境交互。
- **具身智能**是“有身体”的智能体，能与物理世界互动。

形象化理解

- **AI** 是整片森林。
- **大模型**是森林里长出的最高大的几棵关键树木（基础设施）。
- **AIGC** 是利用这些大树结出的“果实”（生成的内容）。
- **智能体**是森林里活跃的、有自主能力的“生物”（软件或实体）。
- **具身智能**就是那些生活在森林里、能跑能跳、能搬东西的“动物型生物”（有物理身体的智能体）。

清华大学新闻与传播/人工智能学院

双聘教授沈阳作序推荐



978-7-115-68074-7
作者：何静
北京航空航天大学

- 本书以AIGC技能应用为抓手，赋能文案、图像、音视频、智能搜索与数据分析、智能化办公、学习生活等场景，全书无代码应用，适配**无计算机基础的人**学习使用。
- 本书案例丰富，设置了**95个**真实场景的AIGC实训案例，配套74个实训教学视频，**3个大型AIGC综合项目实训**。
- 书课一体，本书配套北京航空航天大学**的线上精品课**（何静教授亲自录制），可同步在线学习。

配套视频教程



北航90后副教授何静

bilibili UP主认证：北京航空航天大学高研院副教授

18.6万

粉丝

22

投稿

21.1万

获赞

bilibili

你感兴趣的视频都在B站

保存图片
打开哔哩哔哩APP
扫码查看UP主



推荐图书

生成式人工智能应用实战（慕课版）

购买
渠道

本书即可作为高校课程教材使用，也适合行业读者自学。

更多图书介绍，可查看本书图书主页：<https://www.ryjiaoyu.com/book/details/55443>

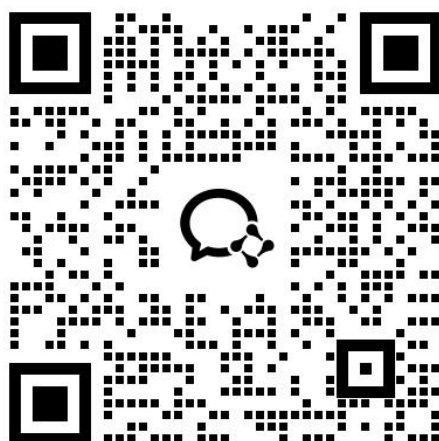


- 认证注册的高校教师，可以扫描下方二维码申请样书。

（仅提供给高校教师身份，需认证！）



扫码申请样书



高校教师服务群（仅限高校教师）

- 行业读者或自学者，可扫描以下二维码购买此书。

