

AI Overview

Xiaohui Huang

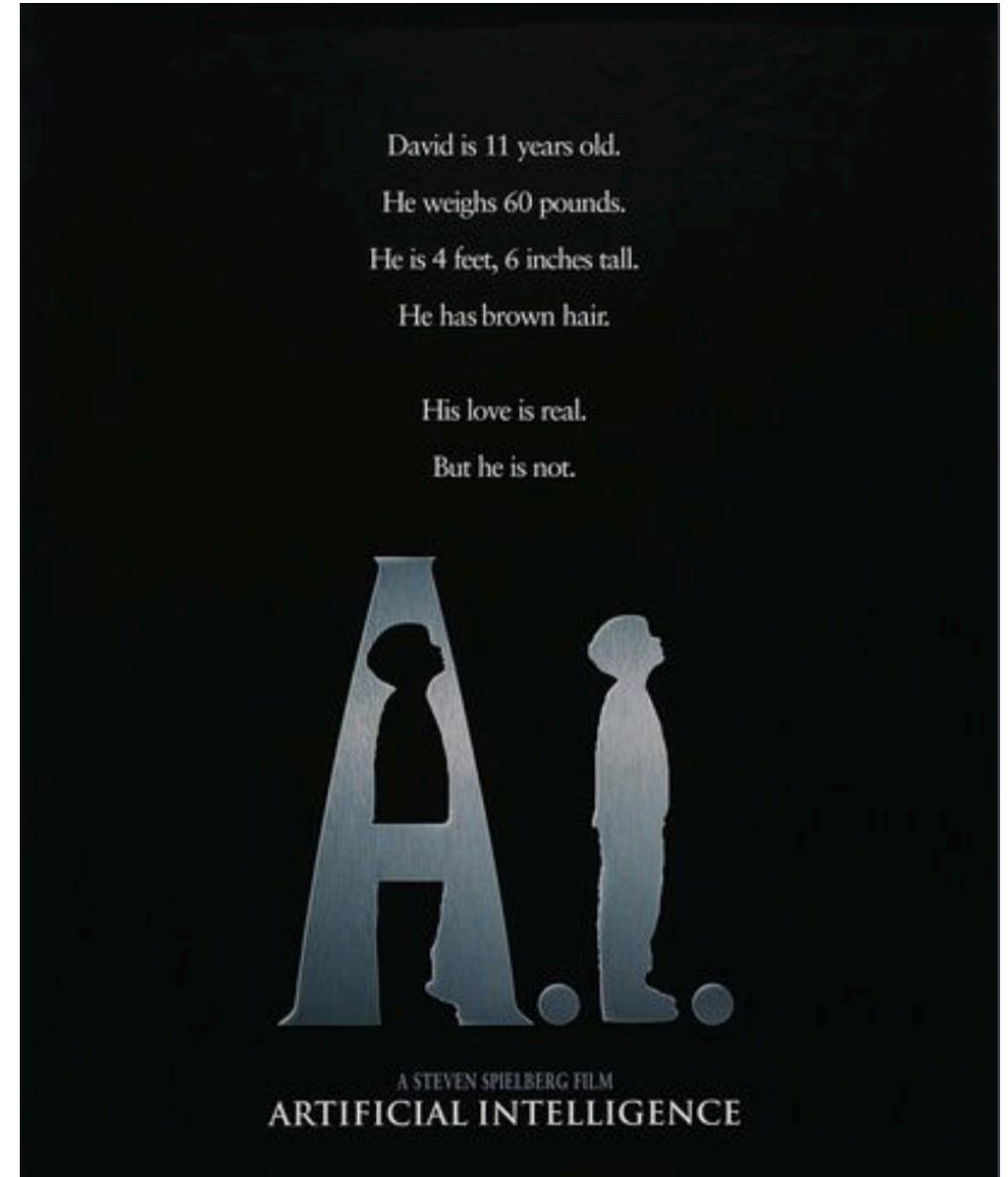
2018.08

"What I cannot **create**, I do not understand." -- Richard Feynman

3 levels of understanding: Read → Tell → Create

Topics

- AI的定义
- 发展史
- AI的学科体系
- 核心思想
- 未来展望
- 总结

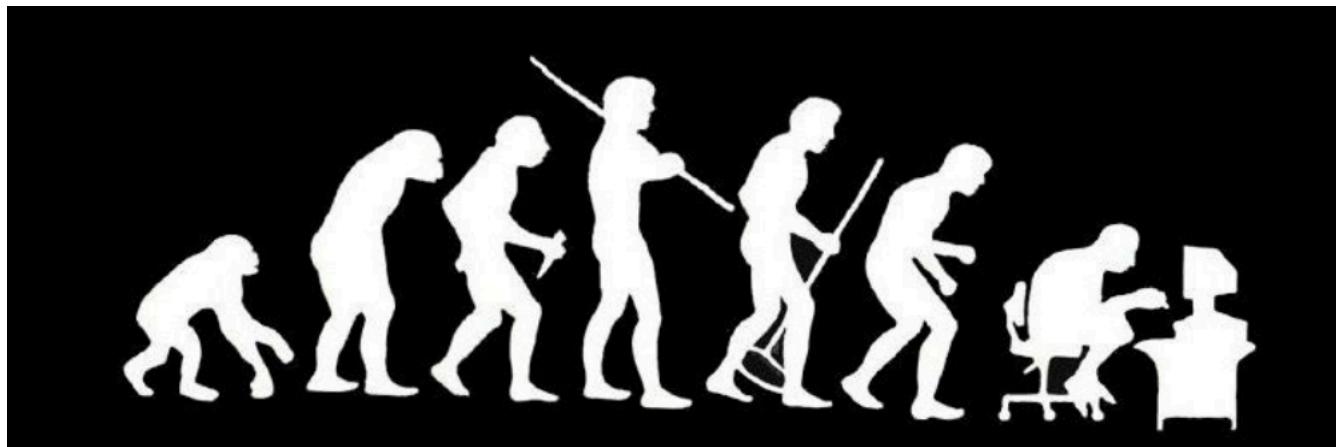


什么是智能？

- Intelligence : A very general mental capability that, among other things, involves the ability to **reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience**. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—"catching on," "making sense" of things, or "figuring out" what to do.
- 根据霍华德·加德纳的**多元智能理论**，人类的智能可以分成七个范畴：1.语言 (Verbal/Linguistic) 2.逻辑 (Logical/Mathematical) 3.空间 (Visual/Spatial) 4.肢体运作 (Bodily/Kinesthetic) 5.音乐 (Musical/Rhythmic) 6.人际 (Interpersonal/Social) 7.内省 (Intra-personal/Introspective)。^[2]
- 智慧 + 能力

为什么会产生智能？

- 由生存本能所驱动，以适应环境为目标，而发展出来的各种能力。
- 智能不能脱离环境而存在。
- 智能必须与环境相适应
- 只有生命才有智能。



智能有大小（高低）吗？

- 智能有很多方面组成，不同的方面可以有大小，但不存在总体智能的可比较性。
- 动物的智能：
- <http://baijiahao.baidu.com/s?id=1573627892561157>



蜜蜂

通过极为复杂的“舞蹈”传递讯息

为什么人的智能看起来比较高级？

- 通过社会分工和协作，来弥补个体的劣势
- 社会分工和协作，需要一套相适应的智能：语言，社交，情感，同理心，领导力等等。（这也是适应环境的必然结果）



智商和智能的关系？

- 智商：先天，基础
- 智能：后天学习和训练

智能有极限吗？

- 思考快与慢：节约资源，提高思维效率
- 最新脑科学的研究发现：人脑已经被100%开发利用

- 如何突破智能瓶颈？
- 基因改造？（有点难）
- 目前主要方式：后天学习，学科分类，专业分工，借助工具-> AI。

什么是AI? – 众说纷纭

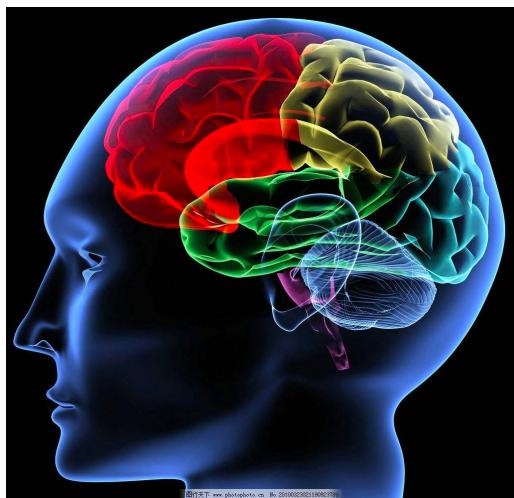
直观的看法	Make computer intelligent
学术界的一种看法	AI = Deep Learning + Reinforcement Learning
从可验证的角度	图灵测试：如果机器能够通过图灵测试，就认为机器是有智能的
从社会现象来看	"AI effect" : AI is whatever hasn't been done yet.

为什么要做AI？

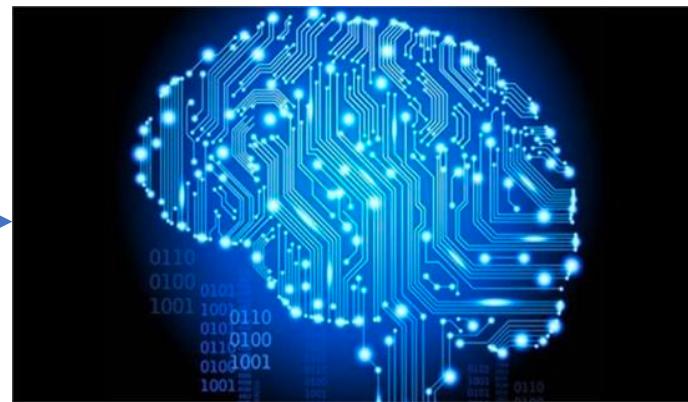
- 提高生产效率，替代人的部分工作？
- 借助外部工具，进一步提高人类的能力
- 人工智能的目的决定了其必定是模拟人或动物的智能的很多方面。但不是全部。
- 自然智能的原理，是实现人工智能的重要的参考依据。

AI的总体思路

- 借鉴自然智能，以问题为导向，结合已有的各学科知识，设计解决问题的方法和步骤，并将这些方法和步骤转化成计算机可执行的指令。



借鉴 →



← 问题与需求



古典AI的设计思路

- 把人类所有的知识，输入到计算机中
- 算术 -> 定理证明 -> 逻辑推理 -> 问题求解（搜索） -> 知识库 -> 规则 -> 领域专家知识



古典AI存在的问题

- 与环境隔离（需要人作为中介），不能适应环境的变化
- 需要持续不断的大量的人工操作
- 只能适用于环境稳定的场景，即：问题所需的前置条件是保持不变的。



现代AI 应该如何借鉴人的智能？

复习下人脑的特点：

先天：

数十亿的进化和自然选择，提供了两样东西：

1. 智能的基本结构和可开发的潜力 -> 智商
2. 生存本能，包括：欲望，需求，好奇心

后天：

1. 由本能所驱使，通过对外部世界的观察和交互，分析现象，不断改进对世界的认识，形成结构化的知识。当遇到新的情况时，做出选择和决策，以期获得最大成果。



现代AI 的设计思路

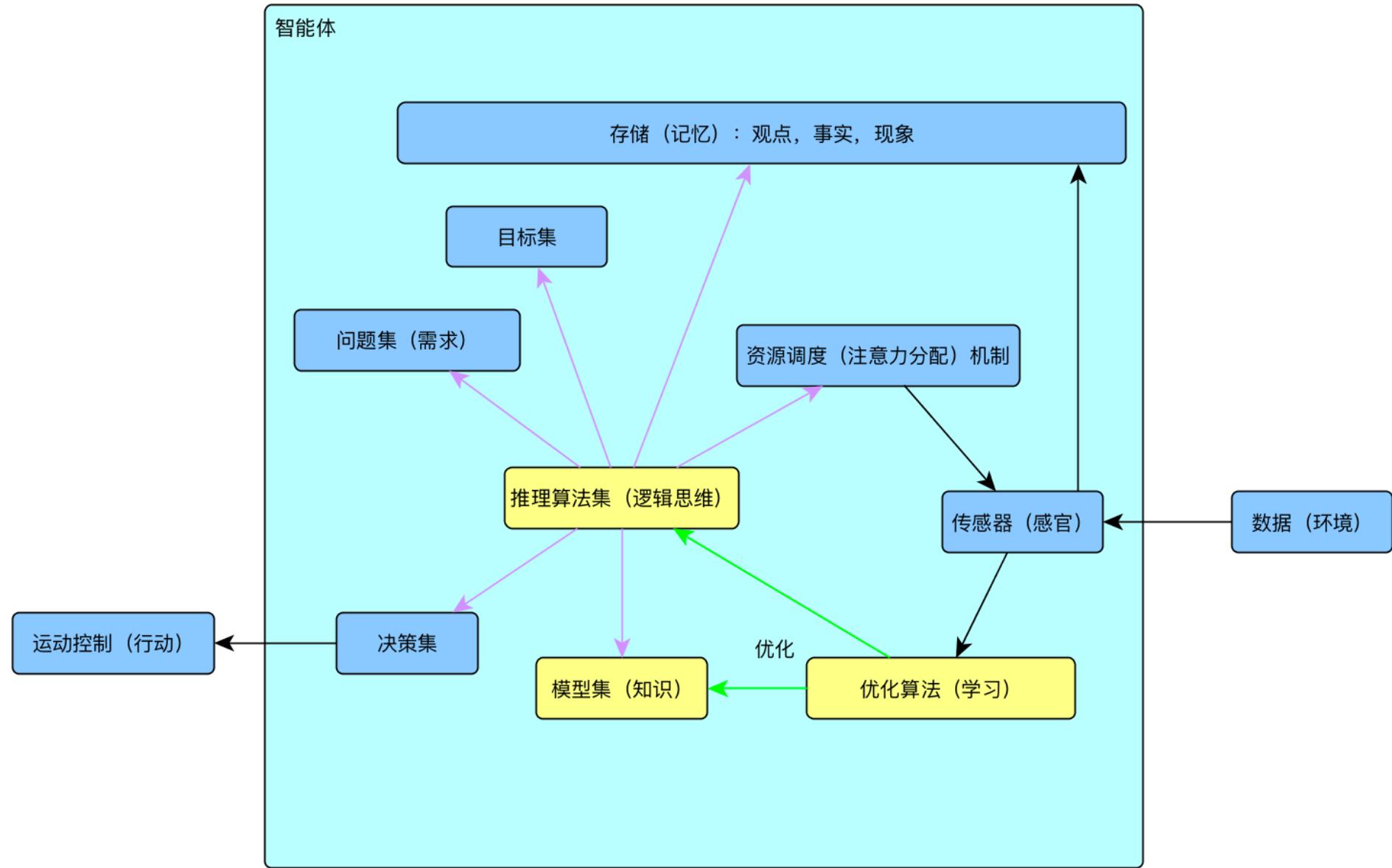
- 1. 先天条件：
 - (1) 计算机硬件（算力、存储） + 模型 + 学习算法
 - (2) 以问题为驱动
- 2. 后天：
 - “通过与外部世界的观察和交互，分析现象，不断改进对世界的认识，形成结构化的知识。当遇到新的情况时，做出选择和决策，以期获得最大成果。”

人的智能	机器智能
观察和交互	传感器，物联网，大数据
分析现象	模式识别，特征提取，表示学习，统计学
改进	最优化理论，贝叶斯定理
世界的认识	模型，数据结构
结构化的知识	图结构，网络，组合泛化
决策	理性决策理论，经济学
以期获得最大成果	最大期望效用理论，概率

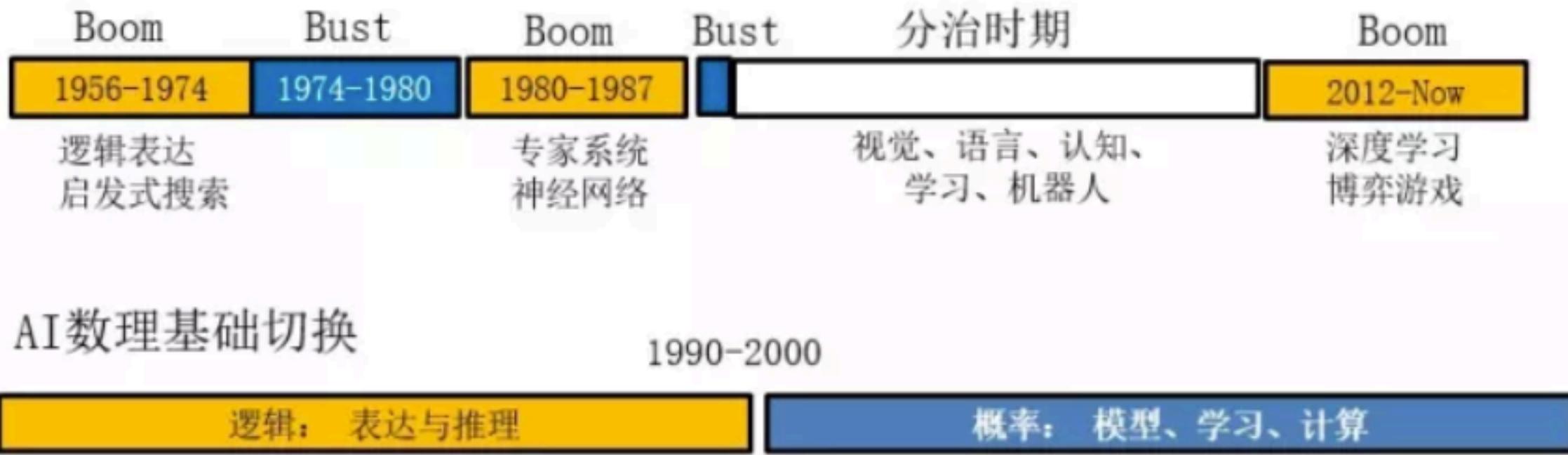
现代AI的难点是什么？

- 不能通过自省来理解学习的运作方式。
 - 原因：人的大脑是如何学习的，是属于无意识层面的机制

AI的组成结构

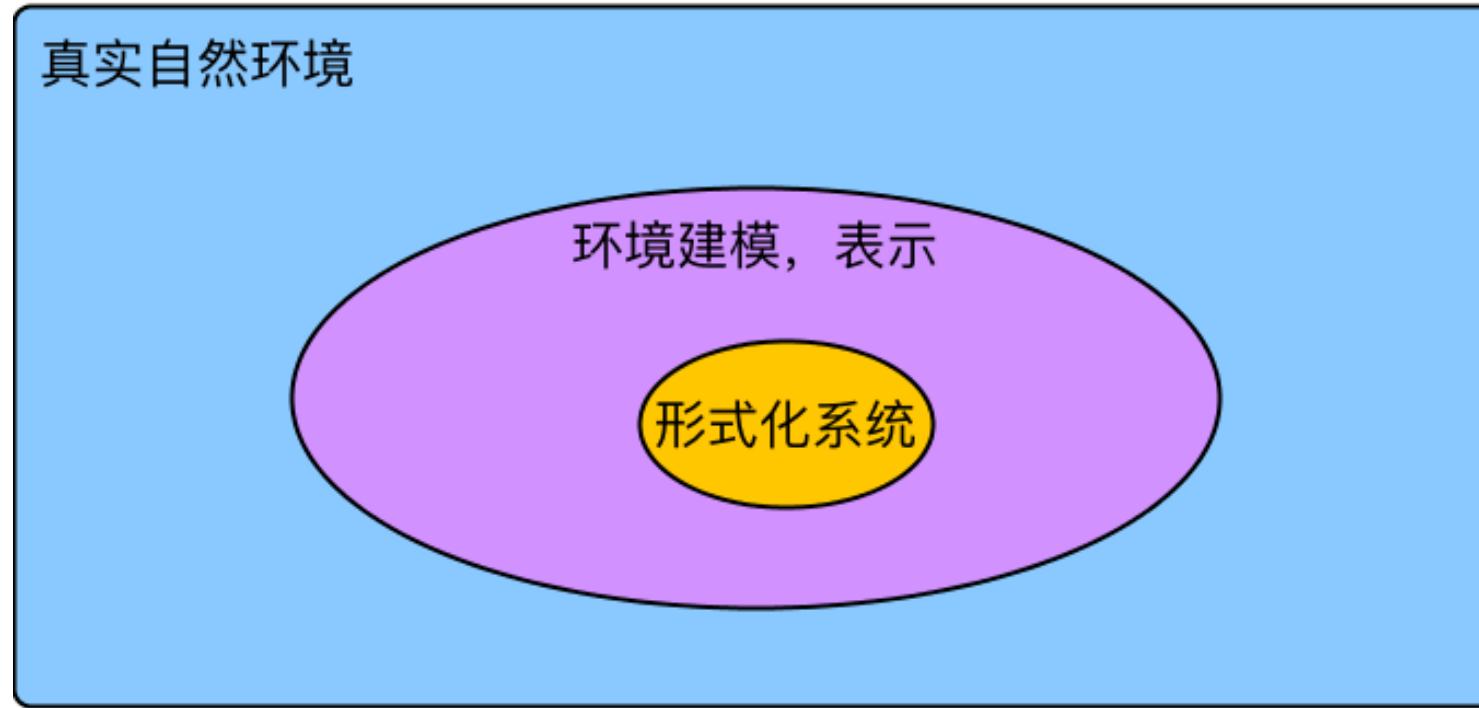


发展史



- **逻辑**：算术，逻辑推理，知识库，定理证明，规则引擎，专家系统，应用程序
- **概率**：统计学习，机器学习，神经网络，强化学习，深度神经网络

为什么这几年AI 变得这么火？



深度学习：大大改善了形式化系统与真实自然环境的交互机制，使得 end-to-end 的AI成为可能。

State of the Art: 深度强化学习

感知 + 认知



AI的学科体系

工程学 (产品化&商业化)

软件工程 机械工程 计算机工程 通信工程 电子工程 材料工程

交叉学科: 整体论

智能科学

计算机视觉 自然语言理解&交流 机器人 认知与推理 博弈&伦理 机器学习

量子计算&通信

认知心理学

认知神经科学

认知科学

物理科学

量子力学 热力学 材料学

生命科学

生理学 遗传学 神经科学 细胞学 动物学

社会科学

心理学 经济学 社会学 语言学 教育学 管理学

形式科学: 模型

密码学

系统论

控制论

信息论

流形

网络科学

决策论

运筹学

最优化理论

数据结构与算法

博弈论

规划

矩阵理论

数值计算

泛函分析

微分几何

随机过程

应用统计

计算理论

图论

概率推理

线性代数

形式语言学

拓扑学

数理逻辑

随机过程

数理统计

布尔代数

数学分析

概率论

应用统计

代数

变化

度量

空间

逻辑

结构

逻辑

度量

空间

哲学

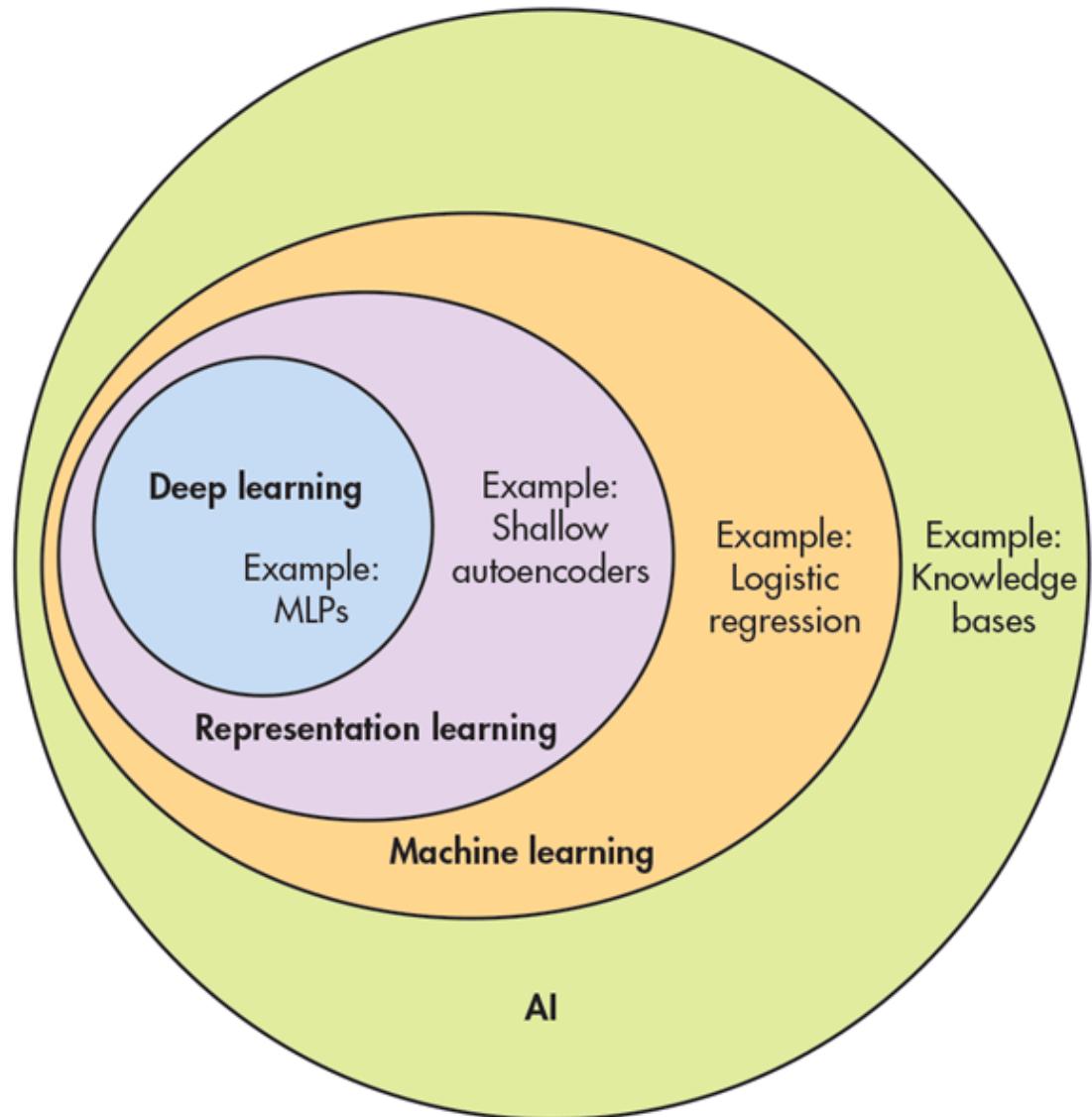
认识论

逻辑学

伦理学

形而上学

相关学科的范畴



核心思想

	思想概述	学科基础
1	“No Free Lunch” 定理	计算理论
2	奥卡姆剃刀	科学思维
3	不确定性的概率表示	概率论, 信息论
4	贝叶斯定理	主观与客观, 认识论
5	所有模型都是错的, 但有些是有用的	世界观, 形式科学, 科学哲学
6	高维空间	线性代数, 几何

未来

- 通用AI (类人)
- 超级AI (超越人)
- 大数据 (机器智能)

通用AI

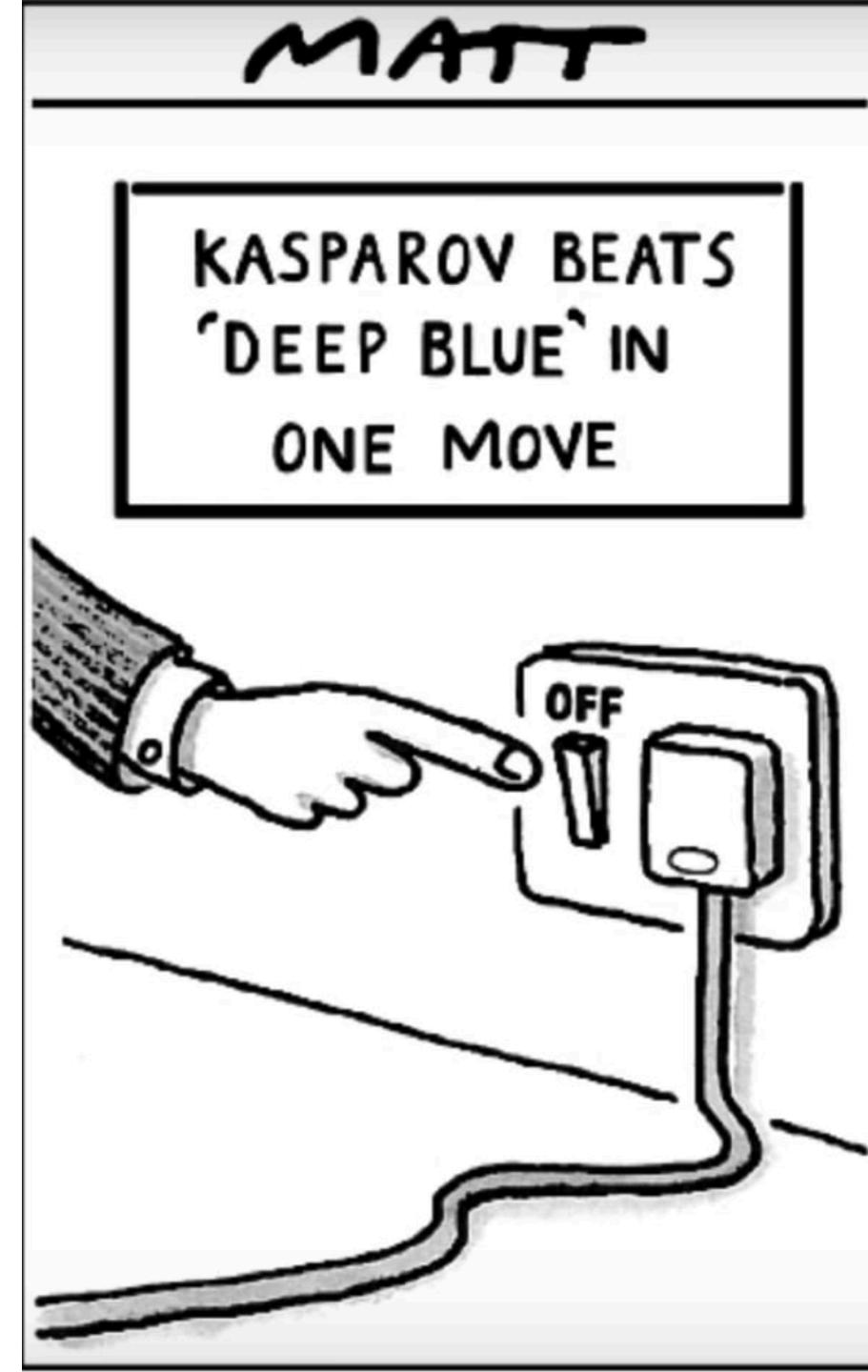
- 学习型（自我更新）
 - 各类模型、算法的集合
 - 自主问题分析与算法调度
-
- 必要性分析：我们并不需要像人一样的通用的智能，我们需要的是在某一方面足够强大，并且具有足够泛化能力的智能。

超级AI

It won't happen...

一篇参考文章 (from 凯文 凯利) : http://www.sohu.com/a/138147780_115035

Ok... If we do succeed,
Don't worry...



大数据（机器智能）：超越个体智能

- 交通路况优化
- 环境监测
- 社会、经济监测
- 智能推荐与个性化定制
- 智能物流

总结

- 智能有很多部分组成，每个部分都体现为对某一类环境的适应能力
- AI当前的首要任务是：在单项能力上达到或超过人的平均水平
- 通用AI、超级AI还只是科学幻想，50年之内都不太可能实现。
- 古典AI不具备泛化（灵活应变）的能力，机器学习致力于解决这个问题

Q&A

附录

- 1. 正式的研究框架介绍
- 2. AI领域相关术语的辨析
- 3. 机器学习的技术体系简介

研究框架

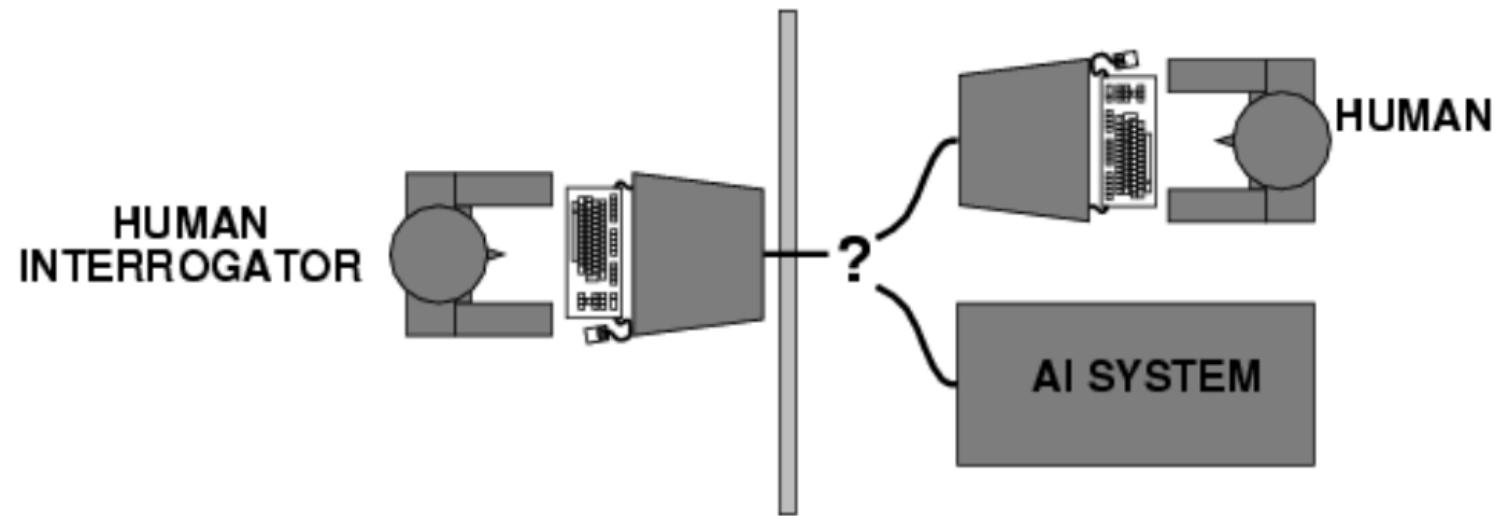
AI的正式的学术定义

- AI is a science concerned with making computers that:

Act humanly	Think rationally
Think humanly	Act rationally

Act humanly

- 图灵测试 (1950)
- “Can machines think?” -> “Can machines behave intelligently?”
- 行为主义 (1900~1950)



- 评论：
- 1. 提供了很好的智能评价框架，但缺乏对工程实践的指导。-> 鸟飞派 VS 空气动力学派
- 2. 随着认知科学、联结主义的兴起，信息论的发展，这一研究思路被抛弃。

Think humanly

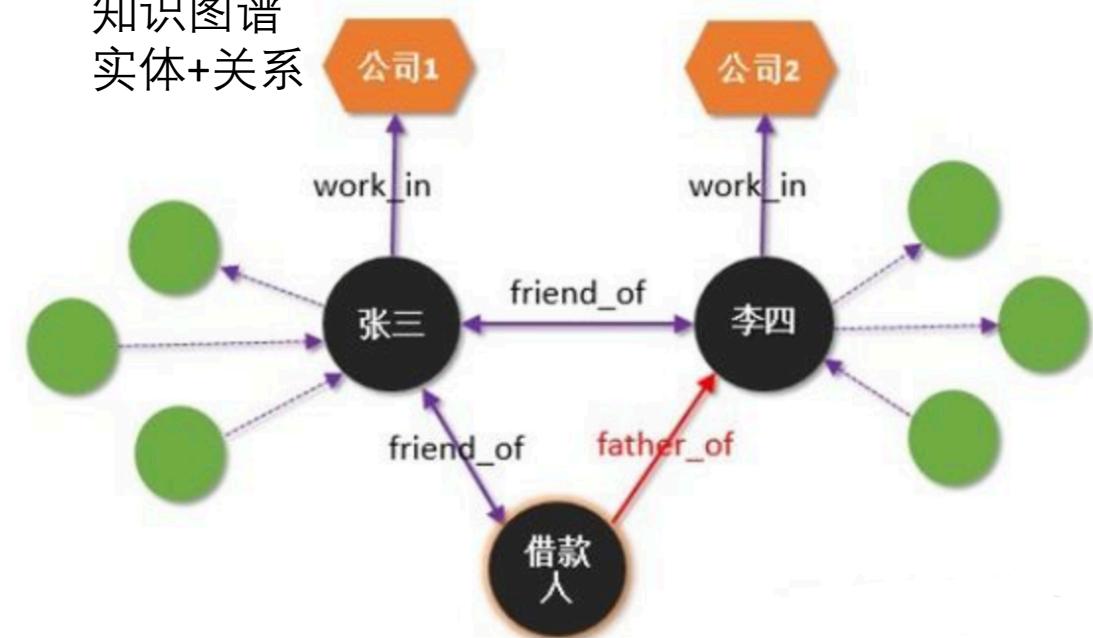
- 研究方法：
 1. 建立认知模型，用信息论的观点来理解大脑的信息处理过程
 2. 通过对人的行为的预测，对大脑活动数据的收集来验证认知模型的正确性
- 优点：人脑代表了最高智能，搞懂了人脑的运行机制，Ai问题就能迎刃而解
- 缺点：
 1. 人脑太复杂，至今未能取得实质性突破
 2. 抽象层次（神经元）比较低，不能够高效地解释人的社会行为
- 研究现状：已经分化成了独立的学科：认知科学，认知神经科学
- 评论：促进了对人工神经网络的早期研究



Think rationally

- 研究方法：
- 1. 研究思维规律，逻辑推理与演绎的正确性和有效性
- 2. 以形式科学（数学）+哲学作为工具，直接构建智能体
- 优点：直接跳过了对人脑的研究，**以人类抽象思维和知识的最高成就，以确定性的指令的形式直接指导机器的行为。**
- 缺点：
- 1. **显式的逻辑思维不能够解释所有的智能现象，例如：人脸识别，语言交流等**
- 2. 抽象层次太高，导致不能做出实质性的、基础性的改进
- 研究成果：定理证明程序，微积分计算，专家系统，知识库
- 研究现状：变成了软件工程，普遍已经不将其归为**AI研究的领域**
- 评论：对智能的理解过于简单，早期取得了一些成果，但很快就遇到瓶颈而停滞不前。

知识图谱
实体+关系



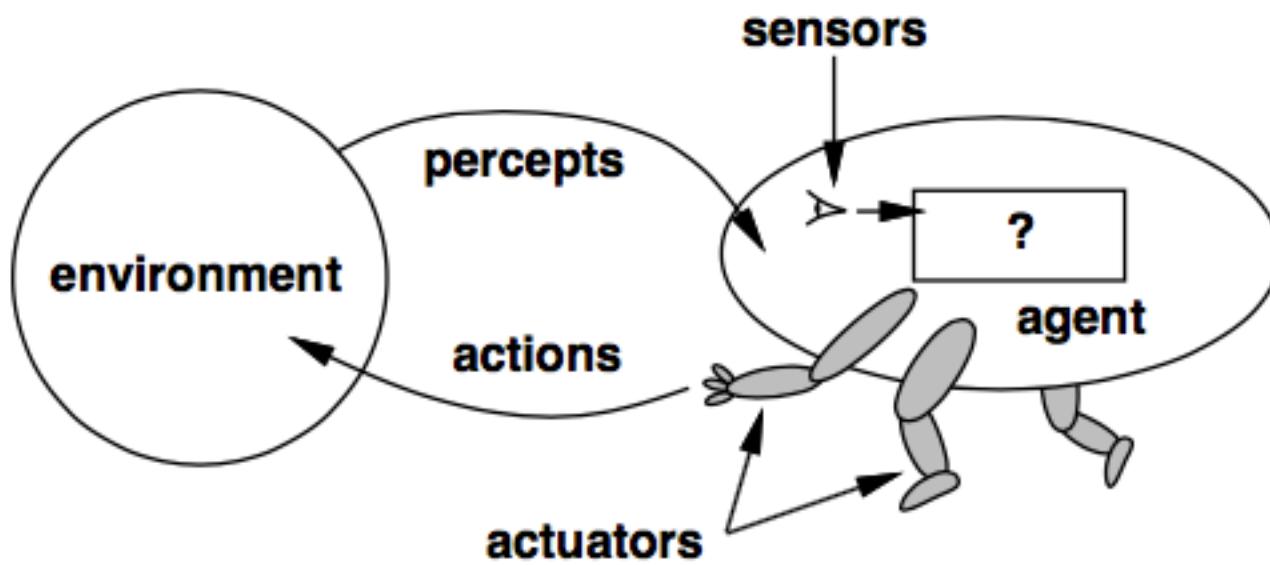
Act rationally

- 理性Agent的假设：做正确的事情，即：在给定可用的信息的条件下，最大化目标的达成。
- 逻辑思考不是必需的，但思考也能帮助产生理性的行为。
- 评论：
- 提供了研究智能的统一框架，即可以是人工智能，也可以是人类（类比于经济学中的“理性人假设”）

基于理性Agent的AI的研究框架

- Agent & Environment
- An intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success
- 表现形式：算法，程序，软件系统，机器人，自然人。

Agents and environments



Agents include humans, robots, softbots, thermostats, etc.

The agent function maps from percept histories to actions:

$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

The agent program runs on the physical architecture to produce f

Environment types

	Solitaire	Backgammon	Internet shopping	Taxi
<u>Observable??</u>	Yes	Yes	No	No
<u>Deterministic??</u>	Yes	No	Partly	No
<u>Episodic??</u>	No	No	No	No
<u>Static??</u>	Yes	Semi	Semi	No
<u>Discrete??</u>	Yes	Yes	Yes	No
<u>Single-agent??</u>	Yes	No	Yes (except auctions)	No

The environment type largely determines the agent design

The real world is (of course) partially observable, stochastic, sequential, dynamic, continuous, multi-agent

- No Silver Bullet!
智能必须和具体的环境、具体的问题、具体的目标相关联！

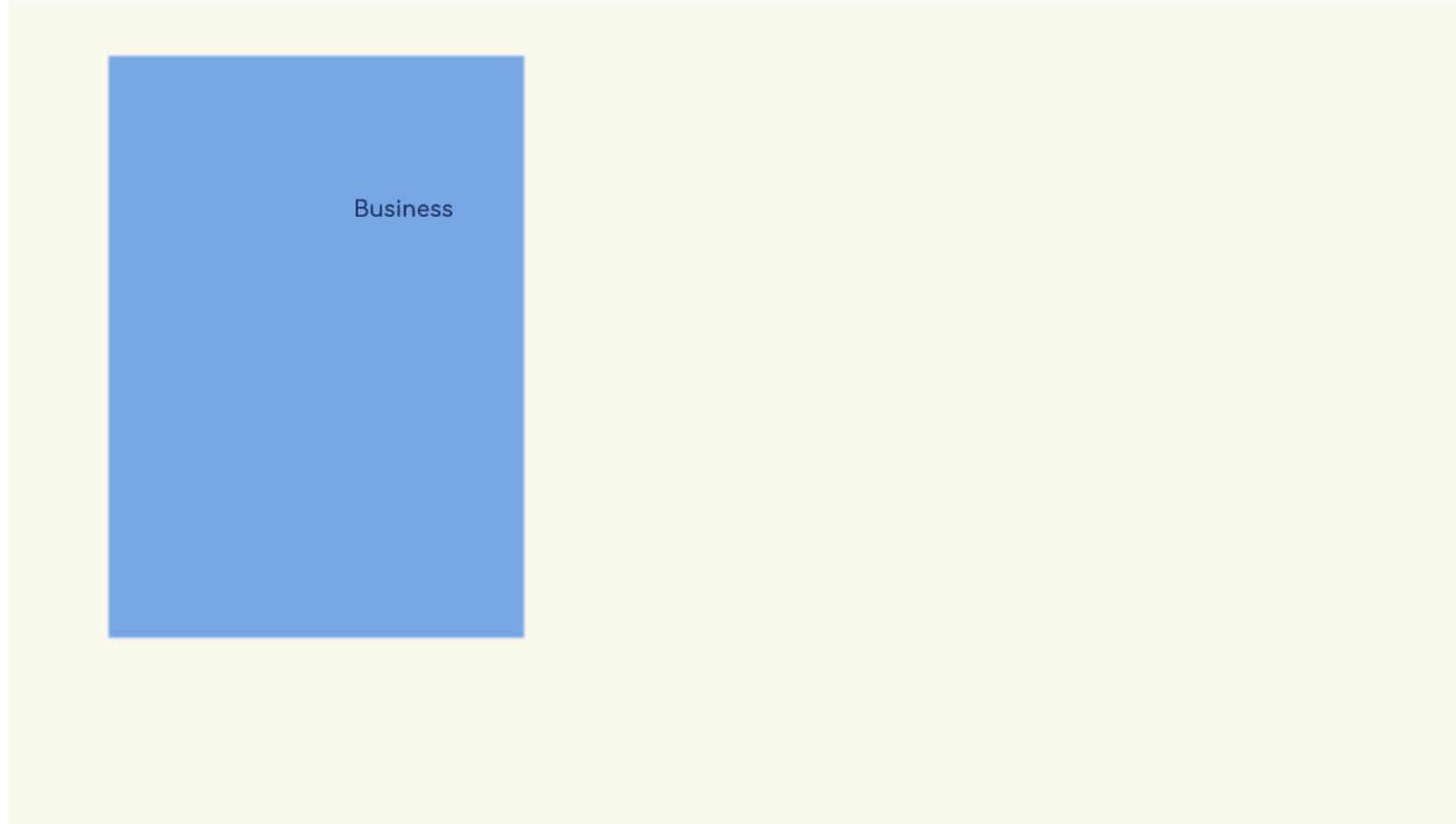
Agent types

Four basic types in order of increasing generality:

- simple reflex agents
- reflex agents with state
- goal-based agents
- utility-based agents

All these can be turned into learning agents

AI相关的领域和术语



<https://www.datascience.com/blog/data-science-vs-business-intelligence-machine-learning-ai>

技术体系

- 以 机器学习 作为 底层支撑技术
- 机器学习三要素：算法 + 模型 + 数据 -> 潜能,智商 + 经验
- 归纳偏置：模型选择， 正则化， 优化算法， 策略。
- 主要模型：图、网络