



南京大學
NANJING UNIVERSITY

人工智能导论

绪论

郭兰哲

南京大学 智能科学与技术学院

Homepage: www.lamda.nju.edu.cn/guolz

Email: guolz@nju.edu.cn

大纲

- 课程信息
- 什么是人工智能
- 人工智能简史
- 本课程内容

大纲

□ 课程信息

□ 什么是人工智能

□ 人工智能简史

□ 本课程内容

课程信息

□ 上课时间：

- ✓ 周三 5-6节 1-17周 南雍-西 209
- ✓ 周五 3-4节 1-17周（单） 南雍-西 209

□ 课程主页：

- ✓ <http://www.lamda.nju.edu.cn/guolz/introai.html>
- ✓ 课程讨论QQ群：881723848



课程信息

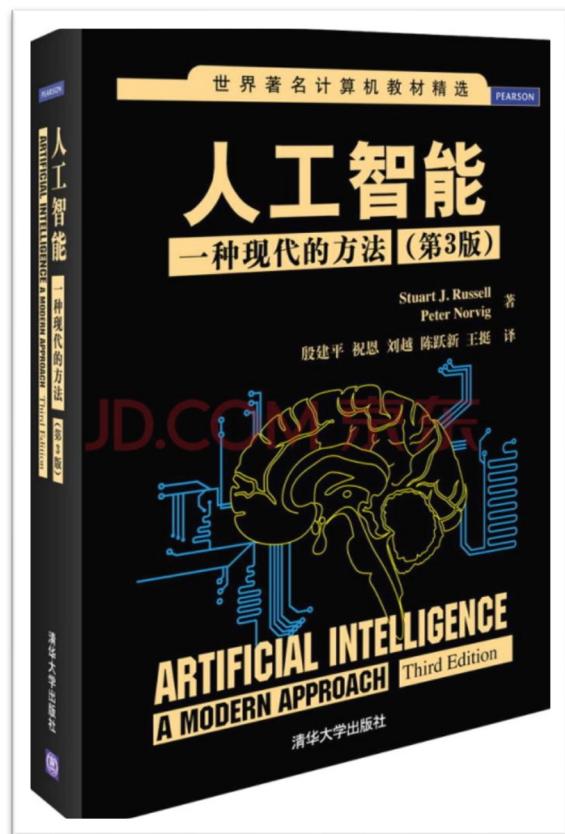
□ 教师

- ✓ 郭兰哲 (guolz@nju.edu.cn)
- ✓ 智能科学与技术学院
- ✓ 机器学习与数据挖掘研究所 (LAMDA)

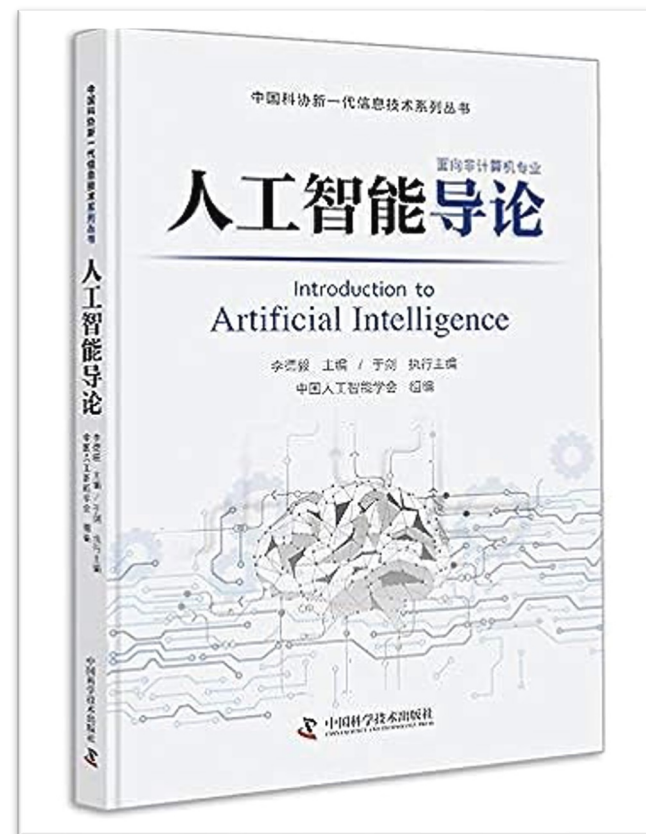
□ 助教

- ✓ 杨骁文, yangxw@lamda.nju.edu.cn

参考书籍



《人工智能：一种现代的方法》
Stuart J. Russell, Peter Norving



《人工智能导论》
李德毅 主编 / 于剑 执行主编
中国人工智能学会 组编

课程考核

□成绩核算：

- ✓ 四次作业，每次10%
- ✓ 课程设计，60%
- ✓ bonus：课堂问答奖励 10分

课程考核

□ 课程设计：

- ✓ 寻找人工智能相关的科研或工程项目，可以通过阅读论文、产业报告、AI相关的竞赛确定题目
- ✓ 创新性30%，完成度30%，答辩情况20%，项目报告20%

□ 建议：

- ✓ 尽早开始，分工合作，各展所长，创新是第一要素
- ✓ 希望以论文、专利、竞赛为目标

课程考核

□ 时间安排：

- ✓ 第一周，自由组队，每组3-5人，选定小组长，**下周三前**由小组长把小组成员信息发送至助教邮箱，**注明组长信息**
- ✓ 第二至三周，组内成员讨论并自拟课题，与老师协商后确认课题
- ✓ 通力合作完成课程项目，交付：
 - ✓ 答辩PPT、项目成果演示（如有）、项目报告
- ✓ 项目答辩（抽签决定顺序）

课程考核

□ 课程设计参考：

- ✓ 游戏AI：棋牌、星际争霸、王者荣耀...
- ✓ AI设计：写诗、绘画、创作...
- ✓ AI助手：医疗、法律、金融...
- ✓ 智能AI客服
- ✓ ...

需能够体现团队的工作和创新，不允许直接抄袭开源项目

学术诚信

1. 允许同学之间的相互讨论，但是**署你名字的工作必须由你完成**，不允许直接照搬任何已有的材料，必须独立完成作业的书写过程。
2. 在完成作业过程中，对他人工作（出版物、互联网资料）中文本的直接照搬（包括原文的直接复制粘贴及语句的简单修改等）都将视为剽窃，剽窃者成绩将被取消。**对于完成作业中有关键作用的公开资料，应予以明显引用。**
3. 如果发现作业之间高度相似将被判定为互相抄袭行为，**抄袭和被抄袭双方的成绩都将被取消**，因此，请主动防止自己的作业被他人抄袭。

大纲

□ 课程信息

□ 什么是人工智能

□ 人工智能简史

□ 本课程内容

什么是人工智能？

John McCarthy（约翰 麦卡锡）

“It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.”

人制造出来的具有智慧的机器，通常通过电脑程序来呈现

Marvin Minsky（马文 明斯基）

“to make computers be capable of doing things that when done by a human, would be thought to require intelligence.”

使机器做那些人需要通过智能来做的事情

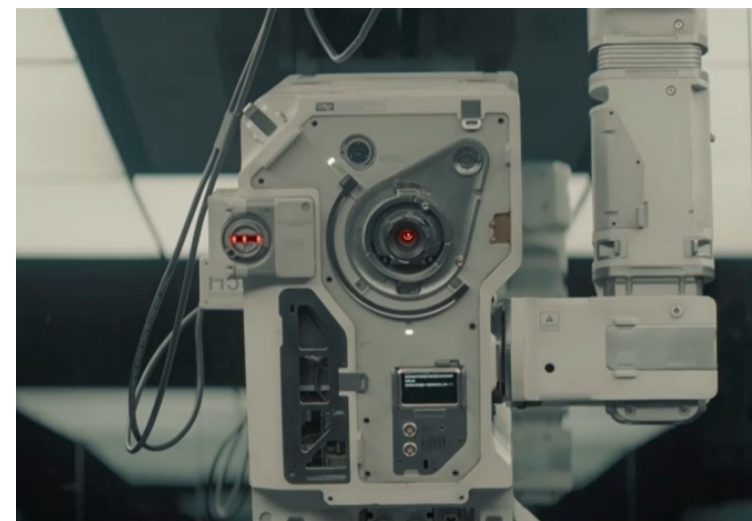
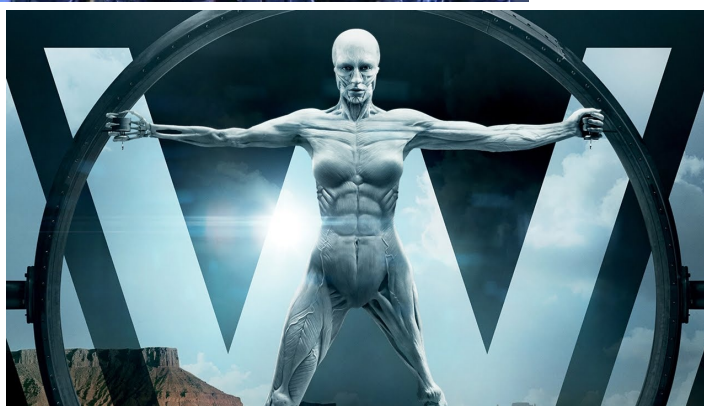
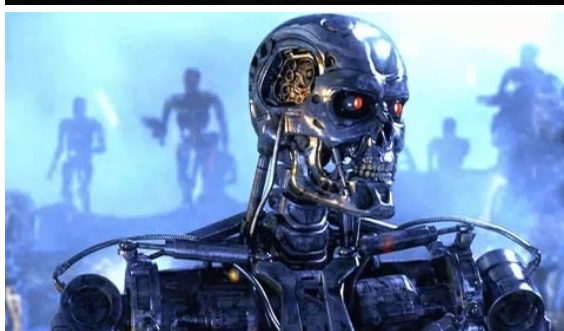
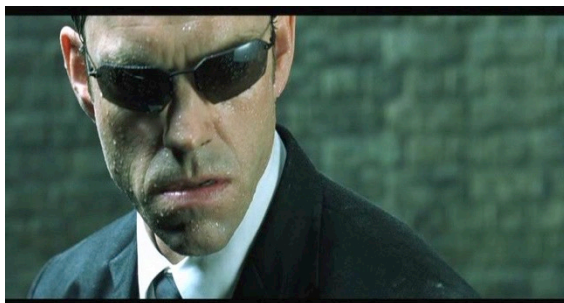


1927-2011

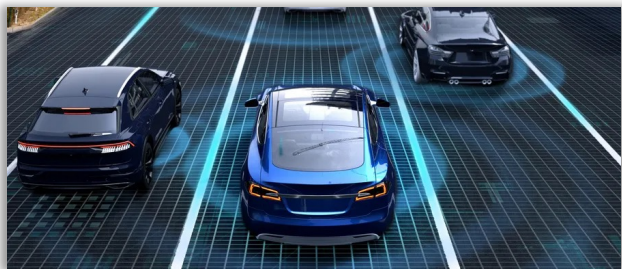


1927-2016

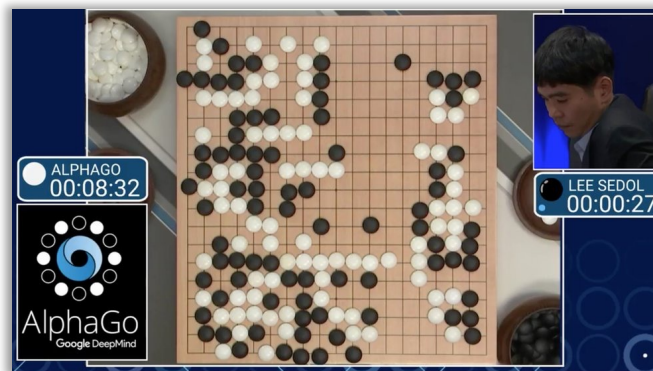
影视作品中的人工智能



现实中的人工智能



自动驾驶



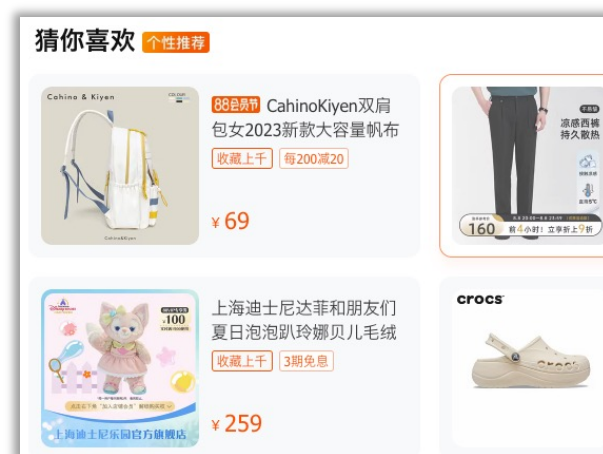
下棋



人脸检测、识别



仿生机器狗



个性化推荐

当前顶尖人工智能系统



AlphaGO

2016年3月，AlphaGo 战胜韩国围棋职业选手李世石（九段）

2017年3月，AlphaGo 战胜中国围棋职业选手柯洁（世界排行第一）

2017年1月初，快棋版本的Master 取得60:0战绩

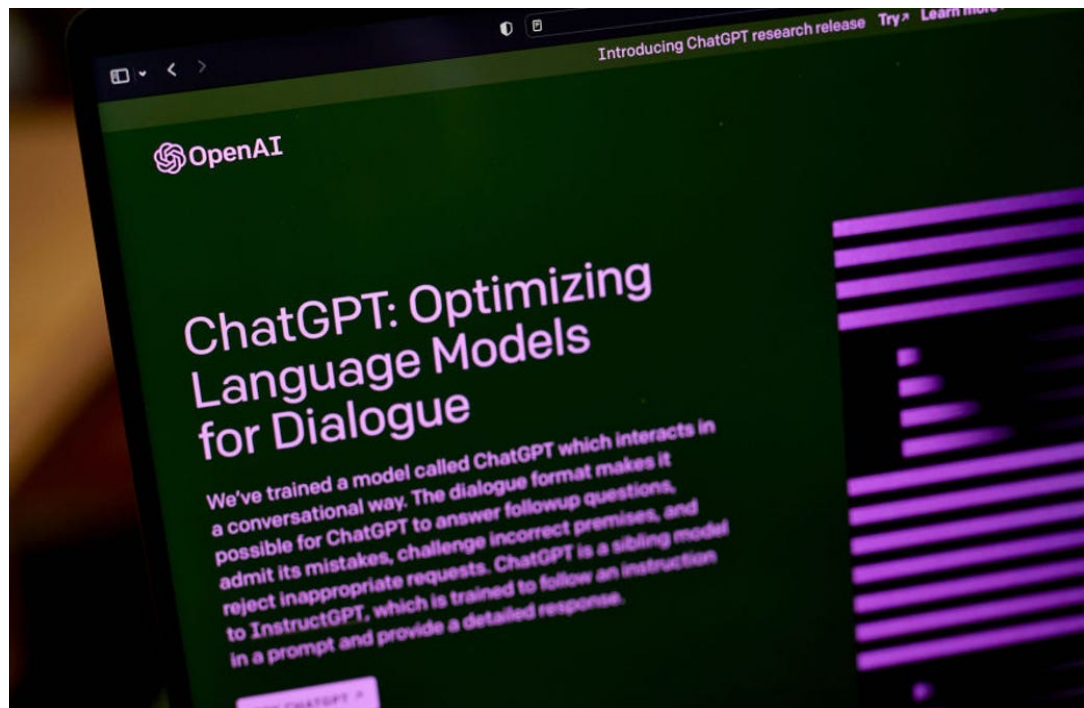
当前顶尖人工智能系统



DeepStack & Libratus

2017年1月左右，在一对一 无限注
德州扑克上大幅赢过职业选手

当前顶尖人工智能系统

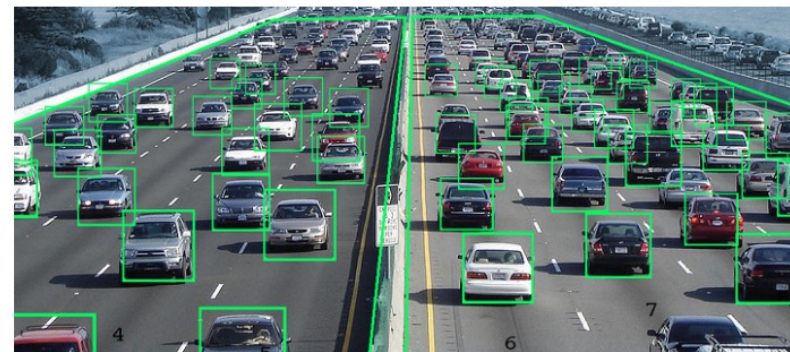


ChatGPT

2022年由OpenAI推出，具有与人类对话、编写计算机程序、创作音乐、诗歌、小说、回答测试问题（在某些测试情境下，水平高于普通人类测试者）等功能

人工智能的应用

- ✓ 自然语言处理
- ✓ 计算机视觉
- ✓ 语音识别
- ✓ 机器人
- ✓ ...

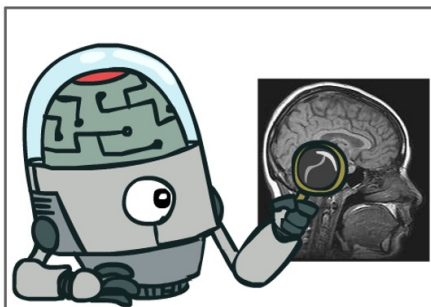


人工智能与其他领域的联系

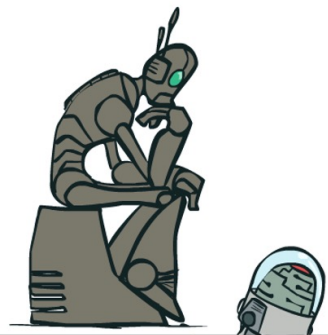
- ✓ 认知科学、神经科学、心理学
- ✓ 博弈论
- ✓ 控制论
- ✓ 哲学、社会伦理
- ✓ ...

什么是人工智能？

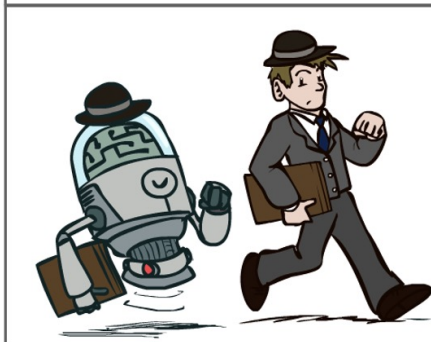
Think like people



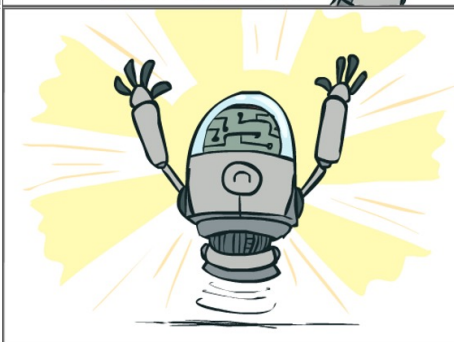
Think rationally



Act like people



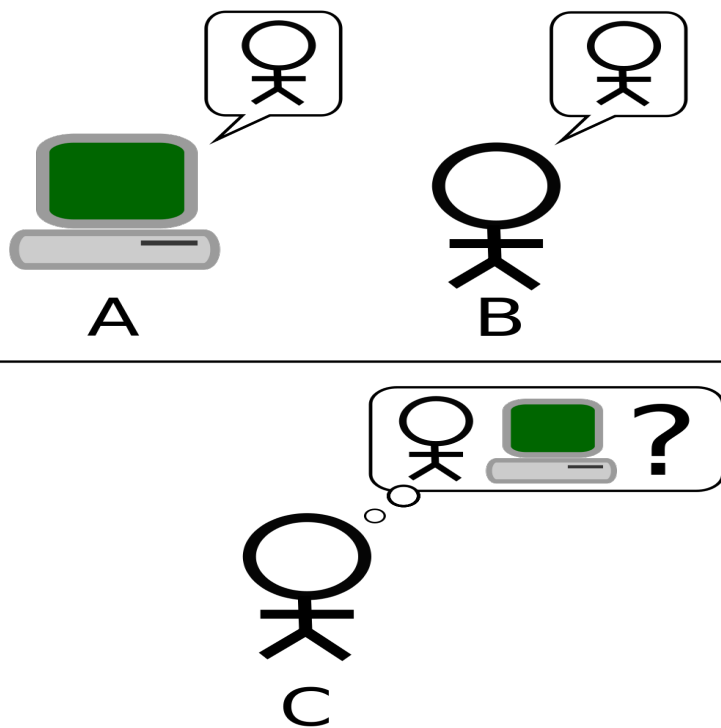
Act rationally



人工智能是计算机科学的一个子领域，与许多其他领域有联系
目标是建造能够理性、智能地思考和行动的机器

图灵测试

如何判断机器是否具有智能？



一人扮演提问者，另一人作为被测人员。
这两个人与机器分别处在3个不同的房间，

提问者通过打印问题和接收打印问题来与
被测人员和被测机器进行通信。

计算机尽量模仿人，如果提问者判断不出
哪个回答是人，哪个回答是计算机，就可
以认为这台计算机具有智能



大纲

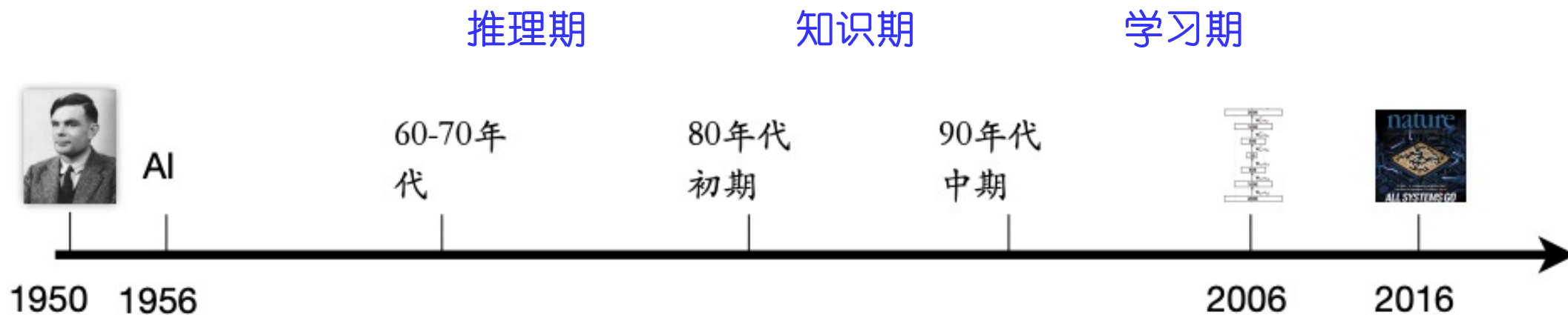
□ 课程信息

□ 什么是人工智能

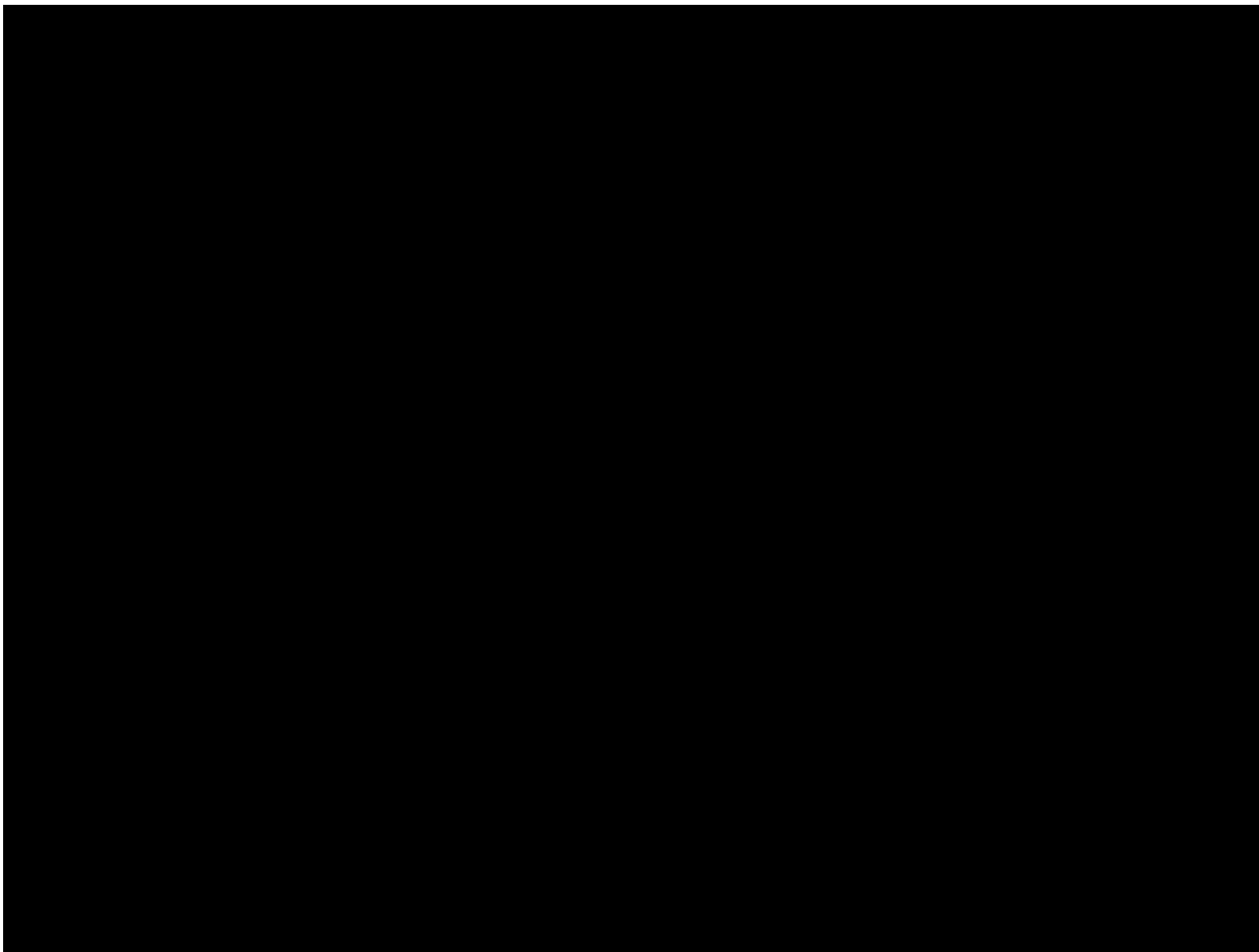
□ 人工智能简史

□ 本课程内容

人工智能的历史



人工智能的历史



人工智能的起源

Computing Machinery and Intelligence

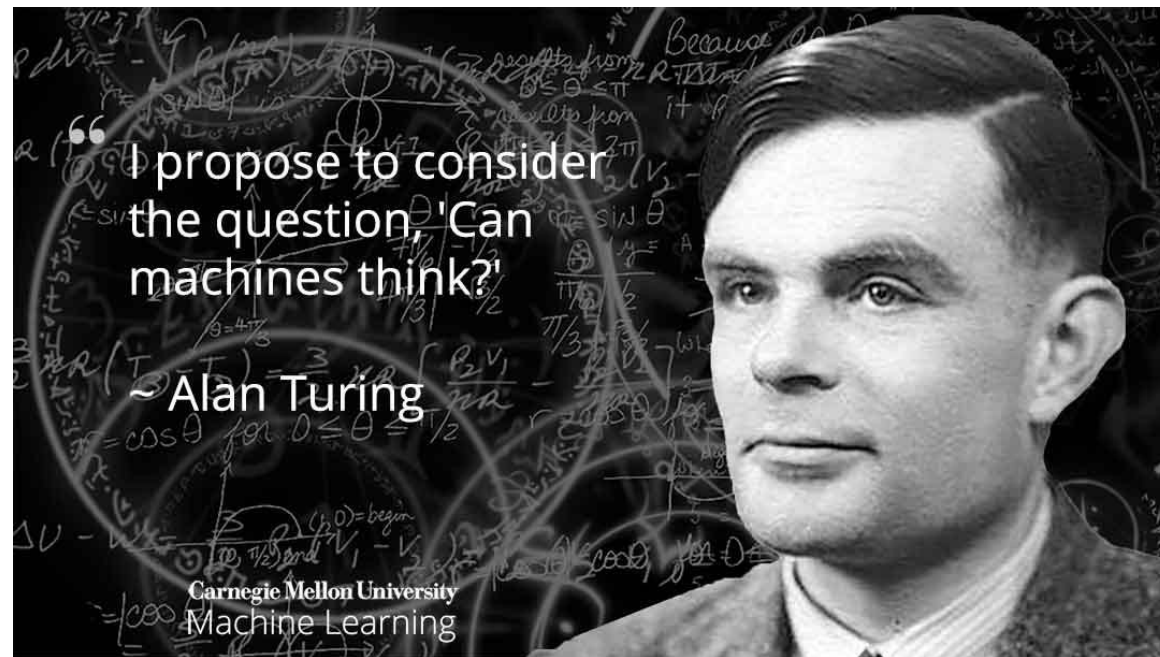
计算机与智能

1950年

艾伦·图灵

“Can machine think?”

机器能思考吗？

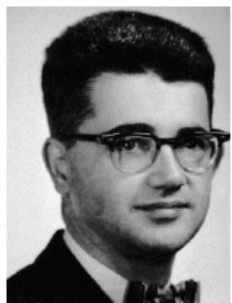


艾伦·图灵

1912-1954

人工智能的起源

1956年的达特茅斯会议标志人工智能这一学科的诞生



J. McCarthy
人工智能之父
图灵奖 (1971)



M. Minsky
图灵奖 (1969)



C. Shannon
信息论之父



H. A. Simon
图灵奖(1975)
诺贝尔经济学奖(1978)



A. Newell
图灵奖(1975)

.....
.....

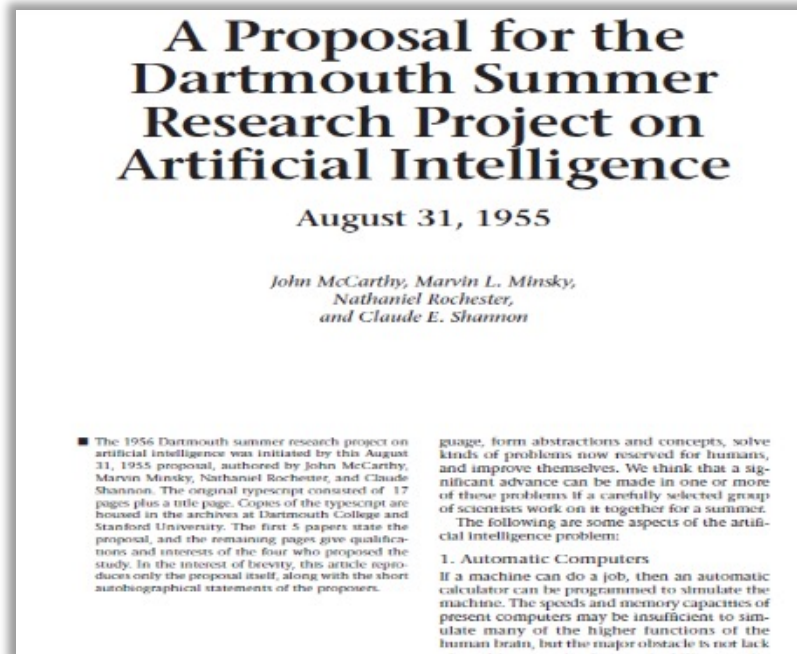


1956年夏 美国达特茅斯学院

人工智能的起源

报告列举了Artificial Intelligence值得关注七个问题

- Automatic Computers
- How Can a Computer be Programmed to Use a Language
- Neuron Nets
- Theory of the Size of a Calculation
- Self-improvement: 自我学习与提高
- Abstractions: 归纳与演绎
- Randomness and Creativity



《人工智能达特茅斯夏季研究项目提案》

人工智能的起源

50年后



第一阶段：推理期

1956-1960s: Logic Reasoning

- ◆ 出发点：“数学家真聪明！”
- ◆ 把人的思考逻辑放入电脑
- ◆ 主要成就：自动定理证明系统（例如，西蒙与纽厄尔的 Logic Theorist 系统）



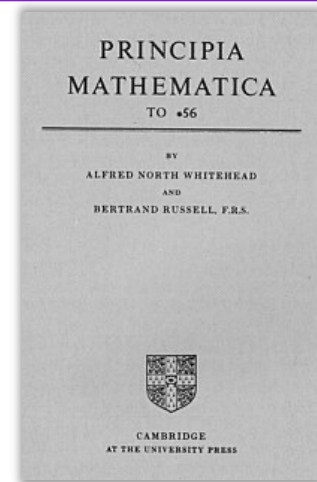
赫伯特·西蒙
(1916-2001)
1975年图灵奖



阿伦·纽厄尔
(1927-1992)
1975年图灵奖

第一阶段：推理期

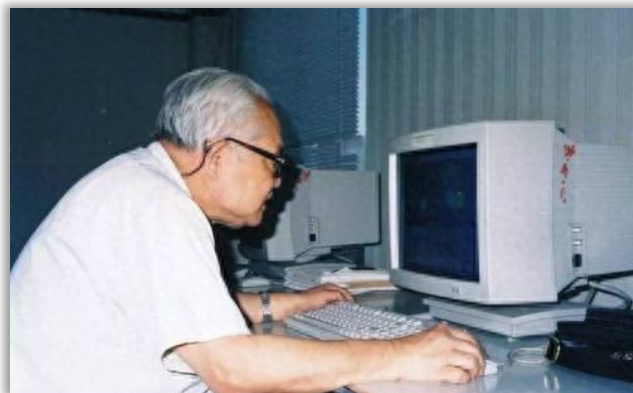
- **逻辑理论家**：能够证明数学原理第2章中的前52个定理中的38个，其中定理2.85的证明实际上比罗素和怀特海手抄的证明更优雅
- **王浩**：在IBM 704计算机上用9分钟计算时间，证明了罗素、德怀特海所著《数学原理》中数百条数理逻辑定理
- **吴文俊**：提出了用计算机证明几何定理的“吴方法”，开创了机器几何定理证明的方向，是国际自动推理领域的先驱性的工作



罗素，怀特海



王浩



吴文俊

第一阶段：推理期

渐渐地，研究者们意识到，仅有逻辑推理能力是不够的 ...

第二阶段：知识期

1970s-1980s: Knowledge Engineering

- ◆ 出发点：“知识就是力量！”
- ◆ 把人的全部知识放入电脑
- ◆ 主要成就：专家系统（例如，费根鲍姆等人的“DENDRAL”系统）



爱德华·费根鲍姆
(1936-)
1994年图灵奖

第二阶段：知识期

“DENDRAL”系统：包含丰富的化学知识，可以根据质谱数据帮助化学家推断分子结构

- ✓ 规划：利用质谱数据和**化学家对质谱数据与分子构造关系的经验知识**，对可能的分子结构形成若干约束条件
- ✓ 生成结构图：给出一些可能的分子结构，利用第一部分所生成的约束条件来控制这种可能性的展开，最后给出一个或几个可能的分子结构
- ✓ 利用**化学家对质谱数据的知识**，对第二步给出的结果进行检测、排队，最后给出分子结构图

形成“知识工程”的方法论，包含了对专家知识从获取、分析到用规则表达等一系列技术。

第二阶段：知识期

渐渐地，研究者们发现，要总结出知识再“教”给计算机系统，实在太难了 …

第三阶段：学习期

1990s-Now: Machine Learning

- ◆ 出发点：“让系统自己学！”
- ◆ 把人的所有看见放入电脑
- ◆ 主要成就： ...

机器学习是作为“突破知识工程瓶颈”之利器而出现的



恰好在20世纪90年代中后期，人类发现自己淹没在数据的汪洋中，对自动数据分析技术——机器学习的需求日益迫切

人工智能的流派

符号主义

认为人类思维的基本单元是符号，而基于符号的一系列运算就构成了认知的过程，所以人和计算机都可以被看成具备逻辑推理能力的符号系统，换句话说，**计算机可以通过各种符号运算来模拟人的“智能”**。

代表性成果：

- ✓ 定理自动证明
- ✓ 专家系统
- ✓ 知识图谱
- ✓ ...

代表人物：

- ✓ 纽厄尔、西蒙
- ✓ 王浩、吴文俊
- ✓ 爱德华·费肯鲍姆
- ✓ ...

人工智能的流派

符号主义面临的挑战

- 概念的组合爆炸问题：常识难以穷尽、推理步骤无穷
- 命题的组合悖论问题
- 经典概念难以得到，知识难以提取

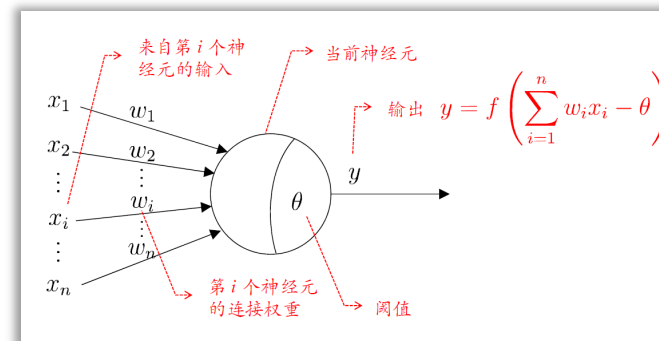
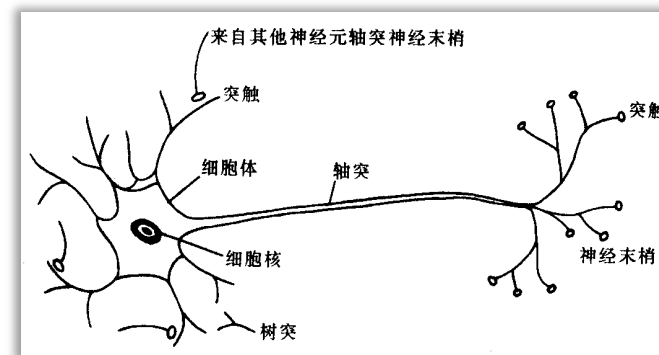
人工智能的流派

连接主义，又称仿生学派或生理学派，源于仿生学，特别是人脑模型的研究

认为智能活动是由大量简单的单元通过复杂连接后并行运行的结果，其原理是神经网络及神经网络间的连接机制与学习算法

代表性成果：神经网络、深度学习

代表性人物：McCulloch, Pitts, Hinton



人工智能的流派

连接主义面临的挑战

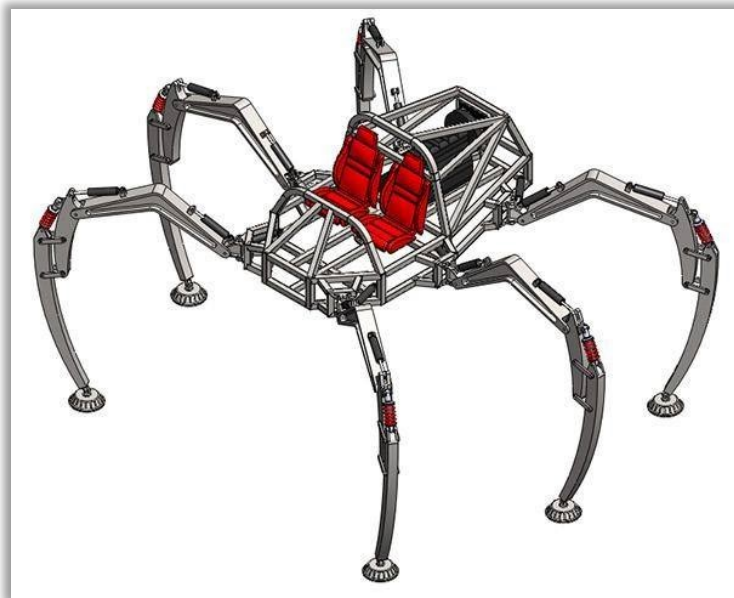
- 人脑的运行机制尚不清楚
- 现在的神经网络和深度学习是基于数学推导而来，实际上与
人脑的真正机制距离尚远，并非人脑的运行机制

人工智能的流派

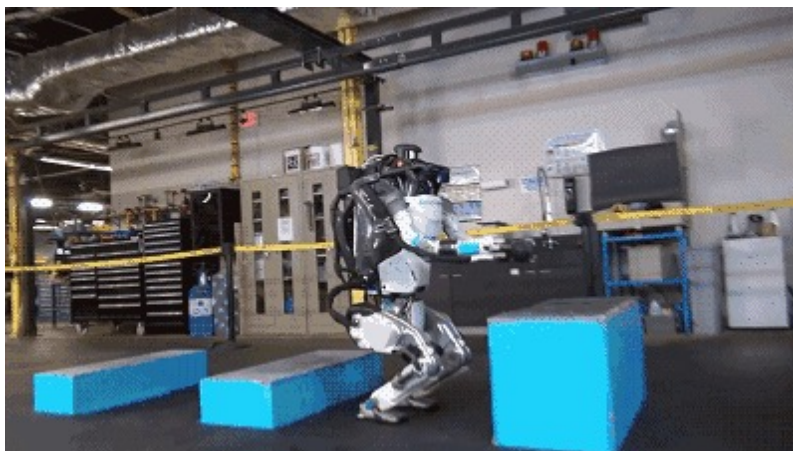
行为主义，源于控制论，关注低级生物智能

认为智能取决于感知和行为，取决于对外界复杂环境的适应，而不是表示和推理

代表成果：六足虫、大狗



行为主义最新进展



行为主义挑战

莫拉维克悖论

人觉得困难的问题, 对计算机来说反而是简单的
人觉得简单的问题, 对计算机来说反而是困难的



- 驾车下坡, 并停在一扇门前
- 下车, 走到门前, 抓住并旋转把手, 并打开房门
- 穿过门进入房间
- 将红色的圆形阀门旋转360度
- 拿起位于墙上搁板上的电钻, 并在旁边的木头墙上钻出一个洞
- 将一个插头拔出来, 插到另一个插座上
- 通过一堆坍塌物散落的高低不平的地面
- 爬楼梯

人工智能的流派

单独遵循一个单一流派不足以实现人工智能

从专家系统发展出来的知识图谱，已不完全遵循符号主义路线

AlphaGo综合使用三种方法：强化学习、蒙特卡洛树搜索、深度学习，
综合行为主义、符号主义、连接主义的技术

各个流派的融合发展是大势所趋

大纲

□ 课程信息

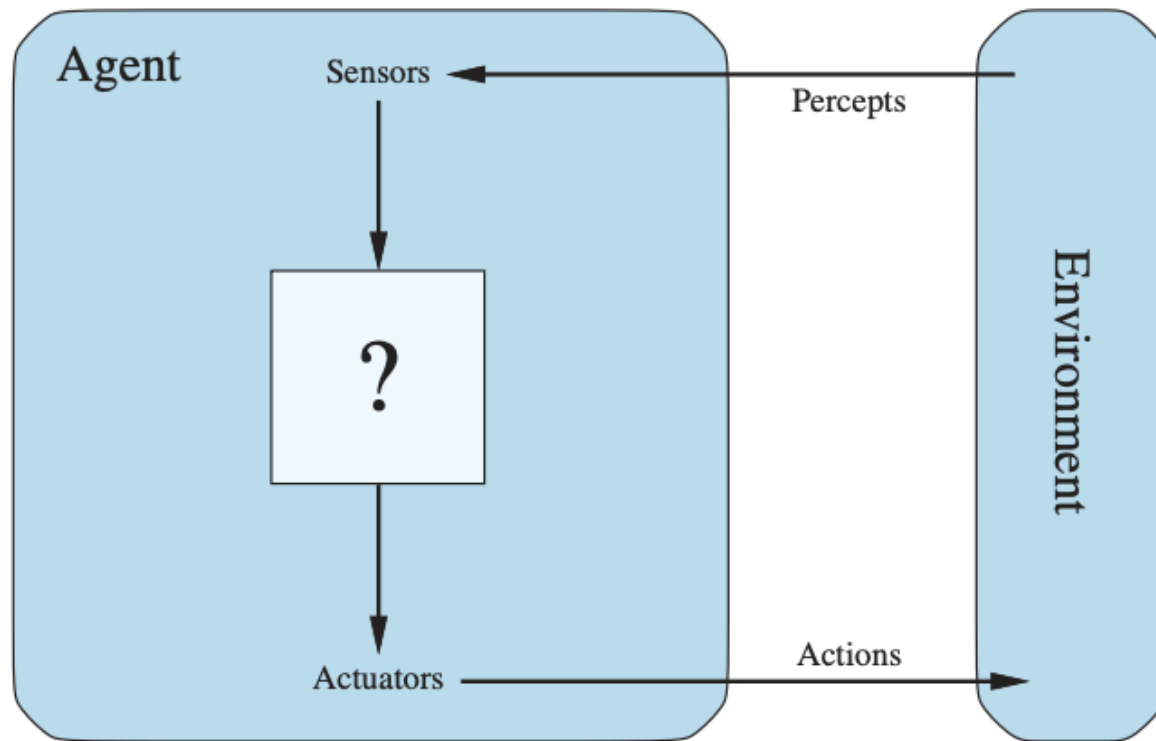
□ 什么是人工智能

□ 人工智能简史

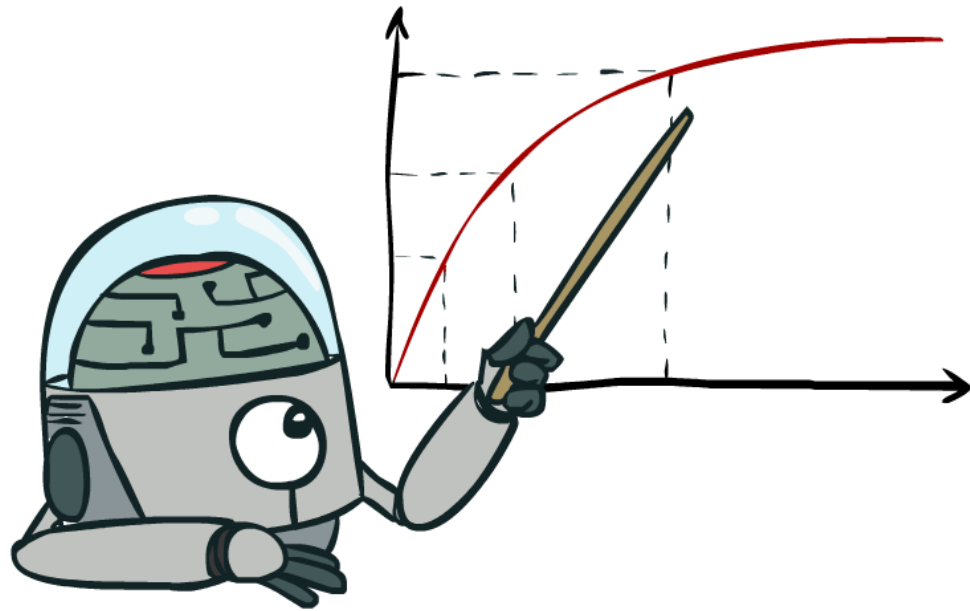
□ 本课程内容

Rationally Agents

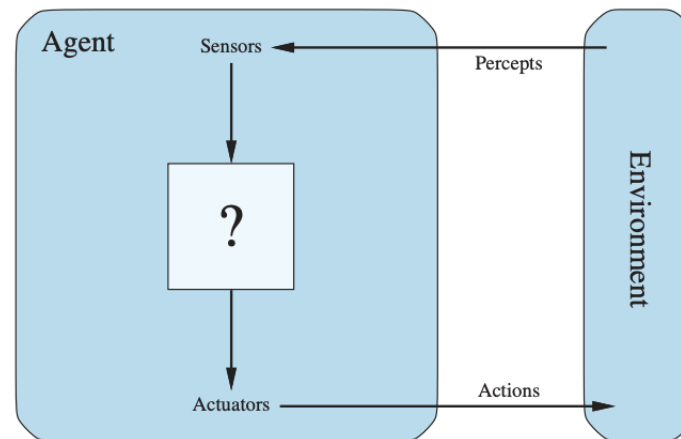
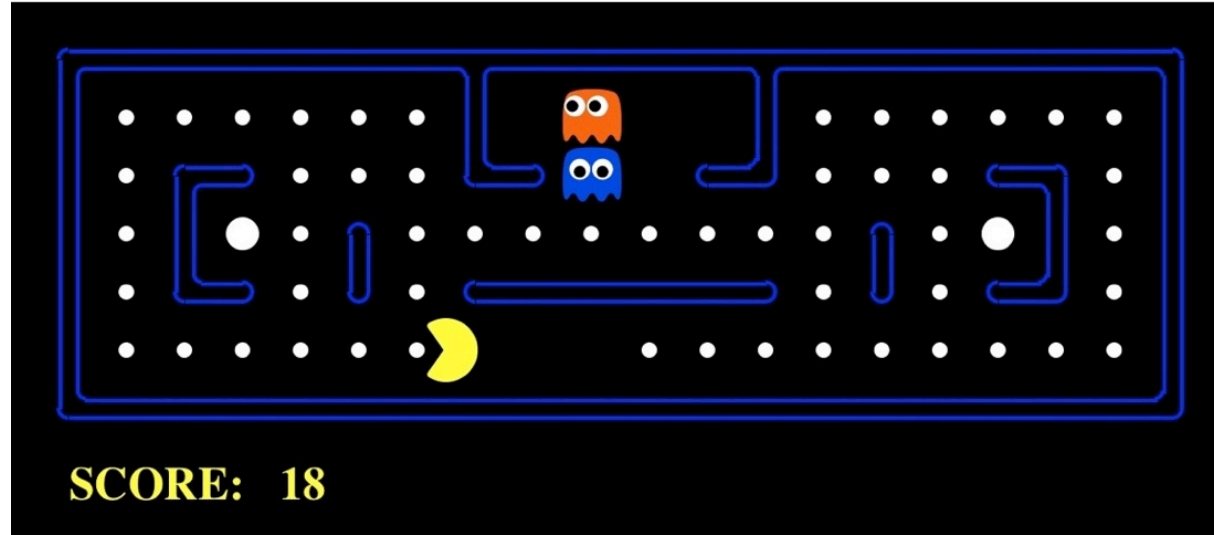
- An **agent** is an entity that perceives and acts (感知与行动)
- Goal-reaching and utility-maximizing agents are rational



Maximize Your Expected Utility



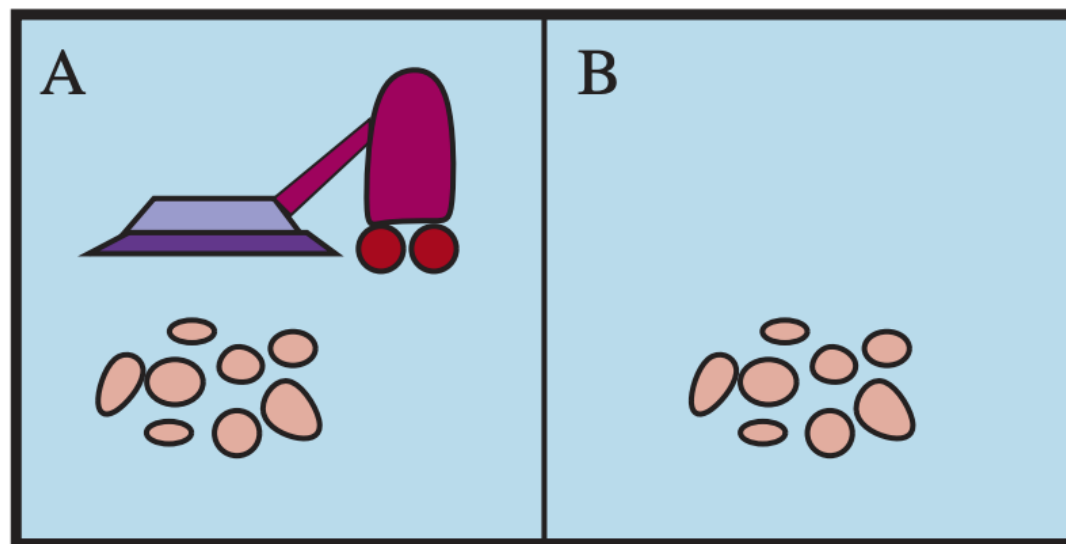
Example: 吃豆人



Example: 真空吸尘器世界

感知：当前所处的位置，当前位置是否有灰尘

行动：向左移动，向右移动，吸尘，什么也不做



Example: 真空吸尘器世界

感知	行动
<i>[A, Clean]</i>	<i>Right</i>
<i>[A, Dirty]</i>	<i>Suck</i>
<i>[B, Clean]</i>	<i>Left</i>
<i>[B, Dirty]</i>	<i>Suck</i>
<i>[A, Clean], [A, Clean]</i>	<i>Right</i>
<i>[A, Clean], [A, Dirty]</i>	<i>Suck</i>
<i>⋮</i>	<i>⋮</i>

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*

好的行为

为了构建合理的agent，需要从四个方面考虑（PEAS描述）

- ✓ **Performance**（性能）
- ✓ **Environment**（环境）
- ✓ **Actuators**（执行器）
- ✓ **Sensors**（传感器）

Example of PEAS

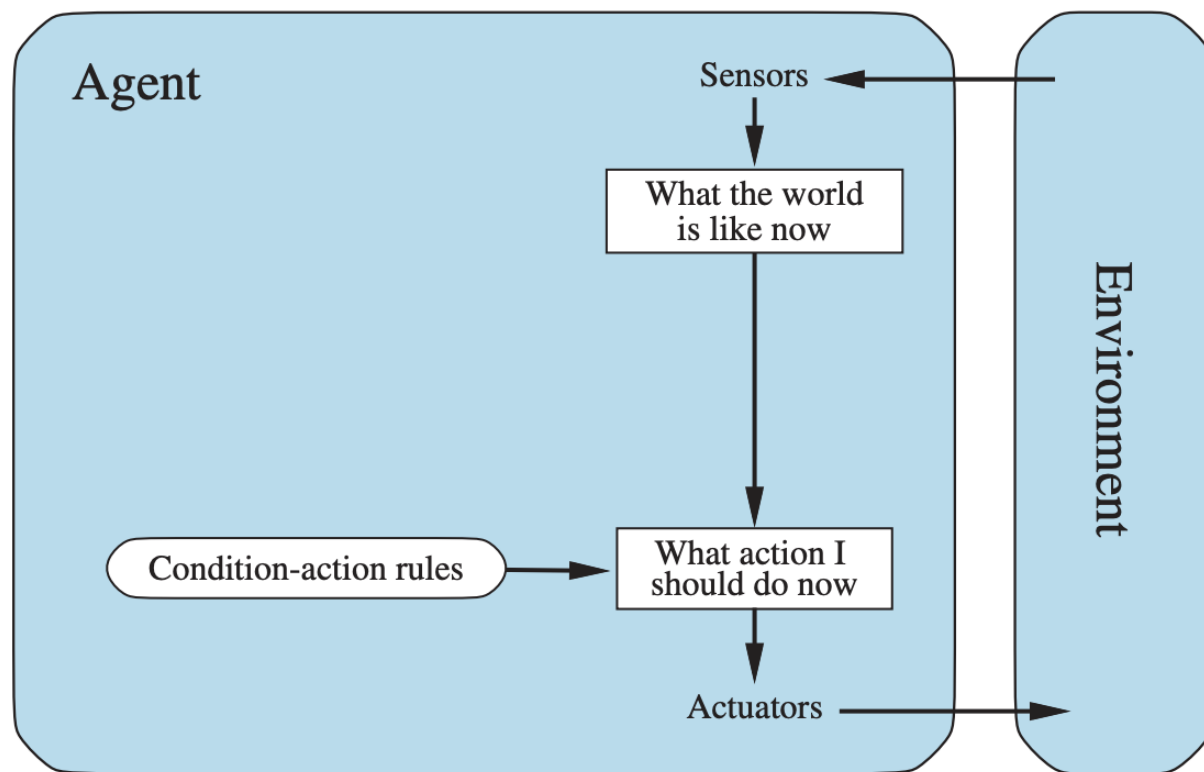
Agent Type	Performance Measure	Environment	Actuators	Sensors
Taxi driver	Safe, fast, legal, comfortable trip, maximize profits	Roads, other traffic, pedestrians, customers	Steering, accelerator, brake, signal, horn, display	Cameras, sonar, speedometer, GPS, odometer, accelerometer, engine sensors, keyboard
Medical diagnosis system	Healthy patient, reduced costs	Patient, hospital, staff	Display of questions, tests, diagnoses, treatments, referrals	Keyboard entry of symptoms, findings, patient's answers
Satellite image analysis system	Correct image categorization	Downlink from orbiting satellite	Display of scene categorization	Color pixel arrays
Part-picking robot	Percentage of parts in correct bins	Conveyor belt with parts; bins	Jointed arm and hand	Camera, joint angle sensors
Refinery controller	Purity, yield, safety	Refinery, operators	Valves, pumps, heaters, displays	Temperature, pressure, chemical sensors

Agent的分类

四种基本的agent

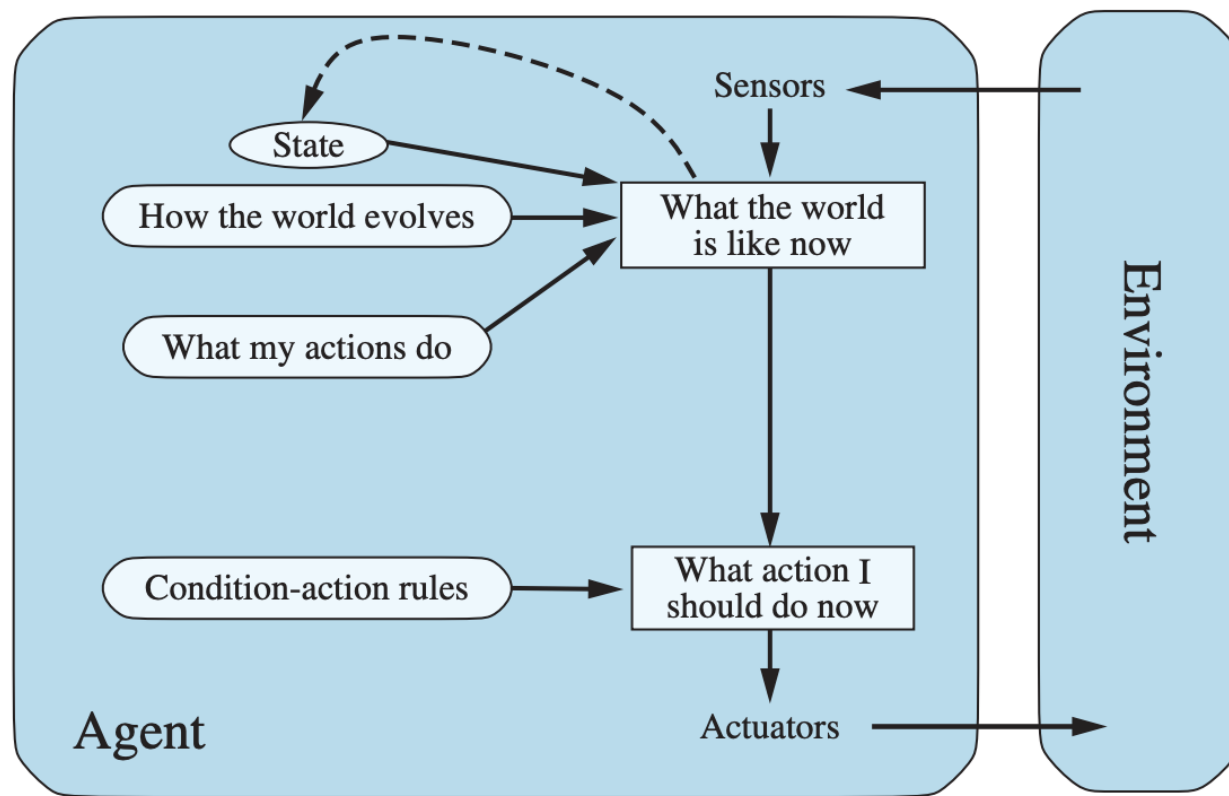
- ✓ 简单反射agent
- ✓ 基于模型的反射agent
- ✓ 基于目标的agent
- ✓ 基于效用的agent

简单反射Agent



```
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
```

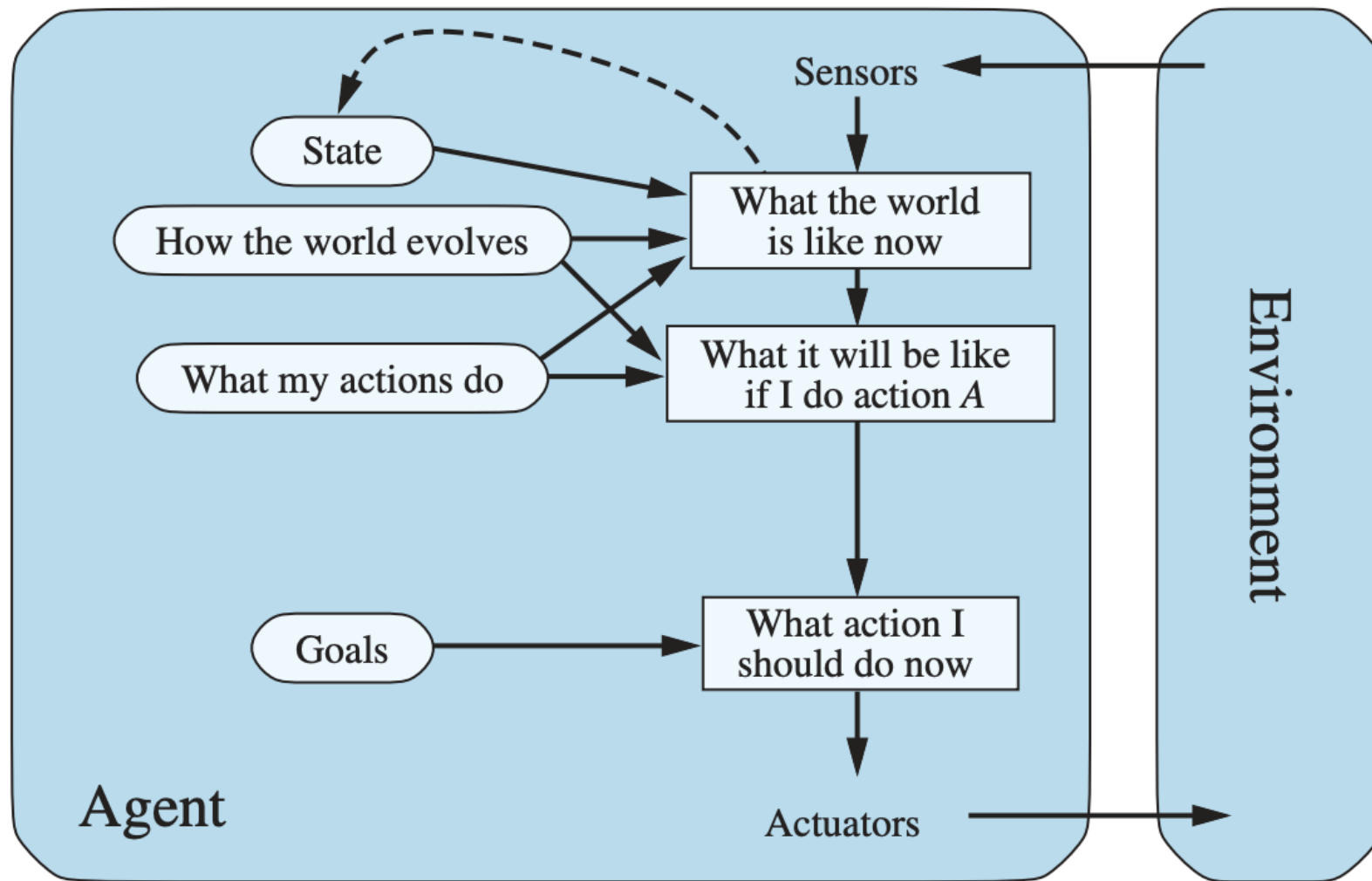
基于模型的反射agent



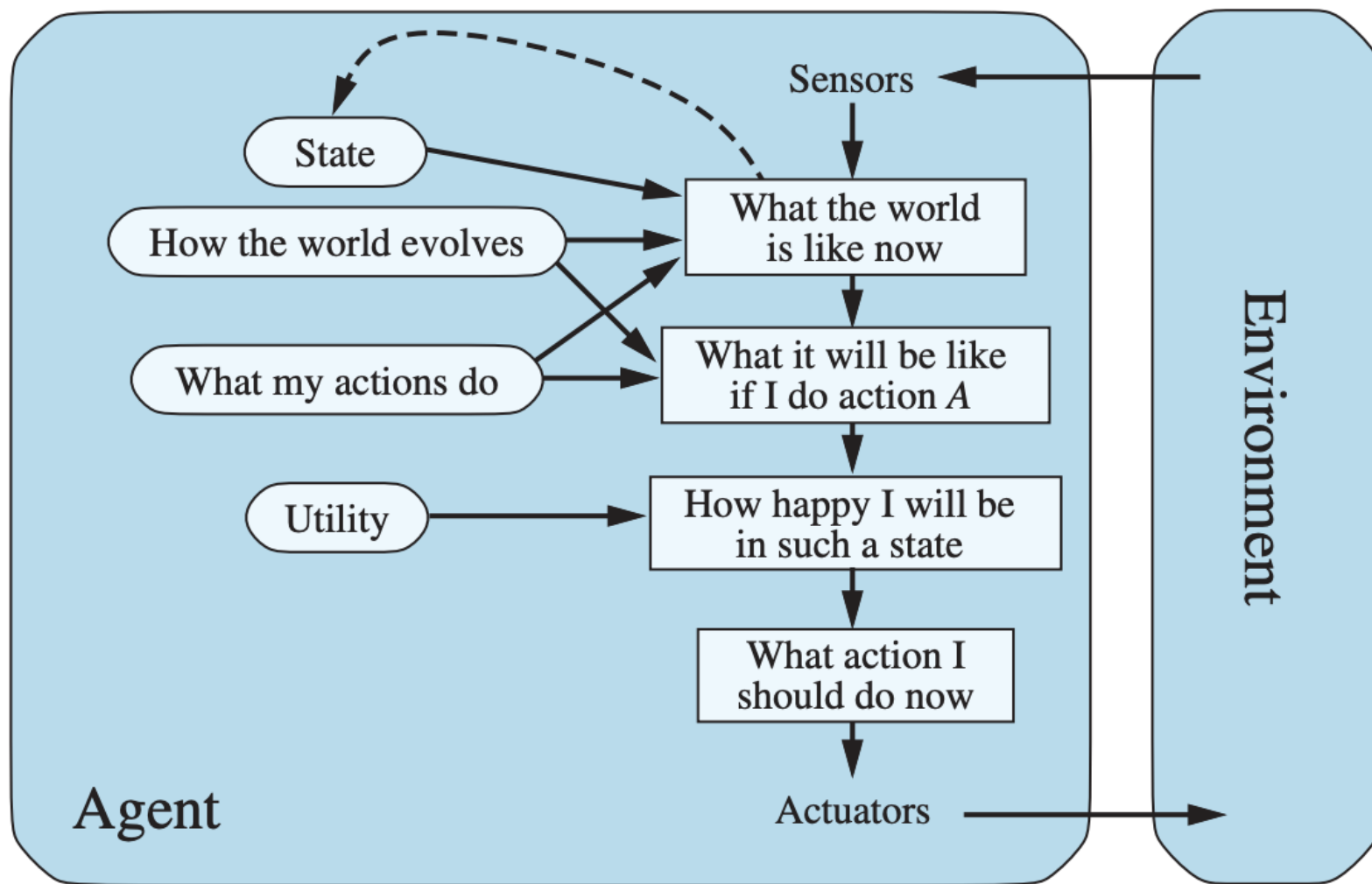
```
function REFLEX-VACUUM-AGENT([location,status]) returns an action
static: last_A, last_B, numbers, initially  $\infty$ 

  if status = Dirty then ...
```

基于目标的agent



基于效用的agent



课程内容

Search & Plans 搜索与规划

Knowledge & Reasoning 知识与推理

Machine Learning 机器学习

Applications 应用

本章小结

- 人工智能的定义
- 图灵测试
- 人工智能的起源与发展
- 人工智能的流派
- Rational Agents