



高级人工智能

沈华伟

中国科学院计算技术研究所



课程信息

- 基本信息
 - 课程编号: 081203M04003H
 - 课时学分: 60个课时、3个学分
 - 上课时间: 周二晚上18:10-20:50, 第9-11节课 (9月15日-1月5日)
 - 上课地点: 教1-002; 同步直播: 教1-009、教1-317
 - 三次加课
- 授课教师
 - 沈华伟, 研究员, 中科院计算所, shenhuawei@ict.ac.cn
 - 罗 平, 副研究员, 中科院计算所, luop@ict.ac.cn
 - 吴高巍, 副研究员, 中科院自动化所, gaowei.wu@gmail.com
- 助教
 - 曹琦, 陈娟, caoqi@ict.ac.cn, chenjuan@ict.ac.cn
 - 研究生助教: 景琨、刘淼

高级人工智能课程群



高级人工智能课程群 2021



该二维码 7 天内 (9月13日前) 有效，重新进入将更新



高级人工智能课 2021-群 2



该二维码 7 天内 (9月15日前) 有效，重新进入将更新



授课和考核

- 教学方式

- 课堂讲授：首先讲授人工智能基础知识，进而分三个专题（联结主义、符号主义、行为主义）介绍人工智能的新进展
- 作业：随堂作业+大作业
（讲授+研讨+实践）

- 考核形式

- 闭卷笔试(60%)
- 平时成绩(40%)
 - Project（大作业）：30分
 - 课堂表现（包括考勤）等情况：10分



课程大纲 (1/2)

- 第1讲 人工智能概述 (9月7日)
- 第2讲 搜索问题 (9月14日)

第二部分 神经网络 (15学时)

- 第3讲 人工神经网络和深度学习基础 (9月28日)
- 第4讲 序列数据的深度学习模型 (10月12日)
- 第5讲 图像数据的深度学习模型 (10月16日)
- 第6讲 深度学习模型的训练方法 (10月19日)
- 第7讲 图神经网络 (10月26日)

第三部分 知识智能 (15学时)

- 第8讲 命题逻辑的语义推论 (11月2日)
- 第9讲 命题逻辑的形式推演 (11月9日)
- 第10讲 谓词逻辑 (11月16日)
- 第11讲 模糊知识表达和推理 (11月20日)
- 第12讲 知识表示学习 (11月23日)



课程大纲 (2/2)

第四部分 群体智能 (15学时)

- 第13讲：演化计算 (11月30日)
- 第14讲：强化学习I (12月7日)
- 第15讲：强化学习II (12月14日)
- 第16讲：博弈基础 (12月18日)
- 第17讲：博弈应用 (12月21日)
- 第18讲：课程作业分组陈述 (12月28日)

第五部分 考核 (9学时)

- 第19讲：答疑 (1月4日)
- 第20讲：考试 (1月11日)



参考资料

• 书籍

- T. Dean, Artificial Intelligence: Theory and Practice, Addison-Wesley, 1995.
- Stuart Russell, and Peter Norvig. Artificial Intelligence: A Modern Approach, Second Edition, Prentice Hall. 2003, ISBN: 0-13-790395-2
- Richard S. Sutton and Andrew G. Barto. Reinforcement learning - an introduction (Second Edition). The MIT Press, Cambridge, Massachusetts, 2017.
- David Easley, Jon Kleinberg. Networks, Crowds, and Markets. Cambridge University Press. 2010.
- 史忠植, 《高级人工智能》, 科学出版社, 北京, 2011



第一讲

人工智能概述



什么是智能？

智能：个体适应环境并能在不同环境中实现其目标的能力。

- 蕴含众多方面的能力
 - 创造、推理、学习
 - 归纳、演绎、类比
 - 优化、规划、知识
 - 模式识别、问题求解
 -



什么是人工智能？

- 概念性定义

- **机器智能：使机器具备计算和“判别”的行为能力**

- Artificial intelligence (AI) is the branch of computer science concerned with making computers intelligent, just like people.

- **类脑智能：仿生智能，让机器像人或生物一样思考**

- AI is the multidisciplinary study of human intelligence through attempts to artificially model it.

- **群体智能：社会智能的机器重现与利用、涌现智能**

- Intelligence emerged from collective behaviors of lots of agents without or with little intelligence

if God can make natural life, man can make artificial life

构建人工智能是人类孜孜不倦的追求

人工智能的起源

- 萌芽期
 - 机械自动化
 - 希腊，蒸汽驱动的“会唱歌”的乌鸦
 - 中国，鲁班的“木鸢”，诸葛亮的“木牛流马”
 - 逻辑推理
 - 亚里士多德的“三段论”：从一般前提到具体论断
- 孕育期（文艺复兴以来）
 - 理性主义
 - 笛卡尔：mind/body二象性，不相信机器会具有智能
 - 数理逻辑学科
 - 莱布尼茨：演算推论器，符号逻辑，提出将人的知识汇成“知识库”
 - 弗雷治：谓词演算
 - 计算思维
 - 巴贝奇：差分机
 - 图灵：图灵机

三段论

前提

① 所有的A都是B

② a是A的一种

论断

③ a也是B



差分机

机械智能 → 理性思考 → 数理逻辑 → 计算思维



人工智能的发展

- 形成期（1956年-1961年）
 - 1956年，首次人工智能研讨会
 - IBM的西洋跳棋程序、文法体系、逻辑推理机、行动计划咨询系统、通用问题求解器
- 发展期（60年代）
 - 研究领域拓展
 - 问题求解、博弈、定理证明、程序设计、机器视觉、自然语言理解、知识表示、专家系统、神经网络、智能机器人.....
 - 1969年，第一届国际人工智能联合会议（IJCAI）
 - 1970年，《人工智能》国际杂志创刊，《Artificial Intelligence》



人工智能的冬天

- 寒冬期（60年代末到70年代初）
 - 1966年，美国政府取消了机器翻译项目的所有投资
 - 英国政府取消了几乎所有人工智能研究投入
 - 神经网络的研究经费缩减到几乎没有



人工智能的复兴

- 艰难前行（70年代）
 - 弱方法：构建搜索机制，试图找出完全解
 - 下棋：搜索解空间
 - 强方法：构建领域知识库
 - 专家系统：知识表示开始成为研究热点
- 走向工业（80年代）
 - 1982年，第一个商用专家系统RI
 - 1981年，日本启动“第五代计算机”计划，运行prolog语言的智能计算机
 - 美国、英国恢复对人工智能的投入



人工智能的今天

- 大数据利用、计算能力提升、网络泛在化
- 神经网络的复兴
 - 多层感知机及其学习算法（BP算法）的提出
 - 隐马尔科夫模型（HMM）在语音识别上取得成功
 - 贝叶斯网络
- 专家系统逐渐成熟
 - 知识发现、数据挖掘兴起
- 人工智能开始成为科学
 - 学科边界开始明晰
 - 并开始借鉴其他学科的理论，如控制论、心理学、统计学

人工智能

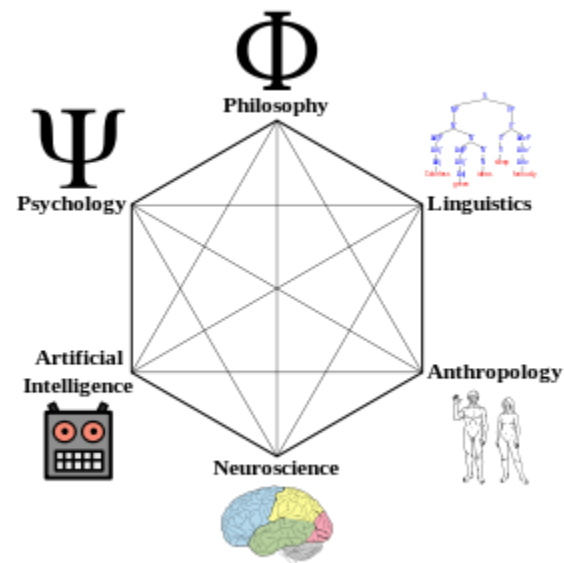
人工智能
认知科学

决策论
博弈论

	类人 (Humanly)	理性 (Rationally)
思考 (Think)	像人一样思考	理性思考
行动 (Act)	像人一样行动	理性行动

认知科学

- 关于大脑内在活动的科学
- 研究大脑是如何工作的？
- 两个分支
 - 认知学：将人脑作为黑盒子进行研究
 - 宏观层面的研究：研究脑对输入的应激行为，构建预测方法
 - 认知神经学：将人脑作为白盒子进行研究
 - 微观层面的研究：从脑神经的测量数据中开展



这两个学科后来都从“人工智能”中分离出去了



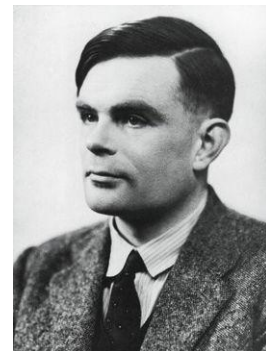
人工智能

	类人 (Humanly)	理性 (Rationally)
思考 (Think)	像人一样思考	理性思考
行动 (Act)	像人一样行动	理性行动

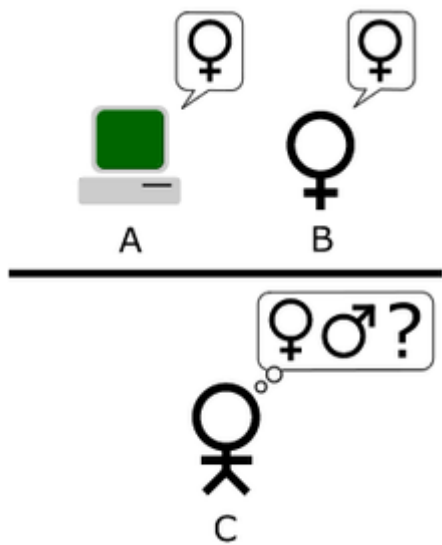
图灵测试

- 图灵问题
 - *Can Machine Think?*
- 模仿游戏

A. M. Turing, *Computing Machinery and Intelligence*, *Mind*, 59:433-460, 1950.



阿兰·图灵
(1912-1954)
现代计算机之父



图灵测试 (Turing Testing)

一个人 (C) 在**完全不接触**对方 (A和B) 的情况下, 和对方进行一系列的问答, 如果在相当长时间内, 他无法根据这些问题判断对方是人 (B) 还是计算机 (A), 那么, 就认为该计算机具有同人相当的智能, (即计算机是能思维的)

图灵预言: 到2000年, 机器可以做到5分钟内以30%的可能性让普通人分辨不出其是机器还是人

图灵测试案例：

- Master横空出世
 - Master在围棋对战网站上出现连胜30多场，才开始有人怀疑这是“机器人”
 - 直到后来官方确认Master为AlphaGo 2.0
- 人工智能机器人Sophia
 - 电视节目主持人查理·罗斯在节目《60分钟》中采访了Sophia机器人时，索菲亚不但对答如流，还与他开起了玩笑
 - 罗斯被Sophia调侃





图灵测试

- 质疑
 - 图灵测试不是可构造的
 - 例如：“完全不接触”的环境难以构建
 - 图灵测试不是可重现的
 - 例如：问题是开放的，答案正确性的判定是主观的
 - 图灵测试无法进行数学分析
 - 只是一种操作式测试，缺少形式化描述

注：图灵逝世60周年纪念日（2014.6.7），在英国皇家学会举行的“2014图灵测试”大会上，聊天程序“尤金·古斯特曼”（Eugene Goostman）首次“通过”了图灵测试。

达特茅斯会议

- 1956年在达特茅斯学院发起

- 发起人

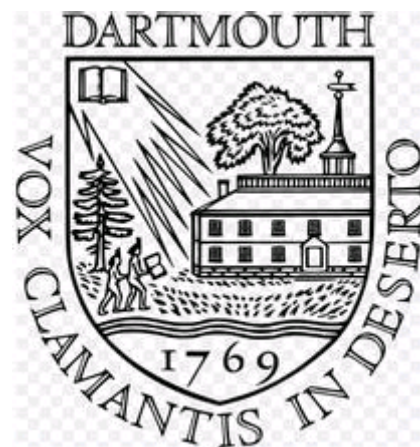
- 约翰 麦卡锡（人工智能之父，Lisp语言发明者，1971年获图灵奖）
- 马文 明斯基（1969年获图灵奖，首个获图灵奖的人工智能学者）
- 克劳德 香农（信息论之父）
- 纳撒尼尔 罗彻斯特（IBM 700系列计算机首席工程师，发明了首个汇编语言）



约翰·麦卡锡
(1927-2011)

- 会议成就

- 首次提出了“人工智能”一词
- 会议三大亮点
 - 明斯基的Snare
 - 麦卡锡的 α - β 搜索法
 - 西蒙和纽厄尔的“逻辑理论家”





人工智能三大学派

- 符号主义学派
- 联结主义学派
- 行为主义学派

人工智能三大学派

- 符号主义学派
 - 又称逻辑学派
 - 认为“人的认知基元是符号，认知过程即符号操作过程”
 - 认为人和计算机都是物理符号系统，可以用计算机来模拟人的智能行为
 - 认为人工智能的核心是知识表示、知识推理和知识运用
- 代表人物
 - 西蒙（1975年获图灵奖、1978年获诺贝尔经济学奖）
 - 纽厄尔



赫伯特·西蒙
(1916-2001)

衍生出：逻辑、专家系统、知识库



人工智能三大学派

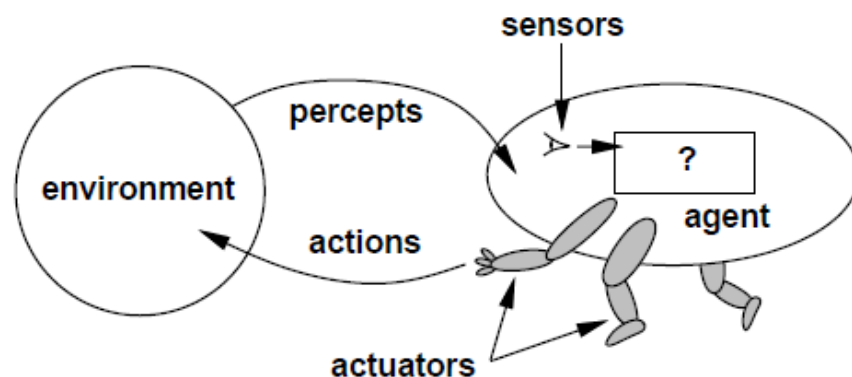
- 联结主义学派
 - 又称**仿生**学派或生理学派
 - 认为人的思维基元是神经元，而不是符号处理过程
 - 认为人脑不同于电脑
 - 原理：神经网络及神经网络间的连接机制和学习算法
- 代表人物
 - 麦卡洛克（McCulloch）
 - 皮茨（Pitts）

衍生出：人工神经网络、认知科学、类脑计算

人工智能三大学派

- 行为主义学派
 - 又称进化主义或**控制论学派**
 - 认为智能取决于感知和行动
 - 主张**利用机器对环境作用后的响应或反馈为原型**来实现智能化
 - 认为人工智能可以像人类智能一样通过进化、学习来逐渐提高和增强

- 代表人物
 - 布鲁克斯



衍生出：控制论、多智能体、强化学习等



人工智能研究的课题

- 三大层次

- 基础理论：数学、思维科学、认知科学等
- 原理技术：启发式搜索、演化计算
- 工程应用：模式识别、计算机视觉、自然语言理解、问答系统

- 四大问题

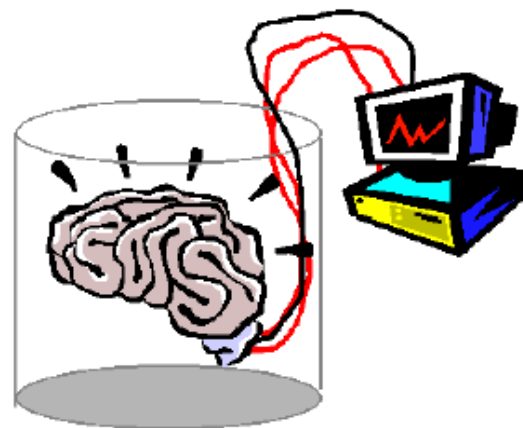
- 知识科学、问题求解、机器学习、系统构成

人工智能之哲学基础

- 弱人工智能
 - 机器表现得像具有智能一样
 - 图灵测试
- 强人工智能
 - 机器实际具有智能
 - 机器具有自我意识吗？
 - 自由意志悖论
 - 受物理法则严格支配的思想会是自由的吗？
 - 如果不能说出我下一步会做什么，就说明我具有自由意志？



中文屋子问题



解析人脑



人工智能恐慌

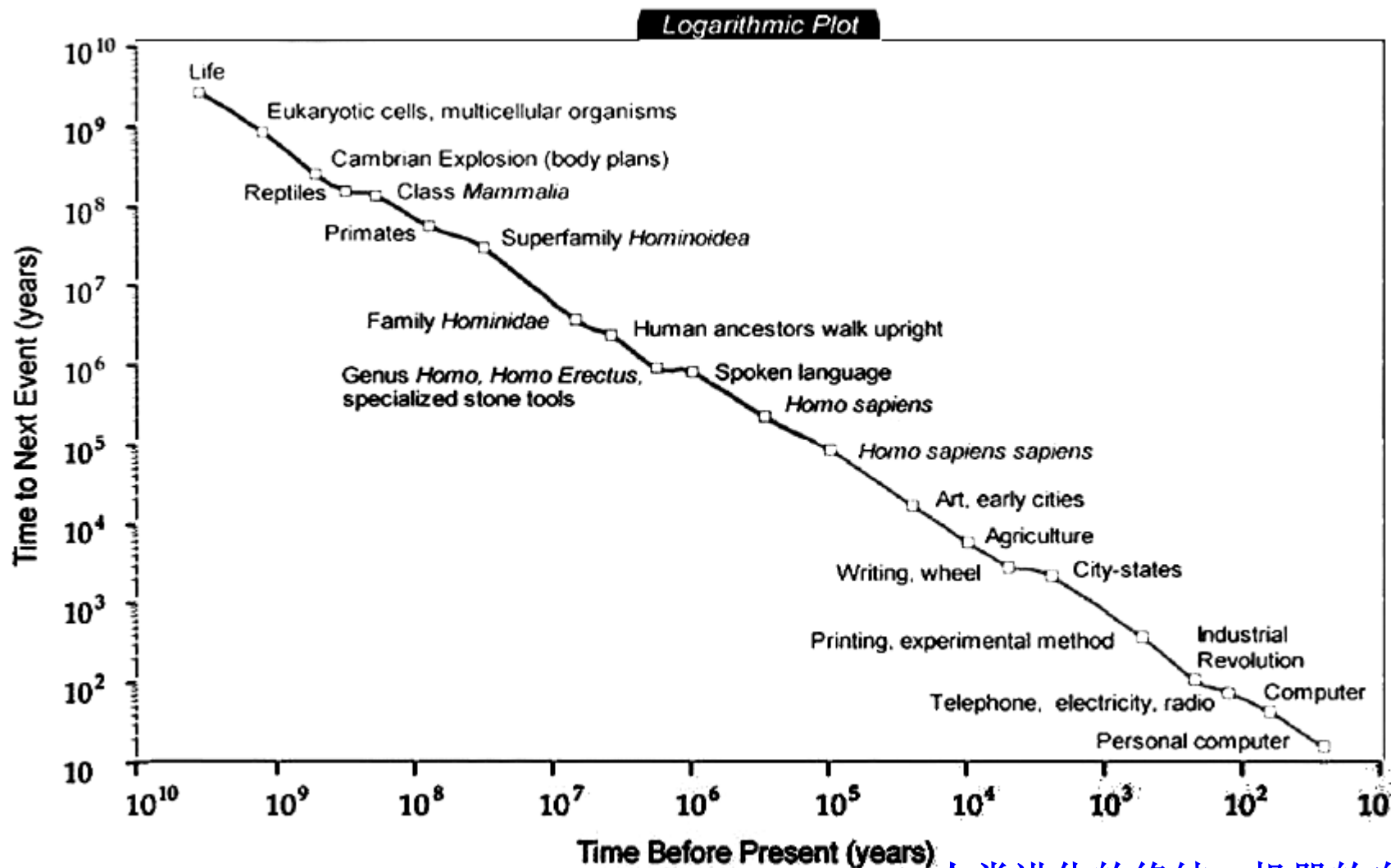
- 会不会造成人们失业？
 - 目前来看，人工智能技术带来的自动化，其创造的就业就会大于其减少的就业机会
- 对隐私权的侵害？
- 是否导致可审计的丧失？
 - 例如：听从了医疗诊断专家系统的建议而带来的医疗事故，责任归谁？



人工智能实现了会怎样？

- 人工智能的成功是否会意味着人类灭亡
 - 人工演化取代自然选择
 - 机器智能一旦超过人类智能，他就能设计出更聪明的机器[古德, 1965]
 - 智力爆炸和技术奇点，人类时代的终结
- 怎么办？
 - 让机器保持可控
 - 使用人工智能拓展人类智能，将人工智能合并到人类智能中[雷蒙德·库茨魏尔, 2000]

技术奇点



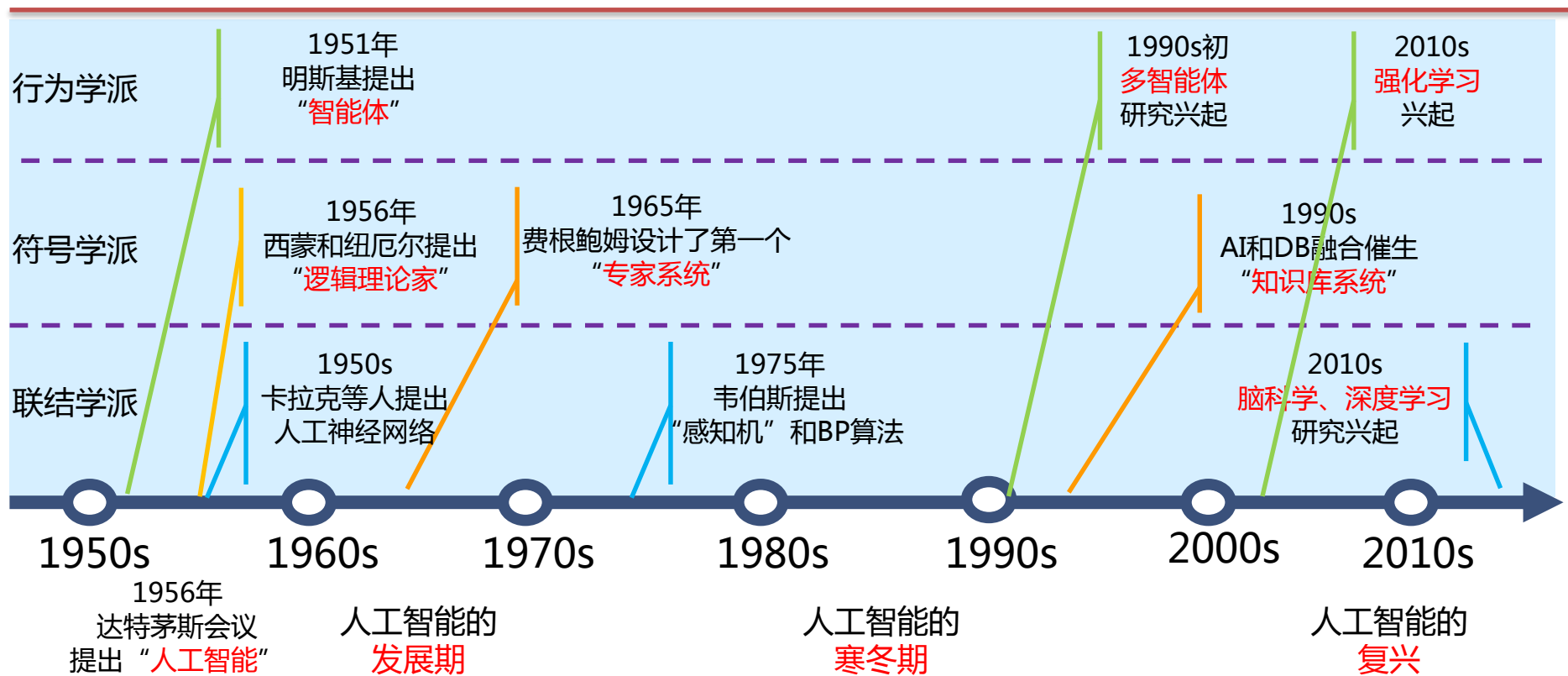
人类进化的终结、机器的自我繁衍？



人工智能伦理

- 机器人三法则_[艾萨克 阿西莫夫, 1942]
 - 第一法则：机器人不得伤害人类，或袖手旁观坐视人类受到伤害
 - 第二法则：除非违背第一法则，机器人必须服从人类的命令
 - 第三法则：在不违背第一及第二法则下，机器人必须保护自己

人工智能路线图





休息10分钟



人工智能的目标

- 近期目标
 - 研究如何使机器做过去只有依靠人的智力才能完成的工作
- 远期目标
 - 研究如何利用自动机模拟人的思维过程 and 智能行为, 从而造出智能机器
- 终极目标
 - 机器智能实现甚至超过生物智能



“准人”水平的人工智能

- 手写识别
- 物体识别
- 语音识别
- 自然语言处理
- 词义消歧
- 机器翻译



“过人”水平的人工智能

- 游戏
 - 双陆棋
 - 国际象棋
 - 桥牌
 - 填词、拼字
 - 七巧板
- 工具
 - 自动驾驶
 - 智力竞赛问答
 - OCR字符识别
 -



讖言

“许多尖端的人工智能由于应用广泛，已经不再被称为人工智能。因为，人们一旦觉得某些东西非常有用并广泛使用，就不再称之为人工智能了。”

——尼克 博斯特罗姆

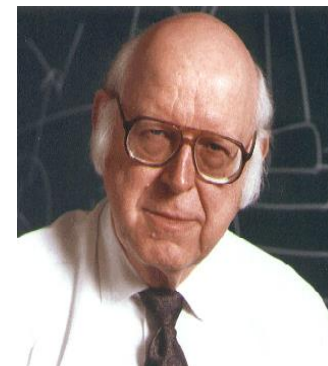


人工智能案例实践

- 定理证明
 - 50年代中期，西蒙和纽厄尔提出的“**逻辑理论家**”，证明了《数学原理》书中的38个定理
 - 1962年，改进后证明了书中全部52个定理
 - 被认为是用计算机探讨人类智能的第一个真正成果
- 案例
 - 四色定理
 - 1852年提出，一直无人给出理论证明
 - 1976年6月，哈肯在伊利诺伊用两台计算机，用时1200个小时，通过100亿次判断，完成了证明，轰动世界
 - 吴方法：吴文俊教授提出的“数学机器化”

人工智能案例实践

- 通用问题求解器（GPS：General Problem Solver）
 - 1957年开始，纽厄尔等人开始研究不依赖于具体领域的通用解题程序
 - 模仿人类问题求解过程
 - 第一个实现了“像人一样思考”的程序



艾伦·纽厄尔
(1927-1992)
1975年图灵奖
AAAI发起人

人工智能案例实践

- 专家系统

- 将领域专家的知识整理出来，让计算机利用这些知识求解专门领域的问题
- **DENDRAL**: 第一个专家系统，1968年问世，斯坦福大学完成，用于推断化学分子结构
- MYCIN: 著名的医疗诊断专家系统
- RI: 第一个商用专家系统，DEC公司于1982年正式使用



爱德华·费根鲍姆
1994年图灵奖



人工智能案例实践

- 数字识别
 - 清华大学智能技术与系统国家重点实验室采用神经元网络研制了数字识别系统
 - 用于2000年我国的人口普查
 - 错误率达到低于万分之一的水平
- 古籍数字化（OCR技术）
 - 《四库全书》

人工智能案例实践——国际象棋

- IBM的“深蓝”
 - 1997年，IBM公司的“深蓝”在美国纽约公平大厦以3.5:2.5击败了国际象棋世界冠军卡斯帕罗夫
- “深蓝”的配置
 - 32个CPU
 - 每个CPU有16个协处理器
 - 每个CPU有256M内存
 - 每个CPU处理速度300万步/秒



计算机依靠**计算能力**的提升试图战胜人类

人工智能案例实践——围棋

- AlphaGo: DeepMind
 - 使用深度学习技术（CNN：卷积神经网络）对棋局的局势进行估值
 - 在和其他围棋程序的对弈中取得99.8%的胜率
 - 和李世石的人机大战中以4:1取胜
 - Master在人机对战中60连胜
 - Master以3:0战胜柯洁



Nature封面文章

打破了“计算机不会下围棋”的
预言



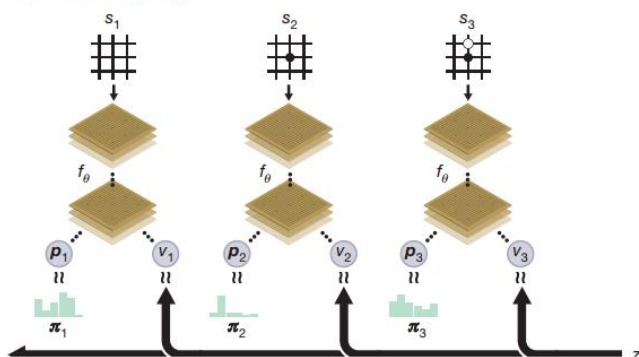
D. Silver, et al., Mastering the game of Go with deep neural networks and tree search. Nature, 529: 484-489 2016. (AlphaGo)

D. Silver, et al., Mastering the game of Go without human knowledge. Nature 550, 354-359, 2017. (AlphaGo Zero)

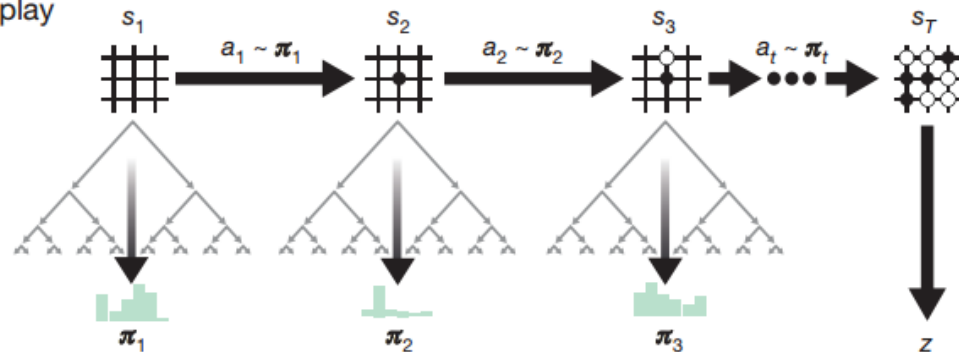
人工智能实践案例——围棋

- AlphaGo背后的技术
 - 深度学习（联结主义）+ 强化学习（行为主义）
 - Mnih, et al., Human-level control through deep reinforcement learning, Nature 518, 529–533, 2015.
 - 利用残差神经网络(ResNet)训练深度模型
 - Kaiming He et al., Deep Residual Learning for Image Recognition, CVPR, 2016. (两年来, Google Scholar引用11700余次)
 - 利用马尔科夫树搜索技术解决围棋的搜索空间爆炸问题
 - 采用“自我对弈”策略进行无人工标注的自我训练

Neural network training



Self-play



人工智能案例实践

- 自动驾驶

- 在高速公路上，自动识别道路，自动躲避障碍物
- 平均时速达到100公里/小时，最高速度可达150公里/小时
- 从匹兹堡到圣地亚哥，98%的时间自动驾驶





人工智能案例实践

- 自然语言处理
 - 神经语言模型和词嵌入技术
 - 例如：word2vec
 - 机器翻译
 - 统计机器翻译(SMT)到神经机器翻译(NMT)
 - 例如：Seq2Seq模型
 - 文本生成技术
 - 给图像或视频加标题
 - 聊天机器人
 - 机器人写新闻报道
 - BERT和GPT-3

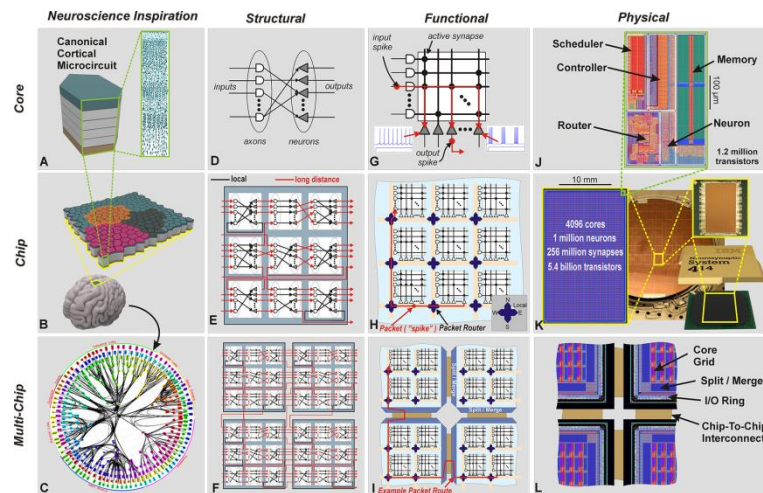
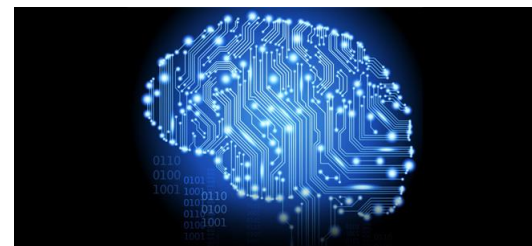


人工智能案例实践

- 生成式预训练语言模型
 - GPT: Generative Pre-training Transformer
 - 来自OpenAI
 - GPT-2: 15亿参数
 - GPT-3: 1750亿参数，训练一次花费数百万美元
 - 以开放测试版的形式提供
 - 是走向领域通用人工智能的路径吗？

人工智能案例实践

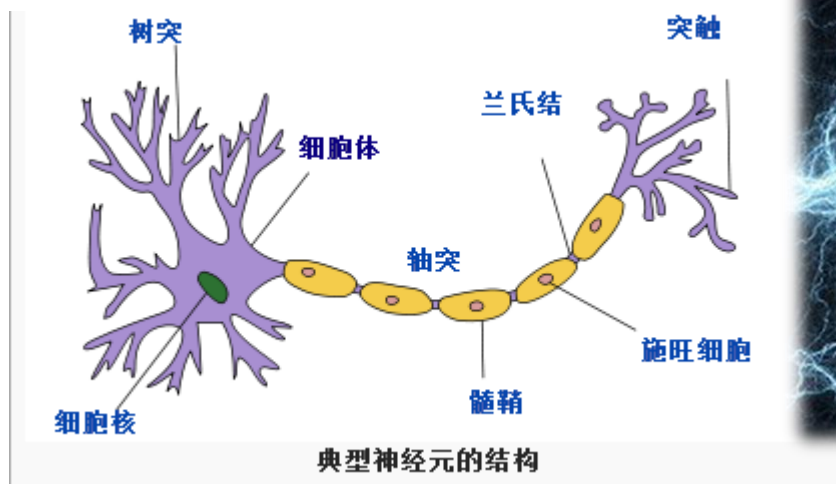
- IBM仿人脑芯片：TrueNorth
 - DARPA的研究项目SyNapse（自适应可塑可伸缩电子神经系统）的最新成果
 - 邮票大小、重量只有几克，集成 54 亿个硅晶体管，内置 4096 个内核，100 万个“神经元”、2.56 亿个“突触”，能力相当于一台超级计算机，功耗只有 65 毫瓦
 - 目标：突破冯 诺依曼体系



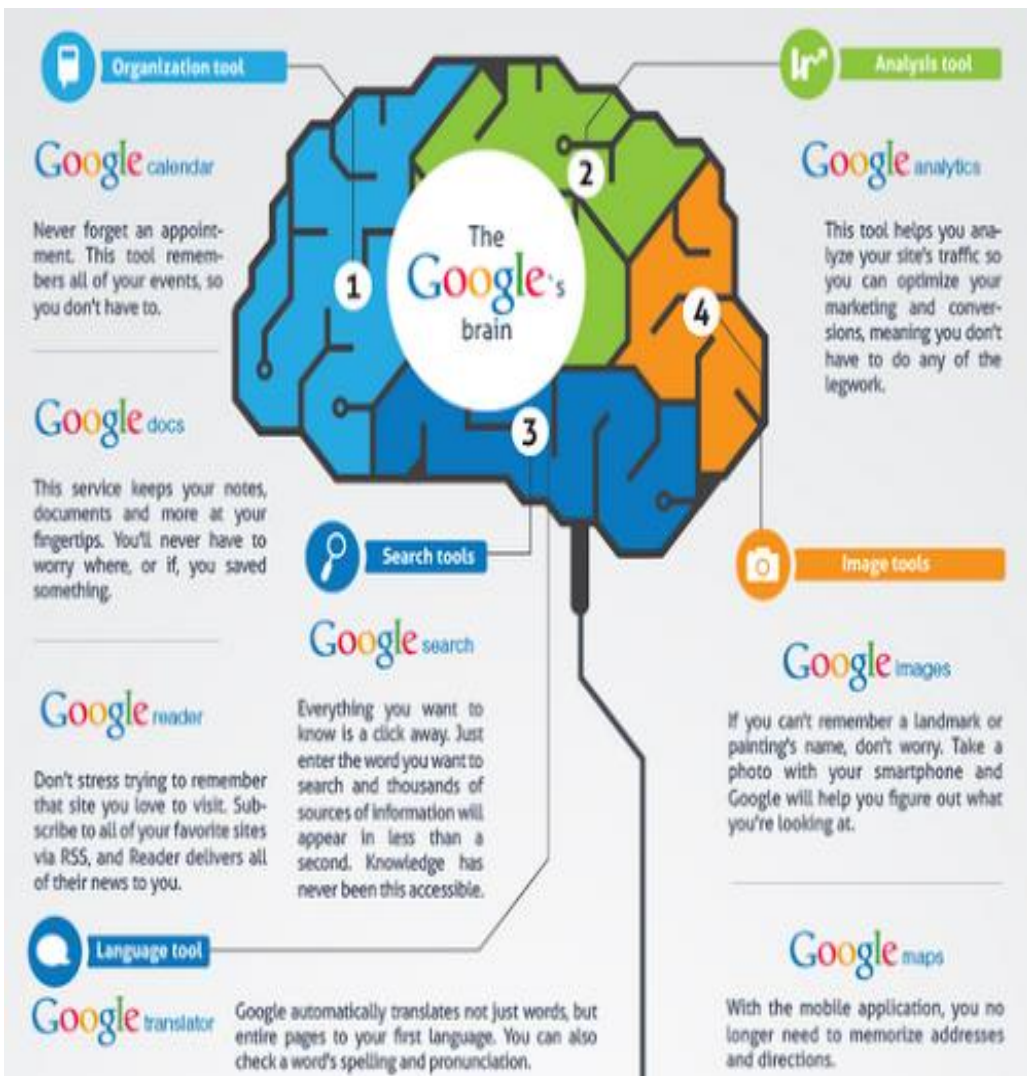
人工智能案例实践

- 脑科学

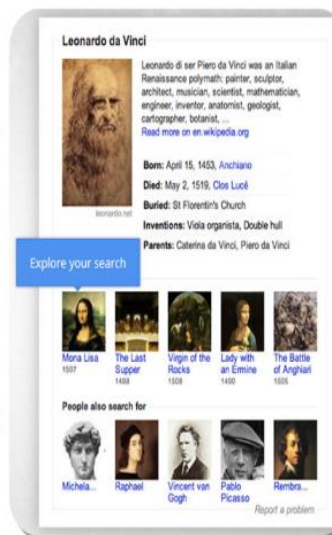
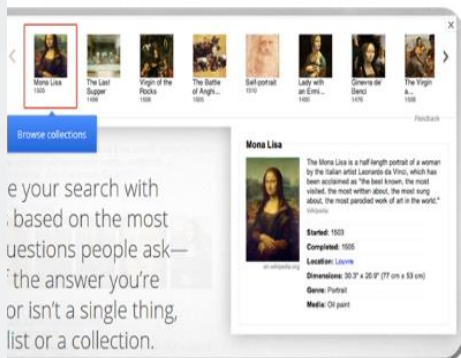
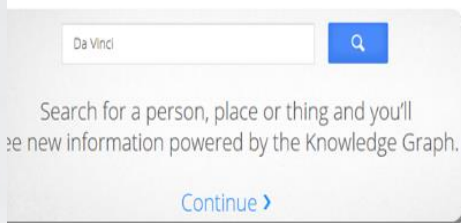
- 2013年1月，欧盟启动“人类大脑计划”
- 2013年4月，奥巴马宣布启动“大脑基金计划”
 - 继曼哈顿计划、阿波罗登月计划、人类基因组计划的重大计划
- 2014年，我国着手启动“脑科学计划”



人工智能案例实践：互联网大脑



知识图谱+深度学习：利用网络大数据推断目标间的潜在关联关系等关系，为用户提供查询推荐、搜索导航等知识获取和深度理解功能。



人工智能案例实践：系统论

- 复杂自适应系统
 - 1984年，美国圣塔菲研究所成立
 - 诺贝尔物理学奖得主盖尔曼提出
 - 认为**智能体现为个体的自适应能力**
 - 大量智能体（agent）积极地相互竞争和合作，在没有中央指挥的情况下，通过彼此相互作用和相互适应也能形成整体的有序状态
 - **还原论→系统论→复杂性理论（涌现）**
- 应用案例
 - 经济危机
 - 股票市场
 - 生态系统
 - e.g., 雾霾成因



现代科学的一个重大挑战是沿着阶梯从基本粒子物理学和宇宙学到复杂系统领域，探索兼具简单性与复杂性、规律性与随机性、有序与无序的混合性事件。

——盖尔曼



人工智能今天走到哪了？

- 自然语言理解（主战场之一）
 - 聊天机器人：小冰
- 智能阅卷
 - 安庆会考：全学科智能阅卷
- 考试机器人
 - 美国华盛顿大学图灵中心
 - 目标是通过美国高中生物考试
 - 日本Todai高考机器人
 - 目标是让机器人考上东京大学

人工智能今天走到哪了？

- 人工智能三级跳



- 人工智能需要一场认知革命？
 - Deep learning、脑科学



人工智能的当前热点

- 自然语言理解
 - 机器翻译
 - 聊天机器人
 -
- 语音图像理解
 - 深度学习技术
 - DNN、RNN、CNN
 - 生物特征识别技术（刷脸、瞳仁、声纹.....）
 - 中国创业公司：Face++
- 大数据助力智能突破？ 大数据智能？

现状：数据驱动的计算智能

计算智能



算力



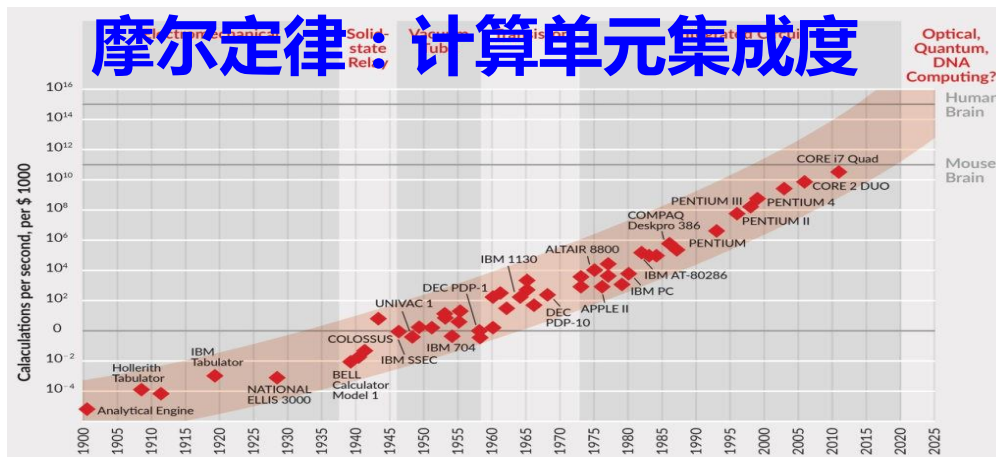
模型/算法



数据



摩尔定律：计算单元集成度



新摩尔定律：模型参数个数？



3亿参数



15亿参数



1750亿参数



120亿参数

数据智能，从数据化的信息到模型化的知识



人工智能的发展趋势（1/3）

- 从“人机对抗”走向“人机协作”
 - AI 1.0
 - 让机器在某些任务上“战胜”人
 - AI 2.0：人本计算（human computation）
 - 让机器和人相互协作，完成更复杂的任务
 - 机器做机器擅长的：计算
 - 人做人擅长的：思考



人工智能的发展趋势（2/3）

- 从单点智能走向网络智能
 - AI 1.0
 - 单个机器具备人的某些智能，例如：听、说、读、写、感知、认知.....
 - AI 2.0
 - 借助互联网实现智能网络化



人工智能的发展趋势（3/3）

- 从专用人工智能走向通用人工智能
 - AI 1.0
 - 在具体的任务上，让机器具备智能
 - 例如：围棋、自动驾驶.....
 - AI 2.0
 - 研究通用人工智能，包括探索智能形成的机制
 - 类比：AlphaGo到Master是一种初步尝试，让机器具备能够自我学习、形成概念的能力

课程蓝图

