

人工智能基础

尹慧琳, yinhuilin@tongji.edu.cn

同济大学 电子与信息工程学院 中德智能中心



第1讲 绪论

1.1 什么是人工智能

1.2 人工智能的基础

1.3 人工智能的历史

1.4 人工智能的判定

1.5 人工智能的层级与流派

1.6 人工智能的最新发展水平

课程教学要求与安排



第1讲 绪论

1.1 什么是人工智能 Artificial Intelligence

► 智能：

► 与智能相关的概念：

- ✓ 知识：是智能行为的基础
- ✓ 智力：是获取知识并应用知识求解问题的能力

——智能是知识与智力的总和。

► 智能的特征：

- ✓ 感知能力：视觉、听觉、触觉、嗅觉等感觉器官感知外部世界的能力
- ✓ 记忆能力：存储由感知器官感知到的外部信息以及由思维所产生的知识
- ✓ 思维能力：逻辑思维（抽象思维）、形象思维（直感思维）、顿悟思维（灵感思维）
- ✓ 学习能力：自觉的 / 不自觉的；有教师指导的 / 通过自己实践的
- ✓ 行为能力：表达能力，是信息的输出



第1讲 绪论

1.1 什么是人工智能

➤ 人工智能：用人工的方法在机器上实现的机器智能

1956年的“达特茅斯夏季人工智能研究计划”会议，是AI赢得其名称、使命和主要参与者的时刻，因此被广泛地认为是AI的诞生标志。

➤ 人工智能的定义：

➤ 从内在智能特征出发

➤ 从外在类人智能表现

➤ 人工智能学科：

人工智能是一门研究如何构造智能机器或智能系统，使它能够模拟、延伸、扩展人类智能的学科。



第1讲 绪论

1.1 什么是人工智能

纵向维度:

- 思考: 关注思维过程与推理
- 行动: 强调行为

横向维度:

- 类人地: 以对人类表现的逼真度来衡量
- 理性地: 用理想的性能表现来衡量

一个系统如果对已知的知识做出正确的动作, 则被称为理性。

	类人地 Humanly	理性地 Rationally
思考	像人一样思考	合理地思考
行动	像人一样行动	合理地行动



第1讲 绪论

1.1 什么是人工智能

1.2 人工智能的基础

1.3 人工智能的历史

1.4 人工智能的判定

1.5 人工智能的层级与流派

1.6 人工智能的最新发展水平



第1讲 绪论

1.2 人工智能的基础

- 哲学 (Philosophy)
- 数学 (Mathematics)
- 经济学 (Economics)
- 神经科学 (Neuroscience)
- 心理学 (Cognitive psychology)
- 计算机科学 (Computer science)
- 控制理论和控制论 (Control theory and cybernetics)
- 语言学 (Linguistics)



第1讲 绪论

1.2 人工智能的基础——数学

➤ 逻辑学：得出正确结论的形式规则是什么？

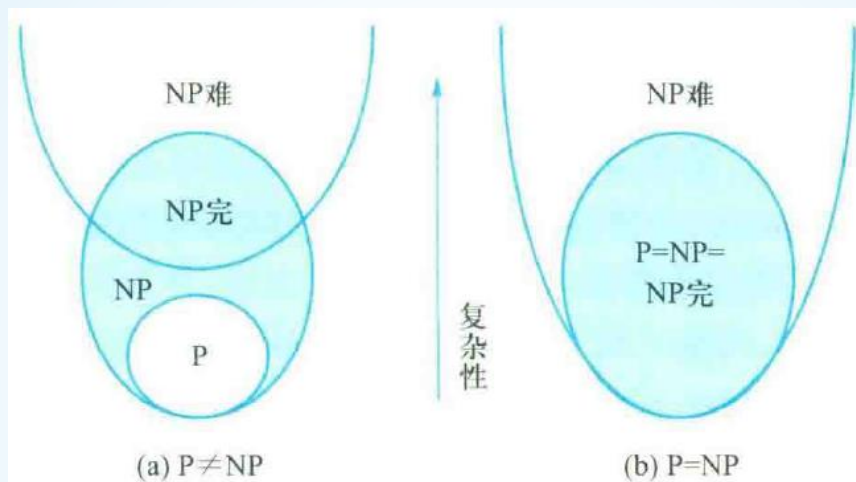
- 命题逻辑 (George Boole, 1847) , 又称布尔逻辑
- 一阶逻辑 (Gottlob Frege, 1879) , 扩展了布尔逻辑, 增加了对象和关系
- 关联理论 (Alfred Tarske, 1902-1983) , 又称指称理论, 揭示如何将逻辑中的对象与现实世界的对象联系起来

➤ 概率论：是一门关于不确定问题的数学理论

- 概率 (Gerolamo Cardano, 1501-1576)
- 统计学方法的引入: James Bernoulli (1654-1705), Pierre Laplace (1749-1827)
- 贝叶斯规则: Thomas Bayes (1702-1761) 提出了根据新证据更新概率的规则, 即贝叶斯规则, 构成了人工智能系统中大多数用于不确定推理的现代方法的基础。

第1讲 绪论

1.2 人工智能的基础——数学



术语	含义
P (Polynomial time)	多项式时间内可计算的问题
NP (Non-deterministic Polynomial time)	无法确定多项式时间内是否可计算
NP完 (NP-complete)	NP中最难的问题
NP难 (NP-hard)	至少是NP中最难的问题

例 3.5 汉诺塔(tower of Hanoi)问题

传说印度的婆罗门庙里有 3 根柱子,第一根柱子上有 64 块金盘,要通过第二根

84 ▶ 第 3 章 搜索问题求解



柱子将这些金盘移动到第三根柱子上,如图 3.8 所示。移动的规则很简单:① 每次仅能移动一块金盘,② 仅能移动最上面的金盘,③ 大的金盘不能放在小的金盘上面。寺庙里的祭司一直在移动这些金盘,据说当移动最后一块金盘时,世界将会毁灭!



图 3.8 汉诺塔问题

我们估算一下移动金盘所需要的时间。设每秒能移动一块金盘,则移动所有金盘所需时间为 $2^{64}-1$ s,约等于 5,850 亿年!

而地球的天文年龄、即地球开始形成到现在,仅有 45.5 亿年。



第1讲 绪论

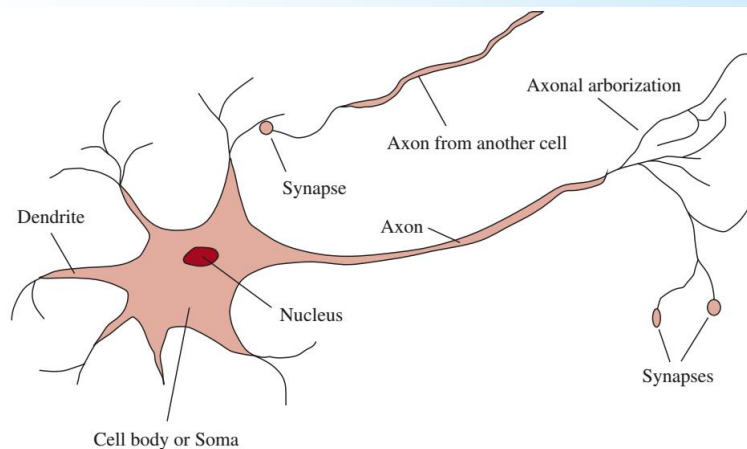
1.2 人工智能的基础——经济学

- 经济学：研究人们如何做出能导致更喜欢的结果的选择
 - 如何决策以便收益最大？
 - 当其他人不合作时我们该如何？
 - 当收益遥遥无期时我们该如何？
- 经济学观点：效用、满意
- 经济学方法：
 - 运筹学：规划论（线性规划、动态规划）、排队论、存储论、图论
 - 决策论：确定型、不确定型、随机型决策，马尔科夫决策过程
 - 博弈论：理性主体之间的冲突与合作

第1讲 绪论

1.2 人工智能的基础——神经科学

- 神经科学：大脑如何处理信息？
 - 神经科学研究神经系统，尤其是大脑
 - 大脑具有决策功能，预测和仿真是决策的关键
- 大脑的结构与连接对构建智能系统的启示
 - 结构要素：神经元具有树突 (Dendrite) 和轴突 (Axon)，它由细胞体 (Soma) 和细胞核 (Nucleus) 组成
 - 工作机制：每个神经元的树突可以接受刺激并传入细胞体，轴突可以把信号传送到另一个神经元或其他组织





第1讲 绪论

1.2 人工智能的基础——认知心理学

- 认知心理学：研究人类如何思维与行动。
 - 注意力 (attention) 机制
 - 记忆 (memory)
 - 感知 (perception)
 - 语言运用 (language use)
 - 问题求解 (problem solving)
 - 创造力 (creativity)
 - 思维 (thinking) 等
- 认知科学：是包括心理学在内的一门交叉学科。



第1讲 绪论

1.2 人工智能的基础——控制论

- 控制理论与控制论：机器如何能在其自身的控制下运行？
- 控制理论 Control theory
 - 工程与数学的交叉学科分支
 - 处理动态系统对输入的行为，以及该行为如何通过反馈进行调整
- 控制论 Cybernetics
 - 跨学科的研究途径，探索调控系统、它们的结构、约束、和可能性
 - 控制论，1948年被定义为“研究动物与机器的控制与通信的科学。”
 - 21世纪，该术语通常被简单通俗地解释为“用技术控制任何系统”



第1讲 绪论

1.1 什么是人工智能

1.2 人工智能的基础

1.3 人工智能的历史

1.4 人工智能的判定

1.5 人工智能的层级与流派

1.6 人工智能的最新发展水平



第1讲 绪论

1.3 人工智能的历史

时间	阶段	重要事件/技术
1950-1956	人工智能的孕育	图灵测试
1956	人工智能正式诞生	达特茅斯会议
1957-1973	人工智能黄金之年	
1974-1979	第一个人工智能之冬	
1980-1986	人工智能的繁荣期	专家系统和神经网络
1987-1992	第二个人工智能之冬	
1993-2005	人工智能的第二次繁荣	互联网普及
2006-现在	人工智能的突破	深度学习和大数据



第1讲 绪论

1.3 人工智能的历史——孕育

➤ 远古的梦想

- 亚里士多德（公元前384~322年）演绎推理
- 列子（公元前450~375年）《列子·汤问》中的偃师献技的寓言典故

➤ 人工智能的曙光

- 数理逻辑的突破
- 计算机的发明
- 控制论、信息论等的建立
- 对人类大脑的探究.....

正是二十世纪30年代到50年代初的一系列科学进展的交汇融合，孕育了人工智能研究

第1讲 绪论

1.3 人工智能的历史——诞生

- ▶ **达特茅斯会议** 1956年夏，围绕机器智能展开两个月的学术研讨



达特茅斯会议（1956）
Dartmouth Conference

- ▶ 正式提出了“人工智能 Artificial Intelligence”这一术语，标志着人工智能作为一个研究领域的诞生。



第1讲 绪论

1.3 人工智能的历史——黄金年代

- 1956, 赫伯特·西蒙和艾伦·纽厄尔演示了第一个AI程序, 名称为逻辑理论家
- 1958, 约翰·麦卡锡发明了著名的Lisp编程语言
- 1960年代, M·马斯特曼与剑桥大学的同事们设计了语义网络, 用于机器翻译
- 1963, 伦纳德·武赫和查尔斯·瓦斯勒发表了关于模式识别的论文, 描述了第一个机器学习程序
- 1960年代, 马文·明斯基指导完成微观世界的有限智能问题
 - SAINT程序 (1963) 求解典型的闭合式微积分问题
 - ANALOGY程序 (1968) 求解智商测试中的几何类推问题
 - STUDENT程序 (1967) 求解代数故事问题
- 1966, 约瑟夫·魏泽鲍姆开发出第一个聊天机器人ELIZA, 用于在临床治疗中模仿心理医生



第1讲 绪论

1.3 人工智能的历史——第一个冬天

现实的困难：

- 早期的程序仅依靠简单的句法处理获得成功，但对其主题一无所知
- 求解能力的限制，包括算力的问题和算法的问题
- 用来产生智能行为的基本结构存在着根本局限

人工智能的第一次寒冬来临

- 1966，机器翻译失败了
- 1970，**连接主义**遭到遗弃
- 1971-1975，美国DARPA（美国高级研究计划局）对卡内基梅隆大学的语音理解研究项目感到沮丧
- 1973，受莱特希尔的“人工智能：综合调查”报告的影响，英国大幅度缩减AI的研究
- 1973-1974，美国DARPA**消减了一般性AI学术研究经费**

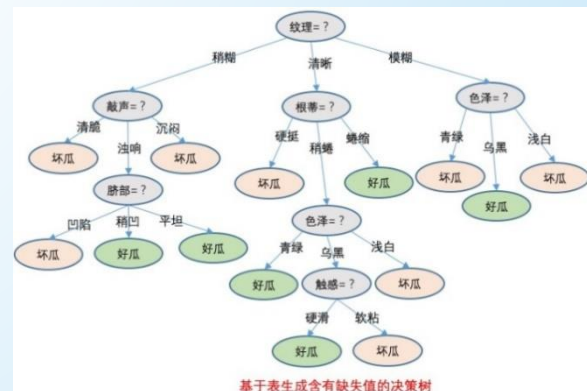
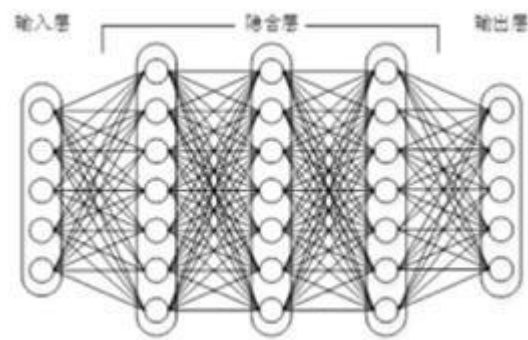
第1讲 绪论

1.3 人工智能的历史——第一次繁荣

- 1980, 兴起了**基于知识的专家系统**
- 1982, 启动了第五代计算机项目用于知识处理

代	年份	特征
第一代	1940年代中期	真空管
第二代	1956	晶体管
第三代	1964	集成电路
第四代	1972	微处理器

- 1980年代中期, 发明了**多层人工神经网络** (ANN) 具有足够多的隐藏层, 一个ANN可以表达任意的功能, 突破了感知的局限
- 1980年代中期, 出现了**机器学习**, 发明了决策树模型并且以软件形式推出。该模型具有可视化、易说明的特点

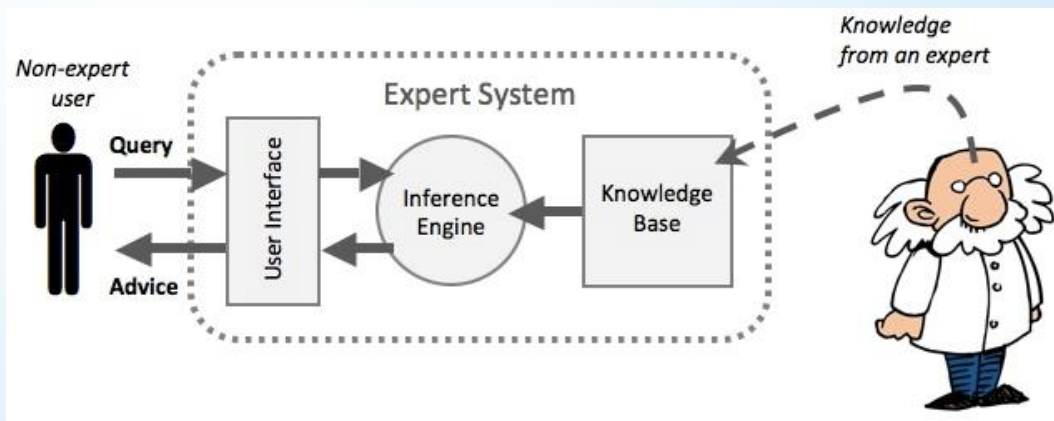


第1讲 绪论

1.3 人工智能的历史——第二个冬天

再次入冬

- 1987年, Apple/IBM台式机性能提升, 昂贵的Lisp机市场出现崩溃;
- 1988年, 专家系统维护成本过高, 美国政府的战略计算促进会取消了新的AI经费。
- 1992年, 日本第五代计算机项目历经十一年后, 悄然退场。
- 1993年, 专家系统缓慢滑向低谷。



第1讲 绪论

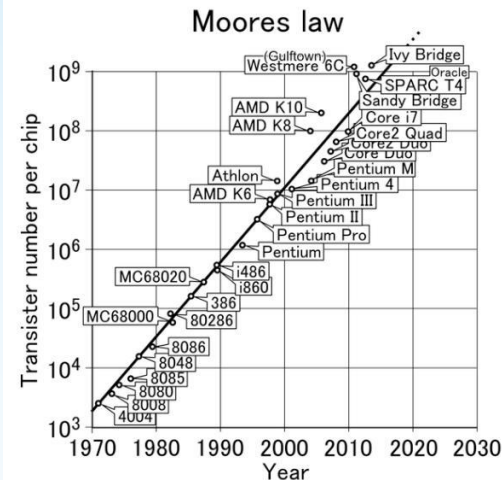
1.3 人工智能的历史——第二次繁荣

新的曙光:

- 新的数学工具
- 新的理论;
- 摩尔定律

代表成果:

- 1997, IBM研制的超级电脑“深蓝”战胜了国际象棋世界冠军加里·卡斯帕罗夫, 震惊世界;
- 2005, 斯坦福的自主机器人车辆Stanley, 赢得了DARPA无人驾驶汽车挑战赛。





第1讲 绪论

1.3 人工智能的历史——突破

- **人机大战**：2011年IBM另一超级电脑Waston在智力竞猜节目“危险边缘”第三场比赛中，以三倍的巨大分数优势力压该竞猜节目有史以来最强的二位选手詹宁斯和鲁特，夺得人机大战的冠军
- **图像识别**：2015年的ImageNet挑战赛中，机器在图像识别准确率上的表现首次超过了人类。（错误率3.5%低于人类的4%左右，2010年机器的错误率是25%）2016年机器的错误率下降到2.99%
- **语音识别**：苹果的“Siri”、微软的“小冰”、百度的“度秘”
- **决策智能**：2016年Google的AlphaGo在人机围棋大赛中以4:1战胜顶级选手李世石。2018年OpenAI在Dota2上战胜人类职业选手。
- **仿生机器人**：美国MIT的猎豹机器人、美国宾夕法尼亚大学的Rhex六足机器人、日本仿人机器人ASIMO、波士顿动力仿人机器人Atlas



第1讲 绪论

1.1 什么是人工智能

1.2 人工智能的基础

1.3 人工智能的历史

1.4 人工智能的判定

1.5 人工智能的层级与流派

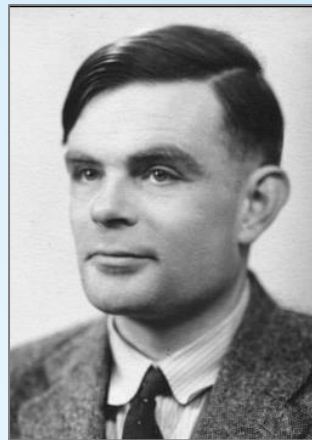
1.6 人工智能的最新发展水平

第1讲 绪论

1.4 人工智能的判定

图灵测试：

1950年艾伦·图灵（计算机之父）发表的《计算机与智能》中设计了一个用以研判“机器能否思维”测试，成为判定人工智能的经典方法。Imitation Game



艾伦·图灵

- 一个人类的裁判与被测试者（包括一个人和一台机器）相互隔离。裁判与被测试的人和机器分别以通讯线路连接起来，并且利用键盘和显示屏进行随意交谈。
- 经过一系列的测试后，如果裁判无法分辨这些回答究竟是来自于人还是机器，则认为该机器通过了测试，具有人类的思维能力。



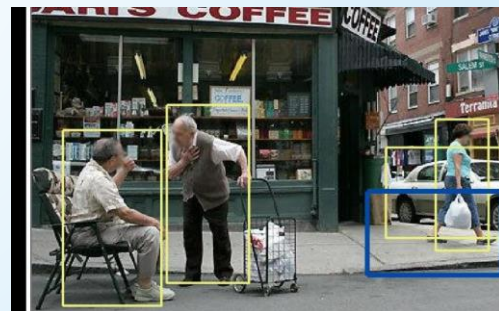
第1讲 绪论

1.4 人工智能的判定

➤ 图灵测试2.0

- 反向图灵测试
- 学科专家图灵测试
- 总体图灵测试
- 最低智能信号测试
- 艾伯特测试
- 视觉图灵测试

- | | |
|--|------------------|
| 1. Q: Is there a person in the designated region? | A: yes |
| 2. Q: Is there a unique person in the designated region? | A: yes (person1) |
| 3. Q: Is person1 carrying something? | A: yes |
| 4. Q: Is person1 female? | A: yes |
| 5. Q: Is person1 walking on a sidewalk? | A: yes |
| 6. Q: Is person1 interacting with any other object? | A: no |



第1讲 绪论

1.4 人工智能的判定

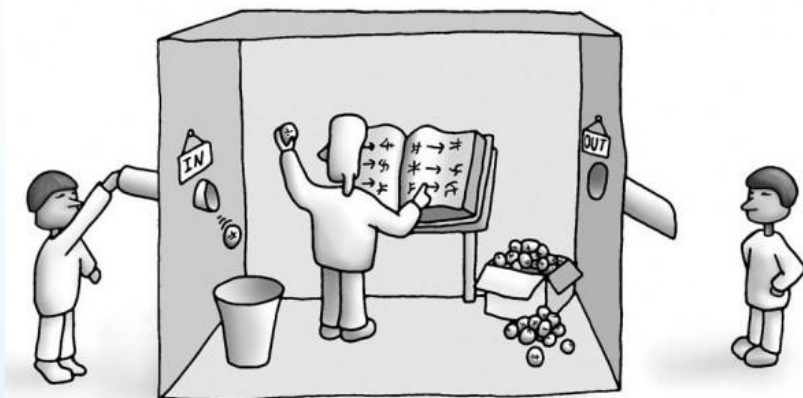
中文屋：

1980年约翰·塞尔发表在《行为与大脑科学》上的论文“智力、头脑与规划”中提出，试图揭示计算机绝不能描述为有“智力”或“知性”，不管它多么智能。



约翰·塞尔

- 一位只会说英语的人独自在屋内，房屋除了门上有孔外，全部都是封闭的。屋里有一本中英对照的翻译指南，以及足够的纸、笔等。
- 写着中文的纸条从一侧的小孔被送入屋内，屋中的人可以使用他的书来翻译这些文字并用中文回复，从而让屋外的人以为屋内有一个懂中文的人。





第1讲 绪论

1.1 什么是人工智能

1.2 人工智能的基础

1.3 人工智能的历史

1.4 人工智能的判定

1.5 人工智能的层级与流派

1.6 人工智能的最新发展水平



第1讲 绪论

1.5 人工智能的层级与流派

► 人工智能的三个层级

- 人工专用智能 (Artificial Narrow Intelligence, ANI)
 - 也被称为**弱人工智能**, 狭义人工智能, 应用人工智能
 - 专注于一个具体的任务 (仅针对一个特定的问题)
- 人工通用智能 (Artificial General Intelligence, AGI)
 - 也被称为**强人工智能**, 广义人工智能, 普适AI, 完全AI
 - 能够将智能应用于任何问题, 而不仅仅是一个特定的任务。意味着机器具有与人类智能相当的处理任何问题的能力, 它是人工智能研究的主要目标
- 人工超级智能 (Artificial Super Intelligence, ASI)
 - 也被称为**超人工智能**, 超智能 (Hyperintelligence)、以及超人智能 (Superhuman intelligence)
 - 是一种假想中的智能体, 拥有远远超过人类大脑的智能
 - 热衷于讨论人工超级智能的主要是**未来学家和预言家**

第1讲 绪论

1.5 人工智能的层级与流派

► 概念的三个功能

- **指物功能**：指向客观世界的对象
- **指心功能**：指向人心世界里的对象
- **指名功能**：指向符号世界表示对象的符号

► 人工智能的三个流派

- **符号主义**：其原理主要为物理符号系统假设和有限合理性原理。即只要在符号计算上实现了相应的功能，那么在现实世界就实现了对应的功能。（**指名功能**）
- **连接主义**：主要关注于**大脑神经元及其连接**机制，试图揭示人类智能的本质机理，进而在机器上实现相应的模拟，如学习算法。（**指心功能**）
- **行为主义**：假设**智能取决于感知和行动**，不需要知识、表示和推理。而且认为人工智能可以像人类智能一样逐步进化。（**指物功能**）



第1讲 绪论

1.1 什么是人工智能

1.2 人工智能的基础

1.3 人工智能的历史

1.4 人工智能的判定

1.5 人工智能的层级与流派

1.6 人工智能的最新发展水平



第1讲 绪论

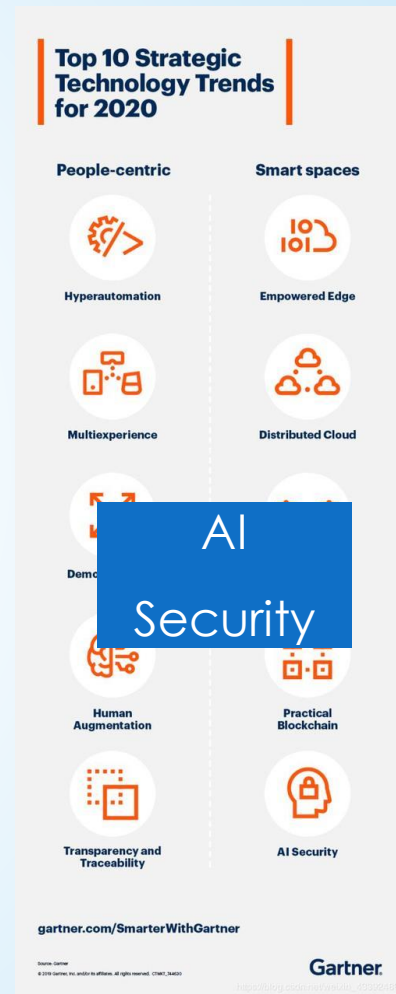
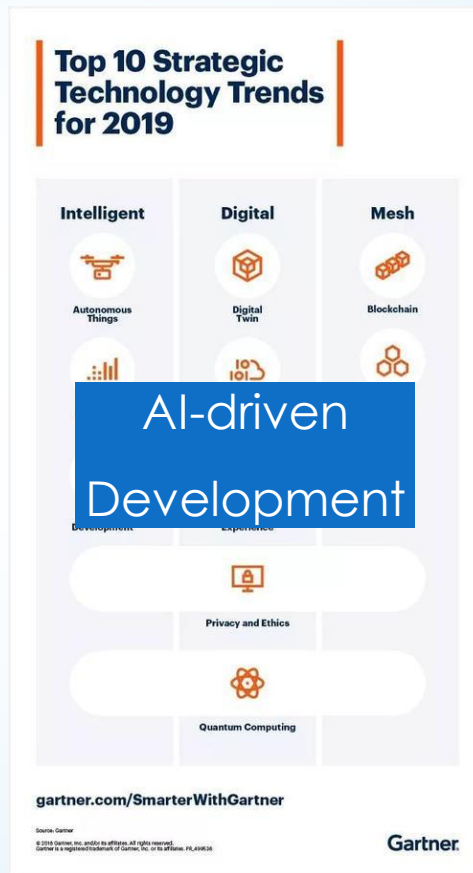
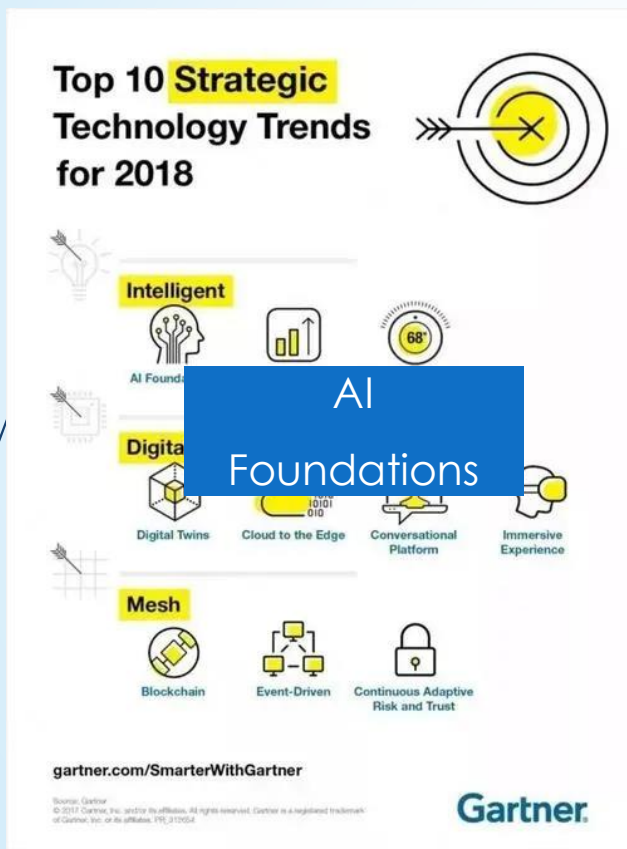
1.6 人工智能的最新发展水平

➡ 人工智能的广泛应用领域

- ➡ 机器人和机器人学
- ➡ 语音识别、机器翻译、自然语言处理
- ➡ 机器视觉、模式识别、图像处理
- ➡ 自主规划、策略规划
- ➡ 智能调度与指挥
- ➡ 智能诊断与智能决策
- ➡ 产品：ChatGPT
- ➡ 产品：文心一言

第1讲 绪论

1.6 人工智能的最新发展水平

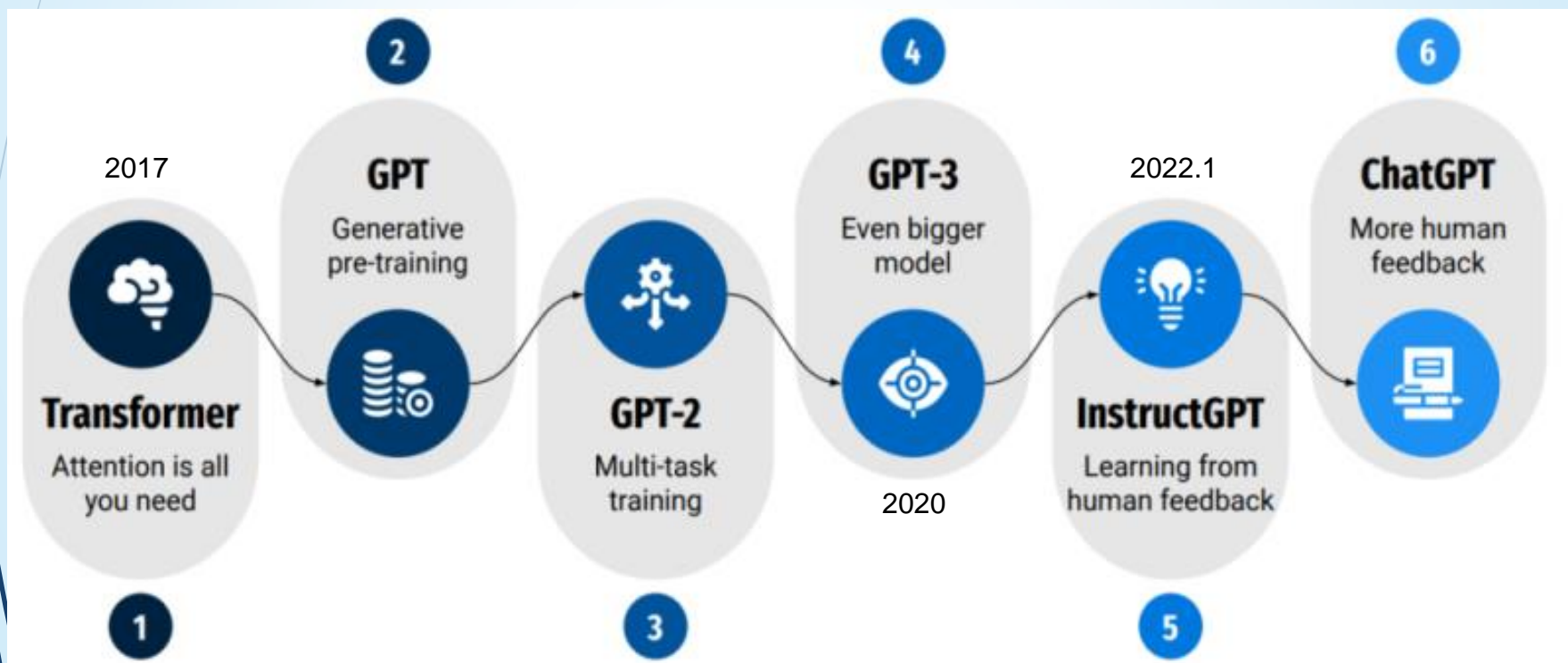




第1讲 绪论

1.6 人工智能的最新发展水平

► ChatGPT的发展历程





第1讲 绪论

1.6 人工智能的最新发展水平

➡ 大模型参数量





第1讲 绪论

1.6 人工智能的最新发展水平

► ChatGPT的广泛应用领域

- 生成代码、生成文本...
- 归纳性的文字类工作
- 智能客服
- 合成人声、仿声
- 人脸生成、人脸替换



第1讲 绪论

1.6 人工智能的最新发展水平

国内大模型应用案例

- 基于电力行业NLP大模型的设备运检知识助手示范应用
- 数字中医大模型示范应用
- 面向建筑领域多模态行业大模型示范应用
- 多模态智慧金融大模型示范应用
- 自动驾驶大模型DriveGPT示范应用

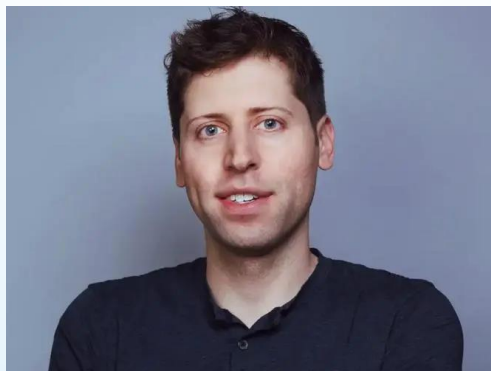


第1讲 绪论

1.6 人工智能的最新发展水平

➡ 大模型风险

- ➡ 数据：恶意数据、偏见、误导数据
- ➡ 模型：不可解释性、知识涌现、测试难以穷举；信息泄露
- ➡ 用户：黑客、误导APP、造假诈骗



阿尔特曼：“我特别担心这些模型可能会被用于大规模的虚假信息，现在他们在编写计算机代码方面做得越来越好，可以用于进攻性网络攻击。”

课程要求与目标

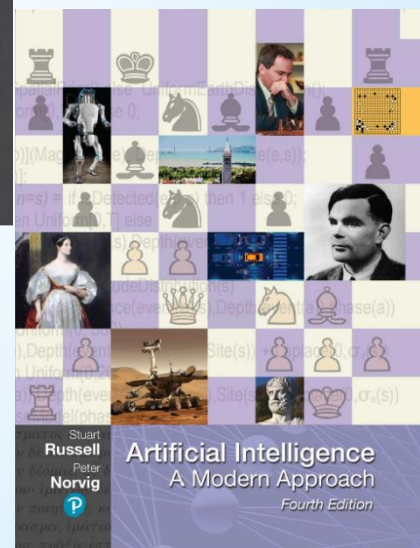
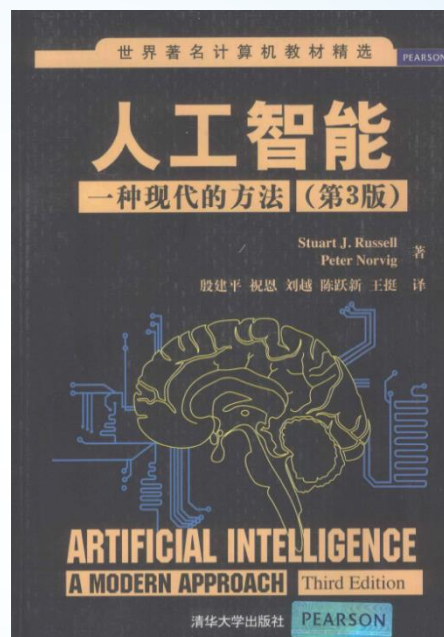
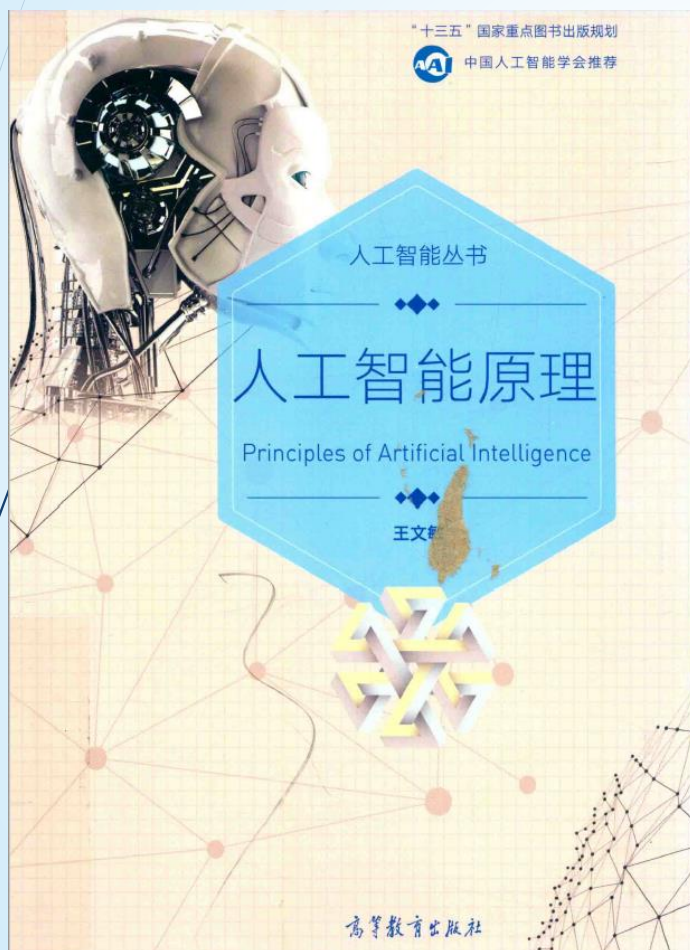
➤ 课程基本要求

面向人工智能和自动化专业高年级本科生，介绍人工智能的基本理论、基本方法 and 应用，使学生了解人工智能各主要分支领域的研究、掌握其核心技术原理与方法，包括**基于搜索的问题求解、基于优化的规划、基于数据的学习、基于知识的推理**等板块。

➤ 课程教学目标

- **了解AI学科范畴，理解AI领域基本概念、原理和算法；**
- 能够结合典型问题，运用人工智能技术进行单元级的**设计与开发**；
- 增强面向问题的文献检索、阅读，进而**跟踪AI前沿进展的能力**；
- 能够认同**AI伦理学**的基本概念、具有综合考虑AI应用中社会、安全、隐私等影响的**意识**。

课程教学教材



课程教学安排

周	内容	学时	周	内容	学时
1	绪论	2	10	学习概述和学习任务	2
2	体系论	2	11	学习范式	2
3	搜索问题求解	2	12	学习框架和人工神经网络	2
4	优化问题求解	2	13	知识表示	2
5	博弈问题求解	2	14	推理机制	2
6	约束问题求解	2	15	不确定性推理	2
7	经典规划	2	16	人工智能应用及AI伦理	2
8	现实世界的规划	2	17	小组大作业交流	2
9	(期中考试)	2	18	(期末考试)	2
					36

课程考核

➡ 课程考核

- ➡ 平时作业（按时+质量）20%，课堂参与评价10%
- ➡ 综合练习（小组大作业）报告与交流讨论20%
- ➡ （期中+）期末考试 50%

➡ 项目综合练习（小组大作业，第5周后布置）

- ➡ 分组（~4人每组）完成；
- ➡ 课堂报告交流-PPT版项目介绍
- ➡ 期末提交大作业报告-word版项目报告
- ➡ 软件实现和系统开发-程序代码、演示和测试

作业

试说明一款现实生活中的AI产品，并分析其智能体现、智能类型、智能提升空间。

