

第 1 章人工智能概论

教材：

王万良《人工智能应用教程》（2版）

清华大学出版社，2025

第1章 人工智能概论

□ 1956 年正式提出人工智能
(artificial intelligence, AI)
这个术语并把它作为一门新兴
科学的名称。

□ 20世纪三大科学技术成就：

{ 空间技术
原子能技术
人工智能



第1章 人工智能概论

- 1.1 你了解人类的智能吗
- 1.2 人工智能的孕育和诞生
- 1.3 人工智能的发展
- 1.4 从两场标志性人机博弈看人工智能的发展
- 1.5 人工智能研究的基本内容
- 1.6 人工智能的三大学派
- 1.7 人工智能的主要应用领域
- 1.8 人工智能伦理

第1章 人工智能概论

- 1.1 你了解人类的智能吗
- 1.2 人工智能的孕育和诞生
- 1.3 人工智能的发展
- 1.4 从两场标志性人机博弈看人工智能的发展
- 1.5 人工智能研究的基本内容
- 1.6 人工智能三大学派
- 1.7 人工智能的主要应用领域
- 1.8 人工智能伦理

1.1 你了解人类的智能吗

- 人类之所以主宰地球，就是因为我们的祖先早就有了比较高级的智能，但谁也不知道我们人类的智能是怎么产生的？
- 智能及智能的本质是古今中外许多哲学家、脑科学家一直在努力探索和研究的问题，但至今仍然没有完全了解。
- 智能的发生与物质的本质、宇宙的起源、生命的本质一起被列为自然界的四大奥秘。

1.1.1 智能的概念

●对智能还没有确切的定义，主要流派有：

（1）思维理论：智能的核心是思维

（2）知识阈值理论：智能取决于知识的数量及一般化程度

（3）进化理论：智能由系统总的行为以及行为与环境的联系所决定。

●智能是知识与智力的总和

知识是一切智能行为的基础

获取知识并应用知识求解问题的能力

1.1.2 智能的特征

1. **感知能力**：通过视觉、听觉、触觉、嗅觉等感觉器官感知外部世界的能力。

80%以上信息通过视觉得到，10%信息通过听觉得到。

2. **记忆与思维能力**

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

对记忆的信息进行处理

1.1.2 智能的特征

(1) 逻辑思维（抽象思维）

- 依靠逻辑进行思维。
- 思维过程是串行的。
- 容易形式化。
- 思维过程具有严密性、可靠性。

(2) 形象思维（直感思维）

- 依据直觉。
- 思维过程是并行协同式的。
- 形式化困难。
- 在信息变形或缺少的情况下仍有可能得到比较满意的结果。

1.1.2 智能的特征

(3) 顿悟思维（灵感思维）

- 2000多年前，希腊希洛王向永垂不朽的神献上金制的皇冠。希洛王让阿基米德鉴别皇冠是否为纯金制成。在那个时代皇冠的体积难以得到。
- 有一天，阿基米德泡入盛满水的澡盆中洗澡时，发现水从澡盆中溢了出来，突然想到盆里溢出来的水的体积，不就是自己的身体浸入水里的那一部分体积吗？
- 阿基米德由此揭开了金冠之谜，断定皇冠是掺了假的。
- 他发现了著名的浮力定律，开创了船舶浮沉理论。

1.1.2 智能的特征

- 在科学发现史上有很多“灵感”故事，例如：
- 1830年奥斯特在指导学生实验时，看见电流能使磁针偏转，从而发现了电磁关系。
- 1928年8月2日，Harold Black(1898-1983)，在前往曼哈顿西街的上班途中，在Hudson河的渡船上灵光一闪，发明了在当今控制理论中占核心地位的负反馈放大器。他将其发明记在了一份纽约时报上。这份早报珍藏在AT&T的档案馆中。

顿悟思维（灵感思维）的特点：

- 不定期的突发性。
- 非线性的独创性及模糊性。
- 穿插于形象思维与逻辑思维之中。

1.1.2 智能的特征

3. 学习能力

学习既可能是自觉的、有意识的，也可能是不自觉的、无意识的；既可以是有教师指导的，也可以是通过自己实践的。

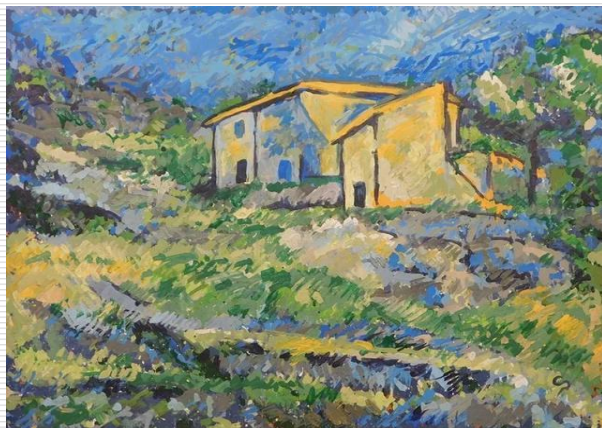
4. 行为能力（表达能力）

人们的感知能力：用于信息的输入。

行为能力：信息的输出。



机器人放牧



机器人作画



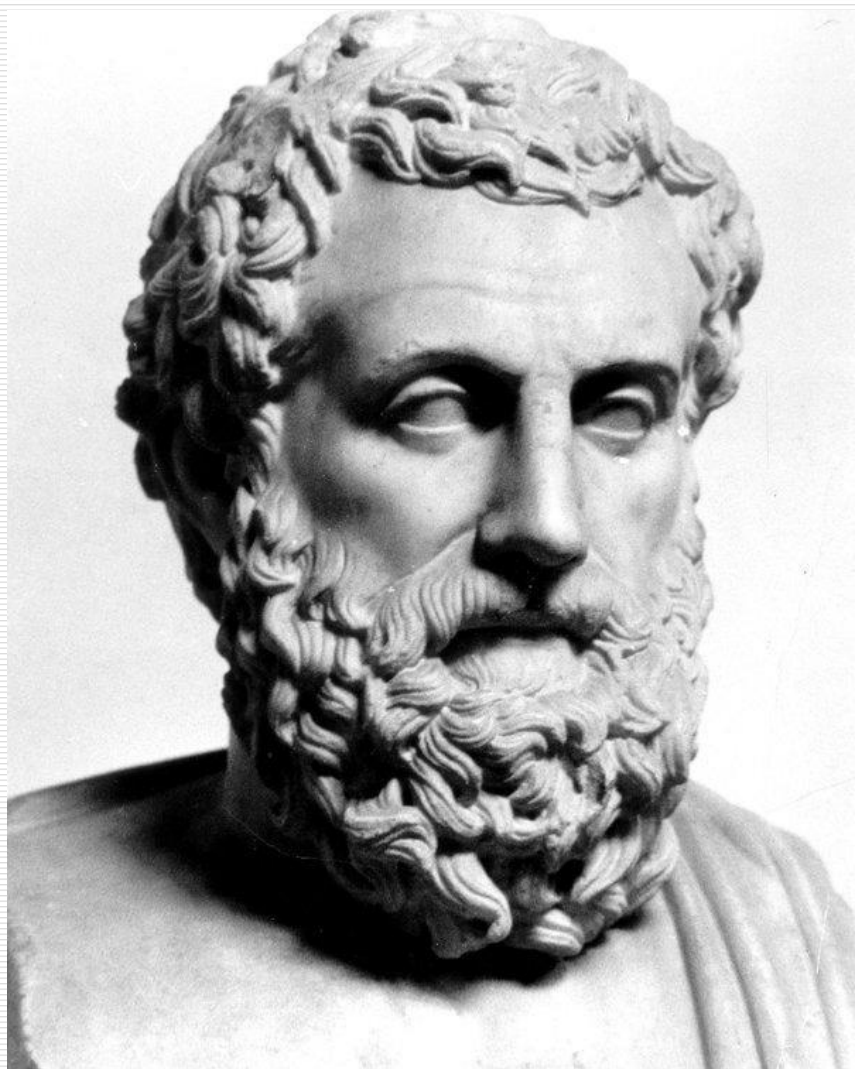
机器人装配

第1章 人工智能概论

- 1.1 你了解人类的智能吗
- 1.2 人工智能的孕育和诞生
- 1.3 人工智能的发展
- 1.4 从两场标志性人机博弈看人工智能的发展
- 1.5 人工智能研究的基本内容
- 1.6 人工智能的三大学派
- 1.7 人工智能的主要应用领域
- 1.8 人工智能伦理

1.2.1 人工智能的孕育期

■早在公元前，伟大的哲学家和思想家亚里斯多德提出的三段论至今仍是演绎推理的基本依据。



1.2.1 人工智能的孕育期

- ✿ 英国哲学家**培根**（**F. Bacon**）：归纳法
- ✿ **莱布尼茨**（**G. W. Leibnitz**）：万能符号、推理计算
- ✿ 英国逻辑学家**布尔**（**G. Boole**）：用符号语言描述思维活动的基本推理法则
- ✿ 英国**图灵**（**A. M. Turing**）：图灵机（**1936年**）
- ✿ 美国神经生理学家**麦克洛奇与匹兹**：**1943年**建成第一个神经网络模型（**M-P模型**），为后来人工神经网络的研究奠定了基础。

1.2.1 人工智能的孕育期

● 现代计算机为AI提供强大的物质基础

- ✿ 美国爱荷华州立大学阿塔纳索夫在1937年至1941年间开发的第一台世界上第一台电子计算机：阿塔纳索夫-贝瑞计算机（ABC）。
- ✿ 不是美国数学家莫克利和埃柯1946年发明的！



阿塔纳索夫



贝瑞 15

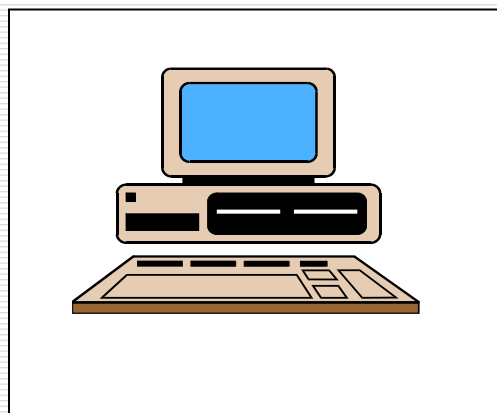
1.2.2 人工智能的诞生——达特茅斯会议

- 1956年夏，当时美国达特茅斯大学数学助教、现任斯坦福大学教授麦卡锡和哈佛大学数学和神经学家、现任MIT教授明斯基、IBM公司信息研究中心负责人洛切斯特、贝尔实验室信息部数学研究员香农共同发起，邀请普林斯顿大学莫尔和IBM公司塞缪尔、MIT的塞尔夫里奇和索罗莫夫以及兰德公司和卡内基—梅隆大学的纽厄尔、西蒙等10名年轻学者在达特莫斯大学召开了两个月的学术研讨会，讨论机器智能问题。
- 会上麦卡锡提议正式采用“人工智能”这一术语，标志着人工智能学科正式诞生。
- 麦卡锡因而被称为人工智能之父。

1.2.3 人工智能的定义与图灵测试

■ **人工智能**：用人工的方法在机器（计算机）上实现的智能；或者说是人们使机器具有类似于人的智能。

■ **图灵测试**：1950年图灵发表的《计算机与智能》中设计了一个测试，用以说明人工智能的概念。



智者



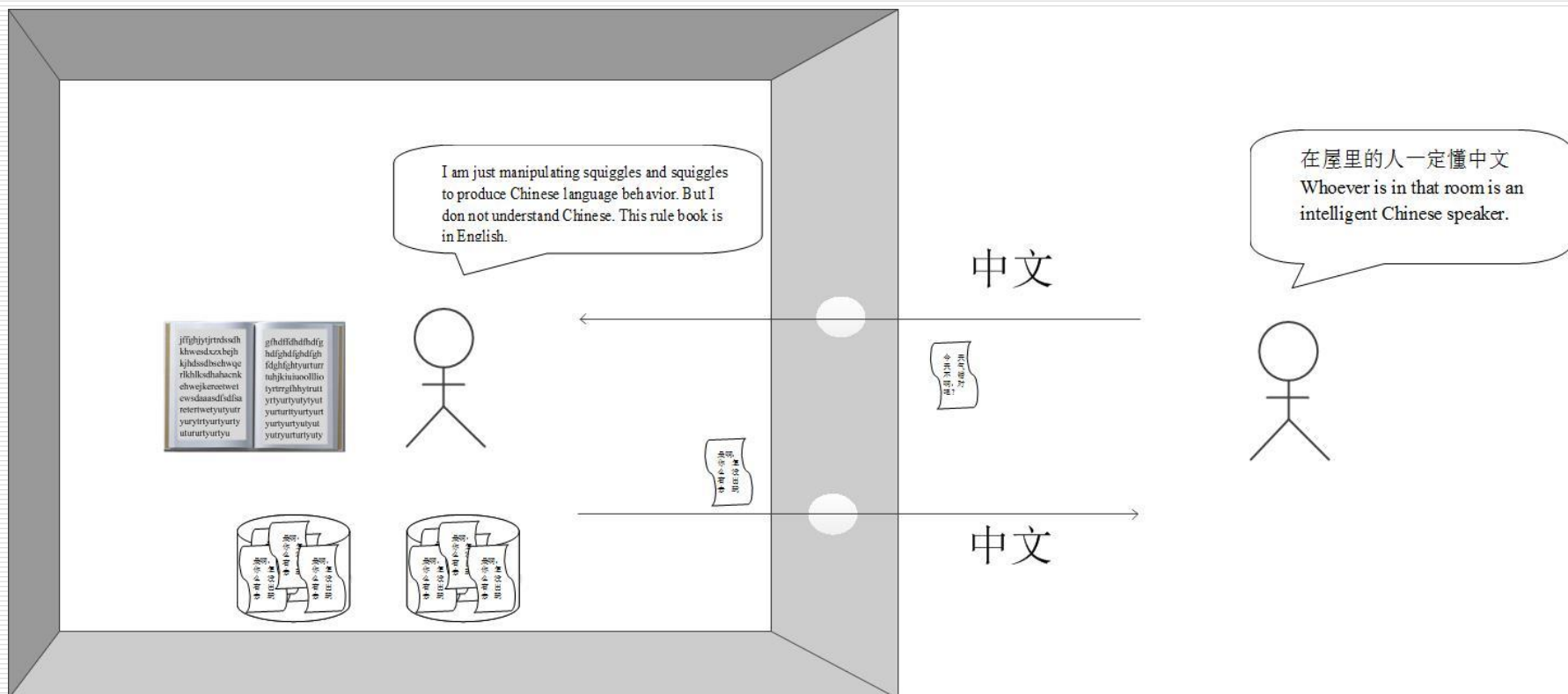
询问者



1.2.3 人工智能的定义与图灵测试

● 中国屋思想实验

语言哲学家约翰.R.塞尔 (John R. Searle, 1980)



第1章 人工智能概论

- 1.1 你了解人类的智能吗
- 1.2 人工智能的孕育和诞生
- 1.3 人工智能的发展
- 1.4 从两场标志性人机博弈看人工智能的发展
- 1.5 人工智能研究的基本内容
- 1.6 人工智能的三大学派
- 1.7 人工智能的主要应用领域
- 1.8 人工智能伦理

1.3.1人工智能的形成期

- 达特茅斯会议之后，美国形成了多个人工智能研究组织，如纽厄尔和西蒙的Carnegie RAND协作组，明斯基和麦卡锡的MIT研究组，塞缪尔的IBM工程研究组等。
- 1956年以后，人工智能的研究在机器学习、定理证明、模式识别、问题求解、专家系统及人工智能语言等方面都取得了许多引人瞩目的成就。
- 机器学习：1957年Rosenblatt研制成功了感知机。
- 定理证明：美籍华人数理逻辑学家王浩1958年在IBM-704机器上用3~5min证明了《数学原理》中有关命题演算的全部定理（220条），并且还证明了谓词演算中150条定理的85%；
- 1965年鲁宾逊（J. A. Robinson）提出了归结原理，为定理的机器证明做出了突破性的贡献。

1.3.1人工智能的形成期

□**专家系统**：美国斯坦福大学的费根鲍姆（E. A. Feigenbaum）1965年开始专家系统DENDRAL的研究，1968年完成并投入使用，已接近于、甚至超过有关化学专家的水平，在美、英等国得到了实际的应用，为以后专家系统的建造树立了榜样，对人工智能的发展产生了深刻的影响，其意义远远超过了系统本身在实用上所创造的价值。

□**人工智能语言**：1960年麦卡锡研制出了人工智能语言LISP，成为建造智能系统的重要工具。

1.3.1 人工智能的形成期

- 1969年，成立了国际人工智能联合会议（International Joint Conferences on Artificial Intelligence, IJCAI）。
- 1970年，创刊了国际性的人工智能杂志（Artificial Intelligence）。
- 1978年，我国开始把“智能模拟”作为国家科技发展规划的主要研究课题。
- 1981年成立了中国人工智能学会（CAAI）。

1.3.2 几起几落的曲折发展期

- ❑ 20世纪60年代末，人工智能研究遇到困难，如机器翻译。1966年美国顾问委员会裁定：还不存在通用的科学文本机器翻译，也没有很近的实现前景。英国、美国中断了大部分机器翻译项目的资助。
- ❑ 1977年，费根鲍姆在第五届国际人工智能联合会议上提出了“知识工程”概念，推动了知识为中心的研究。
- ❑ 1986年之后也称为集成发展时期。计算智能（CI）弥补了人工智能中在数学理论和计算上的不足，使人工智能进入了一个新的发展时期。

1.3.3 大数据驱动的飞速发展期

- 随着大数据、云计算、物联网等信息技术的发展，以及深度学习的提出，人工智能在算法、算力和算料（数据）等“三算”方面取得了重要突破，直接支撑了图像分类、语音识别、知识问答、人机对弈、无人驾驶等人工智能的复杂应用，人工智能进入以深度学习为代表的大数据驱动人工智能发展期。
- 2006年，Hinton等根据生物学的重要发现，提出了著名的深度学习方法。
- 2017年10月，Hinton等进一步提出了胶囊网络。
- 2019年，牛津大学博士生Adam R. Kosiorek等提出了堆叠胶囊自动编码器（SCAE）。深度学习创始人、图灵奖得主Hinton称赞它是一种非常好的胶囊网络新版本。

1.3.3 大数据驱动的飞速发展期

- ❑ 人工智能大体可分为专用人工智能和通用人工智能。
- ❑ 目前主要是面向特定任务的专用人工智能，处理的任务需求明确、应用边界清晰、领域知识丰富，在局部智能水平的单项测试中往往能够超越人类智能。例如，阿尔法狗战胜人类围棋冠军，医学图片识别等达到专业医生水平。
- ❑ 通用人工智能尚处于起步阶段。
- ❑ 人工智能的发展方向应该是从专用智能向通用智能。

第1章 人工智能概论

- 1.1 你了解人类的智能吗
- 1.2 人工智能的孕育和诞生
- 1.3 人工智能的发展
- 1.4 从两场标志性人机博弈看人工智能的发展
- 1.5 人工智能研究的基本内容
- 1.6 人工智能的三大学派
- 1.7 人工智能的主要应用领域
- 1.8 人工智能伦理

1.4.1 人工智能研究中的“小白鼠”

机器博弈是人工智能研究中的最好试验场

- 一是博弈问题非常复杂，能够让人工智能大显身手。
- 二是人工智能算法在博弈问题中容易实现，特别是人工智能算法产生错误不会产生重大损失。

1.4.2 深蓝战胜国际象棋棋王卡斯帕罗夫



此后的十年里，人类与机器在国际象棋比赛上互有胜负，直到**2006**年棋王姆尼克被国际象棋软件深弗里茨击败后，人类再也没有击败过电脑。

1.4.3 AlphaGo无师自通横扫世界围棋大师



阿尔法狗与深蓝等最本质的差异：阿尔法狗不需要师傅，能根据以往的经验来不断优化算法，梳理决策模式，吸取比赛经验，并通过与自己下棋来强化学习。

第1章 人工智能概论

- 1.1 你了解人类的智能吗
- 1.2 人工智能的孕育和诞生
- 1.3 人工智能的发展
- 1.4 从两场标志性人机博弈看人工智能的发展
- 1.5 人工智能研究的基本内容
- 1.6 人工智能的三大学派
- 1.7 人工智能的主要应用领域
- 1.8 人工智能伦理

1.5 人工智能研究的基本内容

1. 知识表示

- 知识表示：将人类知识形式化或者模型化。
- 知识表示方法：符号表示法、连接机制表示法。
 - **符号表示法**：用各种包含具体含义的符号，以各种不同的方式和顺序组合起来表示知识的一类方法。例如，一阶谓词逻辑、产生式等。
 - **连接机制表示法**：把各种物理对象以不同的方式及顺序连接起来，并在其间互相传递及加工各种包含具体意义的信息，以此来表示相关的概念及知识。例如，神经网络等。

1.5 人工智能研究的基本内容

2. 机器感知

- 机器感知：使机器（计算机）具有类似于人的感知能力。以机器视觉与机器听觉为主。

3. 机器思维

- 机器思维：对通过感知得来的外部信息及机器内部的各种工作信息进行有目的的处理。

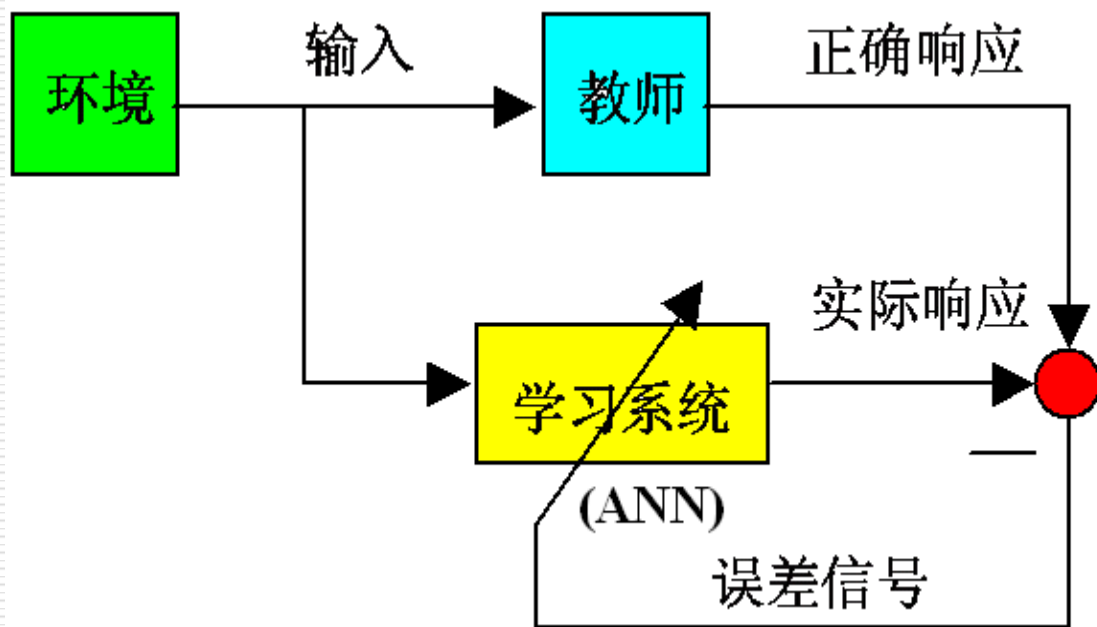
4. 机器学习

- 机器学习（machine learning）：研究如何使计算机具有类似于人的学习能力，使它能通过学习自动地获取知识。

1.5 人工智能研究的基本内容

4. 机器学习

□ 机器学习（machine learning）：研究如何使计算机具有类似于人的学习能力，使它能通过学习自动地获取知识。



监督学习（有教师学习）

1.5 人工智能研究的基本内容

4. 机器学习

- 机器学习（machine learning）：研究如何使计算机具有类似于人的学习能力，使它能通过学习自动地获取知识。

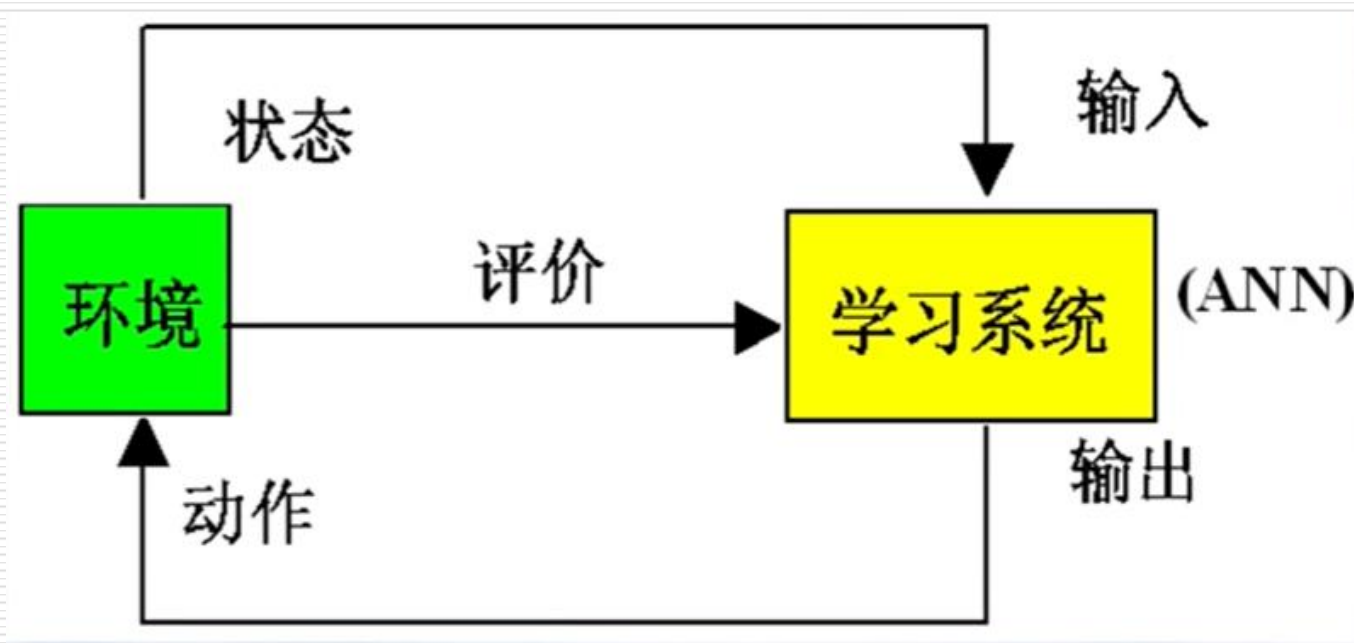


无监督学习（无教师学习）

1.5 人工智能研究的基本内容

4. 机器学习

- 机器学习（machine learning）：研究如何使计算机具有类似于人的学习能力，使它能通过学习自动地获取知识。



强化学习（再励学习或增强学习）

1.5 人工智能研究的基本内容

5. 机器行为

□ 机器行为：计算机的表达能力，即“说”、“写”、“画”等能力。



第1章 人工智能概论

- 1.1 你了解人类的智能吗
- 1.2 人工智能的孕育和诞生
- 1.3 人工智能的发展
- 1.4 从两场标志性人机博弈看人工智能的发展
- 1.5 人工智能研究的基本内容
- 1.6 人工智能的三大学派
- 1.7 人工智能的主要应用领域
- 1.8 人工智能伦理

1.6 人工智能的三大学派

- 不同背景的科学家对人工智能有着不同的认识和理解，在研究工作中采用了不同的方法，逐步形成了人工智能的三大学派：符号主义学派、连接主义学派和行为主义学派。

1.6.1 符号主义

- 符号主义是人工智能最早的主流学派，从模拟人的心智入手，强调知识的表示和推理，经历了从启发式算法到专家系统，再到知识工程，为人工智能发展做出了重要贡献。
- 符号主义起源于数理逻辑，是一种基于逻辑推理的智能模拟方法，所以又称为逻辑主义、心理学派或计算机学派。
- 符号主义的开创性工作自动定理证明，将数学知识表示为谓词公式，然后通过逻辑运算进行推理。
- 符号主义仍是人工智能主要研究领域之一，特别是进入21世纪，知识图谱的提出与广泛应用，使符号法又成为人工智能的热点前沿。

1.6.2 连接主义

- **连接主义学派**：又称仿生学派或生理学派，是一种基于神经网络和网络间的连接机制与学习算法的智能模拟方法。
- 连接主义模拟人脑的结构，采用人工神经网络模拟人脑工作机理，取得突出成果。连接主义认为，大脑神经元机器连接机制是一切智能的基础。
- 1943年，神经生理学家麦克洛克和数学家皮茨，提出了“M-P神经元模型”，开创了人工神经网络方法。
- 连接主义需要极其强大的算力，所以，这个流派到2012年才展示其威力。AI这次大爆发源自连接主义。

1.6.3 行为主义

- **行为主义（Actionism）学派：** 又称进化主义（Evolutionism）或控制论学派，它是控制论向人工智能领域渗透的产物，是一种基于"感知-行动"的行为智能模拟方法。
- 行为主义的试图把神经系统的工作原理与信息论联系在一起，着重研究模拟人在控制过程中的智能行为和作用。
- 行为主义最早来源于20世纪初的一个心理学流派，认为行为是有机体用以适应环境变化的各种身体反应的组合，它的目标在于预测和控制行为。
- **行为主义学派的观点：** 感知周围环境，与现实进行交互的作用，通过进化算法适应环境。智能机器人就是行为主义的典型代表。

第1章 人工智能概论

- 1.1 你了解人类的智能吗
- 1.2 人工智能的孕育和诞生
- 1.3 人工智能的发展
- 1.4 从两场标志性人机博弈看人工智能的发展
- 1.5 人工智能研究的基本内容
- 1.6 人工智能的三大学派
- 1.7 人工智能的主要应用领域
- 1.8 人工智能伦理

1.7 人工智能的主要应用领域

◆自动定理证明

- 定理证明的实质是证明由前提 P 得到结论 Q 的永真性。
- 1958年，王浩证明了有关命题演算的全部定理（220条）、谓词演算中150条定理的85%。
- 1965年鲁宾逊（Robinson）提出了归结原理，使机器定理证明成为现实。
- 我国著名数学家、中国科学院吴文俊院士把几何代数化，建立了一套机器证明方法，被称为“吴方法”。

1.7 人工智能的主要应用领域

◆ 博弈

- 如下棋、打牌、战争等一类竞争性的智能活动称为博弈。
- 1956年，塞缪尔研制出跳棋程序。
- 为什么许多是研究下棋、打牌？

1.7 人工智能的主要应用领域

□ 棋类游戏的计算复杂性

棋局数量

一字棋：9!

西洋跳棋： 10^{78}

国际象棋： 10^{120}

围棋： 10^{761}

国际象棋： 10^{120}

国际象棋：假设每步可以搜索一个棋局，用极限并行速度（ 10^{-104} 年/步）来处理，搜索一遍的全部棋局也得 10^{16} 年即1亿亿年才可以算完！

1.7 人工智能的主要应用领域

◆ 模式识别

- 模式识别（pattern recognition）：研究对象描述和分类方法的学科。分析和识别的模式可以是信号、图象或者普通数据。
- 文字识别：邮政编码、车牌识别、汉字识别。
- 人脸识别：反恐、商业。
- 物体识别：导弹、机器人。

1.7 人工智能的主要应用领域

◆ 模式识别

- **医学影像识别：**基于深度学习等人工智能技术的X光、核磁、CT、超声等医疗影像多模态大数据的分析技术，提取二维或三维医疗影像隐含的疾病特征。
- **黑色素瘤识别：**将1万张有标记的影象交给机器学习，然后让3名医生和 计算机一起看另外的3000账。人的精度84%，计算机97%

1.7 人工智能的主要应用领域

◆ 机器视觉

- **机器视觉**(machine vision)或计算机视觉(computer vision)是用机器代替人眼睛进行测量和判断。
- 机器视觉应用在半导体及电子、汽车、冶金、制药、食品饮料、印刷、包装、零配件装配及制造质量检测等。
- **文字识别**：邮政编码、车牌识别、手写体识别。计算机、手机等输入。
- **人脸识别**：反恐、商业。
- **物体识别**：导弹、机器人。

1.7 人工智能的主要应用领域

◆ 自然语言理解

自然语言理解：研究如何让计算机理解人类自然语言，包括回答问题、生成摘要、翻译等。

□1957年，在苏联人造卫星成功发射的刺激下,美国国家研究会大力支持对俄科技论文的计算机翻译。

1.7 人工智能的主要应用领域

◆ 自然语言理解

□ 人们最初以为机器翻译只要将双向词典及一些词法知识放进计算机就行了。

□ 后来发现有时会出现十分荒谬的错误。

“The spirit is willing but the flesh is weak” 心有余而力不足。

↓
俄语

“The wine is good but the meat is spoiled” 酒是好的但肉变质了。

□ 1966年美国顾问委员会报告：还不存在通用的科学文本机器翻译，也没有很近的实现前景。

□ 英国、美国中断了大部分机器翻译项目的资助。

1.6 人工智能的主要应用领域

□机器听觉

□计算机语音输入：计算机、智能手机等的重要组成

□计算机语音录入、手机语音呼叫、机器人语音控制、语音锁、机器故障诊断等。

□语音识别用语音作为计算机的输入。

□语音识别的主要过程：语音信号采集、预处理（预滤波、采样、预加重、端点检测）、特征参数提取、向量量化、识别。

1.7 人工智能的主要应用领域

◆机器翻译

- 现在，机器翻译已经实用化、商品化。Pilot耳机是世界上首个具备自动翻译的智能耳机，进行实时翻译。
- 2016年谷歌销售语音识别API，将80多种语言转换成文字。
- 2016年谷歌推出商业级神经网络机器翻译，准确率达86%。
- Facebook使用CNN翻译速度比谷歌快9倍。
- 阿里研发NMT，翻译质量大幅度提升。

1.7 人工智能的主要应用领域

◆智能信息检索

□ 智能信息检索系统的功能：

- (1) 能理解自然语言。
- (2) 具有推理能力。
- (3) 系统拥有一定的常识性知识。

◆数据挖掘与知识发现

□ 数据挖掘的目的是从数据库中找出有意义的模式(一组规则、聚类、决策树、依赖网络或其他方式表示的知识)。

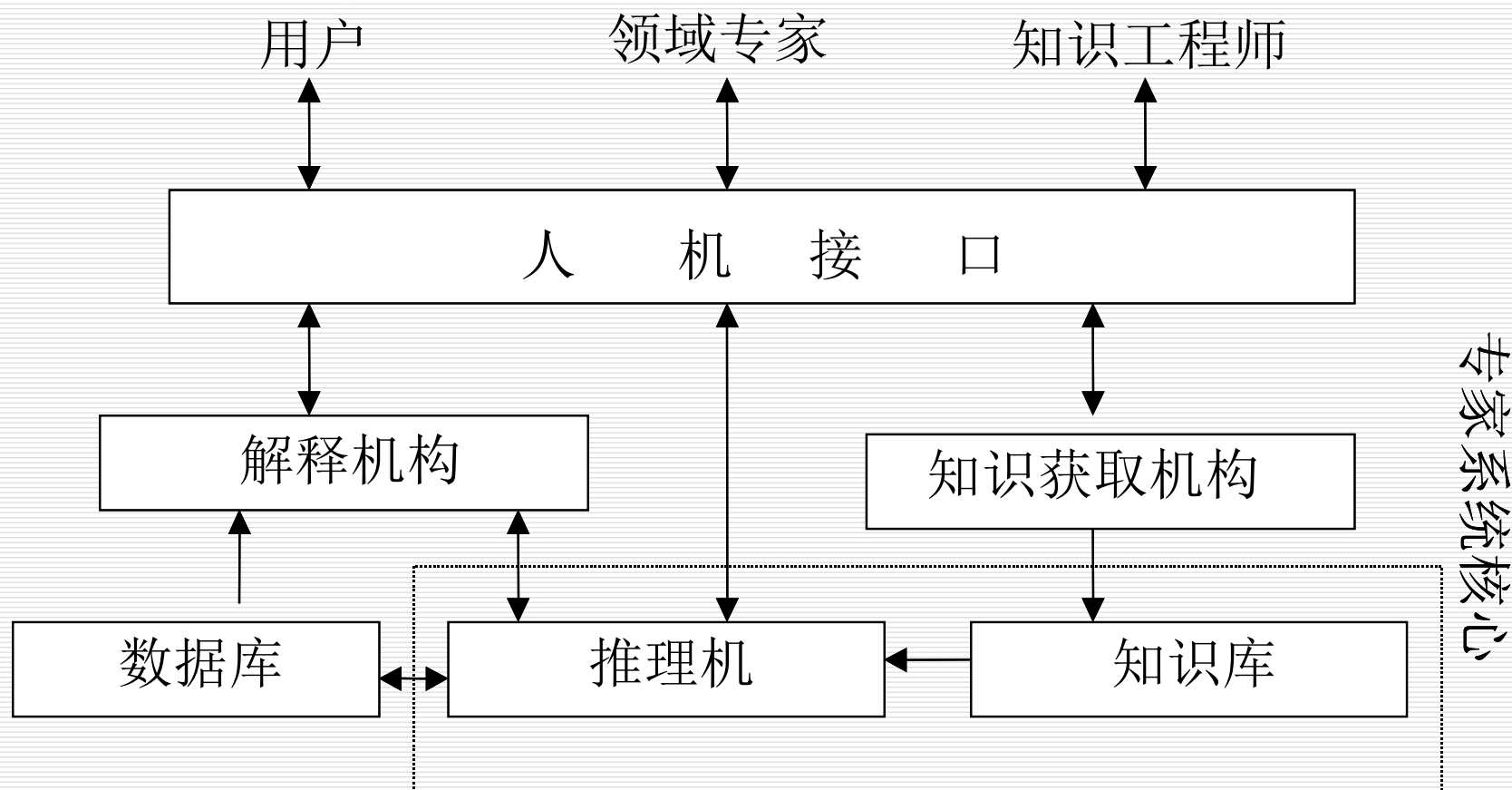
□ 数据挖掘过程：数据预处理、建模、模型评估及模型应用。

1.7 人工智能的主要应用领域

◆ 专家系统

- 专家系统模拟人类专家求解领域内的各种问题。
- 1965年费根鲍姆研究小组开始研制第一个专家系统——分析化合物分子结构的DENDRAL，1968年完成并投入使用。
- 1971年MIT开发成功求解一些数学问题的MYCSYMA专家系统。拉特格尔大学开发的清光眼诊断与治疗的专家系统CASNET。
- 1972年斯坦福大学肖特里菲等人开始研制用于诊断和治疗感染性疾病的专家系统MYCIN。
- 1976年斯坦福研究所开始开发探矿专家系统PROSPECTOR，1980年首次实地分析华盛顿某山区地质资料，发现了一个钼矿。
- 1981年斯坦福大学研制成功专家系统AM，能模拟人类进行概括、抽象和归纳推理，发现某些数论的概念和定理。

1.7 人工智能的主要应用领域



1.7 人工智能的主要应用领域

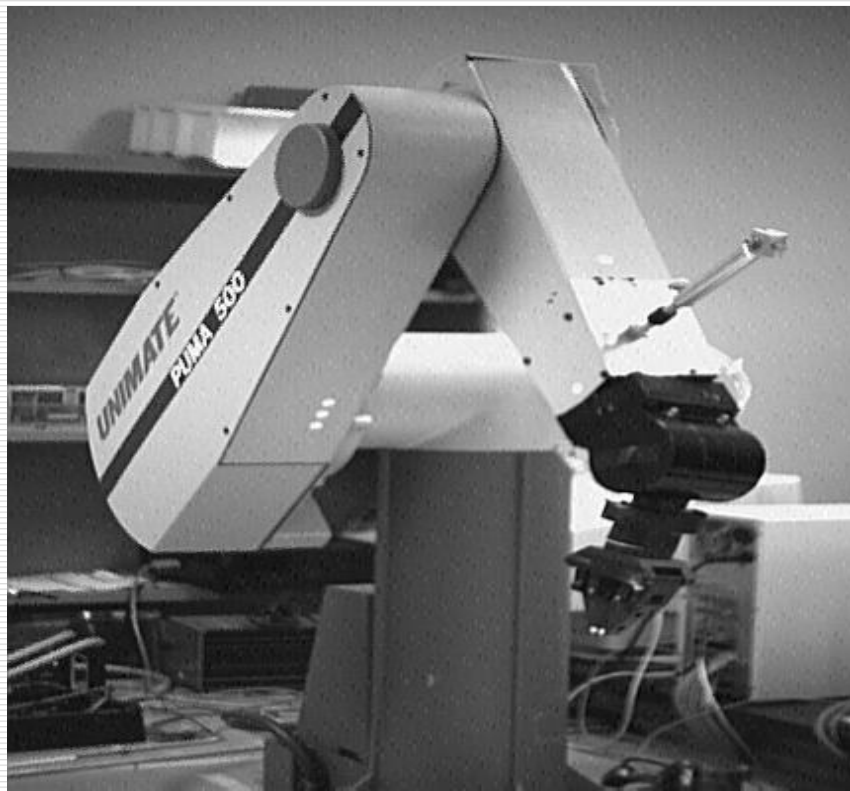
◆ 自动程序设计

- **程序综合**：用户只需要告诉计算机要“做什么”，无须说明“怎么做”，计算机就可自动实现程序的设计。
- **程序验证**：研究出一套理论和方法，通过运用这套理论和方法就可以证明程序的正确性。
- 2014年2月新闻：麻省理工教授 Armando Solar-Lezama开发的一种智能化编程语言“Sketch”，可以自动填补、修正代码内容，在几毫秒内修复代码，让程序员可以忽略许多繁琐的细节。

1.7 人工智能的主要应用领域

◆ 机器人

- 20世纪60年代初，研制出尤尼梅特和沃沙特兰两种机器人。
- 机器人发展：程序控制机器人（第一代）、自适应机器人（第二代）、智能机器人（现代）。



1.7 人工智能的主要应用领域

无人驾驶



- ❑ 机器在感知上比人类强很多
- ❑ 机器比人类精力充沛
- ❑ 机器比人更理性

1.7 人工智能的主要应用领域

无人驾驶商业化的四个关键要素

- **硬件组件**：摄像头、激光雷达、计算处理器等新型传感器和计算组件，发动机、车身等传统汽车组件；
- **软件组件**：无人驾驶操作系统（感知、规划、控制以及汽车互联、数据平台接口等），高精度地图等；
- **整车制造**：超级复杂、重资产、且利润率不高的工程；
- **网络**：类似滴滴、Uber、Lyft这样的出行网络。

1.7 人工智能的主要应用领域

□ 广汽菲亚特白车身生产线



□ 生产线上分布着250个机器人

1.7 人工智能的主要应用领域

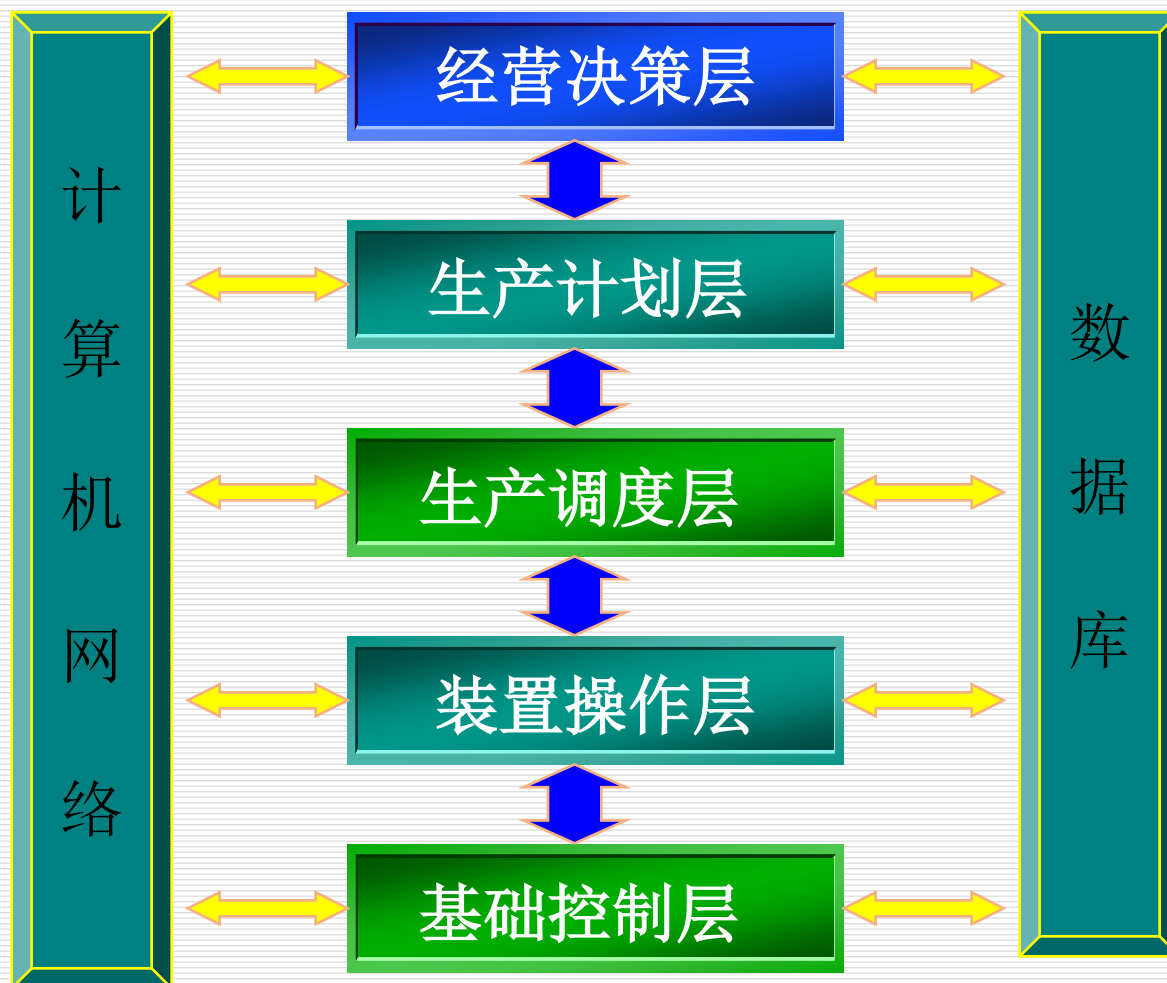
◆ 组合优化问题

□ 组合优化问题：旅行商问题、生产计划与调度、物流中的车辆调度、智能交通、通信中的路由调度、计算机网络信息调度等

□ **NP完全问题**：用目前知道的最好的方法求解，问题求解需要花费的时间是随问题规模增大以指数关系增长。

1.7 人工智能的主要应用领域

CIMS的结构与功能



1.7 人工智能的主要应用领域

智慧物流

- 推广射频识别（RFID）、多维条码、卫星定位、货物跟踪、电子商务等信息技术在物流行业中的应用；
- 加快基于物联网的物流信息平台建设，整合物流资源，实现物流政务服务和物流商务服务的一体化；
- 推动信息化、标准化、智能化的物流企业和物流产业发展。

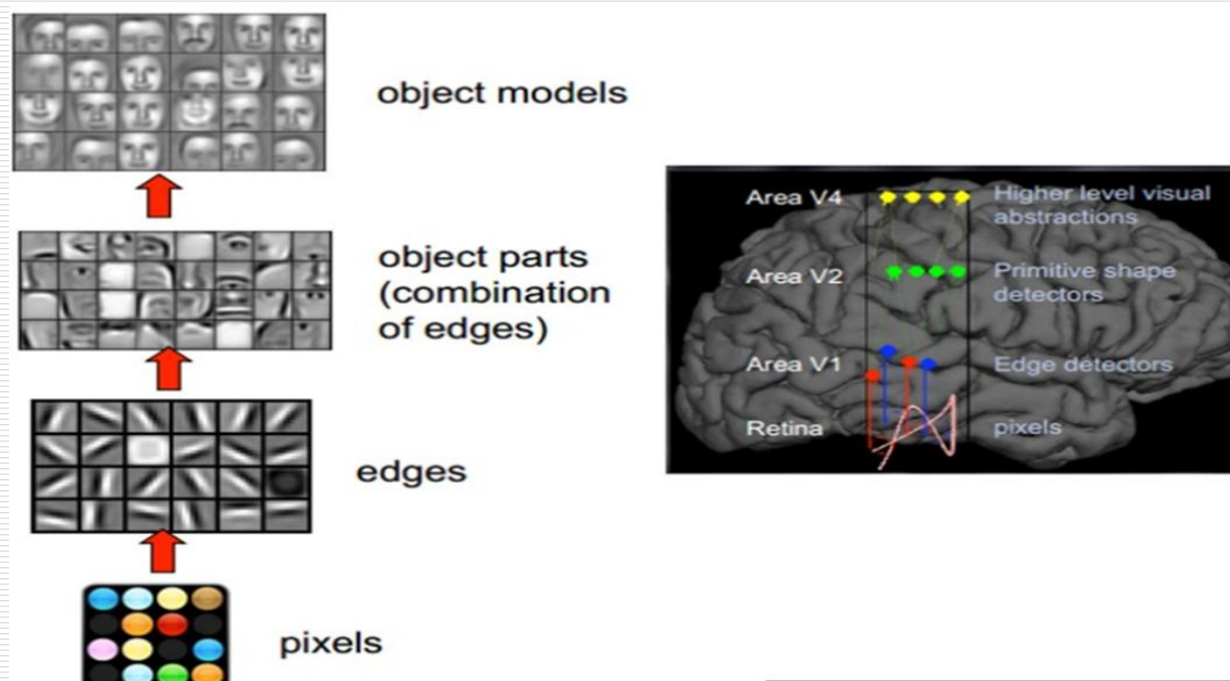
1.7 人工智能的主要应用领域

◆ 人工神经网络

- 人工神经网络：一个用大量简单处理单元经广泛连接而组成的人工网络，用来模拟大脑神经系统的结构和功能。
- 神经元模型的研究（20世纪50年代中期-----）
- 1943年，麦克洛奇和皮兹提出M-P模型，开创了人工神经网络研究。
- 1957年，罗森勃拉特提出感知器模型。
- 1969年，明斯基和佩珀特发表了《Perceptron》，对神经元模型的研究作出了悲观的论断。

1.7 人工智能的主要应用领域

机器学习的第二次浪潮：深度学习



● 诺贝尔医学奖获得者美国神经生物学家**David Hubel**和**Torsten Wiesel**发现：

人的视觉系统的信息处理是分级的。

1.7 人工智能的主要应用领域

◆ 分布式人工智能与多智能体

- 分布式人工智能系统以鲁棒性作为控制系统质量的标准，并具有互操作性，即不同的异构系统在快速变化的环境中，具有交换信息和协同工作的能力。
- 分布式问题求解：把一个具体的求解问题划分为多个相互合作和知识共享的模块或者结点。
- 多智能体系统：研究各智能体之间行为的协调。

1.7 人工智能的主要应用领域

◆ 智能控制

□ 国际知名美籍华裔科学家傅京孙（K S. Fu）在1965年首先把人工智能的启发式推理规则用于学习控制系统。

□ 智能控制的两个显著特点：

第一，智能控制是同时具有知识表示的非数学广义世界模型和传统数学模型混合表示的控制过程。

第二，智能控制的核心在高层控制，其任务在于实际环境或过程进行组织，即决策与规划，以实现广义问题求解。

□ 智能控制的基本类型：

（1）专家智能控制 （2）模糊控制 （3）神经网络控制

1.7 人工智能的主要应用领域

◆ 智能仿真

- 智能仿真是将AI引入仿真领域，建立智能仿真系统。
- 仿真是对动态模型的实验，即行为产生器在规定的实验条件下驱动模型，从而产生模型行为。仿真是在描述性知识、目的性知识及处理知识的基础上产生结论性知识。
- 利用AI对整个仿真过程（建模、实验运行及结果分析）进行指导，在仿真模型中引进知识表示，改善仿真模型的描述能力，为研究面向目标的建模语言打下基础，提高仿真工具面向用户、面向问题的能力，使仿真更有效地用于决策，更好地用于分析、设计及评价知识库系统。

1.7 人工智能的主要应用领域

◆ 智能CAD

□ 智能CAD（简称ICAD）就是把人工智能技术引入计算机辅助设计领域，建立智能CAD系统。AI几乎可以应用到CAD技术的各个方面。从目前发展的趋势来看，至少有下述四个方面：

- （1）设计自动化。
- （2）智能交互。
- （3）智能图形学。
- （4）自动数据采集。

1.7 人工智能的主要应用领域

◆ 智能CAI

□ 智能CAI就是把AI引入计算机辅助教学领域。 ICAI系统一般分成专门知识、教导策略和学生模型和自然语言的智能接口。

□ ICAI应具备下列智能特征：

- (1) 自动生成各种问题与练习。
- (2) 根据学生的学习情况自动选择与调整教学内容与进度。
- (3) 在理解教学内容的基础上自动解决问题生成解答。
- (4) 具有自然语言生成和理解能力。
- (5) 对教学内容有理解咨询能力。
- (6) 能诊断学生错误，分析原因并采取纠正措施。
- (7) 能评价学生的学习行为。
- (8) 能不断地在教学中改善教学策略。

1.7 人工智能的主要应用领域

◆ 智能管理与智能决策

- **智能管理**就是把人工智能技术引入管理领域，建立智能管理系统，研究如何提高计算机管理系统的智能水平，以及智能管理系统的设计理论、方法与实现技术。
- **智能决策**就是把人工智能技术引入决策过程，建立智能决策支持系统。
- **智能决策支持系统**是由传统决策支持系统再加上相应的智能部件就构成了智能决策支持系统。
- **智能部件**可以是专家系统模式、知识库模式等。

1.7 人工智能的主要应用领域

◆ 智能多媒体系统

- 多媒体计算机系统就是能综合处理文字、图形、图像和声音等多种媒体信息的计算机系统。
- 智能多媒体就是将人工智能技术引入多媒体系统，使其功能和性能得到进一步发展和提高。
- 多媒体技术与人工智能所研究的机器感知、机器理解等技术不谋而合。人工智能的计算机视听觉、语音识别与理解、语音对译、信息智能压缩等技术运用于多媒体系统，将会使现在的多媒体系统产生质的飞跃。

1.7 人工智能的主要应用领域

◆ 人工生命

- 人工生命是以计算机为研究工具，模拟自然界的生命现象，生成表现自然生命系统行为特点的仿真系统。
- AL是首先由计算机科学家Christopher Langton在1987年在"生成以及模拟生命系统的国际会议"上提出。
- 研究进化的模式和方式、人工仿生学、进化博弈、分子进化、免疫系统进化、学习等；
- 研究具有自治性、智能性、反应性、预动性和社会性的智能主体的形式化模型、通信方式、协作策略；
- 研究生物感悟的机器人、自治和自适应机器人、进化机器人、人工脑。

1.7 人工智能的主要应用领域

◆ 智能操作系统

□ 智能操作系统的基本模型：以智能机为基础，能支撑外层的AI应用程序，实现多用户的知识处理和并行推理。

□ 智能操作系统三大特点：

并行性：支持多用户、多进程，同时进行逻辑推理等；

分布性：把计算机硬件和软件资源分散而又有联系地组织起来，能支持局域网和远程网处理；

智能性：一是操作系统处理的是知识对象，具有并行推理功能，支持智能应用程序运行；二是操作系统的绝大部分程序使用AI程序编制，充分利用硬件并行推理功能；三是具有较高智能程序的自动管理维护功能，如故障的监控分析等，帮助维护人员决策。

1.7 人工智能的主要应用领域

◆ 智能计算机系统

- 智能计算机系统就是人们正在研制的新一代计算机系统。
- 智能计算机系统从基本元件到体系结构，从处理对象到编程语言，从使用方法到应用范围，同当前的诺依曼型计算机相比，都有质的飞跃和提高，它将全面支持智能应用开发，且自身就具有智能。

1.7 人工智能的主要应用领域

◆ 智能通信

- 智能通信就是把人工智能技术引入通信领域，建立智能通信系统。
- 智能通信就是在通信系统的各个层次和环节上实现智能化。例如在通信网的构建、网管与网控、转接、信息传输与转换等环节，都可实现智能化。这样，网络就可运行在最佳状态，具有自适应、自组织、自学习、自修复等功能。

1.7 人工智能的主要应用领域

◆ 智能网络系统

- 智能网络系统就是将人工智能技术引入计算机网络系统。如在网络构建、网络管理与控制、信息检索与转换、人机接口等环节，运用AI的技术与成果。
- AI的专家系统、模糊技术和神经网络技术可用于网络的连接接纳控制、业务量管制、业务量预测、资源动态分配、业务流量控制、动态路由选择、动态缓冲资源调度等许多方面。

1.7 人工智能的主要应用领域

◆ 云端人工智能

- 云端人工智能（cloud AI）在云端集中使用和共享机器学习工具的技术。
- 云端人工智能系统主要包括人工智能基础设施和人工智能服务两大部分。
- 人工智能基础设施包括各种数据和数据库、大数据平台等。
- 人工智能服务部分提供各种应用程序接口和人机交互接口，用户无需自己创建自定义的机器学习模型。

1.7 人工智能的主要应用领域

◆ 脑机接口

- 脑机接口(brain-computer interface, BCI)是在人或动物脑与外部设备间建立的直接连接通路。
- 脑机接口分为单向脑机接口和双向脑机接口。
- 单向脑机接口只允许脑和外部设备间的单向信息交换。
- 双向脑机接口允许脑和外部设备间的双向信息交换。
- 人工耳蜗是迄今为止最成功、临床应用最普及的恢复损伤的听觉的脑机接口。
- 视觉修复的脑机接口技术尚在研发之中，主要困难在于视觉传递信息量的巨大和外周感觉器官（视网膜）和中枢视觉系统在功能上的相对复杂性。

1.7 人工智能的主要应用领域

- 人工智能应用
如雨后春笋



第1章 人工智能概论

- 1.1 你了解人类的智能吗
- 1.2 人工智能的孕育和诞生
- 1.3 人工智能的发展
- 1.4 从两场标志性人机博弈看人工智能的发展
- 1.5 人工智能研究的基本内容
- 1.6 人工智能的三大学派
- 1.7 人工智能的主要应用领域
- 1.8 人工智能伦理

1.8 人工智能伦理

□在 AI 发达以后，同样不能忽视伦理的约束，这样人类才能在更多地享受到技术革新益处的同时，让冰冷的科技亦呈现温暖的底色，创造有温度的未来。

□人工智能伦理学（the Ethics of Artificial Intelligence）是专门针对人工智能系统的应用伦理学的一个分支。

□人工智能伦理涉及人类设计、制造、使用 and 对待人工智能系统的道德，机器伦理中的机器行为，以及超级人工智能的奇点问题。

1.8.1 人工智能伦理的提出与发展

- ❑ 国际电气和电子工程师协会于2016年4月启动了IEEE 标准协会发布“人工智能设计的伦理准则”白皮书，力图建立社会公认的人工智能伦理标准、职业认证及道德规范。
- ❑ 国际化标准化组织（ISO）于2017年成立人工智能委员会。
- ❑ 牛津大学、剑桥大学和Open AI公司等7家机构于2018年共同发布《人工智能的恶意使用：预测、预防和缓解》。
- ❑ 2021年11月24日，第41届联合国教科文组织大会真实通过《人工智能伦理建议书》。人工智能伦理的首份全球性规范文书，是全球人工智能发展的共同纲领。提出发展和应用人工智能首先要体现出四大价值，规范人工智能技术的10大原则和11个行动领域。

1.8.1 人工智能伦理的提出与发展

- 中国政府高度重视人工智能发展中的伦理问题。
- 2017年7月国务院发布的《新一代人工智能发展规划》中，提出要制定促进人工智能发展的法律法规和伦理规范；开展人工智能行为科学和伦理等问题研究；制定人工智能产品研发设计人员的道德规范和行为守则。
- 2018年9月17日，习近平致信祝贺2018世界人工智能大会指出，处理好人工智能在法律、安全、就业、道德伦理和政府治理等方面提出的新课题，需要各国深化合作、共同探讨。
- 2019年2月，中国国家新一代人工智能治理专业委员会成立，发布《新一代人工智能治理原则——发展负责任的人工智能》。

1.8.2 人工智能伦理的典型案例分析

□1.人工智能伦理典型案例

□案例1：加州禁止执法人员使用面部识别技术。

□2019年9月12日，美国加利福尼亚州议会禁止州和地方执法机构使用面部识别技术。该议案已于2020年1月1日生效。旧金山和奥克兰也已全面禁止面部识别技术的使用。俄勒冈州和新罕布什尔州等也有类似的禁令。

□许多执法团体认为面部识别技术在追踪嫌疑人以及寻找失踪儿童等很多方面都具有重要作用。加州警察局长协会主席罗恩·劳伦斯表示：“这是未来的破案方式，我们需要拥抱而不是逃避技术。”

1.8.2 人工智能伦理的典型案例分析

□案例2：“监测头环”进校园惹争议。

□2019年11月，某小学戴“监测头环”的视频引起广泛的关注与争议。这些头环通过传感器上的三个电极，可检测佩戴者的脑电波，从而评判学生是否集中注意力并进行打分：上课专注亮红灯，上课走神亮蓝灯。

□许多人担忧侵犯学生隐私，以及让学生产生逆反心理。

□脑机接口技术是一种新兴技术，会不容易被人理解，必须慎之又慎地使用。

□对“脑机接口”的使用应完全出于个人意愿，相关数据也绝对地属于个人。意识是个人领地，他人无权侵犯。

1.8.2 人工智能伦理的典型案例分析

□ 案例3：AI 换脸应用引发隐私争议。

□ 2019年9月，一款AI 换脸软件引发严重的隐私争议。用户只需提供一张正面人脸照片上传到换脸软件，就可以把选定视频中的明星面部替换掉，生成以自己为主角的视频片段。该软件被指涉嫌非法收集进而可能滥用用户面部信息。

□ 2019 年 11 月底，国家网信办、文旅部和广电总局等三部门联合发布了《网络音视频信息服务管理规定》：“利用基于深度学习、虚拟现实等的虚假图像、音视频生成技术制作、发布、传播谣言的……应当及时采取相应的辟谣措施”。

1.8.2 人工智能伦理的典型案例分析

□ 案例4：自动驾驶安全事故问责。

□ 2019年3月，50岁的班纳驾驶电动车以每小时109公里的速度与一辆牵引车拖车相撞而身亡；

□ 2018年12月10日，该品牌电动汽车在美国康涅狄格州95号州际公路上追尾了一辆停放在路边的警车；

□ 事故发生时这些汽车的自动驾驶系统都处于开启状态。

□ 德国最新修订的《道路交通法》明文规定自动驾驶事故责任划分：在行驶途中系统不可以完全取代驾驶人，须要有人类司机随时接管车辆，同时其最终责任主要落在驾驶人身上。

1.8.2 人工智能伦理的典型案例分析

□ 案例5：OpenAI 开发出假新闻编写软件。

□ 2019年2月15日，OpenAI开发的GPT-2软件，能编写逼真的假新闻。例如给软件提供信息：“一节装载受控核材料的火车车厢今天在Cincinnati被盗，下落不明。” GPT-2可以写出由7个段落组成的假新闻，甚至还引述政府官员的话语。

□ 虽然GPT-2已经被训练来预测从800万个网页中提取的文本中的下一个单词，但可信度仍相对较低，特别是它生成的文本并不总是一致的，可能包括重复的结构、突然的主题切换和违背常识。例如，GPT-2可以记录发生在水下的火灾，或者某人是他们父亲的父亲等。

1.8.2 人工智能伦理的典型案例分析

□2.人工智能伦理成因分析

□根据人工智能伦理问题的特征和产生方式的不同，可从技术、数据和应用三个层面来分析人工智能伦理的成因。

□技术层面：包括算法及系统的安全问题、可解释性问题、算法歧视问题和算法决策问题。

□算法及系统的安全问题：算法或系统的漏洞被攻击和恶意利用的情形；

□可解释性问题：关注人类的知情利益和主体地位；

□算法歧视问题：算法研究人员的认知偏见或者训练数据集不完美导致的歧视性的决策结果；

□算法决策问题：人工智能自学习导致的不可预见性结果。

1.8.2 人工智能伦理的典型案例分析

□2.人工智能伦理成因分析

□数据层面：人工智能技术在许多领域获得重大进展依赖于大数据驱动。而随着人工智能技术的快速发展以及大量应用案例实施，提供了丰富的个人信息收集场景、范围和数量，数据采集日益成为引发人工智能伦理问题的重要原因之一。

□应用层面：人工智能算法的滥用和误用，人们在社会活动时容易出现使用目的、方式、范围等偏差，从而导致不良的影响和后果，引发的人工智能伦理问题。

□例如：利用人脸识别算法可以提高追查犯罪嫌疑人的效率，但也可能用于跟踪某人的行踪、识别某人的情感，并用于商业行为。

1.8.3 人工智能伦理的治理原则

- 人工智能的发展应遵循以下基本原则：人类根本利益原则、责任原则。
- 1.人类根本利益原则：人工智能以实现人类根本利益为最终目标，体现对人权的尊重，降低技术风险和负面影响价值选择。
- 选择人工智能系统和人工智能方法是否合理的依据：
 - ①选择的人工智能方法对于实现特定合法目标应是适当的；
 - ②选择的人工智能方法不得违背基本价值观，特别是不得侵犯或践踏人权，不得用于社会评分或大规模监控目的；
 - ③在所涉决定具有不可逆转或难以逆转的影响或者在涉及生死抉择的情况下，应由人类作出最终决定。

1.8.3 人工智能伦理的治理原则

□2. 责任原则

□随着人工智能的发展，需要建立明确的责任体系，也就是责任原则。

□基于责任原则，人工智能系统遇到与人类伦理或法律发生冲突时，可以从技术层面对人工智能技术应用部门进行问责。

□在责任原则下，人工智能技术的开发和应用应分别遵循透明度原则和权责一致原则。

□在人工智能系统的整个生命周期内都需要努力提高人工智能系统的透明度和可解释性。

□人工智能系统的可解释性也指各个算法模块的输入、输出和性能的可解释性及其如何促成系统结果。

1.8.4 人工智能伦理的治理措施

□1. 技术改进

□人工智能的许多伦理风险可以通过技术改进予以解决。

□例如，算法可解释性和透明性涉及人类的知情利益和主体地位，是人工智能伦理的重要命题之一。

□要逐步形成算法开发者和使用者信息披露的技术惯例，对算法进行监管，接受公众的审查和质询。

□逐步推进人工智能标准化进程，进而推动人工智能应用领域的不断拓展。

□推动人工智能企业建立伦理委员会，明确扮演人工智能技术落地执行者角色的企业的伦理意识和社会责任。

1.8.4 人工智能伦理的治理措施

□2. 道德规范

□我国《新一代人工智能发展规划》：未来重点开展人工智能行为科学和伦理等问题研究，探索伦理道德多层次判断结构及人机协作的伦理框架。

□通过教育和社会舆论进行适当地引导，倡导正确的价值观，推动人工智能伦理共识的形成，有效地避免一些违背伦理道德却未违反现行法律的恶性事件发生。

□推动科研机构和企业对人工智能伦理风险的认知和实践，通过发布伦理风险分析报告等形式，讨论人工智能伦理风险的应对措施。

1.8.4 人工智能伦理的治理措施

□3. 政策指引

□通过国家政策指引更体现差异化和顺应创新趋势。例如：

□①支持自然资源的保护、监测和管理。

□②支持与气候有关问题的预测、预防、控制和减缓。

□③支持更加高效和可持续的粮食生态系统。

□④支持可持续能源的加速获取和大规模采用。

□⑤促成并推动旨在促进可持续发展的可持续基础设施、可持续商业模式和可持续金融主流化。

□⑥检测污染物或预测污染程度，协助相关利益方确定、规划并实施有针对性的干预措施，防止并减少污染及曝露风险。

1.8.4 人工智能伦理的治理措施

□4. 法律法规

□在AI法律法规中，要在促进发展与风险控制之间探求最佳平衡点，尤其需要把握公权力干预和产业自治之间的平衡。

□在实践相对成熟、已有一定普遍共识的部分，可尝试立法规制，如针对各领域AI应用中均较为普遍的对于个人隐私等人格利益侵犯的风险问题，通过立法的方式进行保障；

□在行业发展需要明确责任的领域启动立法研究，如对自动驾驶等发展相对成熟、迫切需要明确责任的领域，在现行法律之下尝试地方性、试验性的立法，为推进AI立法进程提供经验。

□在实践不足以立法规制的领域，可以通过行业自律、政策引导等方式进行伦理风险控制。

1.8.4 人工智能伦理的治理措施

□5. 人工智能伦理教育

□将AI伦理列入教学、培训等方面教育，促进AI技术技能教育与AI教育的人文、伦理和社会方面的交叉协作，以减轻AI伦理带来的挑战和风险。

□对AI研究开发人员进行AI伦理培训，要求他们将伦理考量纳入设计、产品和出版物中，特别是在分析其使用的数据集、数据集的标注方法以及可能投入应用的成果的质量和范围方面。

□AI应依照相关的个人数据保护标准支持学习过程，既不降低认知能力，也不提取敏感信息。

□在学习者与AI系统互动过程中收集到的为获取知识而提交的数据，不得被滥用、挪用或用于犯罪，包括用于商业目的。



THE END

