

# 人工智能导论

- 第1章 绪论
- 第2章 知识表示
- 第3章 确定性推理
- 第4章 不确定推理
- 第5章 搜索
- 第6章 机器学习

# 第1章 绪论

- 1.1 人工智能的定义、图灵测试、中文屋子
- 1.2 人工智能的诞生与发展
- 1.3 人工智能研究的方法及途径
- 1.4 人工智能的研究及应用领域
- 1.5 人工智能目前存在的问题、未来目标和实现方法

## 1.1 人工智能的定义

### 人类智能

#### 定义1:

- 1) 人类所具有的智力和行为能力
- 2) 这种智力和行为能力是以知识为基础的
- 3) 智力和行为的目的是获取知识并运用知识去求解问题

#### 人类智能的特点:

- 感知、记忆、思维、推理、学习、行为

#### 定义2:

人的智能是人类理解和学习事物的能力。

## 1.1 人工智能的定义

### 智能机器

定义1: 能够在各类环境中自主地或交互地执行各种拟人任务的机器。

定义2: 一种能够呈现出人类智能行为的机器, 这种智能行为是人类用大脑考虑问题或者创造思想。

### 定义1 人工智能

用人工的方法在机器上实现的智能, 或者说是人们使用机器模拟人类的智能。

### 定义2 人工智能(学科)

人工智能(学科)是计算机科学中涉及研究、设计和应用智能机器的一个分支。

它的近期主要目标在于研究用机器来模仿和执行人脑的某些智力功能, 并开发相关理论和技术。

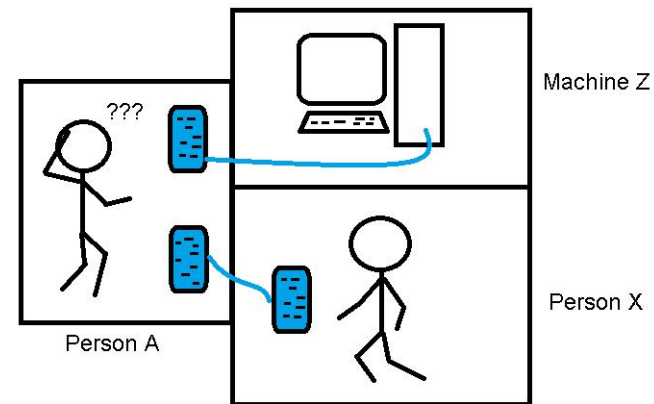
### 定义3 人工智能(能力)

是智能机器所执行的通常与人类智能有关的智能行为, 这些智能行为涉及判断、推理、证明、识别、感知、理解、通信、设计、思考、规划、学习和问题求解等思维活动。

为了更深刻地理解AI，下面综述其它几种关于AI的定义。

- 定义4 是一种使计算机能够思维，使机器具有智力的激动人心的新尝试。
- 定义5 是那些与人的思维、决策、问题求解和学习等有关活动的自动化。
- 定义6 是用计算模型研究智力行为。
- 定义7 是研究那些使理解、推理和行为成为可能的计算。
- 定义8 研究如何使计算机做事让人过得更好。
- 定义9 是一门通过计算过程力图理解和模仿智能行为的学科。
- 定义10 是计算机科学中与智能行为的自动化有关的一个分支。

# 图灵测试



- 图灵肯定机器可以思维的，他提出一个假想：一个人在不接触对方的情况下，通过一种特殊的方式，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还是计算机，那么，就可以认为这个计算机具有同人相当的智力，即这台计算机是能思维的。这就是著名的“图灵测试”（Turing Testing）。
- 他认为，通过键盘等装置，让参加测试的每一个人与另一个人和一台机器进行交流和提问，如果有超过30%的测试者误以为在和自己说话的是人而非计算机，那么这台机器就通过了测试，并被认为是具有人类智能。虽然将30分作为及格线的标准不高，这一测试直到今天仍被作为检验人工智能的标准之一。

## Ex

- 问：你会下国际象棋吗？
- 答：是的。
- 问：你会下国际象棋吗？
- 答：是的，我不是已经说过了吗？
- 问：请再次回答，你会下国际象棋吗？
- 答：你烦不烦，干嘛老提同样的问题。
- 如果再问下去，你也许就会怒砸键盘，你TM要我呢？

- 问：你会下国际象棋吗？
- 答：是的。
- 问：你会下国际象棋吗？
- 答：是的。
- 问：请再次回答，你会下国际象棋吗？
- 答：是的。

如果让你判断，哪一个回答者是正常人类，哪一个是人工智能，通过对比这两组对话，相信你马上就能有答案。

上述两种对话的区别在于，第一种则具有分析综合的能力，回答者知道观察者在反复提出同样的问题；第二种可明显地感到回答者是从知识库中提取简单的答案。



# 图灵测试的局限性

倾向于符号问题求解任务。它不测试需要感知技能或者手工灵巧的能力，即使这些是人的智能的重要成分。

即使某台机器或者软件能够通过图灵测试，我们可以看到结果，但过程却是由人主观评定，而非客观的“量化评定”，这意味着“这台机器的反馈在那一批人看来是恰当的”，如果换一批人来对这台机器进行测试，很可能会得到不一样的结果。

图灵测试没有必要限制机器智能符合人的特性。机器智能可能不同于人的智能。试图按人的条件对它进行评价基本上是一种错误。例如希望机器做计算像人那样缓慢和不精确？

图灵测试是要计算机在**行为**上像人，而主流人工智能是要在解决各种问题的**能力**上像人甚至超过人。这两种研究都有价值且有相互联系，但并不是一回事。这正是主流人工智能不接受图灵测试的根本原因。

# 中文房间

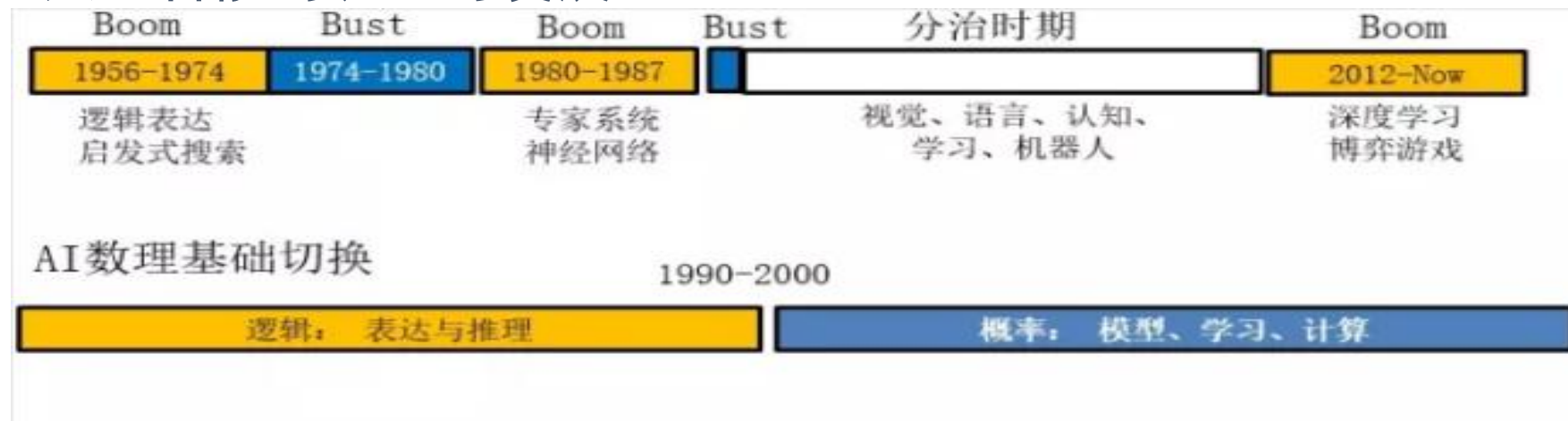
- 中文房间（Chinese room, the Chinese room argument）又称作华语房间，是由美国哲学家约翰·希尔勒（John Searle）在1980年设计的一个思维试验。
- 这个实验要求你想象一位只说英语的人身处一个房间之中，这间房间除了门上有一个小窗口以外，全部都是封闭的。他随身带着一本写有中文翻译程序（中->英，英->中）的书。房间里还有足够的稿纸、铅笔和橱柜。写着中文的纸片通过小窗口被送入房间中。根据Searle，房间中的人可以使用他的书来翻译这些文字并用中文回复。虽然他完全不会中文，Searle认为通过这个过程，房间里的人可以让任何房间外的人以为他会说流利的中文。

## 1.2 人工智能的诞生与发展

以莱布尼兹、维纳、弗雷治、罗素等为代表对发展数理逻辑学科的贡献，丘奇、图灵和其它一些人关于计算本质的思想，控制论、信息论、自动机理论、神经网络模型和电子计算机等方面的创造性贡献奠定了人工智能的理论基础。

1956年美国达特茅斯大学举行了一次会议，讨论用机器模拟人类智能的问题，第一次使用“人工智能”这个术语。这第一次人工智能研讨会标志着人工智能学科的诞生。

## 1.2 人工智能的诞生与发展



- 首先，从表面一层来看。

最早一次的兴起是**1956-1974**，以命题逻辑、谓词逻辑等知识表达、启发式搜索算法为代表。

**1980**年代初又兴起了**第二次**热潮，专家系统、知识工程、医疗诊断等

**第三次**热潮就是最近两年兴起的深度学习推动的

- 其次、从深一层的理论基础看。把人工智能发展的**60**年分为两个阶段。

**第一阶段：**前**30**年以数理逻辑的表达与推理为主。

**第二阶段：**后**30**年以概率统计的建模、学习和计算为主。

## 1.3 人工智能研究的方法及途径

### 人工智能研究的各种学派及其理论

- **符号主义** (Symbolicism)，又称为**逻辑主义** (Logicism)、**心理学派** (Psychologism) 或**计算机学派** (Computerism)。认为人工智能源于**数理逻辑**，研究方法应为**功能模拟方法**。其原理主要为**物理符号系统** (即符号操作系统) 假设。

**认知的基元**是符号，**认知过程**是符号模式的操作处理过程。

以符号处理为核心对人脑功能进行模拟。根据人脑的心理模型，把问题或知识表示为某种逻辑结构，运用符号演算，实现表示、推理和学习等功能。

**定理证明、自动推理、专家系统、自动程序设计、机器博弈、知识图谱**

# 物理符号系统假设

- **物理符号系统**采用模式（符号），将它们组合成结构（表达式），并操纵它们（使用过程）生成新的表达式，系统在物理世界中实现。

- **物理符号系统假设（physical symbol system hyposesis）**是认知心理学建立和发展的重要理论基础，是纽维尔和西蒙提出来的。

物理系统呈现出一般智能行为的**充分必要条件**是它为物理符号系统。

**必要**：一个系统如果能表现出智能的话，就可以被视为一个物理符号系统；

**充分**：任何具有足够尺度的物理符号系统都可以经适当组织之后展现出智能。

## 1.3 人工智能研究的方法及途径

### 人工智能研究的各种学派及其理论

- **联结主义** (Connectionism)，又称为**仿生学派** (Bionicsism) 或**生理学派** (Physiologism)，认为人工智能源于**仿生学**，特别是**人脑模型**的研究，研究方法为**结构模拟**。其原理主要为**神经网络及神经网络间的连接机制与学习算法**。

**认知的基元**是神经元，**认知过程**也不是符号处理。

通过对人脑神经网络、神经元之间的连接以及在神经元间的并行处理的模拟实现对人脑智能的模拟。

**MP模型、BP算法、各种深度神经网络**



## 1.3 人工智能研究的方法及途径

人工智能研究的各种学派及其理论

- **行为主义** (Actionism), 又称**进化主义** (Evolutionism) 或**控制论** 学派 (Cyberneticsism), 认为人工智能源于**控制论**, 研究方法为**行为模拟**。其原理为**控制论**及**感知动作型控制系统**。

智能不取决于符号和神经元, 而取决于感知和行动, 提出智能行为的“感知-动作”模式。认为智能不需要知识、不需要表示和推理, 人工智能可以像人类智能一样逐步进化。

这一学派的代表作首推布鲁克斯 (Brooks) 的六足行走机器人, 它被看做新一代的“控制论动物”, 是一个基于感知-动作模式的模拟昆虫行为的控制系统。



乔瑞





## 1.3 人工智能研究的方法及途径

### 人工智能研究的各种学派及其理论

#### 各人工智能学派的融合：

- 一方面各学派密切合作，取长补短，可把一种方法无法解决的问题转换成另一方法能够解决的问题。
- 另一方面，逐步建立统一的人工智能理论体系和方法。

## 1.4 人工智能的研究及应用领域

人工智能是一个非常广泛的领域。当前人工智能涵盖很多大的学科，把它们归纳为六个：

- （1）**计算机视觉**（模式识别，图像处理等）、
- （2）**自然语言处理与交流**（文本处理，语音识别、合成、对话等）、
- （3）**认知与推理**（包含各种物理和社会常识）、
- （4）**机器人学**（机械、控制、设计、运动规划、任务规划等）、
- （5）**博弈与伦理**（多代理人agents的交互、对抗与合作，机器人与社会融合等议题）。
- （6）**机器学习**（主要是各种统计的建模、分析工具和计算的方法）

这些领域目前还比较散，目前它们正在交叉发展，走向统一的过程中。

## 1.5 人工智能目前存在的问题、未来目标和实现方法

- 能处理的对象的5个限制：
  - 有丰富的数据或者丰富的知识
  - 确定性信息
  - 完全信息
  - 静态的
  - 单任务和有限领域

## 1.5 人工智能目前存在的问题、未来目标和实现方法

- 现有方法及其缺陷：

**知识驱动（符号主义）**：用符号模型来模拟理性行为，符号模型可以准确、严格地在符号空间里表达信息的内容。

1. 缺乏足够的常识

2. 缺乏数学工具

**数据驱动（联结主义）**：用特征空间的向量模拟感性行为，这方面的数学工具非常多，例如优化的工具、概率统计的工具。所以数据驱动方法这几年发展非常快。

1. 缺乏语义

2. 依赖于数据质量和数量

3. 小错不犯，一犯就犯大错

## 1.5 人工智能目前存在的问题、未来目标和实现方法

- 处于弱人工智能阶段，无法处理一些人类觉得容易的事情。

## 目标：带有理解的AI

Ex:

鲨鱼为什么那么可怕？  
因为它们都是食肉动物

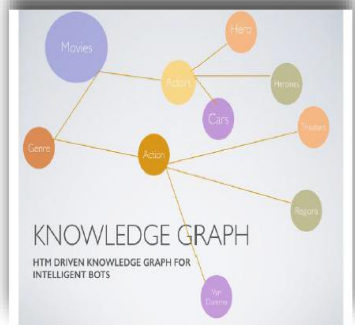
## 概念

鸟儿为何能够飞翔？  
因为它们有翅膀

## 属性

## 鹿晗关晓彤最近为何刷屏？ 因为关晓彤是鹿晗女朋友

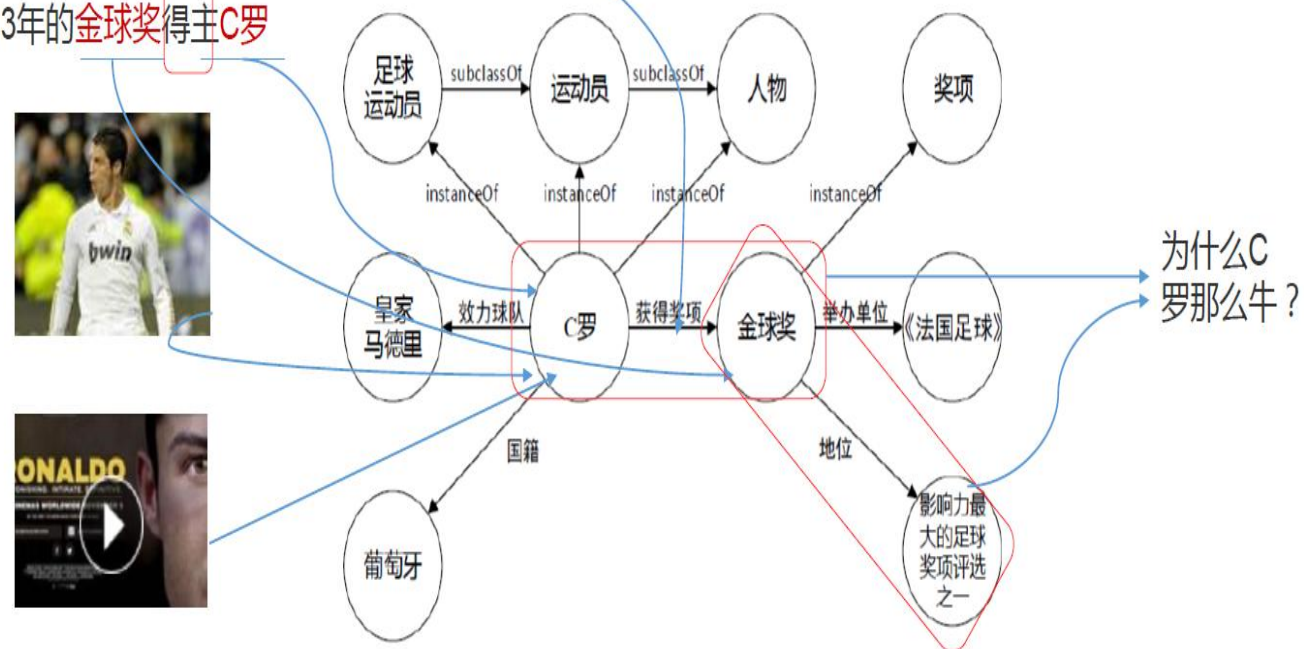
## 关系



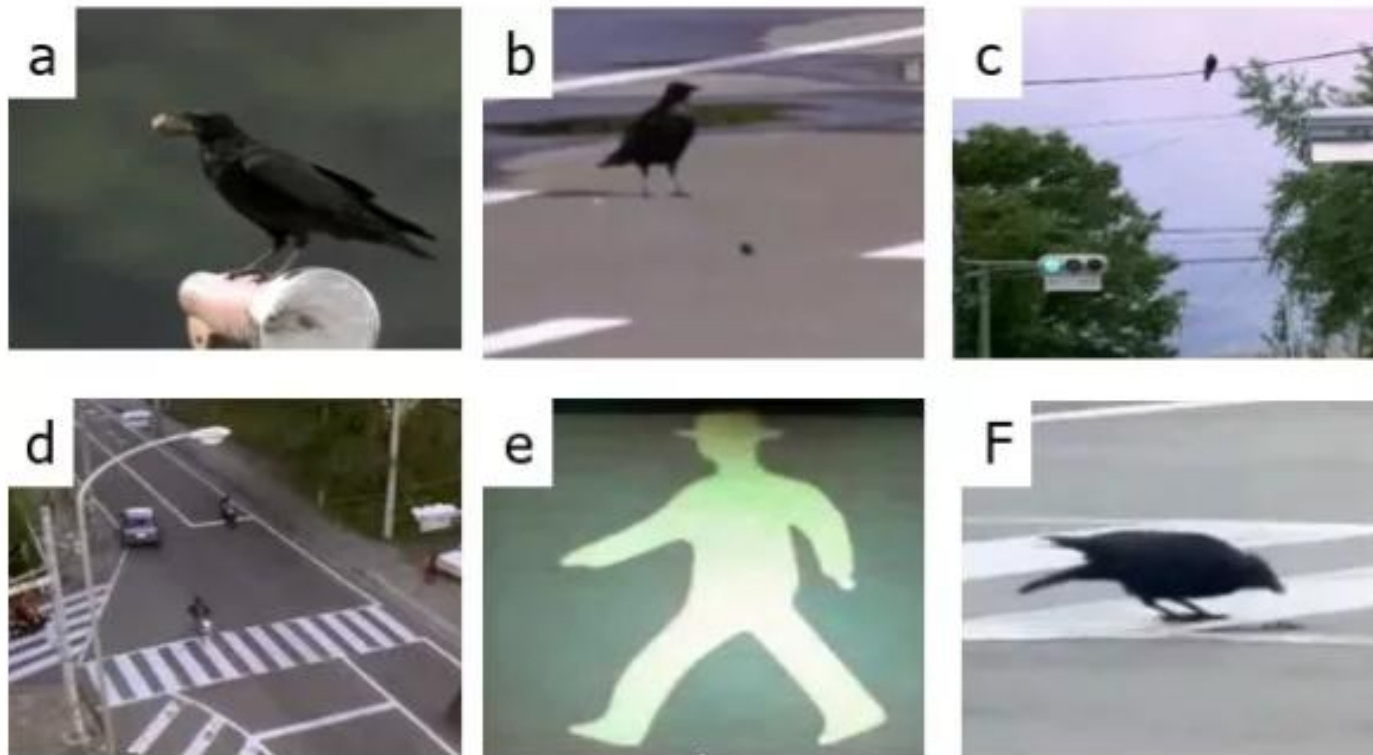
解释取决于人类认知的基本框架；  
概念、属性、关系是认知的基石

"Concepts are the glue that holds our mental world together"  
--Gregory Murphy

2013年的金球奖得主C罗



# 一只乌鸦的启示



- 其一、它是一个完全自主的智能。感知、认知、推理、学习、和执行，它都有。
- 
- 其二、它有大数据学习吗？这个乌鸦有几百万人工标注好的训练数据给它学习吗？没有，它自己把这个事通过少量数据想清楚了，没人教它。
- 
- 其三、乌鸦头有多大？不到人脑的1%大小。人脑功耗大约是10-25瓦，它就只有0.1-0.2瓦，就实现功能了，根本不需要核动力发电。这给硬件芯片设计者也提出了挑战 and 思路。

## 目标： 小数据、大任务范式

- 大数据、小任务范式 => 小数据、大任务范式



- 智能系统的根源可以追溯到两个基本前提条件：
- 一、物理**环境**客观的现实与因果链条。
- 二、智能物种与生俱来的**任务**与价值链条。

机器人的脑、人脑都可以看成一个模型。

什么驱动了模型在空间中的运动，也就是学习的过程？

- 一、 外来的数据。
- 二、 内在的任务。

## 实现目标的方法

- 1.建立关于符号和向量、数据驱动和知识驱动统一的理论。
- 2.物理和生物中相关理论体系的统一
- 3. 脑科学的研究
- ...