

# 《人工智能导论》

宫文娟

软件工程系





中國石油大學(华东)  
CHINA UNIVERSITY OF PETROLEUM

计算机科学与技术学院  
College of Computer Science & Technology

# 第一章 绪论

人工智能课题组

软件工程系





# 目录

## CONTENTS

1.1

人工智能的起源和定义

1.2

人工智能的流派

1.3

人工智能的进展和发展趋势





# 什么是人工智能 (机器智能)?



## 1.1 什么是人工智能-《人工智能导论》20212

- 



## 1.1 人工智能的起源和定义

- 现代人工智能的起源是1956年的达特茅斯会议。
- 参加者：麦卡锡、明斯基、香农、罗切斯特、纽厄尔、西蒙、萨缪尔、伯恩斯坦、摩尔、所罗门诺夫。
- 会议的主要成就就是让人工智能成了一个独立的研究学科。定下来人工智能的英文名称是“Artificial Intelligence”，简称AI。

## 1.1 人工智能的起源和定义

- 现代人工智能的起源是1956年的达特茅斯会议。
- 参加者：麦卡锡、明斯基、香农、罗切斯特、纽厄尔、西蒙、萨缪尔、伯恩斯坦、摩尔、所罗门诺夫。
- 会议的主要成就就是让人工智能成了一个独立的研究学科。定下来人工智能的英文名称是 “**Artificial Intelligence**”，简称AI。



## 1.1 人工智能的起源和定义

- 历史上有很多人工智能的定义，最常见的有两个：
- 明斯基提出：“人工智能是一门科学，是使机器做那些人需要通过智能来做的事情”。







## 1.1 人工智能的起源和定义

- 历史上有很多人工智能的定义，最常见的有两个：
- 明斯基提出：“人工智能是一门科学，是使机器做那些人需要通过智能来做的事情”。





## 1.1 人工智能的起源和定义

- 历史上有很多人工智能的定义，最常见的有两个：
- 明斯基提出：“人工智能是一门科学，是使机器做那些人需要通过智能来做的事情”。
- 尼尔森给出：“人工智能是关于知识的科学”，所谓“知识的科学”就是知识的表示、知识的获取和知识的运用。
- AI的研究是以知识的表示、知识的获取、知识的应用为归依。



# 什么是知识?





## 1.1 什么是知识-《人工智能导论》20212

- 





## 1.1 人工智能的起源和定义

- 知识的定义及否定

- 柏拉图在《泰阿泰德篇》中给出，即“被证实的、真的和被相信的陈述”，简称知识的JTB (Justified True Belief) 条件。
- 然而，这个延续了两千多年的定义在1963年，被哲学家盖梯尔否定了。

- 例子：

史密斯看到琼斯每天开着一辆福特车上班，于是他相信：

c：琼斯拥有一辆福特车。

史密斯有一位朋友叫“布朗”，而他并不知道布朗在什么地方，但他可以从命题c逻辑地推出：

d：琼斯拥有一辆福特车或者布朗在巴塞罗那。

## 1.1 人工智能的起源和定义

### • 知识的定义及否定

- 柏拉图在《泰阿泰德篇》中给出，即“被证实的、真的和被相信的陈述”，简称知识的JTB (Justified True Belief) 条件。
- 然而，这个延续了两千多年的定义在1963年，被哲学家盖梯尔否定了。

### • 例子：

史密斯看到琼斯每天开着一辆福特车上班，于是他相信：

c：琼斯拥有一辆福特车。

史密斯有一位朋友叫“布朗”，而他并不知道布朗在什么地方，但他可以从命题c逻辑地推出：

d：琼斯拥有一辆福特车或者布朗在巴塞罗那。

d满足JTB的三个条件：1、d是真的；2、d是史密斯的一个信念，即史密斯相信d；3、史密斯相信d是有充分理由的，即被证实的。



## 1.1 人工智能的起源和定义

### • 知识的定义及否定

- 柏拉图在《泰阿泰德篇》中给出，即“被证实的、真的和被相信的陈述”，简称知识的JTB (Justified True Belief) 条件。
- 然而，这个延续了两千多年的定义在1963年，被哲学家盖梯尔否定了。

### • 例子：

史密斯看到琼斯每天开着一辆福特车上班，于是他相信：

c: **琼斯拥有一辆福特车。**

史密斯有一位朋友叫“布朗”，而他并不知道布朗在什么地方，但他可以从命题c逻辑地推出：

d: **琼斯拥有一辆福特车**或者**布朗在巴塞罗那。**

这表明，满足JTB不是史密斯知道d的充分条件，因而JBT不是知识的恰当定义。

## 1.1 人工智能的起源和定义

### • 概念的定义

- 经过盖梯尔悖论之后，虽然后来给出了很多替代定义，但是直到现在仍然没有定论。但是至少有一点是明确的：
- 知识的基本单位就是概念。知识核心是概念甚至知识本身就是概念。
- 概念由三部分组成：
- **概念的符号**：即概念的名称，说明这个概念叫什么。
- **概念的内涵**：即命题来表示，命题就是能判断真假的陈述句。
- **概念的外延**：即经典集合来表示，用来说明与概念对应的实际对象是哪些。

例如：

- 概念的符号：素数 (Prime Number)
- 概念的内涵：只能被1和自身整除的自然数
- 概念的外延：{1, 2, 3, 5, 7, 11, 13, 17, ...}





虽然我们没法确定地定义知识，  
但是我们知道知识的基本单位是概念。

## 1.1 人工智能的起源和定义

### 概念的三个功能：指物功能

- 指物功能：即指向客观世界的对象，表示客观世界的对象的可观测性。对象的可观测性是指对象对于人或者仪器的知觉感知特性，不依赖于人的主观感受。
- 例如：《阿Q正传》中的一段话。“那赵家的狗，何以看我两眼呢？” ，句中赵家的狗应该是指现实世界当中一条真实的狗。
- 但概念的指物功能有时不一定能够实现，有些概念及其设想存在对象在现实世界并不存在，例如鬼。

## 1.1 人工智能的起源和定义

### 概念的三个功能：指心功能

- 第二个功能是指心功能，即指向人心智世界里的对象，代表心智世界里的对象表示。
- 鲁迅有一篇著名的文章《论丧家的资本家的乏走狗》，显然这个狗不是现实世界的狗，而是指梁实秋先生。只是他心智世界的狗，即心里的狗。在客观世界，梁实秋先生显然无论如何不是狗。
- 概念的指心功能一定存在，如果对于某一个人，一个概念的指心功能没有实现，则该词对该人不可见，简单的说就是，这个人理解这个概念。

## 1.1 人工智能的起源和定义

### 概念的三个功能：指名功能

- 指名功能既指向认知世界或者符号世界表示对象的符号名称，这些符号名称组成各种语言。最著名的例子是乔姆斯基的“colorless green ideas sleep furiously”这句话翻译过来是“无色的绿色思想在愤怒地休息”。这句话没有什么意思，但是完全符合语法，纯粹是在语义符号世界里，即仅仅指向符号世界而已。
- 也有例外，“鸳鸯两字怎生书”指的是“鸳鸯”两个字组成的名字。
- 一般情况下，概念的指名功能依赖于不同的语言系统或者符号系统，由人类创造，属于认知世界。同一个概念在不同的符号系统里，概念名不一定相同，如汉语的“雨”，英语称“rain”。

## 1.2 人工智能的三个流派

- 在现实生活中，如果你要了解一个概念，就需要知道这个概念的三个功能：
  - 要知道概念的名字
  - 要在自己的心智世界里具有该概念的形象
  - 也要知道概念所指的对象
- 知道了概念的三个功能后，就可以理解人工智能的三个流派以及各流派之间的关系。
- 人工智能也是一个概念，而要使一个概念成为现实，自然要实现概念的三个功能。
  - 专注于实现AI指名功能的人工智能流派，称为**符号主义**
  - 专注于实现AI指心功能的人工智能流派，称为**连接主义**
  - 专注于实现AI指物功能的人工智能流派，称为**行为主义**



## 1.2 人工智能的三个流派

抽象心智世界

符号主义

指名功能

心智世界

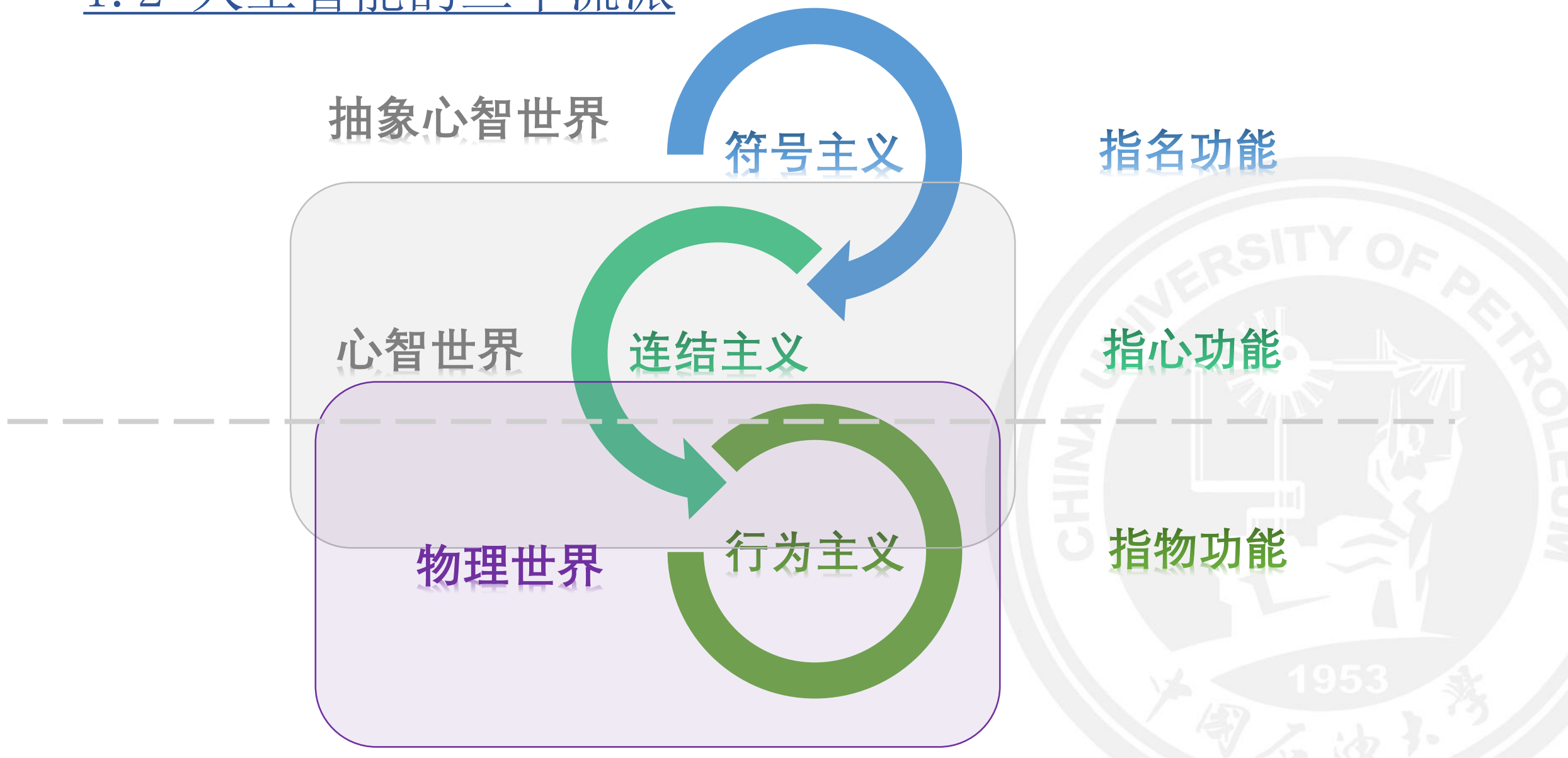
连结主义

指心功能

物理世界

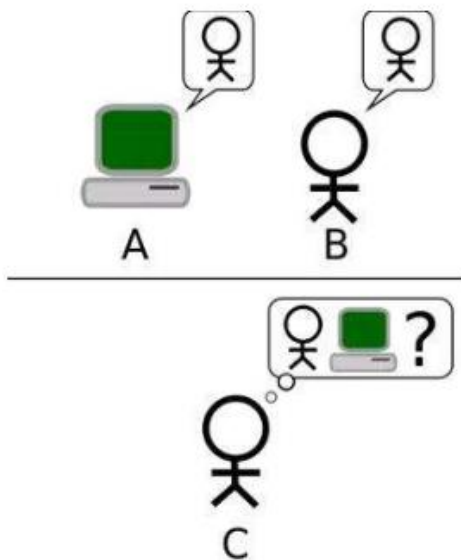
行为主义

指物功能



## 1.2.1 符号主义

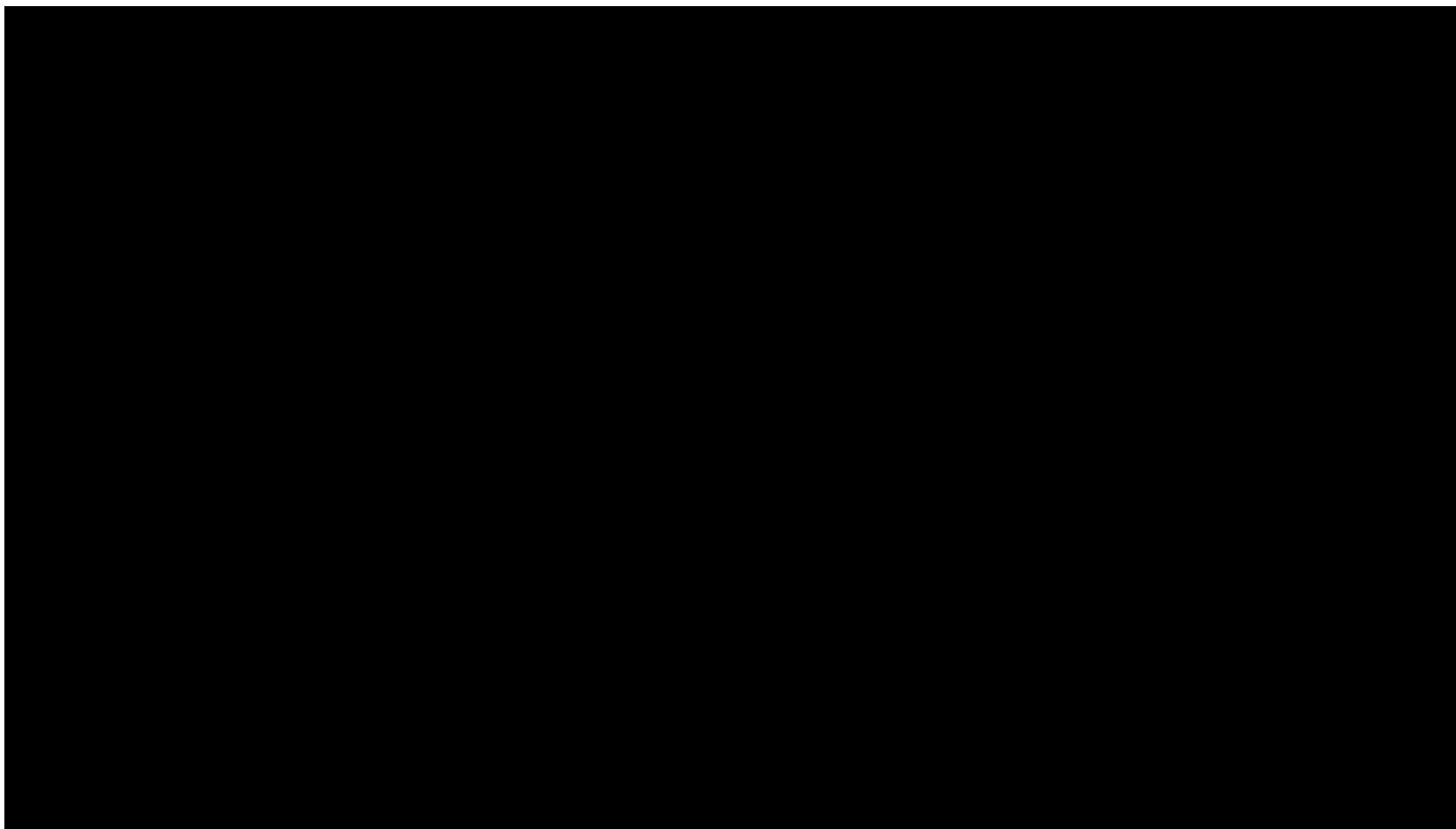
- 符号主义的代表人物是Simon与Newell。他们提出了物理符号系统假设。即只要在符号计算上实现了相应的功能，那么在现实世界就实现了对应的功能，这是智能的充分必要条件。
- 符号主义认为只要在机器上是正确的，现实世界就是正确的。说得更通俗一点，指名对了，指物自然正确。
- 在哲学上，关于物理符号系统假设有一个著名的思想实验图灵测试





## 1.2.1 符号主义

- 在哲学上，关于物理符号系统假设有一个著名的思想实验图灵测试





## 1.2.1 符号主义

### 符号主义面临的挑战

- 图灵测试将智能的表现完全限定在指名功能里，但是指名功能里实现了概念的功能，并不一定实现了概念的指物功能。哲学家Searle专门设计了中文屋实验。



## 1.2.1 符号主义

### 符号主义面临的挑战

- 图灵测试将智能的表现完全限定在指名功能里，但是指名功能里实现了概念的功能，并不一定实现了概念的指物功能。哲学家Searle专门设计了中文屋实验。



## 1.2.1 符号主义

### 符号主义的挑战与限制

- 中文屋实验明确说明：即使符号主义成功了，这全是符号的计算跟现实世界也不一定搭界，即完全实现指名功能也不见得有智能。
- 另外，实现符号主义面临的现实条件主要有三个：
  - 第一个是概念的组合爆炸问题。每个人掌握的基本概念大约有5万个，而形成的组合概念却是无穷的。
  - 第二个是命题的组合悖论问题。两个都是合理的命题和起来，就变成了没法判断真假的句子了。
  - 第三个也是最难的问题是经典概念在现实生活当中是很难得到，知识也很难提取。



只实现符号计算，  
不能实现人工智能。

## 1.2.2 连接主义

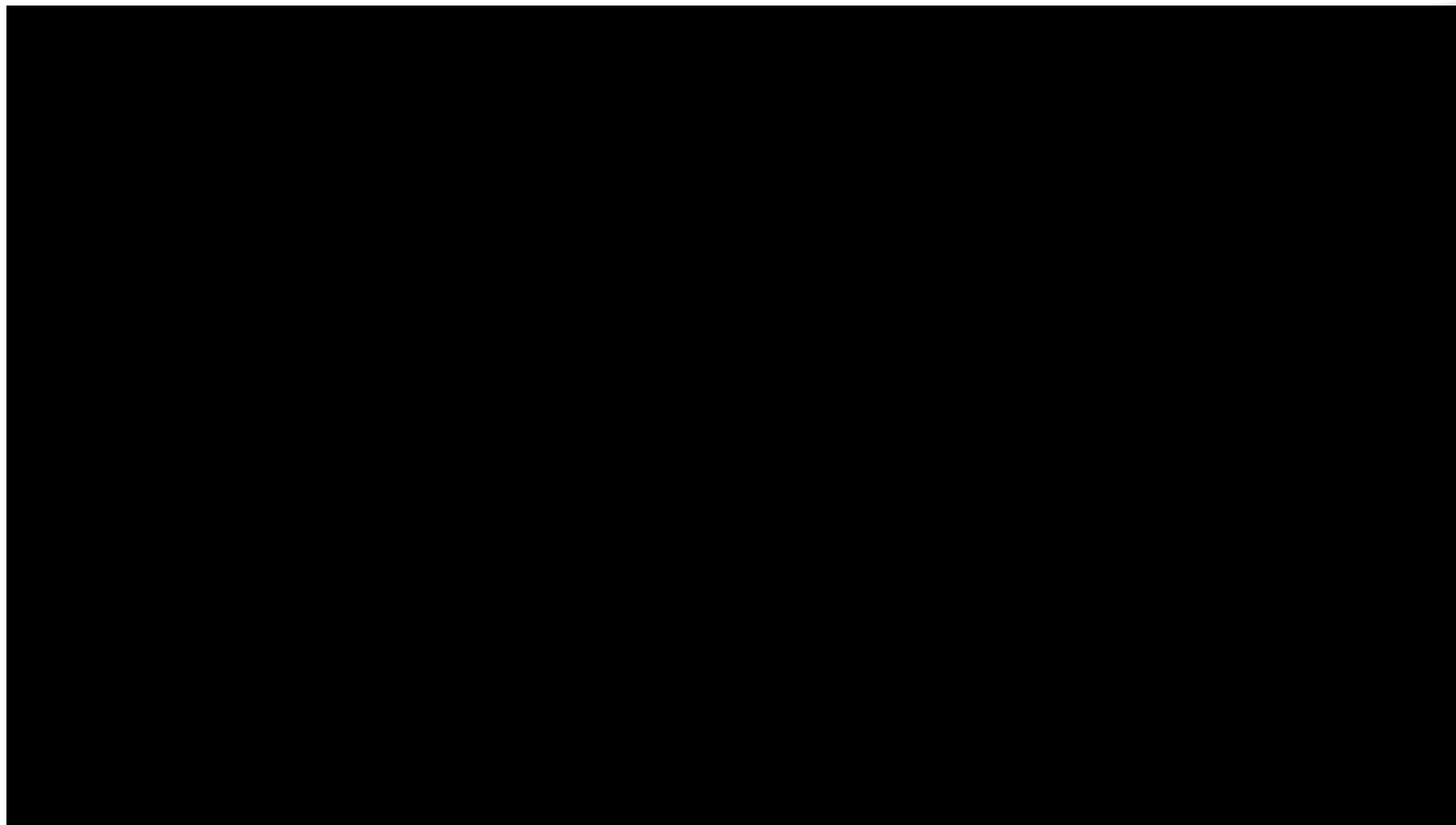
- 连接主义认为大脑是一切职能的基础，主要关注于大脑神经元及其连接机制，试图发现大脑的结构及其处理信息的机制，揭示人类智能的本质机理，进而在机器上实现相应模拟。
- 连接主义学派的早期代表人物有麦克洛克、皮茨、霍普菲尔德等。
- 连接主义认为可以实现完全的人工智能，对此，哲学家普特南设计了著名的“缸中之脑实验”，可以看作是对连接主义的一个哲学批判。





## 1.2.2 连接主义

- 连接主义认为可以实现完全的人工智能，对此，哲学家普特南设计了著名的“缸中之脑实验”，可以看作是对连接主义的一个哲学批判。





## 1.2.2 连接主义

### 连接主义存在的问题

- “缸中之脑”实验说明：即使实现了连接主义，指心没有问题，但指物依然存在问题，因此连接主义实现的人工智能也不等同于人的智能。
- 尽管如此，连接主义仍是目前最为大众所知的一条AI实现路线。
  - 在围棋上，采用了深度学习技术的AlphaGo战胜了李世石，之后又战胜了柯洁。
  - 在机器翻译上，深度学习技术已经超过了人的翻译的水平。
  - 在语音识别和图像深度学习也达到了使用标准。
- 客观地说，深度学习的研究成就已经取得了工业级的进展，但是这并不意味着连接主义就可以实现人工智能，更重要的是，即使要实现完全的连接主义，也面临极大的挑战。
  - 因为到现在为止，人们并不清楚人脑表示概念的机制，也不清楚人脑中概念的具体表示形式，表示方式和组合形式等等。
  - 现在的神经网络和深度学习，实际上与人脑真正的机制距离尚远，并非人脑的运行机制



只模拟大脑功能，  
不能实现人工智能。





## 1.2.3 行为主义

- 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，只需要将智能行为表现出来就好，即只要能够实现指物功能，就可以认为具有智能了。
- 这一学派的早期代表作是Brooks的六足爬行机器人。
- 对此普特南也设计了一个思想实验，可以看作是对行为主义的哲学批判，这就是“完美伪装者和斯巴达人”。
- 完美伪装者可以根据外在的需要进行完美的表演，需要哭的时候就可以哭得撕心裂肺，需要的笑时候就可以笑得兴高采烈，但其内心可能是始终平静如常。斯巴达人则相反，无论其内心是激动万分，还是心冷似铁，总是一副泰山崩于前而色不变的表情。完美的伪装者和斯巴达人的外在表现，都与内心没有联系，这样的智能又如何从外在的行为进行测试呢？因此行为主义实现的人工智能也不等同于智能。



只对环境做出正确反应，  
不能实现人工智能。

## 1.3 人工智能的进展和发展趋势

- 概念的指名、指物与指心功能，在生活中并不等价，单独实现概念的一个功能并不能够保证具有智能，因此单独遵循一个学派不足以实现人工智能。
- 现在的人工智能研究已经不再强调遵循人工智能的单一学派，很多时候会综合各流派的技术，比如：
  - 从专家系统发展起来的知识图谱，已经不完全遵守符号主义的路线了。
  - 在围棋上战胜人类顶尖选手的AlphaGo，综合使用了三种算法：
    - 强化学习（行为主义）
    - 蒙特卡罗树搜索（符号主义）
    - 深度学习（连接主义）
  - 无人驾驶同样是突破的人工智能三大流派限制的综合技术。
- 虽然人工智能发展至今，各个流派依然在发展，也都取得了很好的进展，但是各个流派进行融合已经是大势所趋。



## 本门课组织结构

- 本文课的内容就包括，知识表示、知识获取、知识应用三部分。
- 知识表示包含概念表示、知识表示、知识图谱三部分内容；
- 知识获取包括搜索技术、群智能算法、机器学习、人工神经网络和深度学习四部分内容。
- 知识应用包含计算机视觉、自然语言处理、语音处理、专家系统、规划、多智能体系统和机器人六部分内容。