# 绪论

## 什么是人工智能

From these two dimensions—human vs. rational and thought vs. behavior—there are **four** possible combinations.

GPT属于哪一种？

### 类人行为：图灵测试方法

**图灵测试（Turing test）**是由艾伦·图灵（Alan Turing）提出的（1950），它被设计成一个思维实验，用以回避“机器能思考吗？”这个哲学上模糊的问题。如果人类提问者在提出一些书面问题后无法分辨书面回答是来自人还是来自计算机，那么计算机就能通过测试。

To program a computer to pass a rigorously applied test, the computer would need the following capabilities:

1. 自然语言处理（natural language processing），以使用人类语言成功地交流；
2. 知识表示（knowledge representation），以存储它所知道或听到的内容；
3. 自动推理（automated reasoning），以回答问题并得出新的结论；
4. 机器学习（machine learning），以适应新的环境，并检测和推断模式。

**完全图灵测试（total Turing test）**，该测试需要与真实世界中的对象和人进行交互。为了通过完全图灵测试，机器人还需要具备下列能力：

1. 计算机视觉（computer vision）和语音识别功能，以感知世界；
2. 机器人学（robotics），以操纵对象并行动。

### 类人思考：认知建模方法

我们可以通过3种方式了解人类的思维：

1. 内省（introspection）——试图在自己进行思维活动时捕获思维；
2. 心理实验（psychological experiment）——观察一个人的行为；
3. 大脑成像（brain imaging）——观察大脑的活动。

区分人工智能和认知科学

algorithm performs well on a task≠a good model of human performance

### 理性思考：**“**思维法则**”**方法

1. 任何用逻辑符号描述的可解问题在原则上都可以用程序求解。人工智能中所谓的**逻辑主义（logicism）**传统希望在此类程序的基础上创建智能系统。
2. 逻辑要求关于世界的认知是确定的，而实际上这很难实现，概率（probability）论填补了这一鸿沟，允许我们在掌握不确定信息的情况下进行严格的推理。

### 理性行为：理性智能体方法

1. 理性智能体（rational agent）需要为取得最佳结果或在存在不确定性时取得最佳期望结果而采取行动。
2. The rational-agent approach to AI has **two advantages** over the other approaches. **First,** it is **more general** than the “laws of thought” approach because correct inference is just one of several possible mechanisms for achieving rationality. **Second,** it is more amenable to **scientific development**. The standard of rationality is mathematically well-defined and completely general.

简而言之，人工智能专注于研究和构建做正确的事情的智能体，其中正确的事情是我们提供给智能体的目标定义。这种通用范式非常普遍，以至于我们可以称之为**标准模型（standard model）**。

### Beneficial machines

1. 在真实世界中，我们越来越难以完全正确地指定目标，在我们的真实需求和施加给机器的目标之间达成一致的问题称为价值对齐问题（value alignment problem），即施加给机器的价值或目标必须与人类的一致。
2. 如果我们不能将这些目标完美地传达给机器，就需要一个新的表述，也就是机器正在实现我们的目标，但对于目标是什么则是不确定的。当一台机器意识到它不了解完整的目标时，它就会有谨慎行动的动机，会寻求许可，并通过观察来更多地了解我们的偏好，遵守人为控制。最终，我们想要的是对人类可证益的（provably beneficial）智能体。

## 人工智能基础

与各大学科的联系

## 人工智能历史

总结人工智能历史里程碑的快速方法是列出**图灵奖**得主

### 人工智能的诞生（1943—1956）

人工神经元模型

修改神经元之间连接强度的简单更新规则，现在称为赫布型学习

第一台神经网络计算机——SNARC

西洋跳棋程序

图灵论文“*Computing Machinery and Intelligence*”

“逻辑理论家”（Logic Theorist，LT）的数学定理证明系统

达特茅斯会议

### 早期热情高涨，期望无限（1952—1969）

通用问题求解器（GPS）

物理符号系统（physical symbol system）假说

几何定理证明程序（Geometry Theorem Prover）

亚瑟·萨缪尔对西洋跳棋的研究（强化学习）

人工智能编程语言Lisp

线性自适应神经网络（adaline）

感知机（perceptron）

### 一些现实（1966-1973）

早期系统在更困难的问题上都失败了，这些失败主要出于两个原因：

1. 第一个主要原因是许多早期人工智能系统主要基于人类如何执行任务的“知情内省型”，而不是基于对任务、解的含义以及算法需要做什么才能可靠地产生解的仔细分析。
2. 第二个主要原因是对人工智能要求解的问题的复杂性缺乏认识。大多数早期的问题求解系统都会尝试组合不同的步骤，直到找到解为止。

### 专家系统（1969-1986）

1. DENDRAL程序：第一个成功的知识密集型系统，解决了从质谱仪提供的信息推断分子结构的问题
2. MYCIN系统：诊断血液感染，引入了一种称为确定性因子（certainty factor）的不确定性计算

事实证明，为复杂领域构建和维护专家系统是困难的，一部分原因是系统使用的推理方法在面临不确定性时会崩溃，另一部分原因是系统无法从经验中学习。

### 神经网络的回归（**1986—**现在）

**反向传播（back-propagation）学习算法**

这些所谓的**联结主义（connectionist）模型**被一些人视为纽厄尔和西蒙的符号模型以及麦卡锡和其他人的逻辑主义方法的直接竞争对手。

联结主义模型可能以一种更流畅和不精确的方式形成内部概念，更适配真实世界的混乱。它们还具备从样本中学习的能力，它们可以将它们的预测输出值与问题的真实值进行比较，并修改参数以减少差异，使它们在未来的样本中更有可能表现良好。

### 概率推理和机器学习（1987—现在）

1. 语音识别领域：**隐马尔可夫模型（hidden Markov model，HMM）**，HMM有两个相关的方面。首先，它们基于严格的数学理论。这使得语音研究人员能够在其他领域数十年数学成果的基础上进行开发。其次，它们是在大量真实语音数据的语料库上训练而产生的。这确保了健壮性。
2. 从数据中学习**贝叶斯网络**
3. 强化学习与马尔可夫决策过程联系

### 大数据（**2001—**现在）

大数据的可用性和向机器学习的转变帮助人工智能恢复了商业吸引力。

### 深度学习（**2011—**现在）

深度学习（deep learning）是指使用多层简单的、可调整的计算单元的机器学习。

20世纪90年代：卷积神经网络（convolutional neural network）

直到2011年，深度学习方法才真正开始流行起来，首先是在语音识别领域，然后是在视觉物体识别领域。

ALPHAGO（Silver et al., 2016, 2017,2018）之所以能够战胜人类顶尖的围棋棋手，是因为它使用了深度网络来表示评价函数。

深度学习在很大程度上依赖于强大的硬件，一个标准的计算机CPU每秒可以进行109或1010次运算。运行在特定硬件（例如GPU、TPU或FPGA）上的深度学习算法，每秒可能进行1014～1017次运算，主要是高度并行化的矩阵和向量运算。

### 目前的先进技术

自动驾驶、腿足式机器人、自动规划和调度、机器翻译、语音识别、推荐系统、博弈、图像理解、医学、气候科学

## 人工智能的风险与收益

致命性自主武器、监视和劝诱、有偏决策、就业影响、网络安全、安全关键的应用

通用人工智能（artificial general Intelligence，AGI）

超级人工智能（artificial superintelligence，ASI）