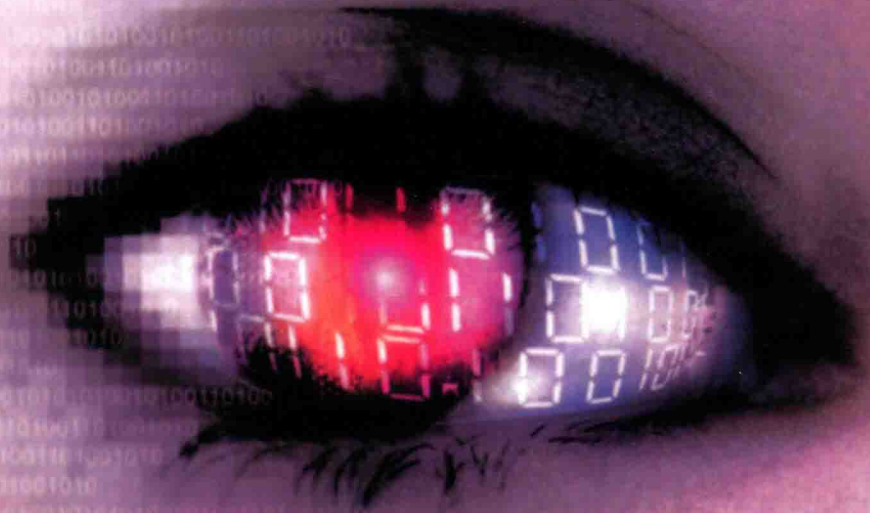


一部人工智能进化史

集人工智能领域顶级大牛、思维与机器研究领域
最杰出的哲学家多年研究之大成

关于人工智能的本质和未来更清晰、简明、切合实际的论述




AI: Its Nature and Future

AI

人工智能的本质与未来

【英】玛格丽特·博登 (Margaret A. Boden) / 著 孙诗惠 / 译

 中国人民大学出版社

AI: Its Nature and Future

AI

人工智能的本质与未来

【英】玛格丽特·博登 (Margaret A. Boden) / 著 孙诗惠 / 译

中国人民大学出版社

• 北京 •

图书在版编目 (CIP) 数据

AI: 人工智能的本质与未来 / (英) 玛格丽特·博登 (Margaret A. Boden) 著; 孙诗惠译. — 北京: 中国人民大学出版社, 2017.6

书名原文: AI: Its Nature and Future

ISBN 978-7-300-24430-3

I. ① A… II. ①玛… ②孙… III. ①人工智能—研究 IV. ① TP18

中国版本图书馆 CIP 数据核字 (2017) 第 116109 号

AI: 人工智能的本质与未来

[英] 玛格丽特·博登 著

孙诗惠 译

AI: Rengong Zhineng de Benzhi yu Weilai

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

邮政编码 100080

电 话 010-62511242 (总编室)

010-62511770 (质管部)

010-82501766 (邮购部)

010-62514148 (门市部)

010-62515195 (发行公司)

010-62515275 (盗版举报)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com> (人大教研网)

经 销 新华书店

印 刷 北京联兴盛业印刷股份有限公司

规 格 148mm × 210mm 32 开本

版 次 2017 年 6 月第 1 版

印 张 6.5 插页 2

印 次 2017 年 6 月第 1 次印刷

字 数 128 000

定 价 55.00 元

版权所有

侵权必究

印装差错

负责调换



阅读成就思想……

Read to Achieve



目 录

01 什么是人工智能

如今，人工智能无处不在。它不仅带来不计其数的技术小发明，还能够对生命科学产生深远的影响。同时，人工智能在给人类带来进步的同时，也向人类发出了挑战——如何看待人性，以及未来在何方。如果未来人工智能将全面超过人类的智能，那还有什么地方能保留人类的尊严和责任？

虚拟机 // 005

人工智能的主要类型 // 008

人工智能的预言 // 009

人工智能的起源 // 010

控制论 // 016

计算机建模者们分道扬镳 // 020

02 强人工智能：人工智能领域的圣杯

随着近来计算机能力的不断增强，强人工智能（AGI）在 21 世纪再次引起人们的兴趣。如果这一目标得以实现，人工智能系统将减少对专用

编程技巧的依赖，而受益于推理和知觉这些通用功能——语言、创造力和情感，然而，这谈何容易。通用智能仍然是一个严峻挑战，让人难以捉摸。强人工智能无疑是人工智能领域的圣杯。

只有超级计算机还远远不够 // 028

启发式搜索 // 029

人工智能领域中的规划 // 031

数学简化 // 034

知识表示 // 035

基于规则的程序 // 036

框架、词向量、脚本、语义网络 // 040

逻辑和语义网 // 043

计算机视觉 // 047

框架问题 // 050

智能体和分布式认知 // 052

机器学习 // 054

通用系统 // 058

梦想复兴 // 063

缺失的方面 // 065

03 语言、创造力和情感

人工智能的一些领域似乎特别具有挑战性，如语言、创造力和情感。如果人工智能不能模拟它们，要实现强人工智能就好似做白日梦。

语言 // 069

创造力 // 080

人工智能与情感 // 085

04 神经网络

神经网络有无数应用，从操控股票市场和监测货币波动到语音或面部识别。但真正有趣的是它们的运行方式。

神经网络更广泛的含义 // 094

分布式并行处理 // 096

神经网络学习 // 099

反向传播、大脑和深度学习 // 102

网络丑闻 // 107

连接不是一切 // 111

混合系统 // 113

05 机器人和人工生命

人工生命模拟生物系统。和人工智能一样，它有着技术和科学双重目的。人工生命对于人工智能而言不可或缺，因为已知的所有智能都可以在生物体上找到。在理解自组织的过程中，生物学对人工智能来说很有用，同样，人工智能对于生物学来说也很有用。

情境机器人和有趣的昆虫 // 119

进化人工智能 // 127

自组织 // 131

06 强人工智能会有真正的智能吗

假设未来的强人工智能系统（银幕上或机器人）能够匹敌人类的表现，那么它们会有真正的智能、理解力和创造力吗？它们会有自我、道德身份和自由选择吗？它们会有意识吗？如果没有意识，它们会有任何其他属性吗？这些显然不是科学问题，而是哲学问题。

- 图灵测试 // 142
- 意识的很多问题 // 144
- 机器意识 // 145
- 人工智能和现象意识 // 151
- 虚拟机和身心问题 // 155
- 意义和理解力 // 157
- 神经蛋白是必要条件吗 // 159
- 不只是大脑，身体也很重要 // 161
- 道德社区 // 162
- 道德、自由和自我 // 163
- 心智和生命 // 167
- 巨大的哲学分歧 // 169

07 奇点

奇点代表人工智能将达到人类水平的智能。不久的将来，强人工智能将变为超人工智能。届时系统将智能化到可以自我复制，从而在数量上超过人类，并且还可以自我提高，从而在思想上超越人类。最重要的问题和决定将交由计算机负责。

这个概念颇具争议性。它是否能够发生、什么时候发生，以及它到底是好事还是坏事，人们对此意见不一。

- 奇点的预言家 // 174
- 竞争的预测 // 176
- 为怀疑论辩护 // 180
- 全脑仿真 // 181
- 我们应该担心什么 // 185
- 我们为此做了些什么 // 191

- 译者后记 // 198



01

什么是人工智能

AI: Its Nature and Future

人工智能（Artificial Intelligence, AI）就是让计算机完成人类心智（mind）能做的各种事情。通常，我们会说有些行为（如推理）是“智能的”，而有些（如视觉）又不是。但是，这些行为都包含能让人类和动物实现目标的心理技能，比如知觉、联想、预测、规划和运动控制。

智能不是一维的，而是结构丰富、层次分明的空间，具备各种信息处理能力。于是，人工智能可以利用多种技术，完成多重任务。

人工智能无处不在。

人工智能的实际应用十分广泛，如家居、汽车（无人驾驶车）、办公室、银行、医院、天空……互联网，包括物联网（连接到小物件、衣服和环境中的快速增多的物理传感器）。地球以外的地方也有人工智能的影子：送至月球和火星的机器人；在太空轨道上运行的卫星。好莱坞动画片、电子游戏、卫星导航系统和谷歌的搜索引擎也都以人工智能技术为基础。金融家们预测股市波动以及各国政府用来指导制定公共医疗和交通决策的各项系统，也是基于人工智

能技术的。还有手机上的应用程序、虚拟现实中的虚拟替身技术，以及为“陪护”机器人建立的各种“试水”情感模型。甚至美术馆也使用人工智能技术，如网页和计算机艺术展览。当然，它还有一些应用不那么让人欢欣鼓舞，如在战场上穿梭的军事无人机——但是，谢天谢地，它也用在机器人扫雷舰上。

人工智能有两大主要目标：一个是技术层面的，利用计算机完成有益的事情（有时候不用心智所使用的方法）；另一个是科学层面的，利用人工智能概念和模型，帮助回答有关人类和其他生物体的问题。大多数人工智能工作者只关注其中一个目标，但有些也同时关注两个目标。

人工智能不仅可以带来不计其数的技术小发明，还能够对生命科学产生深远的影响。某一科学理论的计算机模型可以检验该理论是否清晰连贯，还能生动形象地证明其含义（通常是未知的）。理论是否正确另当别论，但其依据是从相关科学范畴得出的证据。就算我们发现该理论是错误的，结果也能够给人以启迪。

值得一提的是，心理学家和神经学家利用人工智能提出了各种影响深远的心智-大脑理论，如“大脑的运作方式”和“这个大脑在做什么”的模型：它在回答什么样的计算（心理）问题，以及它能采用哪种信息处理形式来达到这一目标等。这两个问题不一样，但都十分重要。还有一些问题尚未回答，因为人工智能本身已经告

诉我们：心智内容十分丰富，远远超出了心理学家们先前的猜想。

生物学家们也用到了人工智能——人工生命（A-Life）。利用这项技术，他们为生物体的不同内部结构建立了计算机模型，以解读不同种类的动物行为、身体的发育、生物进化和生命的本质。

人工智能对哲学也有影响。如今，很多哲学家对心智的解读也基于人工智能概念。例如，他们用人工智能技术来解决众所周知的身心问题、自由意志的难题和很多有关意识的谜题。然而，这些哲学思想都颇具争议。人工智能系统是否拥有“真正的”智能、创造力或生命，人们对此意见不一。

最后，人工智能向我们发出了挑战——如何看待人性，以及未来在何方。的确，有些人会担心我们是否真的有未来，因为他们预言人工智能将全面超过人的智能。虽然他们当中的某些人对这种预想充满了期待，但是大多数人还是会对此感到害怕。他们会问，如果这样，那还有什么地方能保留人类的尊严和责任？

我们将在接下来的几章逐一讨论上述问题。

虚拟机

谈到人工智能，人们可能会说：“那不就是指电脑嘛。”嗯，他们这么说既对也不对。电脑不是重点，重点是电脑做的事情。也就

是说，虽然人工智能离不开物理机（如电脑），但是我们最好把它看作计算机科学家所说的虚拟机。

虚拟机和虚拟现实中所描述的机器不一样，和训练机修工时所使用的模拟汽车引擎也不一样，它是程序员在编程时和人们使用它时所想到的信息处理系统。

让我们拿管弦乐队作类比。首先乐器是不能少的。要想让乐器演奏出美妙的音乐，那么木头、金属、皮革和弦线都必须遵循一定的物理定律。但观众在听音乐会时并不在意这一点，他们感兴趣的是音乐。他们也不在意单个音符，更不用说空气中发声的震动了。他们听的是音符产生的音乐“形状”：旋律与和声、主题与变奏、含混音与切分音。

当我们谈到人工智能时，情况也类似。用户使用设计师设计出来的文字处理器直接处理文字和段落。通常情况下，程序本身既不包含文字，也不包含段落（但有些段落也包含，比如用户可以很容易将版权标示插入到文字中）。神经网络（见第4章）也是并行处理信息，即使它通常是在约翰·冯·诺依曼（John von Neumann）结构计算机上（按顺序）实现的。

当然，这并不是说虚拟机只是杜撰或凭空想象出来的东西。虚拟机是真实存在的。我们不仅可以利用虚拟机完成系统内的任务（如果将其连接到照相机或机器人的手等这样的物理设备上），甚至

还可以做好外部世界的工作。如果程序突发问题，人工智能工作者通常很少去找硬件方面的原因，而是对虚拟机或软件中的事件和因果关系更感兴趣。

编程语言也是虚拟机（它的指令只有翻译成机器码后才能运行）。有些指令用更低级的编程语言进行定义，所以多个层级的指令都需要翻译。否则，要是用机器码的位组合模式处理信息，大多数人将无法思考。如果信息处理过程过于复杂且层级划分过于细化的话，那么也没有人能正常思考。

虚拟机不只是编程语言。虚拟机一般包含各个层级的活动模式（信息处理）。虚拟机也不只是在电脑上运行的虚拟机。在第6章中，我们将看到“人类的心智”也可以被看作在大脑中实现的虚拟机，更确切地说，是并行运行（在不同时间发展和学习得到的）且交互的虚拟机集合。

要实现人工智能领域的进步，我们需要不断完善有趣实用的虚拟机的定义。不断改良物理机（更大、更快）确实有好处，它甚至可能是实现某种虚拟机的必要条件。但是，只有具备海量信息的虚拟机才能在这些物理机上运行，否则后者就算功能再强大也没用（同理，要在神经科学领域取得进步，我们需要清楚了解在神经元上实现什么“心理”虚拟机，详见第7章）。

各类外部世界的信息得到充分利用。所有人工智能系统都需要

输入和输出设备，要是只需要一个键盘和一个屏幕就好了。它通常还需要专用传感器（可能是照相机或压敏晶须）或反应器（可能是供音乐或演讲用的声音合成器或机器人的手）。人工智能程序不仅处理内部信息，还与这些计算机的接口连接，或改变它们。

人工智能程序处理通常包含内部的输入和输出设备，供整个系统内部的虚拟机交互。例如，象棋程序的某一部分可能通过注意其他部分的情况来发现自己所面临的潜在威胁，这时候，它就有可能与那个部分配合，共同阻断本次威胁。

人工智能的主要类型

信息处理的方法取决于其所包含的虚拟机。我们将在后面的章节中看到，这主要有五种处理类型，每种处理类型又都包含很多变体。一种是经典逻辑或符号主义，有时称为有效的老式人工智能（Geod Old-Fashioned AI，以下简称 GOF AI）；另一种是人工神经网络或联结主义。此外，还有进化编程、细胞自动机以及动力系统。

工作者通常只使用一种方法来处理信息，但也存在混合虚拟机。例如，在第 4 章中提到的一个在符号主义处理和联结主义处理之间不断切换的人类行为理论（这解释了为什么有的人在完成计划任务的过程中，会分心去关注环境中与之无关的东西以及这种现象是如何发生的）。第 5 章描述了一款集“情境”机器人学、神经网络

络和进化编程三者于一体的感觉运动装置（在装置的协助下，机器人将纸板三角形用作地标，找到了“回家”的路线）。

除了实际应用外，这些方法能够启发心智、行为和生活。神经网络有助于模拟大脑的内部结构以及进行模式识别和学习。经典逻辑人工智能（特别是与统计学结合时）可以模拟学习、规划和推理。进化编程阐明了生物进化和大脑发育。细胞自动机和动力系统可用来模拟生物体的发育。有些方法更接近于生物学，而不是心理学；有些方法更接近非条件反射行为，而不是慎重思考。要想全面了解心智，除了要用到上述所有方法外，还可能更需要更多别的方法。

许多人工智能工作者并不关心心智的运作方式，他们只注重技术效率，而不追求科学理解。即使人工智能技术起源于心理学，但现在与心理学的联系却很少。然而，我们会发现，如果要想在强人工智能（artificial general intelligence）方面取得进步，我们需要加深理解心智的计算架构。

人工智能的预言

19世纪40年代，埃达·洛夫莱斯（Ada Lovelace）伯爵夫人预言了人工智能。更准确地说，她预言了部分人工智能。她专注于符号和逻辑，从未考虑过神经网络、进化编程和动力系统。她也未考虑过人工智能的心理目标，而纯粹对技术目标感兴趣。例如，她