



《如何创造可信的AI》

[美]盖瑞·马库斯 欧内斯特·戴维斯 著

龙志勇 译

浙江教育出版社

### 【作者简介】

[美]盖瑞·马库斯 (Gary Marcus)

新硅谷机器人创业公司 Robust.AI 首席执行官兼创始人。机器学习公司“几何智能”首席执行官兼创始人,该公司于2016年被优步收购,随后马库斯在优步创立了人工智能实验室。纽约大学心理学和神经科学教授。研究方向跨越人类和动物的行为,涉及神经科学、心理学、人工智能等多个领域。1994年于麻省理工学院博士毕业,师从心理学大师史蒂芬·平克。

[美]欧内斯特·戴维斯 (Ernest Davis)

纽约大学柯朗数学科学研究所计算机科学教授,人工智能领域科学家。

### 科学背景

鉴于这本书主题的重要性,我尝试从三个不同的角度为中国读者提供背景信息与书中核心内容的概述,希望有助于读者更有效地从阅读本书中获益。

我先介绍一下马库斯的个人背景,我认为他特殊的学术和行业履历能帮助读者更加全面地解读书中所表达的主要观点。马库斯在高中时代花了很多精力去开发一套计算系统,目的是把拉丁语自动翻译成现代英语。虽然这个项目没有成功,但整个过程让马库斯学到很多,特别是让他深切感受到要让一个计算系统具备类似人类般的认知能力和语言理解能力,纯粹依赖计算能力是远远不够的。他因此而形成的理念是,一个计算系统具备认知能力和语言理解能力的前提,是该系统必须具有一定程度的内在结构。

为了继续他在科学上的追求,马库斯在大学本科时期打破当时学校的常规,自我设定了认知科学作为主攻专业。在麻省理工学院学习期间,他师从世界著名心理学大师和认知科学家史蒂芬·平克教授。马库斯在他的博士论文工作中做了大量研究儿童学习英语的实验,他的研究成果为现代认知科学发展做出了非常重要的贡献。博士研究生毕业后,马库斯在纽约大学担任心理学教授,推动了一系列有影响力的学术研究,跨越计算机科学、认知科学、语言学 and 心理学等多个领域。

与此同时,马库斯在学术研究过程中形成了越来越完整的学术理念,对认知科学和人工智能的学术研究发展方向也形成了一系列清晰的观念,这一切可以从他所撰写的一系列学术著作中体现出来。

通过这些研究工作,马库斯在认知科学和人工智能等领域的国际学术界也享有了特殊的一席之地,一方面源于他所做的学术研究在这样的地位奠定了重要基础,另一方面源于他杰出的跨学科综合科研能力。他在计算机科学、认知科学、语言学、人

工智能等领域都练就了相当深厚的学术功底,更重要的是,马库斯一生追求科学,通过长期努力打造了一套自己的学术理念和体系。在追求科学真理的道路上,他敢于挑战学术界的主流观点。马库斯学术功底深,思路敏捷,能言善辩,敢于独树一帜。当整个人工智能学术界都在过分乐观地高歌猛进时,他不断撰文和发表演讲来指出许多人工智能核心技术中存在的弊端和局限性。他坚信自己的学术理念,不认同当今学术界的主流方向能把人工智能从今天带到未来。他敢于站出来泼冷水、唱反调,并敢于向人工智能学术界泰斗如杨立昆(Yann LeCun)等人发起多次公开辩论,因为马库斯深信这是追求科学真理所必需的。

更为重要的是,马库斯不仅是独树一帜的科学家,也是一个颇有成就的企业家和创业者,并专注于人工智能技术的开发与应用。他不但在学术上坚持自己的理论和方向,同时也把自己的学术理论和发展方向付诸实践。他在2014年创建了机器学习公司“几何智能”,由于在技术和应用上都做得相当不错,公司于2016年被当时如日中天的优步成功收购,马库斯也由此加入优步成为其第一任首席科学家。离开优步之后,马库斯联合著名机器人专家罗德尼·布鲁克斯等人一同创立了一家名为Robust.AI的新公司,公司专注于新一代的人工智能认知引擎研发,并以机器人为主要应用对象。我们可以看出,马库斯在本书中阐述的核心观念并不是纸上谈兵,而是在实际的创业创新过程中被不断验证和迭代的。

多年前,我在工作中认识马库斯之后,便一直保持联络,我们在人工智能的学术研究和实际应用上有着长期的交流。我也阅读了马库斯过去的著作,听了他的重要的学术演讲。我之所以花时间将马库斯的背景信息介绍给读者,是因为马库斯在本书中所阐述的核心内容是承上启下的。书中的内容既建立在过去累积的学术观念和理论之上,也是他对人工智能的现状和下一步发展方向所做的完整总结与梳理。马库斯以打造“可信的AI”为主题贯穿全书的内容。这里需要指出的是,马库斯所持有的观点在学术界和行业内是不无争议的。这样的争议在人工智能从今天走向未来的过程中,是健康的,也是必要的。重要的是,我们需要让每位读者都有一个全面的背景认知,以便更有效地从本书中获得有价值的信息、观点、方法论和启迪,不论读者是对作者的科学研究背景感兴趣,还是对开发人工智能技术应用有兴趣,不论读者是创业者,还是企业管理人员。

### AI 局限

本书的内容由两大部分组成。第一部分是第1章到第5章,马库斯非常详细和系统化地分析了今天以深度学习为基础的主流人工智能技术所面临的局限性。针对每一类被揭示出来的局限性,马库斯充分发挥了他对自然语言处理、机器人和计算机视觉等领域的科研经验和深刻理解,通

过生动易懂的案例把这些技术局限性的现象和原因清晰地描述给读者。马库斯强调在没有充分理解人工智能技术目前局限性的情况下大量开发人工智能应用将承担相当大的风险。为了帮助读者更方便地理解和判断这些风险,马库斯把这些局限性详细分成了9个类别。

第一部分的中心内容是马库斯对人工智能的核心技术——深度学习所提出的一系列质疑。这里采用了一系列具体案例和推理来强调我们必须在深度学习的基础之上探索新的技术突破。马库斯充分发挥了他在人工智能核心领域以及跨领域科研的实力。从深度学习核心算法的内在体系开始,贯穿多项人工智能发展的前沿应用领域,比如语言理解和机器人等,马库斯深入浅出地分享了对人工智能技术发展的核心观点:如果我们沿用目前以深度学习为主的人工智能技术,而不在其内在结构上探索新的途径,以后将很难建立起具有人类般的认知能力和真正意义上的智能系统。

## 打造可信的 AI

陆奇

奇绩创坛创始人兼 CEO,百度前总裁兼 COO

微软前全球执行副总裁,雅虎前执行副总裁

过生动易懂的案例把这些技术局限性的现象和原因清晰地描述给读者。马库斯强调在没有充分理解人工智能技术目前局限性的情况下大量开发人工智能应用将承担相当大的风险。为了帮助读者更方便地理解和判断这些风险,马库斯把这些局限性详细分成了9个类别。

第一部分的核心内容是马库斯对人工智能的核心技术——深度学习所提出的一系列质疑。这里采用了一系列具体案例和推理来强调我们必须在深度学习的基础之上探索新的技术突破。马库斯充分发挥了他在人工智能核心领域以及跨领域科研的实力。从深度学习核心算法的内在体系开始,贯穿多项人工智能发展的前沿应用领域,比如语言理解和机器人等,马库斯深入浅出地分享了对人工智能技术发展的核心观点:如果我们沿用目前以深度学习为主的人工智能技术,而不在其内在结构上探索新的途径,以后将很难建立起具有人类般的认知能力和真正意义上的智能系统。

在这里尤其要提出的是:在第一部分的开始阶段,马库斯花了大量篇幅来分析今天人工智能技术被媒体所报道的能力与目前人工智能技术能够实际落地的能力之间的对比,指出这两者之间存在严重的脱节。马库斯指出了学术界所存在的浮夸现象以及有些媒体过分乐观的炒作,将这种人工智能报道与现实之间的严重脱节叫作“AI 鸿沟”。他认为如果在推进人工智能科研和应用开发的过程中不能清晰地意识到这个鸿沟的话,我们将会踩很多坑。我们都知人工智能技术已经被广泛地应用,有一些是涉及关键使命的,比如自动驾驶系统,有一些则对社会有着深远的影响,比如信息检索和内容分发等。

特别是在应对目前全球面临的由新冠疫情带来的前所未有的挑战中,人工智能技术将起到越来越关键的作用。在这样的关键时机下,做出正确的判断并避免踩坑将是至关重要的。为此,马库斯追根溯源把“AI 鸿沟”归纳为三种“大坑”。第一是“轻信坑”,这是由于人类进化的现实过程还没有发展出清晰辨别人类与机器之间区别的能力,导致我们往往用基于人类的认知模式去看待机器的能力,从而容易轻信机器拥有类人类的智慧。第二是“虚幻进步坑”,每当AI技术的进展攻克了一类新的问题时,我们往往错误地假设AI技术就能解决以此推及的、现实世界中的类似任务。但是AI学术上的问题往往是定义在狭义而简化的假设下,而现实世界的具体任务都有很大的复杂性和不确定性。第三是“鲁棒坑”,受限于当前深度学习算法和训练数据,对容错性很低特别是使命关键的应用领域比如无人驾驶等,今天的AI还没能达到实际“落地”的能力。马库斯告诫我们必须关注“AI 鸿沟”,因为踩坑的代价是非常高的。

在这里我必须指出,马库斯是人工智能技术的坚信者,并在人工智能学术研究和实际应用上做了很多积极的推动工作。他也认为基于深度学习的人工智能技

术在过去十多年有了长足的进步,为一系列商业化应用做出了重要贡献。马库斯质疑当前主流的人工智能技术,是希望学术界和其他相关行业能一起更有效地克服目前人工智能技术的局限性,探索新技术上的突破,从而真正实现可信的AI。

### 未来核心

本书的第二部分是第6章到第8章,在其中马库斯提出了他认为的能通向未来可信AI的核心路径。这一路径的起点是在研究人类如何获得认知能力时所获得的核心启发,并以此建立下一代人工智能技术中必要的内在结构。这里的核心理念是马库斯在长期跨学科科研探索中逐渐建立起来的,我们可以从库斯马早期的学术著作中看到相应的端倪。马库斯以认知科学、心理学、语言学和哲学为基础提取出11条关键线索,认为这将对未来人工智能技术在发展过程中达到具备人类智能的鲁棒性起到关键作用。这些线索的共同核心是用内在更丰富的结构来表达信息、建立认知、建立起其他智能体系的核心要素,比如因果推理能力等。

在第二部分中,马库斯还强调了常识在实现未来通用人工智能中的重要性。对于如何建立常识,书中的内容充分体现了马库斯作为一个跨学科的学者和跨行业的企业家、创业者所拥有的综合能力,既有核心的科学原则又有务实的系统操作:从表达时间、空间和因果关系开始,建立一个有足够灵活性的框架并以此来连接感知、操作、语言的能力,并能不断地从环境以及与环境交互中学习,同时将先验与学习有机地融合在一起。这将是一个非常艰巨的任务,但马库斯认为这是必需的工作,是建立真正智能体系的必经之路。

最后,马库斯讨论了几个建立可信的AI所需要的实际能力。首先,我们要有能力来工程化地、有效可靠地开发AI系统与应用。正如计算机科学早期发展的历史一样,软件开发工具和开发流程是经过计算机软件工程这门子学科多年的努力才逐渐建立起来的,我们需要同样的努力来逐步打造工程化开发AI系统与应用的能力。其次,确保AI系统的安全性需要全新的开发与运营能力,尤其是关键使命的AI应用。最后,AI系统,比如机器人,需要其创造者赋予正当的道德伦理观念,类似于美国科幻作家艾萨克·阿西莫夫(Isaac Asimov)提出的机器人三定律。只有具备这些实际能力,我们才能真正创造可信的AI。

众所周知,人工智能是目前为止人类历史上创新潜能最大的技术发展浪潮。在未来的几十年里,它不但能诞生全新的大规模基础产业,也能极大地提升和改造所有现有行业。人工智能技术及应用所产生的商业价值和社会价值将是空前的。但要推动人工智能的健康发展,我们需要不断推进人工智能核心技术的进步,以及各种人工智能应用的开发和实践。

(标题为编者所加,有删节)