

人工智能革命

超级智能时代的人类命运

[英] 卡鲁姆·蔡斯 (Calum Chace) 著

张尧然 译

SURVIVING AI

The Promise and Peril of Artificial Intelligence



人工智能很可能导致人类的永生或者灭绝，
而这一切很可能在我们的有生之年发生。

人工智能革命

超级智能时代的人类命运

[英] 卡鲁姆·蔡斯 (Calum Chace) 著

张尧然 译



机械工业出版社

Surviving AI: The Promise and Peril of Artificial Intelligence / By Calum Chace / ISBN: 9780993211621

Copyright © 2015 by Calum Chace

All rights reserved. No part of this book may be reproduced by any means whatsoever without written permission from the publisher.

The Chinese edition Copyright © 2017 by China Machine Press

This title is published in China by China Machine Press with license from Calum Chace. This Chinese simplified language edition is authorized for sale throughout the world. Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书由 Calum Chace 授权机械工业出版社在全球出版与发行。未经许可的出口，视为违反著作权法，将受法律制裁。

北京市版权局著作权合同登记 图字：01-2016-3792 号。

图书在版编目 (CIP) 数据

人工智能革命：超级智能时代的人类命运 / (英) 卡鲁姆·蔡斯 (Calum Chace) 著；张尧然译。—北京：机械工业出版社，2017.5

书名原文：Surviving AI: The promise and peril of artificial intelligence
ISBN 978-7-111-56930-5

I. ①人… II. ①卡… ②张… III. ①人工智能 IV. ①TP18

中国版本图书馆 CIP 数据核字 (2017) 第 115582 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策划编辑：刘本明 责任编辑：刘本明

责任校对：孙丽萍

责任印制：李 昂

三河市宏达印刷有限公司印刷

2017 年 7 月第 1 版第 1 次印刷

145mm × 210mm · 6.25 印张 · 109 千字

标准书号：ISBN 978-7-111-56930-5

定价：49.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务

网络服务

服务咨询热线：010-88361066

机工官网：www.cmpbook.com

读者购书热线：010-68326294

机工官博：weibo.com/cmp1952

010-88379203

金书网：www.golden-book.com

封面无防伪标均为盗版

教育服务网：www.cmpedu.com

权威推荐

本书以严肃的态度、通俗的笔触，审视了人类在超级智能时代的风险与机遇。

——让·塔林 (Jaan Tallinn)，Skype、剑桥大学生存风险研究中心 (Centre for the Study of Existential Risk, CSER)、生命未来研究所 (Future of Life Institute, FLI) 联合创始人

如何理解人工智能的好处和风险，被认为是未来几十年里最大的挑战之一。只要你对这个问题感兴趣，不论你对人工智能的感觉是欢喜还是恐惧，本书都值得一读。

——戴维·舒克曼 (David Shukman)，英国广播公司 (BBC) 科学专栏编辑

随着人工智能越来越快地牵引自动化的发展节奏，现在已经应该深入思考它最终将会怎样改变我们的生活了。

人工智能真的会在所有领域超越我们吗？这幅机会与风险并存的未来场景，令每一个思考此问题的人感到紧张，至于使其怎样发展才最符合人类的利益，那就更令人头疼了。最近，有关该主题的学术分析大量涌现，蔡斯以高质量的工作、针对更广泛的受众，为这场大讨论贡献了一份力量。

——奥布里·德·格雷（Aubrey de Grey），SENS 研究基金会（SENS Research Foundation）首席安全官（CSO），人工智能研究专家

卡鲁姆·蔡斯针对未来广义人工智能及其潜在风险的关键问题，进行了清晰、简洁、启发式的总结。现在该领域众说纷纭，各执一词，新来者可能希望得到一本深入浅出的入门读物，蔡斯的这本书就为你进入该领域的殿堂提供了一个很好的起点。

作为一本可能涉及世界末日的书，本书的难得之处还在于：它既不哗众取宠，也没有把问题过分简单化。

——本·戈策尔（Ben Goertzel），Novamente 公司总裁

卡鲁姆·蔡斯是一位有先见之明的信使，为我们带来了有关人工智能风险与回报的信息。在本书中，他抓住了最根本的问题，并对其进行了深刻睿智的研究，他对这些

问题的思考，对我们探索答案很有帮助。蔡斯在人工智能的憧憬与危机之间敏锐地寻找平衡，对于所有好奇当今世界正在发生什么、我们是怎样走到今天、又将走向何方的人来说，本书都是一本必读书。

——肯尼思·库克耶（Kenneth Cukier），《大数据时代：生活、工作与思维的大变革》作者

如果你现在还没有思考人工智能，那就等于什么都没思考。每一个行业都需要权衡人工智能所带来的颠覆性威胁；每一个政策制定者都应该针对高生产率和颠覆性失业的同时出现做计划；每一个人都应该意识到，我们正在一条迷雾重重的路上飞奔，冲向强大而且无处不在的技术。本书以艺术手法讲述基础科学，其中讲述的案例，既可以让你在鸡尾酒会上谈笑风生，也可以在学术交流中侃侃而谈。

——克里斯·迈耶（Chris Meyer），《站在太阳上》作者

本书条理清晰、态度公允地介绍了有关人工智能的最新探讨，以及有关人工智能对未来人类影响的探讨。

——丹·高曼（Dan Goldman），英国帝国理工学院智能系统与网络集群讲座教授

超级人工智能的出现，可能会成为人类历史上最重要的事件。在本书中，卡鲁姆·蔡斯为读者提供了一个绝妙的指南，引领读者走近围绕人工智能的讨论混沌。总体来说，本书行文简洁流畅，几乎没有使用专业术语，这使得刚接触该领域的读者也能快速入门，了解为何很多声名卓著的思想家不愿对此问题公开发表意见。

——戴维·伍德 (David Wood)，伦敦未来学家协会主席

人工智能是我们这个时代最重要的技术。技术发展导致的失业，迫使我们经济结构进行彻底革新，而超级智能的出现也必将成为人类历史上最重大的事件。本书是了解这些内容的首选之作。

——布拉德·菲尔德 (Brad Feld)，Techstars 创始人

超级机器智能所带来的憧憬与危机，是如今热门的讨论话题。但是尼克·博斯特罗姆 (Nick Bostrom) 等人工智能专家的著作太过深奥，与埃隆·马斯克、比尔·盖茨、史蒂芬·霍金等人撰写的流行读物截然不同。卡鲁姆·蔡斯的这本书完美融合了两者的优势。它言简意赅而又严密地介绍了该领域的所有主要问题。对于新接触

这个主题的人来说，这是一本绝佳的读物。

——约翰·丹纳赫 (John Danaher)，伦理与新兴技术研究院 (Institute for Ethics and Emerging Technologies, IEET)

围绕人工智能的好处与潜在危险这个重要话题，人们的讨论有很大的分歧。卡鲁姆·蔡斯不偏不倚，清晰地介绍了各种主要观点。我们需要这样的讨论，而卡鲁姆的著作则是一本方便有用的指南。

——本·梅德洛克 (Ben Medlock)，SwiftKey 创始人

卡鲁姆·蔡斯这本书的问世让我倍受鼓舞，因为他强调，社会对智能机器（也就是我所说的“人工智能”）不断发展壮大的警惕心，最终将达到我所说的“第三阶段”，或者叫“形成共识阶段”（初始阶段 = 没有意识到，第一阶段 = 零星有人意识到，第二阶段 = 产生行动群体，第三阶段 = 形成共识，第四阶段 = 上升到政治层面）。作为 20 世纪 80 年代“第一阶段”中的少数人，我当时的感受是非常孤独的。看到蔡斯在书中把我所说的“主导物种之辩”介绍给大众读者，我感到很欣慰。很明显，如今人类已经进入了第三阶段。

蔡斯清晰务实的写作风格，将会把我们带入一个非常猛烈的、“超级大片”般的、与真实生活密切相关的第四阶段。如果你想在未来几十年里的第四阶段活下去，那就要早做准备，蔡斯这本书就万万不可错过。

——雨果·德·加里斯 (Hugo de Garis)，厦门大学教授，曾任人工大脑实验室主任，著有《人工智能的战争》

人工智能的发展已经开始影响其他所有领域。如果这种势头能在本世纪催生强大的人工智能，用某领域专家的话来说，那就将是“人类历史上最重大的事件”。因此，现在是思考蔡斯所提出的机会与挑战的最佳时间。

对于这个复杂而又充满争议的话题，本书条理格外清晰。作者进行了很好的研究，围绕主题进行了全方位介绍，确实是一本必不可少的入门读物。

——肖恩·黑吉尔泰恩 (Seán Ó hÉigartaigh)，剑桥大学生存风险研究中心执行主任

前 言

人工智能是人类最强大的技术。如今的软件不但可以解决各种问题，还能把数据转换为知识，它们正在越来越快地改变我们的生活。

对于大多数人来说，最触手可及的人工智能就是我们的智能手机。尽管我们早已对它司空见惯，但是仍然有很多人痴迷于此，因为它把全世界的信息都汇集到我们的指尖，还提供了诸如《愤怒的小鸟》和《植物大战僵尸》等精彩的游戏。

智能手机显然不仅仅是发达国家人民的奢侈品，它们提供的智能支付系统、在线教育和市场信息，让新兴国家的人们也能登上现代世界的舞台参与竞争。

智能手机的发展历程为人工智能的未来发展提供了一个值得借鉴的参考。三十年前，没有人能想象到我们会把强大的人工智能设备揣到口袋里，作为通信工具，用它来交流思想、交换意见、相互指导。我们能够在大型团队或

组织中协同工作，创造经济盈余[⊖]，研发新技术，是因为我们拥有无与伦比的交流能力。正因为如此，我们才能控制这个星球的命运，乃至地球上每一个物种的命运。

虽然人类无法预言未来，但是事后来看，我们最强大的人工智能技术，以通信工具的形式提供给大多数人使用，却是最符合逻辑的。

三十年前，我们还不知道移动通信的市场将会怎样发展。如今，我们也不知道正在改变无数行业的数字革命将会在未来三十年里如何演化。我们不知道，随着人工智能自动化的普及，会不会导致更多的人失业，也不知道人们能否像工业革命以来一样，在一部分人力被替代的同时，又创造出新的工作岗位。数字革命及自动化什么时候才能发展到智能手机这种程度？未来的样子也许和如今大多数人的期望都不一样，但是事后再来看，它一定是自然而然、顺理成章的。

预言未来并不容易。但是本书还是要亮明观点，我们认为：人工智能在给人类带来巨大好处的同时，也伴随着可怕的风险；我们应该密切关注正在发生的变化，并采取适当的策略，以便促成最好的结果。最终结果也许很糟糕，

⊖ 经济盈余 (economic surplus)，是指社会所生产的产品与生产它们的成本之间的差额。——译者注

也许很好，一切皆有可能，而且无法预料。导致这些结果的原因，部分在于运气，部分在于它们的内在逻辑，还有一部分，则在于全社会各阶层都能接受的应对策略。

每一个人都必须做好准备，尽可能灵活地迎接这个快速变化的世界所带来的挑战。各种组织必须积极预测与自身最相关的变化，并相应调整自己的战略战术。政府部门必须制定合理的政策，推动结果向更好的方向发展，防止那些最糟糕的情况发生。从某种程度上说，他们必须充分利用自身掌握的大量人力和物力资源。然而，由于社会各阶层对未来发展的看法还不确定，政府部门又必须慎重使用自己手头的资源。

我们现在已经看到，自动化和超级智能可能会对人类社会产生巨大影响。尽管很多人仍然对它们疑心重重，未来几十年也可能会有其他新力量涌现，但无论如何，本书的主要关注对象就是它们。

自动化可能会导致一个经济奇点。“奇点”一词源于数学和物理学，达到奇点状态时，一般规律将不再适用，对于事件视界[⊖]一侧的人来说，另一侧的世界是无法知晓的。

⊖ 事件视界 (event horizon)，天文学术语，指黑洞的边界。在视界一侧永远无法看到另一侧。——译者注

经济奇点的出现，可能会导致一种反乌托邦的专制政体，那里的精英人群掌握生产方法，压迫其他人。但它也可能导致物质经济的极大丰富，届时人们将不再需要为谋生而工作，我们将拥有完全的自由去追求乐趣，去拓展我们的思维，提升我们的技巧。我希望并且相信，后者是可能实现的，不过我们还是努力让这个实现的过程更加顺畅。

如果超级智能真的实现了，那将代表一个技术上的奇点，也将成为人类历史上最为重要的事件。如何在这个奇点发生之后生存，将是我们这一代和之后几代人面临的最重大挑战。

如果我们不犯错误，它将给我们的生活带来难以想象的改善。超级智能可以迭代改进自身的结构，拓展它的能力，从而解决你能想象到的几乎所有问题。届时死亡将成为一种选项，我们将可以享受永远幸福、永远快乐的生活。但是如果我们误入歧途，也可能招致灭顶之灾。因为这个风险如此巨大，本书将主要围绕超级智能展开，讨论它出现的可能性，以及它造福人类的可能性。

《人工智能革命：超级智能时代的人类命运》是《潘多拉的大脑》一书的姊妹篇，后者是一部有关第一个超级人工智能出现的科幻小说。更多有关这两本书的信息，请参阅网站 www.pandoras-brain.com。

目 录

权威推荐

前 言

第一部分 狭义人工智能 / 1

第 1 章 人工智能的历史及发展现状 / 3

1.1 定义 / 3

1.2 人工智能研究简史 / 7

1.3 人工智能发展现状 / 14

第 2 章 人工智能未来展望 / 31

2.1 未来的一天 / 31

2.2 以不同的速度把信息转化为知识 / 37

第 3 章 从数字化颠覆到经济奇点 / 41

3.1 数字化颠覆 / 42

3.2 杀人机器人 / 48

3.3 经济奇点 / 49

第二部分 广义人工智能 / 61

第4章 我们能造出广义人工智能吗? / 63

- 4.1 原理上是否可行? / 63
- 4.2 全大脑仿真 / 69
- 4.3 在狭义人工智能的基础上创造 / 80
- 4.4 有关思维的综合理论 / 89
- 4.5 结论:我们能否造出大脑? / 90

第5章 什么时候能造出广义人工智能? / 92

- 5.1 专家态度 / 92
- 5.2 摩尔定律与指数增长 / 97
- 5.3 难以预测的突破 / 105
- 5.4 结论 / 105

第三部分 超级人工智能 / 107

第6章 广义人工智能会走向超级人工智能吗? / 109

- 6.1 什么是超级智能? / 109
- 6.2 怎样变得更聪明? / 112
- 6.3 广义人工智能距离超级智能有多远? / 115
- 6.4 结论 / 118

第7章 超级智能会是我们的福音吗? / 119

- 7.1 超级智能的重要性 / 119
- 7.2 乐观场景:不朽与永生 / 120
- 7.3 悲观场景:灭绝,甚至更糟 / 128
- 7.4 乐观主义的理由 / 130
- 7.5 悲观主义的理由 / 137
- 7.6 对初始条件的依赖 / 146
- 7.7 结论 / 148

第四部分 友好人工智能 / 149

第8章 我们能否确保超级智能是安全的? / 151

- 8.1 悬崖勒马 / 151
- 8.2 半人马 / 155
- 8.3 约束友好人工智能的行为 / 157
- 8.4 控制友好人工智能的动机 / 160
- 8.5 研究超级智能危机的机构 / 166
- 8.6 结论 / 168

第9章 总结与建议 / 169

- 9.1 颠覆的好处 / 169
- 9.2 经济奇点 / 170
- 9.3 我们能不能造出超级智能? 若能,什么时候? / 172

人工智能革命:超级智能时代的人类命运

9.4 我们会喜欢它吗? / 173

9.5 永不止步 / 174

9.6 更聪明的哺乳动物 / 176

9.7 是时候发言了 / 177

9.8 人类如何在人工智能时代生存? / 180

致 谢 / 181