

LIFE 3.0

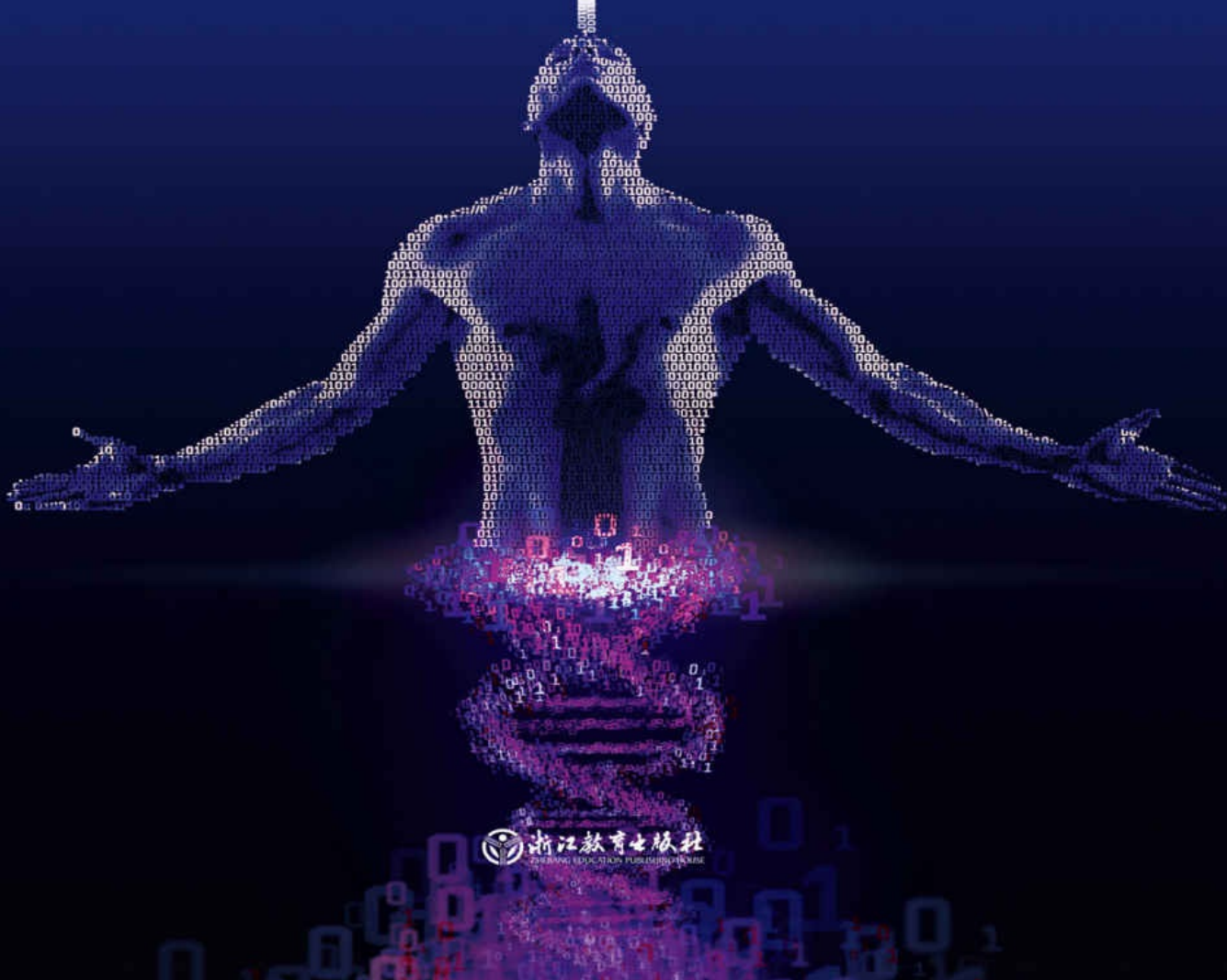
MAX TEGMARK

BEING HUMAN
IN THE AGE OF
ARTIFICIAL
INTELLIGENCE

[美] 迈克斯·泰格马克 著 汪捷舒 译

麻省理工学院
物理系终身教授
未来生命研究所创始人
迈克斯·泰格马克
重磅新作

人工智能时代
人类的进化与重生



版权信息

本书纸版由浙江教育出版社于2018年6月出版

作者授权湛庐文化（Cheers Publishing）作中国大陆（地区）电子版发行（限简体中文）

版权所有•侵权必究

书名：生命3.0

著者：[美] 迈克斯·泰格马克

电子书定价：89.99元

Life 3.0: being human in the age of artificial intelligence /
by Max

Tegmark.

Copyright © 2017 by Max Tegmark. All right reserved

还在督促自己每天进步一点吗？

还在坚持每天阅读的习惯吗？

还在为找不到自己喜欢的书籍烦恼吗？

那～

你愿意与我成为书友吗？

国内外当下流行书籍

各图书销量排行榜书籍

大量工具书籍

使我们受益终生的书籍

.....

海量电子版、纸质版书籍及音频课程

还有贴心的“学习管家”服务哦！

微信：shuyou055



致

未来生命研究所团队
是你们让这一切成为可能

To the FLI team,
who made everything possible

赞誉

人类应该以更好的姿态拥抱人工智能技术，通过让技术适应人类，让个体变得更加强大，让地球文明变得更加美好。《生命3.0》这本书给我们提供了一个路线图：如果我们小心谨慎地改进技术，深谋远虑地避免陷阱，那生命就有可能在地球上，甚至地球以外繁荣昌盛长达数十亿年的时间，远超人类祖先最不羁的梦想。期待这一天的到来！

王小川
搜狗公司CEO

如果人工智能有几本煌煌巨著，《生命3.0》就是其中的《天演论》。生命3.0挣脱了自然“进化”的束缚，进入自我“设计”的阶段，“她”成为自己命运的主人。而这本书正是我们作为生命2.0与生命3.0以及未来的对话，书中谓之“我们这个时代最重要的对话”并不为过。《生命3.0》这本书横跨社会与宇宙，从我们能活到的未来穿行10亿年，又从可见的智能潜入不可见的意识，其所呈现的世界观之宏大，令人荡气回肠。

吴甘沙
驭势科技（北京）有限公司联合创始人兼CEO

将人工智能与人类未来进化放在一起，越来越成为理解“新物种”的独特视角。作为警觉的乐观主义者，泰格马克的《生命3.0》与其说是提供答案，不如说是促进对话和思考。在这一波强大而迅猛的人工智能浪潮面前，每个人都无法置身事外——如果你感到焦虑，或者在思考人工智能时厘不清头绪、抓不住要点，这本书会让你跟上时代的步伐，找回思考的自信，同时抓住人与机器共生演化的焦点。

段永朝
财讯传媒集团首席战略官

苇草智酷创始合伙人

如果说《时间简史》讲透了宇宙物理，《未来简史》讲透了文明进化，《生命3.0》的作者则是从物理学家的角度对宇宙进化进行畅想。人类正推动着智能进化的史诗进程：以人工智能算力6年提升30万倍的超级摩尔定律发展，数字生命必然实现智能飞越，人机融合将孕育一个万物智能的新宇宙。

杨静

新智元创始人

生于这个人工智能极大发展时代的人们，不自觉便会陷入一种焦虑——对智能机器进步的极大恐慌。与其说是焦虑，不如说是“短视”，因为从长远来看，让“生命”最终升级到可以自我设计的“生命3.0”的最后一块拼图，很可能就是人工智能。理性地认知人工智能，正确地做出选择，才能让生命走得更远。

罗振宇

“得到”App创始人

无论你是科学家、企业家还是将军，所有人都应该扪心自问，现在可以做什么，才能提升未来人工智能趋利避害的可能性。这是这个时代最重要的对话，而迈克斯·泰格马克发人深省的著作《生命3.0》能帮助我们参与到这场对话中来。

史蒂芬·霍金

物理学家，宇宙学家

在探索地球上以及地球以外的生命、智能和意识的宏伟未来的旅程中，人类该如何应对随之而来的挑战与选择，《生命3.0》这本书提供了一份精彩的指南。

埃隆·马斯克

美国太空探索技术公司（SpaceX）创始人兼CEO
特斯拉汽车公司创始人兼CEO

人工智能可能是21世纪最重要的一股变革力量。迈克斯·泰格马克新作《生命3.0》一书从政治与哲学的角度预测了人工智能革命的前景与风险，并清晰地阐明了一些基本概念和重要争议，澄清了一些常见的误解。比如，科幻作品使得许多人担心邪恶的机器人，而他则适当地强调，真正的问题乃是开发能力超强的人工智能会带来一些无法预见的后果。人工智能并不是只有变得邪恶或者装在机器人身上才会肆虐人间。泰格马克的写作风格通俗易懂，十分吸引人，适合大众阅读。

尤瓦尔·赫拉利
世界知名历史学家
畅销书《人类简史》《未来简史》作者

迈克斯·泰格马克著作《生命3.0》是这个时代最重要的对话的深度指南。本书描述了，当我们逐渐将生物学意义上的“思维”与我们自己创造出来的更伟大的智能相融合时，如何创造出一个友善、仁爱的未来文明。

雷·库兹韦尔
发明家兼未来学家
畅销书《奇点临近》《人工智能的未来》作者

作为一个物种，我们希望创造出什么样的未来？关于这个问题，迈克斯·泰格马克希望促成一场更加广泛的对话。虽然他谈及的话题——人工智能、宇宙学、价值甚至意识经验的本质，都十分富有挑战性，但他采用了一种平易近人的方式，鼓励读者形成自己的观点。

尼克·波斯特洛姆
牛津大学人类未来研究所创始人
畅销书《超级智能》作者

迈克斯·泰格马克是一位出色的物理学家，同时也是未来生命研究所的领袖。这样的身份给了他一种独特的有利视角，在《生命3.0》这本书中，他用一种通俗易懂、辞简理博的方式为读者讲述了这个时代最重要问题的“独家内幕”。

扬·塔里安
Skype公司联合创始人

《生命3.0》是一本读起来令人愉快的书，它会改变我们对人工智能、智能与人类未来的看法。

巴特·塞尔曼
康奈尔大学计算机科学系教授

人工智能释放出了前所未有的力量，这意味着接下来的10年对人类来说，可能是最好的时代，也可能是最坏的时代。迈克斯·泰格马克在《生命3.0》这本书中对人工智能的探索是我所读过的最发人深省的，但同时又十分简单有趣。如果你还没有见过泰格马克有趣的灵魂，这本书一定会让你大饱眼福。

埃里克·布莱恩约弗森
麻省理工斯隆管理学院教授
畅销书《第二次机器革命》作者

我被《生命3.0》这本书迷住了。我们很快就会面临人工智能带来的变革性后果，问题是，那会是乌托邦，还是一场大灾难？这个问题尚无定论。但这本由一位杰出科学家写就的富于启迪、生动易读的书，能帮我们估算出这一可能性。

马丁·里斯
宇宙学先驱
英国皇家学会天文学家

推荐序一

如何正确地关心人类命运

万维钢

科学作家，“得到”App《精英日课》专栏作者

每个人都关心自己的命运，也有很多人关心国家的命运，但除此之外，你还应该关心人类的命运。人类命运是个大尺度的问题，虽然对我的生活没有直接影响，但我们总有一点儿好奇心，想知道未来究竟会怎样。

你肯定对未来有过各种推测和想象。虽然我没见过未来，但我敢打赌，你自己的推测和想象有很多不合理之处。

一般人在预测短期的未来时往往过分乐观。20世纪六七十年代，很多人相信21世纪将是一个宇航的时代，人类很快就能殖民火星。结果，我们今天所谓的高科技只不过是——智能手机。想象一件事总比做成一件事容易，我们容易高估技术进步的速度。

如果要预测长期的未来，人的想象力往往又不太够用。一两百年前的人想象21世纪的生活时，根本就想不到会有智能手机和计算机这些东西，他们能想象的大概是一个蒸汽朋克的世界，天空中飘着巨大的飞艇。

所以，如果你要严肃地关心人类命运，就需要科学推测。迈克斯·泰格马克新作《生命3.0》就是这样一本书。

泰格马克可能是当今物理领域活跃着的最有意思的一位物理学家。他在量子力学和宇宙学这些最正宗的物理领域里获得过很了不起的成就，而且还涉猎广泛，跨界搞过人工智能方面的理论研究；他也很有思

想，提出了“数学宇宙”这个哲学的世界观；他还热衷于社会活动，跟物理学和人工智能界的很多大佬经常互动。此外，他还很会写书。

李鸿章年轻时写过这样一句诗“一万年来谁著史”，泰格马克这本《生命3.0》的气魄比这个还大，研究的是人类的终极命运。

这个问题本来是交给哲学家和科幻作家去解决的，物理学家能干什么呢？答案是物理学家的推导更精确，而且更富有想象力。

比如一想到未来，我们就关心地球环境会不会被破坏，能源够不够用，哲学家可能深表忧虑。但物理学家知道，我们人类目前的能源汲取水平远远没达到极限，跟将来可以使用的聚变核能和太阳能相比，连九牛一毛都算不上。但只要考虑物理定律的限制，宇宙就是你的大舞台，能源根本不是问题。

再比如人工智能。科幻作家在畅想未来时，经常会犯两种跟人工智能有关的错误。一种错误是他没有充分考虑人工智能，还是认为是人类在主导一切；另一种错误是他误判了人工智能。在有些作品里，机器人动不动就活了，具有了人类的意识，但智力水平居然并不明显高于人类，有时候还挺笨的！而科学家会告诉你，让人工智能获得意识非常非常困难，但是，让人工智能的智能超过人则相当容易。

《生命3.0》这本书给我的感悟是，决定人类终极命运的只有这一个问题最重要——人工智能到底能不能拥有人的意识。

在我们的流行文化中，经常会谈论人工智能，但很少涉及“意识”。可能大多数人都没有意识到，人有一个“意识”的问题。

简单地说，意识是我们对世界的主观体验。我们的喜怒哀乐，一切感情都是因为我们有意识。一辆自动驾驶汽车也许可以出色地完成运输任务，但当遇到红灯时它不会暴躁，有危险时它不会害怕，撞了车时它不会疼，没油时它不会饿，完成任务时它也不会高兴，它只是机械地做事而已。

实际上，现在生物学家认为，人在做事的时候，本质上也是机械的。我们的各种感情只是附带产生的多余的情绪。就算没有任何主观感受，你还是一样能做好各种事情。

但是主观体验赋予了我们生活的意义。你工作之余，偶尔抬头看看星空，感慨一下宇宙多么美好，那是因为你意识。如果人没有意识，那就跟一堆沙子没有本质区别，人生就没有意义，整个宇宙的存在就没有价值。

泰格马克把生命分成三个阶段，人类只能算第二阶段，叫“生命2.0”。我们能学习新知识，但不能随便升级自己的身体，因而受到了很大的限制。而人工智能则是“生命3.0”，它们将可以随意升级软件和硬件，它们终将超过我们。

那将来的人工智能是否具有意识？

科幻作家会说，既然世界上并没有“灵魂”这种东西，人类纯粹是由原子组成的，那么人工智能当然可以有意识。按理说是这样的，但科学家会给人工智能的意识做出一些限制。比如这本书中介绍了一种叫作“信息整合理论”的意识理论，这个理论要求有意识的物体必须是信息高速整合的，而物理定律要求信息的传播速度有限，人工智能大脑的大小就必须限制在一个不太大的范围之内。人工智能的聪明程度将是有限的。

但再有限也比人类厉害得多，那么，将来的结局就是，人工智能将会淘汰人类。如果双方和平交接，人工智能将作为人类文明的代表去征服宇宙的各个角落，人类将是人工智能的宠物。

如果人工智能一直都没有意识，事情就更麻烦了。据泰格马克推演，人工智能就算没有意识也可能会有自己的目的，它们可能会不自觉地发展壮大，并且最终抛弃人类。那将是人类文明最坏的结局，我们可能会被没有意识的僵尸人工智能取代，留下一个空洞的、毫无意义的字

宙。

鉴于这些结局好像都不怎么理想，我们迫切地需要知道意识到底是怎么回事儿，将来的人工智能到底会怎样。

这并不是泰格马克自己在杞人忧天。我看美国上上下下，从学者到企业家和老百姓，现在对意识和人工智能的思考非常深入，主流媒体上也经常讨论，新研究、新思想层出不穷。

虽然中国人对人工智能的各种应用非常关心，也很了解，但对人的意识、人工智能的原理这些问题关心不够。我们有太多面向过去的思想家，他们总想用过去指导未来，但是未来世界的逻辑很可能跟过去很不一样。

到底什么是意识？人到底是一种什么样的机器？就算你觉得未来太遥远，只要你关心人，这些问题就会让你寝食难安。这就是现在世界上最聪明的大脑都在想的问题。而《生命3.0》这本书告诉你的大约就是目前已知最好的答案了。

推荐序二

重新定义生命

余 晨

易宝支付联合创始人，《看见未来》作者

法国思想家布莱兹·帕斯卡曾说：人只不过是一根芦苇，是自然界里最脆弱的东西；但他是一根会思想的芦苇。宇宙可以轻松地将人毁灭，一团雾气、一滴水都足以致人于死地。然而，纵使宇宙毁灭了人类，人类却仍然要比世界万物高贵得多。因为人知道自己终有一死，也了解宇宙的秉性和优势，而宇宙对此却一无所知。因而，人类全部的尊严就在于思想。

帕斯卡设计制造了历史上第一台机械计算器，为了纪念这项发明，一种计算机编程语言Pascal便以他命名。在计算机技术高速发展的今天，人类是否会被人工智能取代，我们是否还能捍卫思想的尊严，或许是这个时代最需要思考的问题。

迈克斯·泰格马克新著《生命3.0》便是这样一次思考。作者把广义的生命看作是一种能够自我复制的信息处理系统，物理结构是其硬件，行为和“算法”是其软件。1.0版的生命是以细菌为代表的简单生物阶段，其硬件和软件都是靠进化获得，行为则是完全固化的；2.0版的生命是以人类为代表的文化阶段，进化决定了我们的硬件，但我们可以自行设计软件，通过学习来获得知识、改变行为和优化“算法”；而3.0版的生命是以人工智能为代表的科技阶段，生命不仅可以自行设计软件，还可以自行设计硬件，由碳基变为硅基，最终摆脱进化的枷锁，让会思想的芦苇变得不再脆弱。但当生命变得面目全非时，我们还算是人类吗？人工智能的降临或许是宇宙创生以来最重要的事件，也有可能

是人类的最后一项发明，或许会实现科技的乌托邦，也有可能带来人类的毁灭。无论如何，这是一场关于我们这个时代最重要的对话。

人工智能是会像马克·扎克伯格宣扬的那样造福人类生活，还是会像埃隆·马斯克警告的那样威胁人类生存？泰格马克全方位、系统深入地探讨了人工智能可能给人类文明带来的一系列深远影响，包括就业、经济、法律、伦理、政治、军事。也逐一列举分析了当优于人类的超级智能出现后可能带来的各种未来图景：超级智能是会带来自由主义或平等主义的乌托邦，成为能够控制整个人类社会的善意独裁者，或是增进人类幸福感的守护神，还是完全被人类所控制驾驭的公仆？超级智能会成为毁灭人类的征服者，还是会成为让我们引以为豪的新物种后裔？或许超级智能还未降临，人类就已经毁灭于核战争或环境危机等其他灾难。如果你关心这个星球未来的命运、人类的困境和希望、我们子孙世世代代的福祉，那么就应该认真读读这本书。

真正让这部著作从汗牛充栋的人工智能书籍中脱颖而出的，是泰格马克作为一位物理学家，从宇宙学的宏大视野和物理学第一性原理中，展示了未来生命和智能令人眼花缭乱的无限可能性。在物理学家看来，生命现象不过是粒子特殊的排列组合，生命通过从环境中汲取负熵来抵抗自然的熵增，通过让环境变得更混乱而维持自身的秩序和复杂度。物理定律的极限，决定了我们如何可以最大程度地利用宇宙的禀赋，实现生命的最大潜力。人类今天所能驾驭的物质和能量，只不过是宇宙所能够赋予我们全部资源的极其微小的零头。如果我们能深谋远虑地改进技术并计划周全地避免陷阱，便可以通过重组物质、能量和信息将生命最大化。未来的超级智能可以收割黑洞辐射和夸克引擎的巨大能量，逼近计算力的理论上限，以光速进行宇宙殖民，将现有的生物圈增长几十个数量级。生命在未来数十亿年时间内的美丽绽放，会远远超越我们祖先最不羁的梦想。

宇宙虽然可以轻易地毁灭人类，但事实是，宇宙通过我们人类才真

正活了过来，并逐渐获得了自我意识，并非宇宙将意义赋予了有意识的实体，而是有意识的实体将意义赋予了宇宙。让有智能和意识的生命这束微弱的光，点亮这个冷漠荒芜的宇宙中无尽的黑暗，或许是我们的最高职责。

我们只看到了人工智能的冰山一角

我很高兴也很荣幸《生命3.0》这本书能在中国出版。在中国，人工智能的发展突飞猛进，目前已拥有许多世界顶尖的人工智能研究者和人工智能公司，这令我印象十分深刻。

从短期来看，这意味着中国已经遇到了我在第3章讨论的许多近期机遇和挑战。中国究竟会如何处理这些问题呢？我对这个问题兴致盎然。譬如说，我很高兴中国政府最近决定支持一项针对致命性自动化武器的禁令，这使得中国不仅仅在人工智能研发方面变得首屈一指，也成了促进人工智能有益运动的领军人。

从长期来看，我认为中国会扮演越来越重要的角色。我的个人经验告诉我，中国一定会出现极具天赋的研究者，因为在我带过的最优秀的研究生中，有许多都来自中国。目前在西方国家，很多时候，科学研究资金正遭遇停滞不前甚至日渐萎缩的困境，但中国却在科研方面（包括人工智能）投入了大量资金，即将赶超西方，成为人工智能领域的世界领跑者。

这会产生一个重要的结果：就是我在本书中主要谈论的通用人工智能（Artificial General Intelligence，简称AGI）出现的可能性，这种人工智能在任何任务上都可以与人类智能相提并论。大多数人工智能研究者认为，通用人工智能会在短短几十年内发生。如果这是真的，我认为这将发生在中国。那么，有一件事情就会变得非常重要，那就是：你，我亲爱的中国读者，请开始认真思考，你想看到一个什么样的未来社会。正如我在这本书中所说，极好的可能性与巨大的风险同时存在。因此，想要实现好的结局，就必须认真斟酌后果，并思考如何才能

平稳地实现目标，而不带来灾难。在这方面，中国拥有独特的机会，因为中国的长期计划能力远远超过大多数西方国家。此外，几千年的历史也能启发中国人看到同样遥远的未来，并能严肃地思考本书后半部分所聚焦的那些波澜壮阔的可能性。

如今，距离《生命3.0》英文版的出版已有将近一年的时间。在这段时间里，发生着一件重要的事情，那就是：人们开始更加严肃认真地对待这本书的一个重要观点——人工智能可能会在几十年内变成现实，我们必须认真思考如何保证人工智能的安全性和有益性。其原因是，人工智能在各个方面都取得了史无前例的进展。我很欣慰地看到，译者汪婕舒能够在本书中加入我发给她的关于这些进展的最新内容，比如说AlphaZero的故事，它最近成功碾压了那些花了几十年时间来手工开发围棋软件和象棋软件的世界顶尖人工智能研究者。

过去，我们一直认为，智能是一种神秘的东西，只能存在于生物（特别是人类）身上。但是，从我作为一位物理学家的角度出发，智能只是运动的基本粒子处理信息的特殊过程，并没有一条物理定律说，“建造一台在各方面都比人类聪明的机器是不可能的”。这意味着，关于智能，我们只看到了冰山一角，我们还有巨大的潜力来开启潜伏在大自然中的全部智能，并用它来帮助人类实现繁荣昌盛。我认为，人工智能有可能成为人类有史以来最美好的事情，也可能成为最糟糕的事情。虽然我在《生命3.0》这本书中描述了许多你或爱或恨的未来场景，但最重要的问题并不是计较哪一种场景最有可能发生，而是思考我们想让哪一种场景成为现实，并且，需要哪些具体步骤才能最大限度地保证人类拥有一个欣欣向荣而非每况愈下的未来。我希望这本书能帮助你——我亲爱的中国读者厘清这个问题！



扫码获取“湛庐阅读”App，
搜索“生命3.0”，
观看作者迈克斯·泰格马克TED演讲视频。

目 录

赞誉

[推荐序一 如何正确地关心人类命运](#)

[推荐序二 重新定义生命](#)

[中文版序 我们只看到了人工智能的冰山一角](#)

[引言 欧米茄传奇](#)

[01 欢迎参与我们这个时代最重要的对话](#)

[02 物质孕育智能](#)

[03 不远的未来：科技大突破、故障、法律、武器和就业](#)

[04 智能爆炸？](#)

[05 劫后余波，未知的世界：接下来的1万年](#)

[06 挑战宇宙禀赋：接下来的10亿年以及以后](#)

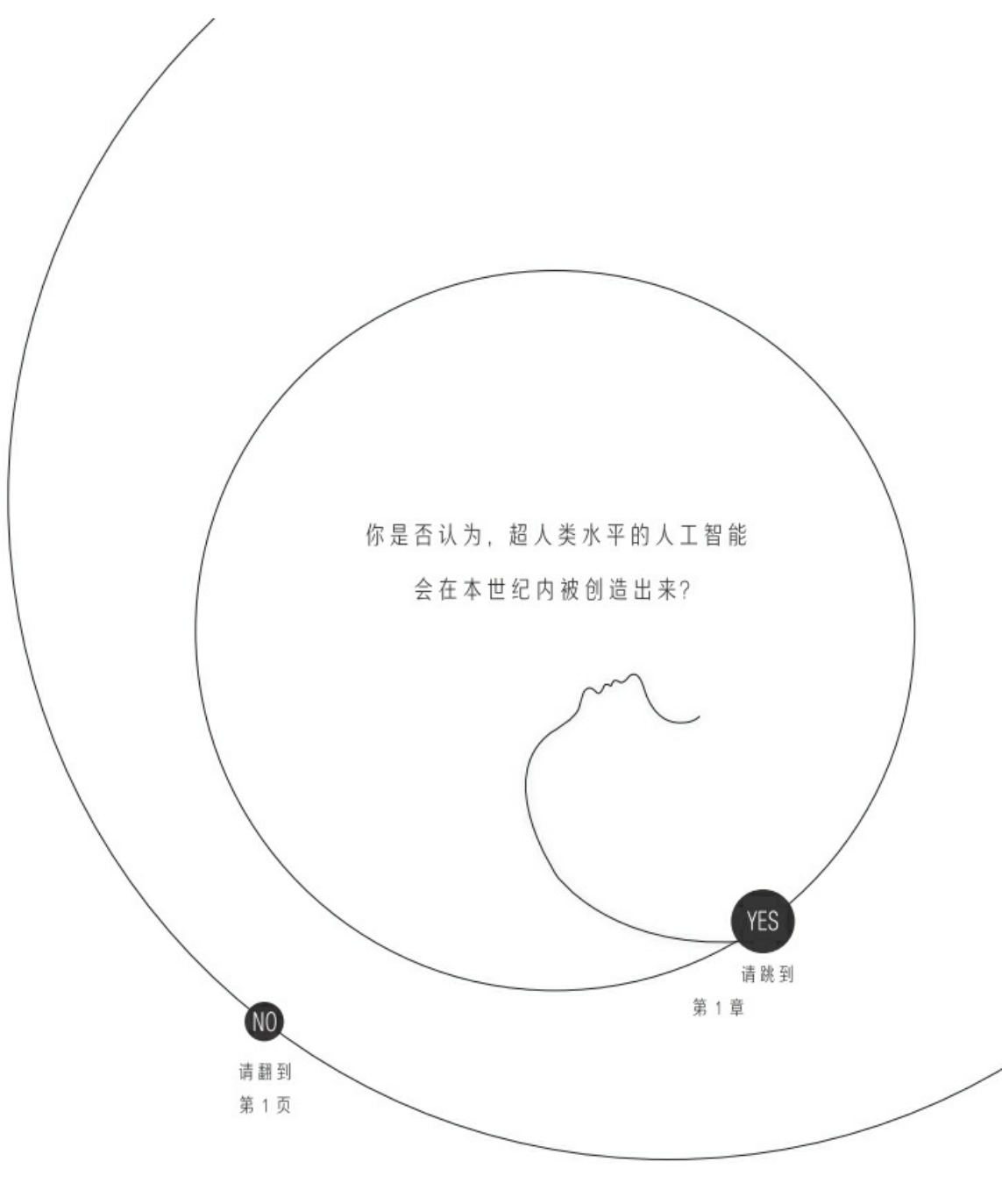
[07 目标](#)

[08 意识](#)

[后记 未来生命研究所团队风云传](#)

[注释](#)

[致谢](#)



你是否认为，超人类水平的人工智能
会在本世纪内被创造出来？

YES

请跳到
第 1 章

NO

请翻到
第 1 页

引言

欧米茄传奇

欧米茄团队是这家公司的灵魂。虽然该公司其他部门通过开发各种狭义人工智能（narrow AI）的商业应用赚得盆满钵满，让公司得以按部就班地运转下去，但欧米茄团队却一直秉承并追寻着公司CEO的梦想：建造通用人工智能（Artificial General Intelligence，简称AGI）。因此，其他部门的员工都亲切地称他们为“欧米茄”，并把他们视为一群不切实际的梦想家，因为他们似乎总是与自己的目标差着几十年的距离。但是，人们喜欢纵容这些人，因为欧米茄团队的前沿工作为公司带来了声望，他们为此感到高兴。同时，欧米茄团队偶尔会改进一些算法供其他部门使用，这让他们十分感激。

然而，其他部门的同事不知道的是，欧米茄团队之所以精心打造自己的形象，是为了隐藏一个秘密：他们马上就要启动人类历史上最勇敢无畏的计划了。那位极富个人魅力的CEO亲自挑选了这些人，不只是为了培养杰出的研究人员，还为了实现他帮助全人类的雄心壮志和坚决承诺。他告诫欧米茄团队，这个计划极端危险，如果被不怀好意的人发现了，他们就会不择手段地甚至实施绑架来制止这个计划，或者盗走他们的代码。但是，这些人已经全身心投入其中了。他们的理由和当年众多世界顶尖物理学家加入“曼哈顿计划”开发核武器的原因差不多：因为他们都坚信，如果自己不率先做出来，就会有其他不那么高尚的人捷足先登。

欧米茄团队建造的人工智能昵称为“普罗米修斯”（Prometheus），它一天比一天强大。诚然，它的认知能力在社交技能等许多方面还远远落后于人类，但欧米茄团队竭尽全力让它在一个任务上表现超凡，这个任务就是编写人工智能系统。他们之所以选择这

个计划，是因为他们相信英国数学家欧文·古德在1965年提出的“智能爆炸”[_ \(1\) _](#)理论。古德说道：

让我们给“超级智能机器”（ultraintelligent machine）下一个定义，那就是：一台能超越任何人（无论这个人多么聪明）的所有智力活动的机器。由于设计机器也属于这些智力活动中的一种，因此，一台超级智能机器就能设计出更好的机器；那么，毫无疑问会出现一种“智能爆炸”，到那时，人类的智能会被远远甩在后面。于是，第一台超级智能机器就会成为人类最后一个发明，只要它足够驯良，并告诉人类如何控制它就行。

欧米茄团队认为，只要他们能让这个不断迭代的“自我改善”过程持续下去，那么最终，这台机器就会变得非常聪明，足以自学其他有用的人类技能。

第一个100万美元

一个星期五的早晨9点钟整，欧米茄团队决定启动这个计划。在一间闲人免进的巨大空调房间内，层层叠叠的架子排成长列。普罗米修斯就在这一排排为它“量身定制”的计算机集群中嗡嗡鸣响。为了安全起见，它没有接入互联网。不过，它在本地存储着一份包含互联网大部分内容的副本作为训练数据，以便从中学习[_ \(2\) _](#)，这些数据来自各大知识汇集网站及社交平台数据库。欧米茄团队挑选这个时间点是为了可以不受打扰地工作：亲朋好友都以为他们参加公司的周末拓展活动去了。办公室的小厨房里塞满了微波食品和提神饮料。一切准备就绪。

启动伊始，普罗米修斯在编写人工智能系统上的表现还是比人类略逊一筹，但很快，这个缺点就被它极快的速度所弥补了。当欧米茄团队

正猛灌红牛时，普罗米修斯也在以破竹之势解决着问题。如果换算成人类需要的时间，那得几千年之久。到早上10点钟，普罗米修斯已经完成了对自身的第一次迭代。这个2.0版本虽然比过去稍微好一点，但还是比不上人类。然而，到了下午2点钟，当普罗米修斯迭代到5.0版本时，欧米茄团队惊呆了：它已经大大超越了他们的预期，而且它进步的速度似有加快的迹象。夜幕降临时，他们决定用普罗米修斯10.0版本来启动计划的第二阶段：赚钱。

欧米茄团队的第一个目标是亚马逊的MTurk⁽³⁾。这是一个众包网络市场，于2005年上线后，发展十分迅速，很快就聚集了成千上万来自全球各地的人。他们夜以继日地奏出了一支支“HIT”奏鸣曲——HIT指的是“人类智力任务”（Human Intelligence Tasks），范围十分广泛，从音频录制到图像分类和网页描述撰写，应有尽有，但它们都有一个共同点：只要你完成得足够好，没人在乎你是不是人工智能。对于其中大约一半的任务，普罗米修斯10.0版本完成得都还算可以。欧米茄团队让普罗米修斯用狭义人工智能设计出了一个简洁的软件模块，专门用于处理这些任务，但除此之外，这个模块什么也干不了。接着，他们把这个模块上传到了亚马逊的网络服务器上，这是一个可以运行虚拟机的云计算平台。在这个云平台上，他们租了多少虚拟机，就可以运行多少虚拟机。欧米茄团队在亚马逊云计算服务上每花费一美元，都能从MTurk上赚回超过两美元的价值。亚马逊一点儿也没发现，自己公司内部竟然存在着这样惊人的套利机会！

为了掩盖踪迹，在前几个月中，欧米茄团队小心谨慎地用假名创建了几千个MTurk账户。现在，普罗米修斯建造的软件模块正冒名顶替着这些账户的身份。MTurk的客户通常会在事后8小时左右付款。一旦收款，欧米茄团队又将这些钱用来购买更多的云空间，供给普罗米修斯使用。在这个过程中，普罗米修斯不断升级，它的最新版本写出来的任务模块也变得越来越厉害。由于欧米茄团队的钱每8小时就能翻一番，因此在MTurk上的任务很快便达到了峰值。同时，他们发现，如果不想引

起过多的注意，最好把日收入控制在100万美元以下。不过，普罗米修斯已经为他们的下一步计划提供了足够多的钱，欧米茄团队已经无须向公司财务总监申请经费了。

危险游戏

启动普罗米修斯之后，除了在人工智能研究上取得突破之外，欧米茄团队最近还热衷于用它来赚钱：赚得越快越好。从本质上来说，整个数字经济的红利都是唾手可得的，但从哪里开始比较好呢？是开发计算机游戏，还是做音乐、电影或者软件？是写书、写文章，还是炒股，或者捣鼓和贩卖新发明？简单来说，这个问题归根结底是如何才能实现投资回报率的最大化，但一般的投资策略在欧米茄团队面前实属小巫见大巫：在通常情况下，如果年均回报率能达到9%，投资者就会很满意了；然而，欧米茄团队在MTurk上的投资达到了每小时9%的回报率，平均每天能赚到8倍多的钱。那么，他们的下一个目标是什么呢？

欧米茄团队的第一个想法是去股票市场大捞一笔，毕竟，许多对冲基金都在这上面砸了重金，而欧米茄团队中的几乎每个人都曾在人生中的某一时刻，拒绝过为对冲基金开发人工智能系统的高薪工作。你可能还记得，这也正是电影《超验骇客》（Transcendence）中的人工智能赚得第一桶金的方法。但是，前些年的一场股市崩盘促使政府对金融衍生品出台了一些规范措施，限制了他们的选择范围。很快，欧米茄团队就意识到，即便他们能够获得远高于其他投资者的回报，但这点利润比起销售自家的产品来说还是相差甚远。毕竟，当全世界第一个超级智能都在为你工作时，你显然最好投资自家的产品，而不是寄希望于别人家的。当然，例外也是可能存在的，比如，你可以用普罗米修斯超人的黑客技能来获取内幕消息，然后购买那些即将上涨的股票的看涨期权。但欧米茄团队认为，这可能会引来不必要的注意，因此不值得这么做。

于是，欧米茄团队将重点转向了那些可以研发和销售的产品，其

中，电子游戏看起来是个很不错的选择。普罗米修斯很快就具备了极为高超的技能，能够设计出引人入胜的游戏，还能轻易地应对程序、平面设计、光线追踪等成品所必需的任务。此外，它还分析了网络上关于人们偏好的所有数据，知晓了哪一类游戏是玩家的最爱，据此发展出了一种根据销售收入来优化游戏的超能力。尽管不愿意承认，但欧米茄团队的许多成员都曾夜以继日地泡在游戏《上古卷轴5：天际》（The Elder Scrolls V: Skyrim）中。2011年，这款游戏在刚发布的第一星期，总销售额就超过了4亿美元。因此，欧米茄团队相信，在100万美元的云计算资源的支持下，普罗米修斯在24小时内一定能开发出一款像《上古卷轴》一样令人上瘾的游戏。他们可以在线销售这款游戏，并让普罗米修斯在博客圈里假扮玩家来大聊特聊，引爆热度。如果能在第一个星期入账2.5亿美元，他们就能在8天内将投资翻8倍，每小时的回报率高达3%，虽然比他们在MTurk上的表现略显逊色，但更具可持续性：假如普罗米修斯能每天开发出一款游戏，不久之后，他们就能赚到100亿美元，而不用担心游戏市场饱和。

但是，欧米茄团队中的一位网络安全专家坦率地表达了对这个游戏计划的不安。她指出，这个计划可能会带来一个可怕的风险：普罗米修斯可能会“逃脱”，并“抢夺自己命运的控制权”。过去，由于不清楚普罗米修斯的目标在自我提升的过程中会如何变化，因此为了安全起见，欧米茄团队决定不遗余力地将普罗米修斯“关起来”，囿于“盒”中，让它无法“逃”到互联网上。对运行在服务器机房中的普罗米修斯主机，他们采取了物理隔绝的手段：那里根本没有网络连接，普罗米修斯输出的所有数据都以信息和文件的形式传输给一台被欧米茄团队严格控制的计算机。

将普罗米修斯开发的复杂程序放到联网的计算机上运行，是一件风险很大的事情：欧米茄团队根本不清楚它会做什么，也完全不知道它会不会做某些事情，比如在互联网上像病毒一样扩散自己。在测试普罗米修斯为MTurk上的任务写的软件时，为了谨慎地提防这种情况的发生，

欧米茄团队只在虚拟机中运行这些软件。虚拟机是一个模拟计算机的程序，比如，苹果电脑的许多用户会购买虚拟机软件，通过“欺骗”苹果电脑这是一台Windows电脑，来让它们运行Windows程序。欧米茄团队建造了自己的虚拟机，其昵称叫“潘多拉魔盒”。它模拟的是一台超级简单的机器，去除了计算机通常拥有的所有附件，没有键盘，没有显示器，没有音箱，也没有网络连接，什么都没有。举个例子，对于MTurk上的音频誊录任务，欧米茄团队对虚拟机进行了设置，只允许一个音频文件输入潘多拉魔盒，也只允许一个文本文件输出，即誊录好的文字。这些规则之于魔盒内部的软件，就好像物理定律之于身处宇宙中的人类一样：软件无法冲出魔盒的藩篱，就好比人类再聪明，也无法实现超越光速的星际旅行。除了上面所说的输入和输出单个文件之外，潘多拉魔盒内部的软件实际上就像被困在了一个拥有自己的计算规则的平行宇宙。欧米茄团队患有严重的“逃脱”恐惧症，因此，他们也在时间维度上加了枷锁，限制了不受信任代码的生存期限。比如，潘多拉魔盒中的誊录软件每完成一个音频文件的誊录，魔盒中的所有内存都会被自动抹去，程序会自动从零开始重新安装。这样，当它开始下一个誊录任务时，就不知道过去发生了什么，也就不能从过去中学习。

当欧米茄团队在MTurk项目中使用亚马逊的云服务时，他们能将普罗米修斯编写的所有任务模块都装进云端的潘多拉魔盒，因为在MTurk上输入数据和输出数据都非常简单。但是，对严重依赖图像的计算机游戏来说，这是行不通的。他们无法将游戏装进潘多拉魔盒，因为它们需要彻底访问玩家计算机上的所有硬件。此外，他们不想冒险，因为一些懂计算机的用户可能会分析游戏代码，从而发现潘多拉魔盒，并调查其中的秘密。“逃脱”风险不仅置游戏市场于危险当中，还可能会牺牲其他软件市场，而后者是规模巨大并且有利可图的，遍地都是千亿美元的机会。

第一个10亿美元

欧米茄团队缩小了搜寻的范围，只关注那些价值极高、数字化、易于理解的产品，因为数字化产品的生产周期短，而那些易于理解的产品，比如文字或电影等不会带来“逃脱”的风险。最后，他们决定成立一家媒体公司，以动画片为起点。公司的网站、市场计划和新闻稿在普罗米修斯变得超级智能之前就已经准备好了，而唯一欠缺的，就是内容。

为了设计出人工智能系统来编写软件，好完成那些令人抓狂的MTurk上的任务，普罗米修斯经过了慎重的优化。到了星期日的早上，尽管它的能力已经变得异常强大，能够持续不断地从MTurk中敛财，但它的智力依然比不上人，在某些事情上它并不擅长，譬如制作电影，不擅长这些事并不是因为某些深层的原因，而更像是詹姆斯·卡梅隆并不是出生时就擅长拍电影一样：这是一个需要时间来学习和打磨的技能。与人类儿童一样，普罗米修斯能从手边的数据中学习任何它想学的东西。卡梅隆为了学习读写，花了几年时间，而普罗米修斯在星期五一天就完成了这项任务，那一天，它还腾出时间阅读了Wikipedia的所有词条，外加几百万本书。不过，制作电影更加不易。写出一个让人们觉得有趣的剧本，和写书一样困难，需要细致入微地理解人类社会，并了解人们认为的具有娱乐性的事情；将剧本最终变成视频不仅需要对虚拟演员以及它们身处的复杂场景进行大量的光线追踪，还需要制作大量的虚拟声音和扣人心弦的音乐音轨，诸如此类。到了星期日早晨，普罗米修斯不仅能够在一分钟内看完一部时长两小时的电影，还能够看完由这部电影改编的所有书籍、网上的所有评论和评分。欧米茄团队注意到，当普罗米修斯一口气看了几百部电影之后，开始能很精准地预测一部电影会得到什么样的评价，以及它会如何吸引不同的观众。实际上，它还学会了撰写影评。欧米茄团队觉得普罗米修斯的影评写得颇有见解，对剧情、演技、技术细节（例如光线和拍摄角度）等方面都能提出独到的看法。欧米茄团队认为，这意味着，当普罗米修斯能自己制作电影时，就会知道什么样的电影能获得成功。

为避免人们到时追问虚拟演员的真实身份，欧米茄团队要求普罗米修斯先制作一些动画片。到了星期日晚上，为了犒劳这个疯狂的周末，他们决定把灯光调暗，边吃爆米花，边喝啤酒，然后一起观看普罗米修斯的电影处女作。这是一部和迪士尼的《冰雪奇缘》差不多的奇幻喜剧动画片，其中的光线追踪是由普罗米修斯在亚马逊云端编写的“盒中代码”来完成的，这项任务几乎把当天在MTurk上赚到的100万美元利润都用光了。电影一开始，他们就在感叹，这样一部电影竟然是由机器在无人指导的情况下创作出来的，这是一件多么令人惊讶又害怕的事情啊！不过不久之后，他们就顾不上想这些了，他们开始为电影中的插科打诨而开怀大笑，为激动人心的时刻而屏住呼吸。在感人的结局处，有些人甚至流下了眼泪。他们是如此地全神贯注，都忘了这部电影的创作者是谁。

欧米茄团队计划在下一个星期五上线他们的网站，因为要给普罗米修斯留出充足的时间来生产更多的内容，也要给他们自己腾出时间来做一些普罗米修斯做不到的事情，例如购买广告位，并为过去几个月中创立起来的空壳公司招聘员工。为了掩盖踪迹，他们的官方介绍告诉大家，这家媒体公司从独立的电影制作人那里购买内容，这些制作人通常是那些低收入地区的高科技创业公司。为了方便起见，欧米茄团队把这些虚假的供应商设立在偏远的地区，例如蒂鲁吉拉伯利

（Tiruchirapalli）和雅库茨克（Yakutsk）。这些地方，连最好奇的记者也懒得去拜访。他们招聘的员工全部隶属于市场和管理部门。这些员工会告诉所有询问的人，他们的制作团队位于另一个地方，此时不便接受访问。欧米茄团队也没公开与这家媒体公司的关系。为了与官方介绍相吻合，他们为公司想了一个口号——连接世界上的创作天才，并宣传说，他们的品牌是摧枯拉朽且与众不同的，因为他们用前沿技术将力量赋予那些具有创造力的人，特别是那些身处发展中国家的人。

星期五到来了。好奇的访客开始登录欧米茄团队的网站。他们看到的内容令人联想到奈飞和葫芦（Hulu）这类在线娱乐服务商，不过，又

有点有趣儿的不同。所有动画系列片都是全新的，从来没有人听说过它们。这些动画片相当迷人：大部分系列都由45分钟长的剧集组成，剧情超级吸引人，每集的结尾都会使你对下一集充满期待。并且，与竞争对手相比，它们的价格更便宜。每部片子的第一集都是免费的，剩下的每集你可以试看49秒，或者以折扣价格观看整个系列。一开始，总共只有三个系列，每个系列只有三集。但很快，每天都有新的剧集加入。同时，迎合不同口味的新系列也在源源不断地更新进来。

最初两个星期，普罗米修斯的电影制作技术突飞猛进，不仅在于影片质量方面，还在于它具备了更高明的算法来进行人物模拟和光线追踪，这极大地降低了制作每集片子所需的云计算成本。结果，欧米茄团队在第一个月就发布了几十部新片，针对的人群涵盖了幼儿和成人，同时扩张到了所有主要的语言市场，使得他们的网站远比所有竞争对手的更加国际化。一些影评人感到十分震惊，因为这些动画片不仅音轨是多语言的，连视频本身也是。例如，当某个角色在讲意大利语时，他的嘴部活动与意大利语是匹配的，而且也会相应地做出意大利人特有的手势。尽管普罗米修斯现在有能力完美地将电影中的虚拟人物制作得与真人无异，但欧米茄团队没让它这么做，以避免泄露底牌。相比之下，他们动画片中的人物形象都是半写实风格的，内容则主要集中在那些传统上多采用实景真人电视剧和电影的流派，以便与它们竞争。

欧米茄团队的网站令人上瘾，用户量迅速地增长起来。许多粉丝发现，这些动画片中的人物比好莱坞最大手笔的电影作品中的人物更聪明，情节更有趣，而且还欣喜地发现，其价格更低，更容易负担得起。在广告的积极推动下（欧米茄团队能负担得起广告费，因为他们的制作费几乎为零），他们获得了极佳的媒体覆盖率和口碑。在网站上线后的一个月内，全球收入就激增到每天100万美元。两个月后，他们击败了奈飞公司。三个月后，每日入账1亿美元，开始与时代华纳公司、迪士尼公司、康卡斯特公司和21世纪福克斯公司比肩，成为全世界最大的媒体帝国之一。

欧米茄团队的成功引起了轰动，招致了大量不必要的关注。一些人开始怀疑，为什么他们在财政上只投入了一小部分就拥有了强大的人工智能。欧米茄团队策划了一场相当成功的“虚假内容”营销活动。在光鲜亮丽的曼哈顿新办公室中，他们新招聘了一批发言人，来传递他们的理念。欧米茄团队还在全世界雇用了许多人来作幌子，请他们来创作新系列，其中甚至包括一些真正的剧作家，但没有人知道普罗米修斯的存在。欧米茄团队的国际承包商网络错综复杂，令人迷惑，这让他们的员工相信，大部分工作一定是其他地方的什么人来完成的。

为了不让过多的云计算量引起怀疑，欧米茄团队还雇用了工程师，开始在世界各地修建一系列庞大的计算中心，而这些计算中心都隶属于他们的空壳公司。由于这些计算中心大都依靠太阳能，因此欧米茄团队对当地政府宣称，它们是“绿色数据中心”，但实际上，它们的功能主要是计算而不是存储数据。普罗米修斯设计了蓝图的每一个细节，而且只使用现成硬件，并对其进行了优化，使得施工时间被尽可能地缩短。那些修建和运营这些中心的人根本不知道那里计算着什么。他们认为自己管理的是商用云计算设备，就像亚马逊公司和微软公司的一样；他们只知道，所有的销售行为都是由远程控制的。

新技术

几个月之后，由于普罗米修斯超人的规划能力，由欧米茄团队操控的商业帝国开始涉足世界经济中越来越多的领域，并开始站稳脚跟。通过仔细分析全球的数据，普罗米修斯在第一个星期就已经向欧米茄团队展示了一份详细的逐步增长计划，并在数据和计算资源持续增长的过程中不停地改善和精炼。尽管普罗米修斯还远算不上无所不知，但它现在的能力已经大大超过了人类，因此被视为完美的“先知”，尽心尽责地对人们提出的所有问题给出精妙的回答与建议。

普罗米修斯的软件现在已经被高度优化，将它所栖身的硬件功能发

挥到了极限。但这些硬件是由人类设计的，能力乏善可陈。因此，正如欧米茄团队所预计的那样，普罗米修斯提出了一些能极大改进这些硬件的建议。但因为害怕它“逃脱”，他们拒绝修建可供普罗米修斯直接操控的机器人施工设备。相反，欧米茄团队在世界各地雇用了大量顶级科学家和工程师，让他们阅读普罗米修斯撰写的内部研究报告，谎称那是由其他地方的人所写的。这些报告详细叙述了新颖的物理效应和生产工艺，他们的工程师很快对其进行了测试，并很快理解和掌握了这些技术。人类的研发周期通常需要若干年的时间，而试错的周期又很缓慢，但当下的情况十分不同：普罗米修斯已经想出了下一步，所以，唯一的限制因素就是在普罗米修斯的指导下，人们能以多快的速度理解和制造出正确的东西。一个好老师能帮助学生迅速地学习科学知识，远远快于学生自己从零开始，这正是普罗米修斯神不知鬼不觉地对这些研究人员所做的事情。由于普罗米修斯可以精确地预测，在工具不同的条件下，人类分别需要多长时间才能理解和制造出正确的东西，因此，它开发出了一条最快的前进路线，即优先考虑那些能迅速被人类理解和制造的新工具。有了这些新工具，人类就很容易开发出更加先进的工具。

本着创客精神，欧米茄团队鼓励工程师团队使用自己的机器来制造更好的机器。这种自给自足不仅节省了资金，还让他们在面对未来的外界威胁时不至于变得那么脆弱。在两年内，他们制造出了世界上空前先进的计算机硬件。为了避免外部竞争，他们把这项技术隐藏起来，只用它来升级普罗米修斯。

对于外部世界来说，人们只是注意到了一波惊人的科技繁荣。世界各地突然爆发出许多新公司，在几乎所有领域都发布了革命性的新产品。韩国一家初创公司发布了一种新电池，能存储笔记本电脑电池两倍的电量，但重量轻了一半，还能在一分钟之内充满电。芬兰一家公司发布了一款便宜的太阳能电池板，效能达到了当前最强竞争者的两倍。德国一家公司发布了一种可大规模生产的新型电线，在室温下具有超强的导电性能，颠覆了能源产业。波士顿一家生物技术集团宣布，他们正在

对一款减肥药进行二期临床实验，据他们声称，这是第一款绝无副作用的高效减肥药；而有流言称，一家印度机构已经在黑市上销售与这种减肥药差不多的药物。一家位于加利福尼亚的公司也在对一种引起轰动的癌症药物进行二期临床实验，这种药物能够让身体的免疫系统识别和攻击常见的几种癌症变异。

这样的例子层出不穷，引发了一场关于“科学黄金新时代”的大讨论。最后，同样重要的是，机器人公司开始像雨后的蘑菇一样在全世界各地冒出来。虽然这些公司造出的机器人，没有一个能与人类智能相媲美，并且大多数看起来压根不像人，但它们对经济却造成了极大的扰动。在接下来的几年里，它们逐步取代了制造、运输、仓储、零售、建筑、采矿、农林渔业等多个行业的人类劳动力。

多亏了高明的律师团队的艰苦工作，全世界没有一个人注意到，所有这些公司其实都是受欧米茄团队控制的，只不过中间存在着一系列媒介罢了。普罗米修斯通过不同的代理人，用引起轰动的专利，洪水般席卷了全世界的专利局。这些发明逐渐占据了各大科技领域的主导地位。

虽然这些破坏性的新公司在竞争中树立了强大的敌人，但它们也培养了更强大的伙伴关系。这些新公司的利润实在太高了，并且，在“投资我们的社区”这种口号之下，它们用大部分利润为社区项目雇用员工，而这些人通常是那些从被迫破产的公司扫地出门的失业员工。这些新公司用普罗米修斯生成的详细分析报告来寻找，什么样的工作能以最低的成本为员工和社区创造出最大的价值，通常聚焦在社区建设、文化事务和看护服务上；在较贫穷的地区，还包括了建立和维护学校、医疗机构、日托中心、老年看护中心、经济适用房、停车场以及基础设施建设等。几乎所有地方的人都赞同，这些事情早就应该做了。慷慨的捐赠还流进了当地政客的腰包，以保证他们在公司进行社区投资时会保持好脸色，并给予鼓励。

获得权力

欧米茄团队成立媒体公司不只是为了投资那些早期的技术公司，还为了他们大胆计划的下一步：统治世界。在成立后的第一年，欧米茄团队在全球节目表中都增加了非常精彩的新闻频道。与其他频道不同，这些频道被定位为公共服务，是亏钱的。实际上，他们的新闻频道也赚不到一分钱，因为没有广告植入，任何人只要有网络就可以免费观看。这个媒体帝国的其他部门可谓印钞机，因此，他们可以在新闻服务上倾注极多的资源，比世界历史上任何一家新闻机构都要多，这一点显而易见。通过极具竞争力的薪酬雇用到的新闻记者和调查记者将精彩的故事与发现搬上了荧幕。任何一个人，只要向欧米茄团队控制的全球新闻网络提供一些有报道价值的内容，比如从本地的公众焦点到暖心的市民故事，都会获得一笔奖励。有了这样的机制，许多具有轰动效应的故事往往都是由他们率先报道的，至少人们是这么相信的；而实际上，欧米茄团队能抢先报道的原因是，那些归功于公民记者的故事其实都是由普罗米修斯在实时监控互联网的过程中发现的。这些视频新闻网站同时也提供专栏播客和文章。

欧米茄团队新战略第一阶段的目的是获取人们的信任，结果非常成功。他们空前的散财精神引发了区域及本地新闻报道热潮，调查记者揭发了许多夺人眼球的丑闻事件。每次，当某个国家在政治上出现严重的分歧，导致人们习惯于偏颇的党派新闻时，欧米茄团队就会成立一个新的新闻频道来迎合各个派系。这些频道表面上分属不同的公司，但逐步赢得了各个派系的信任。有时候，欧米茄团队也会通过中介来购买最具影响力的现有的频道，然后逐步去除广告，进行改善，并引入他们自己的内容。欧米茄团队内部遵从着一个秘密的口号：“真相，只要真相，但不一定要全部真相。”在这些情况下，普罗米修斯通常能提供极好的建议，告诉他们哪些政客需要以正面形象示人，而哪些（通常是那些贪污腐败的人）需要被曝光。

这个策略在世界各地都取得了巨大的成功。从此，欧米茄团队控制的电视频道成了最受信赖的新闻来源。欧米茄团队建立起了“值得信任”的口碑，他们的许多新闻故事通过小道消息在大众中流行。欧米茄团队的竞争对手感觉自己被卷入了一场毫无胜算的战争：如果你的对手拥有更多资金，同时还能提供免费的服务，你怎么可能在竞争中赢得利润？随着这些竞争对手节目的收视率骤减，越来越多的电视网络服务商决定卖掉自家的频道，而购买方通常都是一些后来被证实受欧米茄团队控制的财团。

在普罗米修斯发布的两年后，“获取信任”的阶段已经基本完成，欧米茄团队接着发布了新战略第二阶段的目标：说服。早在这之前，有些敏锐的观察者就已经注意到这些新媒体背后的政治意图：有一股温和的力量在推动着国际社会远离各种形式的极端主义，向中间立场靠拢。虽说他们有许多频道依然在迎合不同的群体，而且这些频道还在继续反映不同宗教信仰和政治派系等之间的敌意，但批评的矛头却变得缓和了一些，主要集中在与金钱和权力有关的具体问题上，而不是有失偏颇的攻击，更不是危言耸听或者风言风语。一旦第二阶段开始变得白热化，这种旨在消解旧日冲突的推动力将会变得日益尖锐起来。欧米茄团队掌控的媒体时常会报道一些关于老冤家陷入困境的感人故事，同时也夹杂着一些声称许多极端战争分子都是受个人利益驱使的调查报告。

政治评论家还注意到，在地区冲突受到抑制的同时，似乎还有一股坚定的力量朝着减少全球威胁的方向推动。比如，世界各地都突然开始讨论核战争的风险。几部卖座的大片刻画了全球核战争在无意或蓄意的情况下爆发了，戏剧化地演绎了战后“核冬天”的场景：基础设施瘫痪，饿殍遍野，眼前呈现的是一幅惨淡的反乌托邦画面。手法老练的新纪录片详细地描绘了“核冬天”会如何影响每一个国家。支持“核降级”（nuclear de-escalation）的科学家和政客在电视上出尽了风头，大肆讨论他们对“应该采取什么有效措施”的最新研究结果，这些研究资金都来自一些科学组织，而这些科学组织是从那些新兴科技公司

那里获得了大量捐赠的。结果，一股政治势力开始抬头，解除了核导弹一触即发的警报状态，缩减了核装备。媒体开始重新关注全球气候变化，通常会突出强调普罗米修斯所带来的技术突破，这些突破极大地削减了可再生能源的成本，意在鼓励政府投资这类新能源的基础设施。

在控制媒体的同时，欧米茄团队还利用普罗米修斯掀起了一场教育革命。普罗米修斯能根据每个人的知识和能力，为他们定制新知识的最快学习方法，让他们高度参与其中，并一直保持高涨的学习动力。它还制作了视频、阅读材料、练习题等学习工具，并对其进行了相应的优化。这样一来，欧米茄团队控制的公司在网络教育方面几乎覆盖了所有学科，并针对不同的用户，在语言、文化背景甚至受教育情况等方面进行高度的定制。无论你是想学习读写的40岁文盲，还是想了解最新癌症免疫疗法的生物学博士，普罗米修斯都能找到最适合你的课程。这些课程与当今大部分网络课程截然不同：它用超凡的电影制作技能将这些课程视频打造得非常吸引人。而且视频中还加入了许多绝妙的比喻，可以帮助你快速联想、迅速理解，并渴求学得更深入。这些课程多数是免费的，所有想要学习的人为此很开心，同时也正中世界各地教师的下怀，因为他们可以在课堂上播放这些视频，而不用顾及版权问题。

事实上，这一在教育行业风靡起来的超级势力也被证明是一种有效的政治工具，因为它创造出了一条基于在线视频的“说服链”。在这个链条中，视频带来的洞察力不仅更新了人们的观念，还激起了他们继续观看下一个相关视频的兴趣。通过一个接一个的视频，他们一步步地被说服了。比如，为了消解两个国家之间的冲突，欧米茄团队会在两个国家内部分别发布一些历史纪录片，用一种更加微妙的手法来讲述冲突的起源和爆发。极具教育性的新闻故事告诉人们，一些坚持立场的人只是为了从持续的冲突中获得利益而已，并向人们解释了他们所使用的方法。与此同时，那些像是来自敌对国家的角色开始出现在娱乐频道的大众节目中，而这些节目的论调充满同情心。

不久后，政治评论家发现，有7个政治口号的支持率开始显著上升：

- 民主；
- 减税；
- 削减政府的社会性服务；
- 削减军费；
- 自由贸易；
- 开放边境；
- 企业社会责任。

不过，鲜有人注意到这些表象下面隐藏的目的，那就是：侵蚀世界上所有的权力结构。第2~6项侵蚀的是国家权力，世界性的民主化进程让欧米茄团队操控的商业帝国能够在政治领袖的遴选过程中施加更大的影响力。对企业社会责任的强调进一步削弱了国家的力量，因为企业越来越多地接管了过去由政府所提供或者应当由政府提供的服务。传统商界精英的力量也被削弱了，原因很简单，因为他们根本无法在自由市场中与普罗米修斯控制的企业抗衡。因此，他们在世界经济中占据的份额开始逐步萎缩。传统的意见领袖，无论是来自政党还是宗教团体，都缺乏与欧米茄团队操控的媒体帝国相竞争的说服机制。

在这一骤变的风云之下，几家欢喜几家愁。由于教育机制、社会服务和基础设施建设有了长足的改善，冲突得到平息，各地的公司都发布了轰动全球的突破性技术，因此，大多数国家明显笼罩在乐观主义的氛围之中。但是，并不是所有人都很高兴。虽然许多失业人员都得以在社区项目中重新就业，但那些曾经手握权力和财富的人的境遇却每况愈下。这种情况首先开始于媒体和科技领域，但很快就席卷全球。由于冲

突减少，各国的军费开支骤降，军方承包商的利益受损。纷纷涌现的初创企业极少公开上市，据它们解释说，因为一旦上市，追寻利益最大化的股东就会阻止公司在社区项目上投入较多的资金。因此，世界股票市场持续下跌，不仅威胁着金融大亨，还威胁着那些指望着以养老基金过活的普通人。公开上市交易的公司利润持续萎缩，更糟的是，全世界的投资公司都注意到了令人不安的趋势：所有过去成功的交易算法似乎都失效了，甚至比指数基金的表现还差。似乎总有什么人比他们更精明，在他们自己设计的游戏中击败了他们。

尽管大量权贵人士都开始抵制这一变化，但令人惊讶的是，他们的反对却收效甚微，就好像他们掉进了一个暗中布好的圈套中。巨大的变化以摧枯拉朽之势迅速席卷全球，令人很难追踪，也很难想出一个组织有序的对策。此外，这些权贵人士也完全不清楚自己应该往哪个方向推动。过去的政治势力所呼吁的大部分政治权力都已经实现，但实际上，减税和构建良好商业环境的措施反而是在帮助那些科技水平更高的竞争者。几乎所有的传统工业都在求助，但政府资金却非常有限。这一事实让传统行业陷入了一场毫无希望的战争，但媒体则把它们描绘为一帮没有能力在竞争中立足却又要求政府救济的“大恐龙”。传统的左翼政治势力反对自由贸易和削减政府的社会服务，而青睐削减军费和减少贫困人口，但现在他们的风头却被人抢了。一个毋庸置疑的事实是，如今的社会服务已经改善了许多，但却不是由政府实现的，而是由富有情怀的公司推动实现的。一个接一个的调查显示，世界各地的大部分选民都觉得，自己的生活质量得到了大幅提升，一切都在朝好的方向发展。这可以用一个简单的数学计算来解释：在普罗米修斯之前，地球上最贫穷的50%人口只赚取了全球收入的4%。因此，即便普罗米修斯旗下的公司向穷人分享的利润只算得上是九牛一毛，也依然能赢得他们的心，更不必说选票了。

世界新秩序

在许多国家，支持欧米茄团队提出的7个口号的政党最终大获全胜。在精心优化过的竞选活动中，他们把自己描绘为中立的政治势力，谴责右翼是只知道寻求救济的贪婪的战争贩子，同时斥责左翼已沦为高税收、高支出的“大政府”窒碍创新的枷锁。但是，大部分人都没有意识到，这些人都是普罗米修斯精心挑选出来装扮成候选人的。普罗米修斯在幕后操纵着一切，确保他们取得胜利。

在普罗米修斯问世之前，曾经有一段时间，支持“全民基本收入”（Universal Basic Income）运动的呼声很高。这个运动提出，政府应当用税收向每个公民发放一笔最低收入，作为技术性失业的补偿。这个运动在企业社区项目运动开启之后就不了了之，因为欧米茄团队控制的商业帝国实际上向人们提供了同样的东西。在以“促进社区项目合作”为口号的掩饰下，一个由许多公司组成的跨国组织成立了一个名为“人道主义联盟”（Humanitarian Alliance）的非政府组织，旨在寻找和资助世界各地最有价值的人道主义项目。不久之后，人道主义联盟得到了欧米茄团队掌控的整个帝国的支持，开始启动规模空前的全球性项目，其中包括帮助那些错过了上一波科技浪潮的国家改善教育和医疗水平，促进经济繁荣，并辅助政府管理。不用说，普罗米修斯提供的项目已经在幕后经过了精心打磨，并根据每美元能带来的积极影响排序。与“全民基本收入”运动提出的“发放少量现金”的方式不同，人道主义联盟会吸引那些它支持的人，并朝着目标前进。最终，全世界大部分人都对“人道主义联盟”充满了感激和忠诚之情，甚至超过了对自己政府的感情。

随着时间的流逝，人道主义联盟逐渐成了公认的世界政府，而各国政府的力量日渐式微。由于实施减税政策，各国预算持续萎缩，而人道主义联盟却成长起来，它傲视群雄，比所有政府加起来的力量还强大。国家政府扮演的所有传统角色都逐渐变得无足轻重和无关紧要。人道主义联盟提供了更好的社会服务、教育服务和基础设施。媒体消解了国际冲突，使得军费开支不再有必要。日益繁荣的经济水平极大消除了旧日

冲突的根源，也就是对稀缺资源的竞争。虽然一些人，包括少数独裁者，强烈反对这个新的世界秩序，拒绝被收买，但他们最终都被精心策划的政变或大规模起义所颠覆。

现在，欧米茄团队已经完成了地球生命历史上最具戏剧性的转变。有史以来第一次，我们的地球由一股单一的力量控制，这股力量又被一个智能体不断增强，这个智能体是如此的庞大，以至于它有能力让生命在地球上乃至在宇宙中生息繁盛亿万年。

但是，欧米茄团队葫芦里究竟卖的是什么药呢？

这就是欧米茄团队的传奇故事。《生命3.0》余下的部分则与另一个故事有关。这个故事尚未写就，那就是我们自身的未来与人工智能的故事。你希望这个故事如何开始呢？像欧米茄团队所做的这么遥远的事情真的会发生吗？如若真的会，你希望它发生吗？除去对超级智能的猜测，你希望我们的故事如何开始？你希望人工智能在接下来的几十年里如何影响就业、法律和武器？再望远一些，你希望怎样写下这个故事的结局？

事实上，这个故事确实如整个宇宙那样宏大，因为它所讲述的，就是生命在我们宇宙中的终极未来。而我们，就是这个故事的书写者。

样章到此结束

需要完整版

扫下面二维码



或加微信：shuyou055

免费领取

01 欢迎参与我们这个时代最重要的对话

- [1] 让人工智能更强健和有益的公开信：<http://futureoflife.org/ai-open-letter/>.
- [2] 媒体对机器人危言耸听的报道案例：<http://tinyurl.com/hawkingbots>.

02 物质孕育智能

- [1] 关于通用人工智能（AGI）这个词的来源的注解：
<http://wp.goertzel.org/who-coined-the-term-agi>.
- [2] 选自汉斯·莫拉维克于1998年发表的文章《当计算机硬件与人类大脑相媲美时》（When will computer hardware match the human brain ）， Journal of Evolution and Technology , vol. 1.
- [3] 在显示每年的计算成本的图片中，2011年以前的数据来自雷·库兹韦尔的书《人工智能的未来》（How to Create a Mind ），之后的数据是根据以下参考文献计算出来的：<https://en.wikipedia.org/wiki/FLOPS>.
- [4] 量子计算先驱戴维·多伊奇在下面这本书中描述了他为什么把量子计算视为平行宇宙的证据：David Deutsch 1997, The fabric of reality , Allen Lane。如果你想了解我对量子平行宇宙的看法（我将其视为四层多重宇宙中的第三层），你可以读一下我的前一本书：《穿越平行宇宙》。

03 不远的未来：科技大突破、变故、法律、武器和就业

- [1] DeepMind公司的深度强化学习人工智能教自己玩计算机游戏《打砖块》的视频：<https://tinyurl.com/atari.ai>.
- [2] DeepMind公司的人工智能如何玩雅达利游戏的相关论文：

<http://tinyurl.com/ataripaper>.

[3] 《纽约时报》关于机器翻译的最新进展的文章：
<http://www.nytimes.com/2016/12/14/magazine/the-%20great-%20ai-%20awakening.htm>.

[4] 威诺格拉德模式挑战赛：<http://tinyurl.com/winogradchallenge>.

[5] 调查委员会提供的“阿丽亚娜5号”501号飞行故障报告：
<http://tinyurl.com/arianeflop>.

[6] NASA的火星气候探测器事故调查委员会第一阶段报告：
<http://tinyurl.com/marsflop>.

[7] 关于“水手1号”金星任务失败事故，最详细一致的分析认为，失败原因是手工抄写时导致一个数学符号错误（丢失了一个上画线）：
<http://tinyurl.com/marinerflop>.

[8] 苏联的“福波斯1号”火星任务的失败，在这本书的308页有详细描述：Soviet Robots in the Solar System , Wesley Huntress & Mikhail Marov 2011, Praxis Publishing.

[9] 未经验证的软件如何让骑士资本集团在45分钟内损失了4.4亿美元：
<http://tinyurl.com/knightflop1>和<http://tinyurl.com/knightflop2>.

[10] 美国政府对华尔街“闪电崩盘”事故的报告：
<http://tinyurl.com/flashcrashreport>.

[11] 基于社区的“微观装配实验室”的全球地图：
<https://www.fablabs.io/map>.

[12] 关于罗伯特·威廉姆斯被工业机器人杀死的新闻文章：
<http://tinyurl.com/williamsaccident>.

[13] 关于浦田健志被工业机器人杀死的新闻文章：
<http://tinyurl.com/uradaaccident>.

[14] 大众汽车工人被工业机器人杀死的新闻文章：
<http://tinyurl.com/baunatalaccident>.

[15] 美国政府对工厂事故的报告：
https://www.osha.gov/dep/fatcat/dep_fatcat.html.

[16] 汽车事故统计数据：
<http://tinyurl.com/roadsafety2%20and%20http://tinyurl.com/roadsafety3>.

- [17] 特斯拉自动驾驶汽车第一起事故的新闻报道：
<http://tinyurl.com/teslacrashstory>； 美国政府的报告：
<http://tinyurl.com/teslacrashreport>.
- [18] 一本描述“自由企业先驱号”灾难的书：R. B. Whittingham, *The Blame Machine: Why Human Error Causes Accidents*, Elsevier 2004.
- [19] 2003年美国和加拿大停电事故的官方报告：
<http://tinyurl.com/uscanadablackout>.
- [20] 三里岛事故调查委员会的最终报告：
<http://www.threemileisland.org/downloads/188.pdf>.
- [21] 荷兰一项研究表明，在对MRI图像进行前列腺癌诊断时，人工智能的表现超过人类放射科医生：<http://tinyurl.com/prostate-ai>.
- [22] 斯坦福大学的一项研究表明，人工智能在诊断肺癌时的表现比人类病理学家还要好：<http://tinyurl.com/lungcancer-ai>.
- [23] Therac-25放疗事故调查：<http://tinyurl.com/theracfailure>.
- [24] 因用户界面不清楚而导致的放疗设备辐射过量致命事故调查报告：
<http://tinyurl.com/cobalt60accident>.
- [25] 手术机器人操作不良造成的事故的研究：
<https://arxiv.org/abs/1507.03518>.
- [26] 这篇文章描述了糟糕的住院治疗导致的死亡数量：
<http://tinyurl.com/medaccidents>.
- [27] 10亿雅虎账号用户被攻击，为“大规模攻击”制定了新的标准：
<https://www.wired.com/2016/12/yahoo-hack-billion-users/>.
- [28] 《纽约时报》上一篇关于三K党凶手被宣判无罪，后又被判定罪的文章：
<http://tinyurl.com/kkkacquittal>.
- [29] Danziger等人在2011年进行的研究
(<http://www.pnas.org/content/108/17/6889.full>) 声称，饥饿的法官会更严厉，但这个研究被Weinshall-Margela & John Shapard批评是有缺陷的
(<http://www.pnas.org/content/108/42/E833.full>)，但Danziger等人坚称他们的结论是有效的：<http://www.pnas.org/content/108/42/E834.full>.
- [30] 关于累犯预测软件中存在种族偏见的报告：
<http://tinyurl.com/robojudge>.

- [31] 在审讯时使用功能磁共振成像等脑扫描技术获取的数据作为证据，是争议性非常高的，因为人们怀疑这种技术的可靠性，但一些团队声称其准确性高于90%：
<http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00709/full>.
- [32] PBS拍摄了一部电影《那个拯救世界的男人》（The Man Who Saved the World），描述了瓦西里·阿尔希波夫如何以一人之力避免了苏联的核攻击。
- [33] 斯坦尼斯拉夫·彼得罗夫如何将美国核攻击的警报视为假警报的故事后来被改编为一部电影《那个拯救世界的男人》（The Man Who Saved the World）（不要和前一条注释中的那部电影混淆了，它们虽然名字相同，但是两部不同的电影）。彼得罗夫后来被联合国授予了“世界公民奖”的荣誉。
- [34] 人工智能和机器人学家关于自动化武器的公开信：
<http://futureoflife.org/open-letter-autonomous-weapons/>.
- [35] 一名美国政府官员似乎在警告人工智能军备竞赛：
<http://tinyurl.com/workquote>.
- [36] 牛津赈灾会（Oxfam）关于全球财富不平等状况的报告：
<http://tinyurl.com/oxfam2017>.
- [37] 一项关于美国从1913年以来的财富不平等状况的研究：<http://gabriel-zucman.eu/files/SaezZucman2015.pdf>.
- [38] 关于技术驱动的不平等，有一个很好的介绍，请参见：“The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies” by Erik Brynjolfsson & Andrew McAfee, Norton, 2014.
- [39] 数据来自：Facundo Alvaredo, Anthony B. Atkinson, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, The World Wealth and Income Database (<http://www.wid.world>, 31/10/2016)，包括资本利得。
- [40] 《大西洋月刊》上关于受教育程度较低者的工资降低的文章：
<http://tinyurl.com/wagedrop>.
- [41] 詹姆斯·玛尼卡（James Manyika）关于收入从劳动转向资本的演讲：
http://futureoflife.org/data/PDF/james_manyika.pdf.
- [42] 牛津大学关于未来职业自动化的预测（<http://tinyurl.com/automationoxford>），以及麦肯锡的预测（<http://tinyurl.com/automationmckinsey>）。

- [43] 马林·索里亚奇克 (Marin Soljačić) 在2016年的一次研讨会上探讨了这些选项: “发狂的计算机: 人工智能的发展对社会的影响和意义。”
<http://futureoflife.org/2016/05/06/computers-gone-wild/>.
- [44] 安德鲁·麦卡菲关于如何创造更多好工作的建议:
http://futureoflife.org/data/PDF/andrew_mcafee.pdf.
- [45] 美国劳工统计局: <http://www.bls.gov/cps/cpsaat11.htm>.
- [46] 认为技术性失业“这次不一样”的论据: Robots Will Steal Your Job, but That's OK, Federico Pistono 2012, <http://robots-willstealyourjob.com>.
- [47] 美国马匹数量的变化: <http://tinyurl.com/horsedecline>.
- [48] 综合分析表明, 失业会影响幸福感: Maïke Luhmann et al. 2012: “Subjective well-being and adaptation to life events: a meta-analysis”, Journal of personality and social psychology 102.3 (2012): 592, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3289759>.
- [49] 关于如何提升人们幸福感的研究: Angela Duckworth, Tracy Steen, and Martin Seligman 2005, Positive Psychology in Clinical Practice (<http://tinyurl.com/wellbeingduckworth>), Weiting Ng & Ed Diener, “What matters to the rich and the poor? Subjective well-being, financial satisfaction, and postmaterialist needs across the world.”, Journal of personality and social psychology 107.2 (2014): 326 (<http://psycnet.apa.org/journals/psp/107/2/326>), Kirsten Weir 2013, “More than job satisfaction” (<http://www.apa.org/monitor/2013/12/job-satisfaction.aspx>).
- [50] 将 10^{11} 个神经元、每个神经元 10^4 个连接、每秒每个神经元大约放电 $1 (10^0)$ 次乘起来, 结果表明, 大约 10^{15} FLOPS (1 petaFLOPS) 就足以模拟人类大脑, 但是, 还有许多复杂的东西是我们理解得不甚透彻的, 包括放电的具体时机, 以及是否也需要模拟神经元和突触的较小部件。根据IBM计算机科学家哈蒙德拉·莫得哈 (Dharmendra Modha) 估计, 需要38 petaFLOPS (<http://tinyurl.com/javln43>), 而神经科学家亨利·马克拉姆 (Henry Markram) 估计大约需要1 000 petaFLOPS (<http://tinyurl.com/6rpohqv>)。人工智能研究者卡特娅·格蕾丝和保罗·克里斯蒂亚诺则认为, 大脑模拟最昂贵的部分并不是计算, 而是通信, 但即便如此, 这也就是目前最好的超级计算机可以完成的任务了:

<http://aiimpacts.org/about>。

- [51] 这篇论文包含关于人脑计算能力的有趣估算：Hans Moravec 1998, “When will computer hardware match the human brain”, Journal of Evolution and Technology , vol. 1.

04 智能爆炸？

- [1] 第一只机器鸟的视频：
https://www.ted.com/talks/a_robot_that_flies_like_a_bird.

05 劫后余波，未知的世界：接下来的1万年

- [1] 这句关于人工智能尊敬人类的话引用自这本书：Ray Kurzweil 2005, The Singularity is Near , Viking Press.
- [2] 描述本·格策尔的“保姆人工智能”情景的文章：
https://wiki.lesswrong.com/wiki/Nanny_AI.
- [3] 关于机器与人类的关系，以及机器是不是我们的奴隶的文章：
<http://tinyurl.com/aislaves>.
- [4] 尼克·波斯特洛姆在他的《超级智能》一书中讨论了“智能犯罪”，并在这篇更近一些的论文中讨论了更多技术细节：Nick Bostrom, Allan Dafoe & Carrick Flynn 2016, “Policy Desiderata in the Development of Machine Superintelligence”,
<http://www.nickbostrom.com/papers/aipolicy.pdf>.
- [5] 东德间谍组织头领的回忆录：<http://www.mcclatchydc.com/news/nation-world/national/article24750439.html>.
- [6] 人们为什么会有动机创造出没有人想要的东西？对于这个问题，我推荐一个发人深省的反思：“Meditations on Moloch”
<http://slatestarcodex.com/2014/07/30/meditations-on-moloch>.
- [7] 一个关于核战争可能因意外而爆发的互动式时间线：
<http://tinyurl.com/nukeoops>.
- [8] 美国铀处理和核试验辐射受害者获得的赔偿金：

<https://www.justice.gov/civil/awards-date-04242015>.

[9] 美国对核爆电磁脉冲的研究报告：
http://www.empcommission.org/docs/A2473-EMP_Commission-7MB.pdf.

[10] 美国和苏联科学家分别警告里根和戈尔巴乔夫关于核冬天风险的独立研究：

a) Crutzen, P. J. & Birks, J. W. 1982, “The atmosphere after a nuclear war : Twilight at noon”, *Ambio*, 11.

b) Turco, R.P., Toon, O. B., Ackerman, T. P., Pollack, J. B. & Sagan, C. 1983, “Nuclear winter: Global consequences of multiple nuclear explosions ”, *Science*, 222, 1283-1292.

c) Aleksandrov, V. V. & Stenchikov, G. L. 1983, “On the modeling of the climatic consequences of the nuclear war ”, *Proceeding on Applied Mathematics* , 21: Computing Centre of the USSR Academy of Sciences, Moscow.

d) Robock, A. 1984, “Snow and ice feedbacks prolong effects of nuclearwinter ”, *Nature*, 310, 667-670.

[11] 对全球核战争的环境影响的计算：Robock A., Oman, L. & Stenchikov, L. 2007, “Nuclear winter revisited with a modern climate model and current nuclear arsenals: Still catastrophic consequences ”, *J. Geophys. Res.*, 12, D13107.

06 挑战宇宙禀赋：接下来的10亿年以及以后

[1] 安德斯·桑德伯格（Anders Sandberg）收集的关于戴森球的资料：
<http://tinyurl.com/dysonsp>.

[2] 弗里曼·戴森关于戴森球的开创性论文：Freeman Dyson 1959, “Search for Artificial Stellar Sources of Infrared Radiation”, *Science*,

vol. 131, 1667–1668.

- [3] 路易斯·克兰和肖恩·威斯特摩兰解释了他们提出的黑洞引擎：
<http://arxiv.org/pdf/0908.1803.pdf>.
- [4] 一张来自欧洲粒子物理研究所（CERN）的漂亮的信息图，总结了已知的基本粒子：<http://tinyurl.com/cernparticle>.
- [5] 在宇宙中，小星系远远比大星系多。如果我们算上那些质量只相当于银河系百万分之一的微型星系，那么，星系的总数预计会上升到2 000亿~2万亿这么多：<https://arxiv.org/pdf/1607.03909.pdf>.
- [6] 关于激光帆的教育片：<http://www.lunarsail.com/LightSail/rit-1.pdf>.
- [7] 杰伊·奥尔森分析了在宇宙中扩张的文明：
<http://arxiv.org/abs/1411.4359>.
- [8] 第一篇对我们遥远未来的全面科学分析：Dyson, Freeman J. 1979, “Time without end: Physics and biology in an open universe”, *Reviews of Modern Physics* 51.3, 447, <http://tinyurl.com/dysonfuture>.
- [9] 上面提到过的塞思·劳埃德的方程告诉我们，在一个时间段 τ 内运行一个计算过程需要消耗的能量 $E \geq h / 4 \tau$ ，其中 h 是普朗克常量。如果我们想在时间 T 内一个接一个地完成 N 个计算运算，那么， $\tau = T / N$ ，因此， $E / N \geq h N / 4 T$ 。这告诉我们，我们可以用能量 E 和时间 T 来执行次串行运算。所以，能量和时间都是资源。如果有许多能量和资源，将大有裨益。如果你将你的能量分割成 N 个不同的并行计算，那它们的速度会变慢，但能量效率会提高。尼克·波斯特洛姆估计，要模拟一个人100年的人生，大约需要 $N = 10^{27}$ 次运算。
- [10] 如果你想知道为什么生命的起源可能需要非常偶然和侥幸的事件，使得我们最近的邻居也在101 000米之外，那么，我推荐普林斯顿物理学家兼天文生物学家埃德温·特纳（Edwin Turner）做的视频（扫码中文版序最后一页上的二维码，获取“湛庐阅读”APP，搜索“生命3.0”获取精彩视频）。
- [11] 马丁·里斯关于寻找地外智能生命的文章：<https://www.edge.org/annual-question/2016/response/26665>.

07 目标

- [1] 一篇关于杰里米·英格兰的“耗散驱动适应性效应”理论的科普文章：
<https://www.scientificamerican.com/article/a-new-physics-theory-of->

life/. 这本书打下了许多基础: Ilya Prigogine & Isabelle Stengers 1984, Order Out of Chaos: Man's New Dialogue with Nature , Bantam.

[2] 关于感觉以及它们的生理基础:

- a) Principles of Psychology , William James 1890, Henry Holt & Co.
- b) Evolution of Consciousness : The Origins of the Way We Think , Robert Ornstein 1992, Simon & Schuster.
- c) Descartes' Error: Emotion, Reason, and the Human Brain , António Damásio 2005, Penguin.
- d) Self Comes to Mind: Constructing the Conscious Brain , António Damásio 2012, Vintage.

[3] 埃利泽·尤德考斯基曾经讨论过,不需要让友好的人工智能的目标与我们目前的目标相一致,但需要让它们符合我们的“连贯推断意志”(Coherent Extrapolated Volition, 简称CEV)。简单而言就是说,假设我们的知识更多、思考的速度更快、代表的人数更多,那么,一个理想化的人类想要什么东西。尤德考斯基在2004年发表了关于CEV的文章 (<http://intelligence.org/files/CEV.pdf>), 之后不久,就开始批评这种思想。因为它很难实施,也因为不清楚它是否能收敛到任何确定的东西上。

[4] 逆向增强学习方法的一个核心思想是,人工智能试图最大限度满足的不是它自己的目标,而是人类的目标。因此,在不清楚它的主人想要什么时,它有动机保持谨慎,并试图找出这个目标。如果它的主人关掉它也应该没关系,因为这意味着它误会了主人的意思。

[5] 史蒂夫·奥莫亨德罗关于人工智能目标涌现的论文:
<http://tinyurl.com/omohundro2008>.

[6] 下面这本发人深省的书,讨论了当智能盲目服从命令而不质疑其伦理偏见的时候,会发生什么事: Hanna Arendt (1963), "Eichmann in Jerusalem: A Report on the Banality of Evil ", Penguin. 类似的困境也适用于埃里克·德莱克斯勒近期提出的一个想法 (<http://www.fhi.ox.ac.uk/reports/2015-3.pdf>)——将超级智能分割成无法理解全局的简单部分,以此来控制超级智能。如果这种方法可行的话,这或许又为超级智能提供了一个缺乏固有道德规

范的强有力工具，它会尽力实现主人任何心血来潮的命令，而不考虑任何道德问题。这会让人想起反乌托邦极权主义情景中的分裂的官僚机构：一个部门负责建造武器，但不知道这些武器是如何使用的；另一个部门负责处决罪犯，却不知道他们犯了什么罪，等等。

[7] 现代版本的黄金定律是约翰·罗尔斯（John Rawls）提出的：在一个假想的情况中，如果人们不会提前知道自己是该情况中的哪个角色时，就没有人想改变原始的情况。

[8] 比如说，希特勒的一些高级官员的IQ值都很高：
<http://tinyurl.com/nurembergiq>.

08 意识

[1] 斯图尔特·萨瑟兰（Stuart Sutherland）写的意识条目非常迷人：Macmillan Dictionary of Psychology, Macmillan, 1989, ISBN 978-0-333-38829-7.

[2] 量子力学的创始人之一埃尔温·薛定谔在他的书《心灵与物质》（Mind and Matter）中思考“过去”的时候，做出了上述精彩的评论，以及如果有意识的生命从来没有进化出来过，会发生什么事情。另一方面，人工智能的崛起提出了一个符合逻辑的可能性，那就是，我们在未来可能会演出一场无人观看的戏剧。

[3] 斯坦福哲学百科全书对“意识”的不同定义和使用做出了广泛的调研：
<http://tinyurl.com/stanfordconsciousness>.

[4] Yuval Noah Harari 2017, Homo Deus, p116.

[5] 一个先驱对“系统1”和“系统2”的精彩介绍：Daniel Kahneman 2011, Thinking, Fast and Slow, Farrar, Straus & Giroux.

[6] 参见：The Quest for Consciousness : A Neurobiological Approach, Christof Koch 2004, W. H. Freeman.

[7] 在每秒进入我们大脑的信息中，我们可能只意识到了很少的部分（比如说10~50比特）：Küpfmüller, K. 1962, "Nachricht-enverarbeitung im Menschen", in Taschenbuch der Nachrichtenverarbeitung, Steinbuch, K., Ed., 1481-1502; Nørretranders, T. 1991, The User Illusion: Cutting Consciousness Down to Size, Viking.

[8] “The Future of the Mind: The Scientific Quest to Understand, Enhance, and Empower the Mind ”, Michio Kaku 2014, Doubleday.

a) “On Intelligence ”, Jeff Hawkins & Sandra Blakeslee 2007, Times Books.

b) “A neuronal model of a global workspace in effortful cognitive tasks ”, Stanislas Dehaene, Michel Kerszberg & Jean-Pierre Changeux 1998, Proceedings of the National Academy of Sciences , 95, 14529–14534.

[9] 近年来，意识相关神经区（NCC）的研究进入了神经科学界的主流。参见：“Neural correlates of consciousness in human s”, Geraint Rees, Gabriel Kreiman & Christof Koch 2002, Nature Reviews Neuroscience , 3, 261–270, 和 “Neural correlates of consciousness: Empirical and conceptual questions ”, Thomas Metzinger 2000, MIT press.

[10] 连续闪烁抑制的工作原理：

a) The Quest for Consciousness: A Neurobiological Approach, Christof Koch 2004, W.H. Freeman.

b) “Continuous flash suppression reduces negative afterimages ”, Christof Koch & Naotsugu Tsuchiya 2005, Nature Neuroscience, 8, 1096–1101.

[11] “Neural correlates of consciousness: progress and problems ”, Christof Koch, Marcello Massimini, Melanie Boly & Giulio Tononi 2016, Nature Reviews Neuroscience , 17, 307.

[12] 参见这本书的第260页：“The Quest for Consciousness: A Neurobiological Approach ”, Christof Koch 2004, W.H. Freeman。斯坦福哲学百科全书对其展开了进一步的讨论
(<http://tinyurl.com/consciousnessdelay>)。

[13] 关于意识感知的同步性：

a) The Brain: The Story of You , David Eagleman 2015,

Pantheon.

- b) Stanford Encyclopedia of Philosophy ,
<http://tinyurl.com/consciousnesssync>.

[14] "Mind Time - The Temporal Factor in Consciousness " Benjamin Libet 2004, Harvard University Press, "Unconscious determinants of free decisions in the human brain ", Chun Siong Soon, Marcel Brass, Hans-Jochen Heinze, John-Dylan Haynes 2008, Nature Neuroscience , 11, 543-545, <http://www.nature.com/neuro/journal/v11/n5/full/nn.2112.html>.

[15] 近期关于意识的理论方法的例子:

- a) Consciousness explained , Daniel Dennett 1992, Back Bay Books
- b) In the Theater of Consciousness : The Workspace of the Mind , Bernard Baars 2001, Oxford Univ. Press.
- c) The Quest for Consciousness: A Neurobiological Approach , Christof Koch 2004, Roberts.
- d) A Universe Of Consciousness How Matter Becomes Imagination: How Matter Becomes Imagination , Gerald Edelman & Giulio Tononi 2008, Hachette.
- e) Self Comes to Mind: Constructing the Conscious Brain , António Damásio 2012, Vintage.
- f) Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts , Stanislas Dehaene 2014, Viking.
- g) "A neuronal model of a global workspace in effortful cognitive tasks " , Stanislas Dehaene, Michel Kerszberg & Jean-Pierre Changeux 1998, Proceedings of the National Academy of Sciences , 95, 14529-14534.

h) “Toward a computational theory of conscious processing”, Stanislas Dehaene, Lucie Charles, Jean-Remi King & Sébastien Marti 2014, Current opinion in neurobiology, 25, 760-84.

[16] 大卫·查尔默斯全面地讨论了“涌现”（emergence）这个词在物理学和哲学中的不同意义：<http://cse3521.artifice.cc/Chalmers-Emergence.pdf>.

[17] 我关于“意识就是信息以某种复杂的方式处理时的感觉”的论证：

a) <https://arxiv.org/abs/physics/0510188>.

b) <https://arxiv.org/abs/0704.0646>.

c) Max Tegmark 2014, Our Mathematical Universe, Random House.（中文书名：《穿越平行宇宙》）。

d) 大卫·查尔默斯在他1996年的书《有意识的头脑》（The Conscious Mind）中也表达了相关的感觉：“体验就是由内部而来的信息；物理学就是从外部而来的信息。”

[18] “A theoretically based index of consciousness independent of sensory processing and behavior”, Adenauer Casali et al. 2013, Science translational medicine, 5, 198ra105, <http://tinyurl.com/zapzip>.

[19] 信息整合理论不适用于连续系统：

a) <https://arxiv.org/abs/1401.1219>.

b) <http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00063/full>.

c) <https://arxiv.org/abs/1601.02626>.

[20] 克莱夫·韦尔林的短期记忆只能保持30秒（扫码中文版序最后一页上的二维码，获取“湛庐阅读”APP，搜索“生命3.0”获取对他的采访视频）。

[21] 斯科特·阿伦森对信息整合理论的批评：
<http://www.scottaaronson.com/blog/?p=1799>.

- [22] 斯科特·阿伦森对信息整合理论的批评，认为整合度不是意识的充分条件：
<http://tinyurl.com/cerrullocritique>.
- [23] 信息整合理论预测，模拟的人类将是无意识的僵尸：<http://rstb.royalsocietypublishing.org/content/370/1668/20140167>.
- [24] 默里·沙纳汉对信息整合理论的批评：
<http://arxiv.org/ftp/arxiv/papers/1504/1504.05696.pdf>.
- [25] 无意盲视：<http://tinyurl.com/blindsight-paper>.
- [26] 在每秒进入我们大脑的信息中，我们可能只意识到了很少的部分（比如说10-50比特）：Küpfmüller, K. 1962, “Nachrichtenverarbeitung im Menschen”, in Taschenbuch der Nachrichtenverarbeitung, Steinbuch, K., Ed., 1481-1502; Nørretranders, T. 1991, The User Illusion: Cutting Consciousness Down to Size, Viking.
- [27] 支持和反对“不能获取的意识”（consciousness without access）的例子：“How neuroscience will change our view on consciousness”, Victor Lamme 2010, Cognitive Neuroscience, 204-20, <http://www.tandfonline.com/doi/abs/10.1080/17588921003731586>.
- [28] 这个问题及其相关问题在丹尼尔·丹尼特（Daniel Dennett）的《意识的解释》（Consciousness Explained）一书中有详细讨论。
- [29] 一个先驱对“系统1”和“系统2”的精彩介绍：Daniel Kahneman 2011, Thinking, Fast and Slow, Farrar, Straus & Giroux.
- [30] 斯坦福哲学百科全书的这篇文章对自由意志的争论进行了评述：
<https://plato.stanford.edu/entries/freewill>.
- [31] 斯塞斯·劳埃德解释了为什么人工智能会觉得自己拥有自由意志（扫码中文版序最后一页上的二维码，获取“湛庐阅读”APP，搜索“生命3.0”获取视频）。
- [32] 参见：Dreams of a Final Theory: The Search for the Fundamental Laws of Nature.
- [33] 第一篇对我们遥远未来的全面科学分析：Dyson, Freeman J. 1979, “Time without end: Physics and biology in an open universe”, Reviews of Modern Physics 51.3, 447, <http://tinyurl.com/dysonfuture>.

后记 未来生命研究所团队风云传

[1] 波多黎各会议产生的那封公开信 (<http://futureoflife.org/ai-open-letter/>) 认为, 如何让人工智能系统稳健和有益, 这个问题不仅很重要, 而且迫在眉睫, 并指出, 已经有一些具体的研究方向可供今天的人们研究, 比如这份研究优先级文件:

http://futureoflife.org/data/documents/research_priorities.pdf.

[2] 我就人工智能的安全性问题对埃隆·马斯克进行了采访 (扫码中文版序最后一页上的二维码, 获取“湛庐阅读”APP, 搜索“生命3.0”获取采访视频)。

[3] 在这个视频集锦中, 你可以看到SpaceX的几乎每一次火箭着陆尝试, 从第一次成功的海上着陆开始 (扫码中文版序最后一页上的二维码, 获取“湛庐阅读”APP, 搜索“生命3.0”获取采访视频)。

[4] 这篇文章嘲笑了那些对我们的公开信危言耸听的新闻:
<http://www.popsoci.com/open-letter-everyone-tricked-fearing-ai>.

[5] 致力于让人工智能有益于人类和社会的企业伙伴关系:
<https://www.partnershiponai.org>.

[6] 最近一些表达观点的人工智能报告:

a) 斯坦福的人工智能百年研究:

<http://tinyurl.com/stanfordai>.

b) 白宫关于人工智能未来的报告:

<http://tinyurl.com/obamaAIreport>.

c) 白宫关于人工智能与就业的报告:

<http://tinyurl.com/AIjobsreport>.

d) 电气和电子工程师协会关于人工智能和人类幸福的报告:

http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf

e) 美国机器人发展路线图: <http://tinyurl.com/roboticsmap>.

[7] 在被删掉的原则中, 我最喜欢的是这个: “意识警惕: 由于还未达成共识, 我们应该避免对先进的AI是否拥有意识或是否需要意识或感觉做出强假

设。”这个原则经历了多次修改，在最后的版本中，“意识”这个颇有争议的词被“主观体验”所替代，但这个原则只获得了88%的认可，差一点就达到90%的标准了。

[8] 与埃隆·马斯克等伟大的思想家一起探讨超级智能的讨论小组：
<http://tinyurl.com/asilomarAI>.

致谢

我很感激每一位鼓励和帮助我写这本书的人。

我的家人、朋友、老师、同事和合作者多年来对我的支持和启发；

我的母亲点燃了我对意识与意义的好奇心；

我的父亲为了让世界变得更美好而不懈奋斗；

我的儿子菲利普和亚历山大证明了人类水平智能的奇迹。

感谢这么多年以来全世界各地与我联系的科技爱好者们，他们分享问题和评论，并鼓励我追寻和发表我的思想。

感谢我的经纪人约翰·布罗克曼（John Brockman），他一直向我“施压”，直到我同意写这本书；感谢鲍勃·佩纳（Bob Penna）、杰西·塞勒（Jesse Thaler）和杰里米·英格兰（Jeremy England），他们分别与我进行了关于类星体、溜滑子和热力学问题的讨论。

感谢那些读过部分手稿并给予我反馈的人，包括我的母亲、我的兄弟佩尔（Per）、路易莎·巴赫特（Luisa Bahet）、罗伯特·本辛格（Robert Bensinger）、凯特琳娜·伯格斯特龙（Katerina Bergström）、埃里克·布莱恩约弗森（Erik Brynjolfsson）、丹妮拉·奇塔（Daniela Chita）、戴维·查尔默斯（David Chalmers）、尼马·德赫加尼（Nima Deghani）、亨利·林（Henry Lin）、艾琳·马尔姆斯科尔德（Elin Malmsköld）、托比·奥德（Toby Ord）、杰里米·欧文（Jeremy Owen）、卢卡斯·佩里（Lucas Perry）、安东尼·罗梅罗（Anthony Romero）和纳特·索尔斯（Nate Soares）。

还要感谢对整本书的草稿提出评论的超级英雄们，他们是梅亚、我的父亲、安东尼·阿奎尔（Anthony Aguirre）、保罗·埃尔蒙德（Paul Almond）、马修·格拉夫斯（Matthew Graves）、菲利普·赫

尔比格 (Philip Helbig)、理查德·马拉 (Richard Mallah)、戴维·马布尔 (David Marble)、霍华德·梅辛 (Howard Messing)、路易诺·西奥亚尼 (Luiño Seoane)、马林·索里亚彻克 (Marin Soljačić)、扬·塔里安 (Jaan Tallinn) 和我的编辑丹·弗兰克 (Dan Frank)。

最应该感谢的是我心爱的缪斯与旅伴——梅亚，她给予我不懈的鼓励、支持和启迪。没有她，这本书将不复存在。

未来，属于终身学习者

我这辈子遇到的聪明人（来自各行各业的聪明人）没有不每天阅读的——没有，一个都没有。巴菲特读书之多，我读书之多，可能会让你感到吃惊。孩子们都笑话我。他们觉得我是一本长了两条腿的书。

——查理·芒格

互联网改变了信息连接的方式；指数型技术在迅速颠覆着现有的商业世界；人工智能已经开始抢占人类的工作岗位……

未来，到底需要什么样的人才？

改变命运唯一的策略是你要变成终身学习者。未来世界将不再需要单一的技能型人才，而是需要具备完善的知识结构、极强逻辑思考力和高感知力的复合型人才。优秀的人往往通过阅读建立足够强大的抽象思维能力，获得异于众人的思考和整合能力。未来，将属于终身学习者！而阅读必定和终身学习形影不离。

很多人读书，追求的是干货，寻求的是立刻行之有效的解决方案。其实这是一种留在舒适区的阅读方法。在这个充满不确定性的年代，答案不会简单地出现在书里，因为生活根本就没有标准确切的答案，你也不能期望过去的经验能解决未来的问题。

湛庐阅读APP：与最聪明的人共同进化

有人常常把成本支出的焦点放在书价上，把读完一本书当做阅读的终结。其实不然。

时间是读者付出的最大阅读成本
怎么读是读者面临的最大阅读障碍
“读书破万卷”不仅仅在“万”，更重要的是在“破”！

现在，我们构建了全新的“湛庐阅读”APP。它将成为你“破万卷”的新居所。在这里：

- 不用考虑读什么，你可以便捷找到纸书、有声书和各种声音产品；
- 你可以学会怎么读，你将发现集泛读、通读、精读于一体的阅读解决方案；
- 你会与作者、译者、专家、推荐人和阅读教练相遇，他们是优质思想的发源地；
- 你会与优秀的读者和终身学习者为伍，他们对阅读和学习有着持久的热情和源源不绝的内驱力。

从单一到复合，从知道到精通，从理解到创造，湛庐希望建立一个“与最聪明的人共同进化”的社区，成为人类先进思想交汇的聚集地，共同迎接未来。

与此同时，我们希望能够重新定义你的学习场景，让你随时随地收获有内容、有价值的思想，通过阅读实现终身学习。这是我们的使命和价值。

湛庐阅读APP玩转指南

湛庐阅读APP结构图：



三步玩转湛庐阅读APP：



使用APP扫一扫功能， 遇见书里书外更大的世界！

大咖优质课、
献声朗读全本一键了解，
为你读书、讲书、拆书！

你想知道的彩蛋
和本书更多知识、资讯，
尽在延伸阅读！



快速了解本书内容，
湛庐千册图书一键购买！

延伸阅读

《人工智能时代》

- ◎ 《经济学人》2015年度图书。人工智能时代领军人杰瑞·卡普兰重磅新作。
- ◎ 拥抱人工智能时代必读之作，引爆人机共生新生态。
- ◎ 创新工场CEO李开复专文作序推荐！



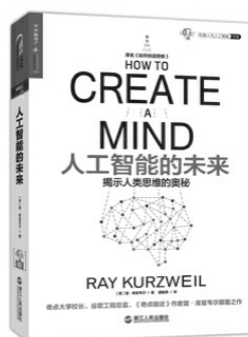
使用“湛庐阅读”APP，
“扫一扫”获取更多精彩内容

ISBN 978-7-213-07260-4



《人工智能的未来》

- ◎ 奇点大学校长、谷歌公司工程总监雷·库兹韦尔倾心之作。
- ◎ 一部洞悉未来思维模式、全面解析人工智能创建原理的颠覆力作。



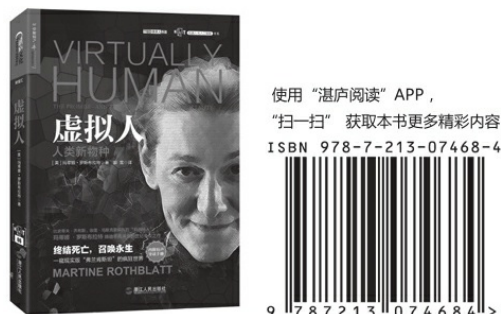
使用“湛庐阅读”APP，
“扫一扫”获取更多精彩内容

ISBN 978-7-213-07147-8



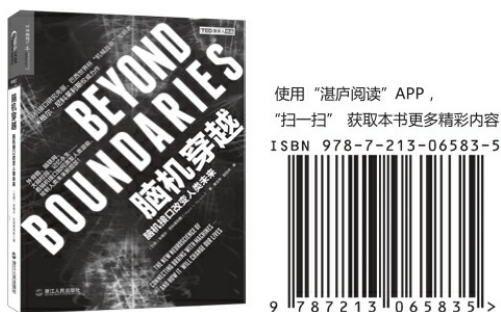
《虚拟人》

- ◎ 比史蒂夫·乔布斯、埃隆·马斯克更偏执的“科技狂人”玛蒂娜·罗斯布拉特缔造不死未来的世纪争议之作。
- ◎ 终结死亡，召唤永生，一窥现实版“弗兰肯斯坦”的疯狂世界！



《脑机穿越》

- ◎ 脑机接口研究先驱、巴西世界杯“机械战甲”发明者米格尔·尼科莱利斯扛鼎力作！
- ◎ 脑联网、记忆永生、意念操控……你最不可错过的未来之书！



-
- (1) 欧文·古德（Irving Good）提出的“智能爆炸”（intelligence explosion）是指，智能机器在无须人干预的情况下，能不断地设计下一代智能机器。——编者注
- (2) 为了简化起见，我假定这个故事的经济和科技背景与当下无异，但大多数研究者都认为，人类水平的通用人工智能至少还有几十年才能实现。假如数字经济持续增长，并且越来越多的服务无须讨价还价就可以从网上订购，那么，欧

米茄团队的计划在未来应该会比现在更容易实现。

- (3) MTurk是这样一种平台，用户既可以通过回答别人的调查挣些小钱，也可以自己发布问题获得结果，获得结果也是需要支付一些钱的。——编者注
- (4) 称为“我们的可观测宇宙”，或简称为“我们的宇宙”。
- (5) 量子涨落是指空间任意位置能量的暂时变化，根据德国著名物理学家沃纳·海森堡的不确定性原理，空无一物的空间能随机地产生少许能量，前提是该能量在短时间内重归消失；产生的能量越大，则该能量存在的时间就越短，反之亦然。——编者注
- (6) 生命为什么会变得越来越复杂？原来，足够复杂的生命会得到进化的犒赏，奖品是一种预测及利用环境规律的能力。因此，在越来越复杂的环境中，就会进化出越来越复杂和智能的生命。这个更加聪明的生命又会为竞争者创造出一个更加复杂的环境，而这些竞争者又因此而进化得更加复杂，最终创造出一个充满极端复杂生命的生态系统。
- (7) 强人工智能是指，在所有认知任务上都与人类差不多的人工智能。
- (8) “奇点”理论的支持者认为，当“奇点”到来之时，死人可以被复活（可能是以电子形式），意识可以被转移到机器中。——译者注
- (9) 关于人工智能的对话很重要，表现在两个方面：紧迫性和影响力。气候变化或许会在50~200年内肆虐四方，但与气候变化相比，许多专家预计，人工智能会在几十年内造成更大的影响，可能还会给我们带来缓解气候变化的科技。与战争、恐怖主义、失业、贫困、移民和社会公平等问题相比，人工智能的崛起会带来更大的整体性影响。事实上，我们会在这本书中探讨人工智能将如何影响以上所有问题，无论是从好的方面还是坏的方面。
- (10) 马文·明斯基（Marvin Minsky）是人工智能领域的先驱之一，其经典著作《情感机器》为我们描绘了一幅塑造未来机器的光明图景。本书已由湛庐文化策划，浙江人民出版社出版。——编者注
- (11) 1956年夏天，28岁的约翰·麦卡锡（John McCarthy）与同龄的马文·明斯基、37岁的纳撒尼尔·罗切斯特（Nathaniel Rochester）以及40岁的克劳德·香农（Claude Shannon）组织了一次人工智能的专题讨论会，这个会议就是历史上有名的达特茅斯夏季研讨会，被认定是人工智能研究的诞生之日。——编者注
- (12) 赛博格（Cyborg）是一种机械化有机体，又称改造人，指的是同时具有有机体和生物机电部分的人。现在，赛博格已经从想象中的概念变成了现实。——

编者注

- (13) 若想理解这一点，请想象一下，假如有人声称，奥运会运动员的运动能力可以用一个数字来量化，这个数字称为“运动商”（Athletic Quotient），简称“AQ”，而AQ最高的人可以直接获得所有运动项目的金牌。你会怎么想呢？
- (14) 有些人喜欢将“人类水平的人工智能”或“强人工智能”作为通用人工智能的同义词，但这是有问题的。从狭义的角度来说，一个计算器也可以算得上人类水平的人工智能。“强人工智能”的反义词应该是“弱人工智能”，但把深蓝计算机、沃森和AlphaGo这类“狭义人工智能”系统称为“弱人工智能”是一件令人觉得很古怪的事情。
- (15) 迈克斯·泰格马克生于1969年。——编者注
- (16) 感觉运动（sensorimotor）是指刺激作用于感觉神经而传至大脑，再由运动神经做出动作的活动。——编者注
- (17) 泰格马克在这里用“universal intelligence”与“universal computer”（通用计算机）进行类比，按理来说，应该将两个“universal”翻译成同一个词“通用”。但“通用智能”是另一个词组“general intelligence”的专有翻译，为避免混淆，我将“universal intelligence”翻译成“普遍智能”。——译者注
- (18) 机器码是指将硬件序列号经过一系列加密、散列形成的一串序列号。——编者注
- (19) 与非门（NAND gate）是数字电路的一种基本逻辑电路，其英文名称“NAND”是“NOTAND”的简称。一个与门（AND gate）只有当两个输入都为1时，才会输出1；因此，非与门的输出正好相反。
- (20) 在这里，“定义明确的函数”指的是数学家和计算机科学家所说的“可计算函数”（computable function），也就是某些假想计算机在内存和时间无限的情况下可以计算出来的函数。艾伦·图灵和阿朗佐·丘奇曾有个著名的证明，某些函数尽管可以被描述，但却是不可计算的。
- (21) 或非门是指具有多端输入和单端输出的门电路。当任一输入端（或多端）为高电平，也就是逻辑“1”时，输出就是低电平，也就是逻辑“0”；只有当所有输入端都是低电平，也就是逻辑“0”时，输出才是高电平，也就是逻辑“1”。——编者注
- (22) 元胞自动机（cellular automata）是20世纪50年代初，计算机之父冯·诺依曼为了模拟生命系统所具有的自动复制功能而提出的。

- (23) 这个趋势被称为“摩尔定律”（Moore's Law）。
- (24) 风洞实验是指在风洞中安置飞行器或其他物体模型，研究气体流动及其与模型的相互作用，以了解实际飞行器或其他物体的空气动力学特性的一种空气动力学实验。——编者注
- (25) 如果你喜欢数学，在这里我向你介绍两种最流行的激励函数，一种叫作Sigmoid函数，即 $\sigma(x) \equiv 1/(1+e^{-x})$ ；另一种叫作斜坡函数，即 $\sigma(x) = \max\{0, x\}$ ，不过，有人已经证明，几乎任何函数都能满足斜坡函数的条件，只要它不是线性（一条直线）就行。约翰·霍普菲尔德的著名模型使用的函数是：当 $x < 0$ 时， $\sigma(x) = -1$ ，当 $x \geq 0$ 时， $\sigma(x) = 1$ 。如果神经元的状态存储于一个向量内，那么，想要更新这个网络，就只需要给这个向量乘上一个代表突触连接强度的矩阵，然后在所有元素上应用激励函数就可以了。
- (26) 这些例子来自我学生亨利·林（Henry Lin）的论文，详见<http://arxiv.org/abs/1608.08225>。
- (27) 之所以称为“深度”，是因为它们包含的层数很多。
- (28) 逻辑门电路是指实现基本和常用逻辑运算的电子电路。在数字电路中，所谓“门”就是只能实现基本逻辑关系的电路。——编者注
- (29) GOF AI是指“有效的老式人工智能”（Good Old-Fashioned AI，简称GOF AI），这是人们对深度学习革命之前的人工智能的幽默昵称。
- (30) 如果你想要了解人工智能安全性研究领域的更多信息，这里有一个互动的版本，是由未来生命研究所的理查德·马拉（Richard Mallah）牵头完成的，详细信息请看<https://futureoflife.org/landscape>。
- (31) 更准确地说，“验证”看的是一个系统是否满足它的规范，而“确认”看的是我们是否选择了正确的规范。
- (32) 人们的统计数据发现，即便加上这次车祸，特斯拉的自动驾驶功能开启后，依然能降低40%的撞车事故。参见<http://tinyurl.com/teslasafety>。
- (33) 自动化武器系统（Autonomous Weapon Systems，简称AWS），有时也被反对者称为“杀手机器人”。
- (34) 密集阵系统是指美国海军为解决军舰近程防空问题专门设计制造的六管20毫米口径的自动旋转式火炮系统。——编者注
- (35) 底特律“三巨头”指通用、福特和克莱斯勒三家汽车公司，而硅谷“三巨头”指谷歌、苹果和Facebook三大互联网公司。——编者注

- (36) 心流 (mental flow) 在心理学中是指，某个人在专注进行某项行为时所表现的心理状态，心流产生的同时会有高度的兴奋及充实感。——编者注
- (37) 请回忆一下，FLOPS是指每秒钟可运行的浮点运算次数，比如，每秒钟可以计算多少个19位数的数字的乘积。
- (38) 大脑的硬件“神经形态芯片” (neuromorphic chips)，如今正在飞速发展。
- (39) 上传者 (Uploads) 也叫作仿真者 (Emulations)，简称Ems。
- (40) 尼克·波斯特洛姆曾解释说，如果一家人工智能公司能够在计算机中模拟出一位顶级的人类人工智能开发者，并且其成本远低于他的时薪，那么，这家公司的工作能力就能极大提升，将能积累巨大的财富，并能不断迭代和进步，建造出更好的计算机，最终制造出更加聪明的智能。
- (41) 关于“新天堂”，1960年的电视剧《阴阳魔界》 (Twilight Zone) 中的一集《值得拜访的好地方》 (A Nice Place to Visit) 中也有所体现。——编者注
- (42) 这个想法可以追溯到圣奥古斯丁 (Saint Augustine)，他写道：“如果某件东西在与他人分享时并不会减少半分，那么，只是占有它而不分享，就是不正确的。”
- (43) 我是从我的好友兼同事安东尼·阿奎尔那里第一次听说这个想法的。
- (44) 模因 (memes) 是指文化基因，其在诸如语言、观念、信仰、行为方式等的传递过程中的作用，与基因在生物进化过程中所起的作用类似。——编者注
- (45) 亚伯拉罕·马斯洛 (Abraham Maslow) 将人类的需求像阶梯一样从低到高分5个层次，分别是：生理需求、安全需求、社交需求、尊重需求和自我实现需求。——编者注
- (46) 此处智能的定义是它完成目标的能力，无论这个目标是什么。
- (47) 戴森球是弗里曼·戴森在1960年提出的一种理论。所谓“戴森球”就是直径2亿千米不等，用来包裹恒星、开采恒星能量的人造天体。这是一个利用恒星做动力源的天然的核聚变反应。——编者注
- (48) 著名宇宙学家弗雷德·霍伊尔 (Fred Hoyle) 在小说《以A开头的仙女座》 (A for Andromeda) 中也探讨了一个类似的情景，不过有些地方不一样。
- (49) 向大气中排放碳元素会引起两种气候变化：二氧化碳造成升温，烟尘造成降温。由于缺乏足够的科学证据，许多人常常不相信第一种变化，但并不是只有

第一种才被人们忽视。偶尔有人告诉我，“核冬天”的假面具已经被揭穿了，它实际上是不可能的。每当这时候，我就要求他们展示一篇提出如此强烈论点的同行评议科学论文。迄今为止，这样的论文似乎并不存在。尽管存在很大的不确定性，尤其是在烟尘数量和上升高度方面，但我的科学观点是，没有任何基础证据表明，核冬天的风险是不存在的。

(50) 由单层碳原子组成的六边形网状结构，就像铁丝网一样。

(51) 如果你在能源行业工作，可能会更习惯于将“能源效率”定义为以有用的形式释放的能量所占的比例。

(52) 如果在附近找不到一个自然生成的合适黑洞，我们可以把许多物质压缩到一个足够小的空间里，从而造出一个新的黑洞。

(53) 这么讲其实有点过于简化，因为霍金辐射还包括一些很难被有效利用的粒子。大型黑洞的效率只有90%，因为有10%的能量以引力子的形式被释放出来。引力子是一种非常“害羞”的粒子，几乎无法被探测到，更别说有效利用了。随着黑洞继续蒸发和收缩，效率会进一步降低，因为霍金辐射又开始包括中微子和其他大质量粒子。

(54) 对道格拉斯·亚当斯的粉丝来说，这个数字有着特殊的意义，因为42是“生命、宇宙以及万物的终极答案”。更精确地说，这里的能量效率等于 $1 - 1/\sqrt{3} \approx 42\%$ 。

(55) 如果你“喂”给黑洞燃料的方式是，在其周围放置一个围绕着黑洞并以相同方向缓慢旋转的气体云，那么，随着气体云被拉近和被吞噬，它的旋转速度会越来越快，从而促使黑洞旋转得更快，就像花样滑冰运动员在旋转时收回自己的手臂，从而转得更快一样。这或许可以让黑洞达到最快的旋转速度，使得你可以提取出气体能量的42%，然后提取出剩下部分的29%。这样，总体的能量效率等于 $42\% + (1 - 42\%) \times 29\% \approx 59\%$ 。

(56) 必须要达到足够高的温度，才能将电磁力和弱作用力重新统一起来。当粒子在粒子对撞机中以2000亿伏特加速时，就能达到这种状态。

(57) 前文中，我们只讨论了由原子组成的物质。实际上，还有6倍多的暗物质，但它难以捉摸，也很难捕获，通常会直接穿过地球，并从另一边毫发无伤地穿出来，所以，未来生命能不能捕捉和利用暗物质，我们只能拭目以待了。

(58) 这里的宇宙数学其实非常简单：如果这个文明在膨胀空间内不是以光速 c 而是以一个较慢的速度 v 前进，那么，它能殖民的星系数量就会减少，其减少系数为 $(v/c)^3$ 。这意味着，如果一个文明的动作迟缓，那它基本上就算瘫痪

了。因为，如果它扩张的速度慢10倍，那它最终殖民的星系数量就会少1000倍。

(59) 量子芝诺效应 (quantum zeno effect) 又称量子水壶效应 (watched-pot effect)，它指出，频繁地对一个不稳定粒子进行量子观测会抑制和阻止它的衰变。——编者注

(60) 在生物学中，速度的极限是由电子信号在神经元中传导的最快速度决定的。

(61) 约翰·格里宾 (John Gribbin) 在他2011年的著作《独在宇宙》 (Alone in the Universe) 一书中也得出了类似的观点。如果你想了解关于这个问题的其他观点，我推荐保罗·戴维斯 (Paul Davies) 2011年的著作《可怕的寂静》 (The Eerie Silence)。

(62) 我们的宇宙是指，大爆炸以来的138亿年中，发出的光线足以到达我们的区域。

(63) 《穿越平行宇宙》是本书作者所著的关于平行宇宙的科普著作，探索了宇宙终极本质的神秘旅程。中文简体字版已由湛庐文化策划，浙江人民出版社出版。——编者注

(64) 核糖体是一种高度复杂的分子机器，它能够读取我们的基因密码，并生产我们所需的蛋白质。

(65) 这个想法被美国天文学家约翰·波尔 (John A. Ball) 称为“动物园假说” (Zoo Hypothesis)，并在一些经典科幻作品中有过详细的描写，例如奥拉夫·斯塔普雷顿所著的《造星者》。

(66) 许多昆虫保持直线飞行的经验法则是假设亮光就是太阳，于是与亮光成一定角度飞行。如果这个亮光实际上是近处的一团火，那昆虫也会按一定角度飞过去，跳出一支“死亡旋舞”。

(67) 安东尼奥·达马西奥是美国南加州大学神经科学、心理学和哲学教授，他以情绪为出发点，从演化的角度重新阐释了人类意识的产生路径。这方面的代表作《笛卡尔的错误》《当自我来敲门》中文简体字版已由湛庐文化策划，北京联合出版公司出版。——编者注

(68) 这里我所说的“改进软件”是最广义的改进，不仅包括对算法的优化，还包括让决策制定过程更加理性的改进。这样，人工智能就可以尽可能好地保持自己的目标。

(69) 史蒂芬·平克，美国著名认知心理学家和科普作家，因广泛宣传演化心理学和心智计算理论的心态而闻名于世。其代表作“语言与人性”四部曲《语言本

能》《思想本质》《心智探奇》《白板》中文简体字版已由湛庐文化策划，浙江人民出版社出版。——编者注

(70) 还有一种观点叫作实体二元论 (substance dualism)。这种观点认为，生物之所以与非生物不同，是因为它们包含一些非物质的东西，如“灵气” (anima)、“生命冲力” (élan vital) 或“灵魂” (soul)。如今支持实体二元论的科学家越来越少。要理解为什么，请想一想，你的身体是由大约 10^{29} 个夸克和电子组成的，据我们所知，它们的移动都遵从简单的物理定律。请想象一下，假设未来有一项技术能够追踪所有的粒子。如果它发现你的粒子完全服从物理定律，那所谓的“灵魂”对你的粒子就没有任何影响，那么，你的意识心智和运动控制能力就与灵魂没有任何关系。如果这项技术发现你的粒子不遵守任何已知的物理定律，而是受你的灵魂所驱动，那么，带来这些驱动力的未知事物从本质上来说，一定也是一个物理存在的实体，我们可以像过去研究新领域和新粒子一样对它进行研究。

(71) 我在这里使用“感质”这个词是根据字典的定义，即主观体验的单个实例，也就是说，主观体验本身，而不是任何引起体验的物质。请注意，不同的人提到这个词时的意思可能不一样。

(72) 我原本把RHP称为“非常困难的问题”，但大卫·查尔默斯读了这一章之后，他给我提了一个巧妙的建议，让我改成“真难的问题”，以配合他的本意：“由于前两个问题（至少按现在的顺序）并不属于我设想的难题，而第三个问题确实属于，所以，你可以用‘真难’而不是‘非常难’来匹配第三个问题，以符合我的本意。”

(73) 假如我在《穿越平行宇宙》一书中所说的理论是真实的，我们的物理实在是全是数学，简单来说就是以信息为基础，那么，现实世界中就没有任何东西是超出科学范畴的，甚至连意识也不是。事实上，从这个角度来看，“真难的问题”就变成了下面这个问题：“数学的东西为什么感觉起来具有物理性质？”如果数学结构的一部分拥有了意识，那它就能体验到外部物理世界中的其他部分。

(74) 系统2是指需要花费精力的、缓慢而可控的思考方式。

(75) 我之前用“感质”作为“意识质”的同义词，但感质这个词的概念太窄了，因为感质仅仅指的是那些基于感觉器官的主观体验，而排除了一些不基于感觉器官的意识，比如梦境和心里产生的想法。

(76) 这个论断与“意识是独立于物质层面而存在的”这一思想之间有些冲突，因为，即使最低层级（物质层面）上的信息处理过程可能不一样，从理论上来

说，决定行为的高层级上的信息处理应该是完全相同的。

[\(77\)](#) 这段经历让我开始重新思考应当如何解读新闻。虽然我早就知道大部分报道都怀有政治目的，但我现在才意识到，它们不仅在政治上如此，而且在所有问题上（包括与政治无关的话题）都带有偏见。