"A fascinating, well-written, and important book." —YUVAL NOAH HARARI

TECHNOLOGY, POWER, AND THE 21ST CENTURY'S GREATEST DILEMMA

THE COMING WAVE

MUSTAFA SULEYMAN

CO-FOUNDER OF DEEPMIND AND INFLECTION AI

WITH MICHAEL BHASKAR

I'm sorry, but it seems that the source text you provided is em

即将到来的浪潮

源文本: 翻译文本:

“*The Coming Wave*”是一本引人入胜、写得很好且重要的书。它探讨了人工智能和生物技术对人类构成的生存威胁，并提供了如何遏制这一威胁的实际解决方案。即将到来的技术浪潮有望赋予人类神一般的创造力，但如果我们未能明智地管理它，它可能会毁灭我们。

—尤瓦尔·诺亚·赫拉利，*New York Times* 畅销书作者 *Sapiens*

“这本来自未来的警钟警告了即将到来的事情，以及可能的全球经济和政治影响。真正非凡、雄心勃勃且无法忽视，这本书是由一位行业领先专家提出的令人信服的力作，将塑造你对未来的看法，并重塑你对现在的理解。”

—纽约大学名誉教授努里尔·鲁比尼

“Mustafa Suleyman 作为技术专家、企业家和远见者的洞察力至关重要。这本书经过深入研究，与时俱进，为我们这个时代一些最重要的挑战提供了引人入胜的见解。”

—美国前副总统阿尔·戈尔

“在这本大胆的书中，Mustafa Suleyman，科技界真正的内部人士之一，探讨了我们这个时代最重要的悖论：我们必须控制无法控制的技术。正如他所解释的，生成式人工智能、合成生物学、机器人技术和其他创新正在改进和传播。

迅速地。它们带来了巨大的好处，但也带来了真实且日益增长的风险。Suleyman足够聪明，知道没有简单的三点计划可以管理这些风险，并且足够勇敢地告诉我们这一点。这本书诚实、充满激情，并且无畏地面对显然是我们这个世纪将面临的重大挑战之一。多亏了Suleyman，我们知道了情况如何以及我们的选择是什么。现在就看我们如何行动了。”

—麻省理工学院斯隆管理学院首席研究科学家，
*The Geek Way*的作者安德鲁·麦卡菲

“AI革命正在进行中，但我们对它的理解到底有多深刻？*The Coming Wave* 提供了一本博学且清晰的指南，既涉及激进技术变革的历史，也涉及未来深刻的政治挑战。”

——安妮·阿普尔鲍姆，普利策奖获奖历史学家

“当这本书出现在我的收件箱时，我清空了日程表开始阅读。这是一本非凡且必要的书；令人敬畏的想法是，在*twenty years*中，它几乎会像是对未来的保守愿景，而现在，阅读它时，不得不每隔几页停下来想：这是真的吗？这本书的天才之处在于，它冷静而温和地解释了，是的，这一切都会成真——以及为什么和如何。语气温和且善解人意，体贴读者的震惊感。书中有令人恐惧的时刻，因为当人们意识到大多数熟悉的事物即将被改变时，应该如此。但最终，人们会感到充满活力，并为现在的生活感到兴奋。浪潮即将来临，而这就是预报。”

——阿兰·德波顿，哲学家和畅销书作家

“*The Coming Wave* 提供了急需的具体性、现实性和清晰性，关于人工智能、合成生物学和其他先进技术可能带来的意外且灾难性的后果。这本重要的书是一本生动且有说服力的路线图，展示了人类如何引导技术创新，而不是被其控制。”

—Martha Minow, Harvard professor, former dean of Harvard Law School

“没有人比穆斯塔法·苏莱曼更接近正在展开的人工智能革命，也没有人比他更适合概述当前巨大技术变革的风险和回报。这是一本关于人类历史上这一独特时刻的非凡且绝对不容错过的指南。”

——埃里克·施密特，谷歌前首席执行官，*The Age* 的合著者 *of AI*

“在 *The Coming Wave,* 中，Mustafa Suleyman 提出了一个有力的论点，即当今爆炸性的技术革命正处于独特的颠覆性边缘。阅读这本必读书籍，以了解这些技术的速度和规模——它们将如何在我们的社会中迅速扩散，以及它们挑战组织我们世界的机构结构的潜力。”

—伊恩·布雷默，欧亚集团创始人，畅销书作者 *The Power of Crisis*

“这本重要的书同时令人振奋和恐惧。对于那些不理解我们正在经历的技术革命的人来说，这是一个关键的教育，而对于那些理解的人来说，这是一个正面的挑战。这本书关乎我们所有人的未来：我们需要阅读它并采取行动。”

—大卫·米利班德，前英国外交大臣

“穆斯塔法·苏莱曼对人工智能的危险和奇迹进行了严峻评估，提出了一项紧急行动议程，政府必须立即采取行动，以限制这一革命性挑战中最可能带来灾难性后果的应用。”

—哈佛大学教授、畅销书作家格雷厄姆·艾利森 *Destined for War*

“指数技术的快速发展以其力量和危险让我们感到不知所措。穆斯塔法·苏莱曼在追溯工业发展的历史到最近技术的令人眼花缭乱的加速过程中，

进步，以冷静、务实且深具伦理的散文为我们提供了更大的视野。他的个人旅程和经历增强了*The Coming Wave*，使其成为所有想要从每日的科技新闻浪潮中抽身的人都能享受的引人入胜的读物。”

—安吉拉·凯恩，前联合国副秘书长兼裁军事务高级代表

“从终极内部人士的视角，深入了解当前AI的发展及其指数级的未来……如果你真的想了解社会如何安全地驾驭这一改变世界的技术，请阅读这本书。”

—布鲁斯·施奈尔，网络安全专家，*A Hacker's* 的作者 *Mind*

“即将到来的人工智能和合成生物学浪潮将使下一个十年成为人类历史上最好的十年。或者是最糟糕的。没有人比穆斯塔法·苏莱曼更能识别和解释未来的史诗般挑战。这本书发人深省、紧迫，并以强大且高度易懂的散文写成，是任何对了解这些技术的惊人力量感兴趣的人必读的书。”

—埃里克·布林约尔松，斯坦福大学以人为本人工智能教授

“世界面临的最大挑战之一是设计出能够利用人工智能和生物技术的好处，同时避免其灾难性风险的治理形式。本书深入思考了这两种技术的‘遏制挑战’。它经过精心研究，充满了原创见解，并为政策制定者和安全专家提供了建设性的建议。”

—杰森·马西尼，RAND公司首席执行官，前国家情报助理主任，前IARPA主任

“如果你想理解即将到来的变革性技术浪潮的意义、承诺和威胁，这些技术甚至现在正在膨胀并

在主流上汇聚，然后这本由人工智能关键先驱之一穆斯塔法·苏莱曼撰写的深具回报且始终令人惊叹的书绝对是必读之作。”

—斯蒂芬·弗莱，演员、主持人及畅销书作家

“这本重要的书是一个生动的警钟。它仔细地概述了近年来令人振奋的科学进步所带来的威胁和机遇。*The Coming Wave* 富含有趣的事实、引人注目的论点和引人入胜的观察；这是必读之作。”

—丹尼尔·卡尼曼，诺贝尔奖得主，畅销书作者
Thinking Fast and Slow

“*The Coming Wave* 是一本极其清晰、充满活力、研究透彻且易读的书，来自我们时代最伟大技术革命的前线。它将个人和技术故事无缝结合，并展示了为什么对极其强大的技术进行更好的治理既如此重要又如此困难。”

——伦敦大学学院教授杰夫·穆尔根爵士

“迄今为止对人工智能对人类未来意义的最佳分析……穆斯塔法·苏莱曼作为两家主要当代人工智能公司的联合创始人是独一无二的。他是一位极具天赋的企业家，一个深刻的思想家，也是即将到来的将塑造我们世界的技术浪潮中最重要的声音之一。”

—Reid Hoffman，LinkedIn 和 Inflection 的联合创始人

“技术正在迅速改变社会，因此看到技术行业中的人以如此诚实和严谨的态度写作比以往任何时候都更为重要。从最早的工具到当前人工智能能力和研究爆炸的核心，这本书是全景式的调查，也是一个不容忽视的行动号召。每个人都应该阅读它。”

——李飞飞，斯坦福大学计算机科学教授，人类中心人工智能研究所联合主任

“ *The Coming Wave* 提出了一个令人耳目一新且令人信服的论点，即先进技术正在重塑社会的各个方面：权力、财富、战争、工作，甚至人际关系。我们能否在这些新技术控制我们之前控制它们？作为人工智能领域的世界领袖以及长期倡导政府、大型科技公司和公民社会为共同利益采取行动的Mustafa Suleyman，是解答这一关键问题的理想向导。 ”

—杰弗里·D·萨克斯，哥伦比亚大学大学教授，联合国可持续发展解决方案网络主席

“ 一个尖锐、富有同情心且毫不妥协地框定我们时代最重要问题的作品，*The Coming Wave* 是技术从业者必读之作，但更重要的是，它坚定地呼吁我们所有人参与这一最重要的讨论。 ”

—奇鲁，MiraclePlus首席执行官，前百度首席运营官，前执行副总裁
微软必应

“ 苏莱曼在表达人工智能和合成生物学不受限制的发展可能带来的严重后果——地缘政治动荡、战争、国家主权的侵蚀——方面具有独特的优势，而在我们最需要这个信息的时候。幸运的是，对于读者来说，他也深入思考了需要做些什么来确保新兴技术用于人类的利益，提出了一系列逐步的努力，如果共同实施，可以改变这些技术开发和传播的环境，为维护那个更光明的未来打开大门。这本书是必读之作。 ”

—梅根·L·奥沙利文，哈佛大学肯尼迪政府学院贝尔弗科学与国际事务中心主任

“ 一个勇敢的警钟，我们都需要在为时已晚之前作出回应……Mustafa Suleyman 以清晰和精确的方式解释了所带来的风险

失控技术和人类面临的挑战……不可或缺的读物。”

—Tristan Harris，联合创始人兼执行董事
人性科技中心

“关于我们这个时代最重要问题的实用且乐观的行动路线图：如何保留对比我们自身更强大实体的控制权。”

——加州大学伯克利分校计算机科学教授斯图尔特·拉塞尔

“*The Coming Wave* 是一幅现实的、深刻知情的且高度易懂的地图，展示了人工智能和合成生物学带来的前所未有的治理和国家安全挑战。Suleyman 的这本非凡且在某种意义上令人恐惧的书展示了必须采取的措施来遏制这些看似无法遏制的技术。”

—哈佛大学法学院的杰克·戈德史密斯，伦德·汉德法学教授
大学

“精彩而吸引人，复杂而清晰，紧迫而平静，*The Coming Wave* 引导我们所有人去理解和面对可能是我们这个世纪最关键的问题：我们如何确保即将到来的令人惊叹、快速发展的技术革命——人工智能、合成生物学等——创造出我们想要的世界？这并不容易，但 Suleyman 奠定了坚实的基础。关心未来的每个人都应该读这本书。”

—埃里克·兰德，麻省理工学院和哈佛大学布罗德研究所
创始主任，前白宫科学顾问

“*The Coming Wave* 清晰而平衡地描述了我们当前的技术困境，阐明了我们这个时代的定义性挑战。它将实用主义与谦逊相结合，提醒我们没有明显的二元对立或简单的答案：技术为我们带来了指数级的福祉改善，但它的加速速度比

机构可以适应。人工智能和合成生物学的进步解锁了科幻小说未曾想象的能力，而由此产生的力量扩散威胁着我们所建造的一切。为了保持生存，我们必须在触手可及的灾难的斯库拉和无处不在的监视的卡律布狄斯之间航行。每翻过一页，我们的几率就会提高。”

—凯文·埃斯维尔特，生物学家，麻省理工学院媒体实验室副教授

实验室

THE COMING WAVE

TECHNOLOGY, POWER, AND THE TWENTY-FIRST CENTURY'S GREATEST DILEMMA

MUSTAFA SULEYMAN

WITH MICHAEL BHASKAR

CROWN
NEW YORK

版权所有 © 2023 Mustafa Suleyman 和 Michael Bhaskar 保留所有权利。
由皇冠出版集团的皇冠印记在美国出版，隶属于企鹅兰登书屋有限责任公司，纽约。C 和皇冠标志是企鹅兰登书屋有限责任公司的注册商标。精装书 ISBN　9780593593950 国际版 ISBN　9780593728178 电子书 ISBN　9780593593967 crownpublishing.com *Book design by Barbara M. Bachman, adapted for ebook* 封面设计：Christopher Brand 和 Oliver Munday

ep_prh_6.1_144835715_c0_r0

目录

Cover
Title Page
Copyright
Glossary of Key Terms

序章 第1章：无法遏制

第I部分：*Homo Technologicus* 第2章：无尽的扩散 第3章：遏制问题

P第二部分：下一波浪潮第4章：智能技术第5章：生命技术第6章：更广泛的浪潮第7章：即将到来的浪潮的四个特征第8章：不可阻挡的激励

P第III部分：失败状态 第9章：大交易 第10章：脆弱性放大器 第11章：国家的未来 第12章：困境

第四部分：穿越波浪

Chapter 13: Containment Must Be Possible
Chapter 14: Ten Steps Toward Containment

人类世之后的生活

Acknowledgments
Notes
Index
About the Authors

术语表

AI、AGI 和 ACI：人工智能（AI）是教授机器学习类人能力的科学。人工通用智能（AGI）是指当人工智能能够比最聪明的人类更好地执行所有人类认知技能的时刻。ACI，即人工能力智能，是介于 AI 和 AGI 之间的一个快速接近的点：ACI 可以完成广泛的复杂任务，但距离完全通用仍有很长的路要走。

即将到来的浪潮：一个以人工智能和合成生物学为中心的新兴相关技术集群，其变革性应用将同时赋予人类力量并带来前所未有的风险。

遏制：监控、限制、控制甚至关闭技术的能力。

遏制问题：技术倾向于以波浪形式广泛扩散，并产生无法预测或控制的新兴影响，包括负面和意想不到的后果。

困境：新技术及其缺失可能导致灾难性和/或反乌托邦结果的可能性日益增加。
四个特征：即将到来的浪潮的独特特征，加剧了遏制的挑战。它们是非对称性、超进化、全用途和自主性。

脆弱性放大器：即将到来的技术浪潮的应用及其对已经脆弱的民族国家基础的影响。

大交易：作为对使用武力垄断权的交换，公民期望民族国家维护秩序并提供公共服务。

服务，包括利用新技术，同时尽量减少有害的副作用。

THE NARROW PATH: 人类在即将到来的技术浪潮中，在开放性和封闭性之间取得平衡的潜力，以避免灾难性或反乌托邦的结果。

PESSIMISM AVERSION: 人们，特别是精英，倾向于忽视、淡化或拒绝他们认为过于消极的叙述。这是一种乐观偏见的变体，它影响了关于未来的许多辩论，尤其是在技术圈。**SYNTHETIC BIOLOGY:** 设计和工程新生物或重新设计现有生物系统的能力。**TECHNOLOGY:** 应用科学知识（以最广泛的意义）来生产工具或实际成果。**WAVES:** 一代技术的全球扩散或传播，这些技术以一种新的通用技术为基础。

序幕

源文本: T翻译文本:

看到它

Q: 即将到来的技术浪潮对人类意味着什么？

在人类历史的编年史中，有些时刻作为转折点而突出，在这些时刻，人类的命运悬而未决。火的发现、轮子的发明、电力的利用——所有这些都是改变人类文明的时刻，永远改变了历史的进程。

现在，我们正处于另一个这样的时刻的边缘，因为我们面临着包括先进人工智能和生物技术在内的新一波技术的兴起。我们从未见过具有如此变革潜力的技术，承诺以既令人敬畏又令人生畏的方式重塑我们的世界。

一方面，这些技术的潜在好处是巨大而深远的。通过人工智能，我们可以揭开宇宙的秘密，治愈长期困扰我们的疾病，并创造出超越想象界限的新艺术和文化形式。通过生物技术，我们可以设计生命来应对疾病并改造农业，创造一个更健康、更可持续的世界。

但是，另一方面，这些技术的潜在危险同样巨大而深远。通过人工智能，我们可能会创建超出我们控制的系统，并发现自己受制于我们不理解的算法。通过生物技术，我们可能会操控生命的基本构成部分，可能为个人和整个生态系统带来意想不到的后果。

在这个转折点上，我们面临着一个选择——一个在无与伦比的可能性和难以想象的危险之间的选择。人类的命运悬而未决，我们在未来几年和几十年中所做的决定将决定我们是迎接这些技术的挑战，还是成为其危险的受害者。

但在这个不确定的时刻，有一件事是确定的：先进技术的时代已经来临，我们必须做好准备，直面其挑战。

源文本：—翻译文本：

由 AI 创造。其余部分尚未如此，尽管很快可能会是。这就是即将到来的。

AI

CHAPTER 1

无法进行CONT　NMENT

波浪

A

在古印度文本中，我们宇宙中的第一个人，摩奴，被警告即将到来的洪水，并成为其唯一的幸存者。*Epic of Gilgamesh* 记录了神恩利尔在一场巨大的洪水中毁灭世界的故事，这个故事会让任何熟悉旧约诺亚方舟故事的人产生共鸣。柏拉图谈到了失落的亚特兰蒂斯城，被巨大的洪流冲走。贯穿人类口头传统和古代著作的是一个巨浪席卷一切的想法，使世界重塑和重生。

洪水在字面意义上也标志着历史——世界大河的季节性泛滥，冰河时代结束后海洋的上升，海啸在地平线上毫无预警地出现的罕见冲击。杀死恐龙的小行星创造了一个高达一英里的巨浪，改变了进化的进程。这些波涛的巨大力量已经深深烙印在我们的集体意识中：水墙，不可阻挡，不可控制，不可遏制。这些是地球上最强大的力量之一。它们塑造了大陆，灌溉了世界的农作物，并滋养了文明的成长。

其他类型的浪潮同样具有变革性。再看看历史，你会发现它被一系列隐喻性的浪潮所标记：帝国和宗教的兴衰，以及商业的爆发。想想基督教或伊斯兰教，这些宗教最初是小波纹，然后在地球的广阔区域上建立并崩溃。这样的浪潮是反复出现的。

主题，构成历史的潮起潮落、大国斗争以及经济繁荣与萧条的框架。

技术的兴起和传播也呈现出改变世界的浪潮形式。自从人类发现火和石器以来，一种单一的主导趋势经受住了时间的考验，这是我们物种最早掌握的技术。几乎每一种曾经发明的基础技术，从镐到犁，从陶器到摄影，从电话到飞机，以及介于两者之间的一切，都遵循着一条看似不可改变的法则：它变得更便宜、更易于使用，最终广泛传播。

这种技术的浪潮式扩散是*Homo technologicus*的故事——是技术动物的故事。人类为了改善——我们自己、我们的命运、我们的能力以及我们对环境的影响——而进行的探索推动了思想和创造的不断演变。发明是一个展开的、广泛的、涌现的过程，由自组织且高度竞争的发明家、学者、企业家和领导者推动，每个人都带着自己的动机向前推进。这个发明生态系统默认是扩张的。这是技术的内在本质。

问题是，从这里开始会发生什么？在接下来的页面中，我将告诉你历史上下一个伟大浪潮的故事。

源文本：—翻译文本：

源文本：　.

你看到了什么？家具？建筑物？手机？食物？一个景观公园？几乎你视线中的每一个物体都很可能是由人类智慧创造或改变的。语言——我们社会互动、文化、政治组织的基础，或许也是人类意义所在——是我们智慧的另一种产物和推动力。每一个原则和抽象概念，每一个小的创造性努力或项目，你生活中的每一次遭遇，都是通过我们物种独特且无比复杂的想象力、创造力和理性能力来调解的。人类的创造力是一个令人惊叹的事物。

在这幅图景中，只有另一种力量是如此无处不在：生物生命本身。在现代之前，除了少数岩石和矿物，大多数人类制品——从木屋到棉衣再到煤火——都来自曾经活着的东西。从那时起进入世界的一切都源于我们，源于我们是生物体这一事实。

毫不夸张地说，人类世界的全部依赖于生命系统或我们的智慧。然而，这两者现在正处于前所未有的指数级创新和动荡时刻，这种无与伦比的增强将几乎改变一切。在我们周围开始崩溃的是一股新的技术浪潮。这股浪潮正在释放出工程这两个普遍基础的力量：一股不亚于智慧和生命的浪潮。

即将到来的浪潮由两项核心技术定义：人工智能（AI）和合成生物学。它们将共同为人类带来新的曙光，创造前所未有的财富和盈余。然而，它们的快速扩散也威胁着赋予各种不良行为者以释放破坏、不稳定，甚至是难以想象的灾难的能力。这股浪潮带来了一个巨大的挑战，将定义二十一世纪：我们的未来既依赖于这些技术，也受到它们的威胁。

从我们今天所处的位置来看，似乎遏制这一浪潮——即控制、抑制甚至阻止它——是不可能的。这本书探讨了为什么这可能是真的，以及如果确实如此，这意味着什么。这些问题的影响最终将影响到所有在世的人和我们之后的每一代人。

我相信即将到来的这波技术浪潮正在将人类历史推向一个转折点。如果无法遏制，其对我们物种的影响将是巨大的，甚至可能是可怕的。同样，没有它的成果，我们将面临暴露和不稳定的局面。这是我在过去十年里多次在闭门会议上提出的论点，但随着影响变得越来越不可忽视，现在是时候公开提出这个观点了。

THE DILEMMA

源文本: 翻译文本:

思考人类智慧的深刻力量让我提出了一个简单的问题，一个自此以来一直困扰我生活的问题：如果我们能够将使我们人类如此高效和有能力的本质提炼成软件，变成一个算法，会怎样？找到答案可能会解锁难以想象的强大工具，帮助解决我们最棘手的问题。这里可能有一个工具，一个不可能但非凡的工具，帮助我们度过未来几十年的巨大挑战，从气候变化到人口老龄化再到可持续食品。

考虑到这一点，在一个俯瞰伦敦罗素广场的古雅的摄政时代办公室里，我与两位朋友 Demis Hassabis 和 Shane Legg 于 2010 年夏天共同创立了一家公司，名为 DeepMind。这是我们的目标，一个在回顾时仍然感觉像当时一样雄心勃勃、疯狂和充满希望的目标：复制使我们作为一个物种独特的东西——我们的智能。

为了实现这一目标，我们需要创建一个能够模仿并最终超越所有人类认知能力的系统，从视觉和语音到规划和想象，最终到同理心和创造力。由于这样的系统将受益于超级计算机的大规模并行处理以及来自开放网络的庞大新数据源的爆炸式增长，我们知道，即使在这一目标上取得适度的进展，也会对社会产生深远的影响。

当时确实感觉相当超前。那时，人工智能的广泛应用还是白日梦的产物，更像是幻想而非现实，只是少数隐居学者和狂热科幻迷的领域。然而，当我写下这些文字并回顾过去十年时，人工智能的进步可谓惊人。DeepMind 成为世界领先的 AI 公司之一，取得了一系列突破。即使是我们这些最接近前沿的人，也对这场新革命的速度和力量感到惊讶。在撰写这本书的过程中，AI 的进步速度令人叹为观止，新模型和新

每周都有产品推出，有时甚至每天都有。显然，这股浪潮正在加速。

如今，AI系统几乎可以完美地识别面部和物体。我们理所当然地认为语音转文本转录和即时语言翻译是理所当然的。AI能够很好地导航道路和交通，在某些环境中可以自主驾驶。基于一些简单的提示，新一代AI模型可以生成新颖的图像，并以非凡的细节和连贯性撰写文本。AI系统可以生成具有惊人真实感的合成语音，并创作出令人惊叹的美丽音乐。即使在更具挑战性的领域中，那些长期被认为是人类能力独有的领域，如长期规划、想象力和复杂思想的模拟，进展也在飞速前进。

几十年来，人工智能一直在攀登认知能力的阶梯，现在看来将在未来三年内在非常广泛的任务中达到人类水平的表现。这是一个很大的主张，但即使我接近正确，其影响也是非常深远的。当我们创立DeepMind时，曾经感觉像是堂吉诃德式的事情，现在不仅变得可行，而且似乎不可避免。

从一开始，我就清楚地知道，人工智能将是一个极其强大的工具，可以带来非凡的好处，但像大多数形式的权力一样，它也充满了巨大的危险和伦理困境。我长期以来一直担心的不仅仅是人工智能发展的后果，还有整个技术生态系统的发展方向。除了人工智能之外，一场更广泛的革命正在进行中，人工智能正在推动一代强大的新兴基因技术和机器人技术的发展。在一个混乱且相互催化的过程中，一个领域的进一步进展加速了其他领域的发展，超出了任何人的直接控制。很明显，如果我们或其他人在复制人类智能方面取得成功，这不仅仅是像往常一样的盈利业务，而是人类的一个重大转变，开启了一个前所未有的机遇与前所未有的风险并存的时代。

随着这些年来技术的进步，我的担忧也在增加。如果这波浪潮实际上是一场海啸怎么办？

源文本：—翻译文本：

2010年，没有人认真谈论人工智能。然而，曾经看似小众的研究人员和企业家的任务现在已成为一项庞大的全球事业。人工智能无处不在，出现在新闻中和你的智能手机中，进行股票交易和网站建设。世界上许多最大的公司和最富有的国家正在大力推进，开发尖端的人工智能模型和基因工程技术，这一切都由数百亿美元的投资推动。

一旦成熟，这些新兴技术将迅速传播，变得更便宜、更易获得，并在整个社会广泛普及。它们将提供非凡的新医疗进展和清洁能源突破，不仅创造新业务，还将创造新产业，并在几乎所有可以想象的领域改善生活质量。

然而，与这些好处并存的是，人工智能、合成生物学和其他先进形式的技术在一个令人深感担忧的规模上产生尾部风险。它们可能对国家构成生存威胁——这些风险如此深远，以至于可能扰乱甚至颠覆当前的地缘政治秩序。它们开启了通往巨大人工智能赋能的网络攻击、可能摧毁国家的自动化战争、人工制造的流行病以及一个受无法解释但似乎无所不能的力量支配的世界的途径。每种情况的可能性可能很小，但可能的后果却是巨大的。即使是这些结果的微小可能性也需要紧急关注。

一些国家将以一种技术驱动的威权主义形式对这种灾难性风险的可能性作出反应，以减缓这些新力量的传播。这将需要巨大的监控水平以及对我们私人生活的大规模侵入。对技术保持严格控制可能会成为一种趋势，即所有事物和每个人在任何时候都受到监视，在一个反乌托邦的全球监控系统中，这种行为被合理化为防范最极端可能结果的愿望。

同样可能的是卢德派的反应。禁令、抵制和暂停将随之而来。是否有可能停止开发新技术并引入一系列暂停？不太可能。

由于其巨大的地缘战略和商业价值，很难想象国家或公司会被说服单方面放弃这些突破所释放的变革力量。此外，试图禁止新技术的发展本身就是一种风险：技术停滞的社会在历史上是不稳定的，容易崩溃。最终，它们失去了解决问题和进步的能力。

无论是追求还是不追求新技术，从这里开始都充满了风险。随着技术变得更便宜、更强大、更普及，风险不断累积，通过“狭窄路径”混过去并避免两种结果——一方面是技术威权主义的反乌托邦，另一方面是开放引发的灾难——的机会随着时间的推移而变得越来越小。然而，退缩也不是一个选项。即使我们担心它们的风险，我们比以往任何时候都更需要即将到来的技术浪潮所带来的巨大好处。这是核心困境：迟早，一代强大的技术将人类引向灾难性或反乌托邦的结果。我相信这是二十一世纪的重大元问题。

这本书详细说明了为什么这种可怕的困境正在变得不可避免，并探讨了我们如何可能面对它。不知何故，我们需要充分利用技术，这对于应对一系列令人生畏的全球挑战至关重要，同时也要摆脱困境。目前围绕技术伦理和安全的讨论是不足的。尽管有许多关于技术的书籍、辩论、博客文章和推文风暴，但你很少听到关于*containing*的任何信息。我将其视为一套相互关联的技术、社会和法律机制，在每个可能的层面上限制和控制技术：理论上是一种逃避困境的方法。然而，即使是技术最严厉的批评者也倾向于回避这种严格控制的语言。

这需要改变；我希望这本书能说明原因，并暗示如何改变。

THE TRAP

源文本: 翻译文本:

在我们创立DeepMind几年后，我制作了一份关于人工智能长期经济和社会影响的幻灯片。在西海岸一个时尚的会议室里，我向十几位科技行业最有影响力的创始人、CEO和技术专家展示，论证人工智能带来了需要积极应对的一系列威胁。它可能导致大规模的隐私入侵或引发信息误导的灾难。它可能被武器化，创造出一套致命的新型网络武器，为我们的网络世界引入新的漏洞。

我还强调了人工智能可能导致大量人失业的潜力。我请在座的人考虑自动化和机械化在历史上取代劳动力的长久历程。首先是更高效的完成特定任务的方法，然后整个角色变得多余，很快整个行业需要的工人数量就大幅减少。在接下来的几十年里，我认为，人工智能系统将以类似的方式取代“智力体力劳动”，而且肯定会远早于机器人取代体力劳动。在过去，新工作在旧工作被淘汰的同时被创造出来，但如果人工智能也能简单地完成大多数这些工作呢？我提出，几乎没有先例可以应对即将到来的新形式的集中权力。即使它们看起来很遥远，潜在的严重威胁也正在向社会逼近。

在最后的幻灯片中，我展示了来自 *The Simpsons*. 的一个静止画面。在场景中，斯普林菲尔德的市民们已经起来了，熟悉的角色们手持棍棒和火炬向前冲去。信息很明确，但我还是把它说出来了。“干草叉来了，”我说。是来找我们的，技术的制造者。我们有责任确保未来比这更好。

在桌子周围，我遇到了茫然的目光。房间里的人无动于衷。信息没有传达。反驳接踵而至。为什么经济指标没有显示出我所说的任何迹象？人工智能将刺激新的需求，从而创造新的工作岗位。它将增强和赋予人们更高的生产力。也许确实存在一些风险，他们承认，但并不算太糟。人们很聪明。解决方案已经

一直被发现。*No worries,* 他们似乎认为，*on to the next presentation.*

几年后，在 COVID-19 大流行开始前不久，我参加了在一所知名大学举办的关于技术风险的研讨会。布置类似：又是一张大桌子，又是一场高谈阔论的讨论。一天中，一系列令人毛骨悚然的风险在咖啡、饼干和 PowerPoint 演示中被提出。

有一个引人注目。演示者展示了DNA合成器的价格如何迅速下降，这些合成器可以打印定制的DNA链。它们的价格仅需几万美元，小到可以放在你车库的工作台上，让人们合成——即*manufacture*——DNA。而现在，所有这些对于任何拥有生物学研究生水平培训或对在线自主学习充满热情的人来说都是可能的。

鉴于工具的日益普及，演讲者描绘了一个令人不安的愿景：有人可能很快就能创造出比自然界中任何已知病原体更具传播性和致命性的全新病原体。这些合成病原体可能会逃避已知的对策，无症状传播，或对治疗具有内在的抗性。如果需要，有人可以通过在线订购的DNA来补充自制实验，并在家中重新组装。末日，通过邮件订购。

这不是科幻小说，主持人辩称，这位受人尊敬的教授有着超过二十年的经验；这现在是一个现实的风险。他们以一个令人震惊的想法结束：今天一个人很可能“有能力杀死十亿人。”所需的只是动机。

与会者不安地挪动着。人们扭动着身体，咳嗽着。然后抱怨和推诿开始了。没有人愿意相信这可能是真的。肯定不是这样的，肯定有一些有效的控制机制，肯定这些疾病很难制造，肯定数据库可以被锁定，肯定硬件可以被保护。等等。

在研讨会上的集体反应不仅仅是轻蔑。人们干脆拒绝接受演讲者的愿景。没有人愿意面对他们所听到的硬性事实和冷酷概率的影响。我保持沉默，坦率地说，我感到震惊。很快，研讨会结束了。那

晚上我们都出去吃晚饭，像往常一样继续聊天。我们刚刚度过了一整天讨论世界末日的话题，但仍然有披萨要吃，有笑话要讲，还有办公室要回去，况且，总会有事情发生，或者某些论点肯定是错的。我也加入了。

但那次演讲在此后的几个月里一直困扰着我。为什么我，为什么我们所有人，都没有更认真地对待它？为什么我们尴尬地回避进一步的讨论？为什么有些人会变得讽刺，并指责提出这些问题的人是在危言耸听或“忽视了技术的惊人好处”？我观察到的这种普遍的情绪反应是我称之为悲观厌恶陷阱的东西：当你因害怕面对潜在的黑暗现实而不知所措时产生的错误分析，以及由此导致的视而不见的倾向。

几乎每个人对这种反应都有某种版本，其结果是这导致我们忽视了一些在我们眼前展开的重要趋势。这几乎是一种天生的生理反应。我们这个物种并没有真正准备好应对如此规模的变革，更不用说技术可能以这种方式让我们失望的可能性。在我的职业生涯中，我一直有这种感觉，我也看到过许多人有同样的本能反应。面对这种感觉是本书的目的之一。即使感到不适，也要冷静而严肃地看待事实。

正确应对这一浪潮，控制技术，并确保其始终服务于人类，这意味着要克服对悲观主义的厌恶。这意味着要直面即将到来的现实。

源文本：—翻译文本：

尝试做到这一点。承认并阐明即将到来的浪潮的轮廓。探讨是否可以遏制。将事情置于历史背景中，通过远离围绕技术的日常喧嚣来看到更广阔的图景。我的目标是直面这一困境，并理解推动这一出现的潜在过程。

科学技术。我想尽可能清晰地向尽可能广泛的受众展示这些想法。我以开放和探究的精神写下这些内容：进行观察，追随其影响，但也保持开放态度，接受反驳和更好的解释。我最希望的就是在这里被证明是错误的，而不是遏制变得轻而易举。

有些人可能会理所当然地期望从像我这样的人——两家人工智能公司的创始人——那里得到一本更具科技乌托邦色彩的书。作为一名技术专家和企业家，我天生就是个乐观主义者。作为一个年轻的青少年，我记得第一次在我的Packard Bell 486 PC上安装Netscape时完全被迷住了。我被呼呼作响的风扇和56 Kbps拨号调制解调器的失真哨声所吸引，它伸出手去连接万维网，把我连接到论坛和聊天室，这些给了我自由并教会了我很多。我热爱技术。它一直是*the*进步的引擎，也是我们为人类成就感到自豪和兴奋的原因。

但我也相信，那些推动技术创造的人必须有勇气预测——并对其可能在未来几十年带我们去的地方负责。我们必须开始提出建议，如果看起来技术确实有失败的风险，该怎么办。所需的是社会和政治的回应，而不仅仅是个人的努力，但这需要从我和我的同行开始。

有些人会认为这完全是夸大其词。变化其实是更加渐进的。这只是炒作周期的又一次转折。应对危机和变化的系统实际上相当稳健。我对人性的看法过于悲观。人类的记录，到目前为止，还算不错。历史上充满了被证明错误的假先知和末日预言者。为什么这次会有所不同？

悲观厌恶是一种情绪反应，是一种根深蒂固的本能拒绝，拒绝接受可能导致严重不稳定结果的可能性。它往往来自于那些处于安全和强大地位的人，这些人有着根深蒂固的世界观，他们可以表面上应对变化，但难以接受对其世界秩序的任何真正挑战。我指责许多陷入悲观厌恶陷阱的人完全接受日益增长的批评。

技术。但他们只是点头附和，却没有实际采取任何行动。他们说，我们会应对的，我们总是这样。

在科技或政策圈子里待一段时间，很快就会发现鸵鸟心态是默认的意识形态。相信并采取其他行动的风险在于，可能会因对巨大、不可阻挡的力量的恐惧和愤怒而变得如此瘫痪，以至于一切都显得徒劳无功。因此，这种对悲观主义的奇怪的知识分子半世界继续存在。我应该知道，我在其中困了太久。

自从我们创立DeepMind以及那些演讲以来，这种讨论在某种程度上发生了变化。关于工作自动化的辩论已经被反复讨论了无数次。一场全球大流行病展示了合成生物学的风险和潜力。某种程度上的“ 科技反感 ”出现了，批评者在专栏文章和书籍中抨击科技和科技公司，在华盛顿、布鲁塞尔和北京的监管中心也是如此。以前关于技术的小众恐惧爆发到了主流，公众对技术的怀疑增加，来自学术界、民间社会和政治的批评更加尖锐。

然而，面对即将到来的浪潮和巨大的困境，以及面对厌恶悲观主义的技术精英，这一切都不够。

论点

源文本: 翻译文本:

波浪无处不在于人类生活中。这只是最新的一个。人们常常认为它还很遥远，如此未来主义和荒谬，以至于它只是少数书呆子和边缘思想家的领域，更多的夸张，更多的技术术语，更多的吹捧。这是一个错误。这是真实的，就像从开阔的蓝色海洋中出现的海啸一样真实。

这不仅仅是幻想或是一个沉思的智力练习。即使你不同意我的框架，认为这些都不太可能，我还是敦促你继续阅读。是的，我有人工智能背景，并且习惯于通过技术的视角看待世界。我在这个问题上是否重要的问题上有偏见。然而，近距离接触这一点

在过去的十五年中，革命正在展开，我确信我们正处于我们一生中最重要的转型的风口浪尖。

作为这些技术的构建者，我相信它们可以带来极大的好处，改善无数人的生活，并解决根本性挑战，从帮助解锁下一代清洁能源到为我们最棘手的医疗状况生产廉价且有效的治疗方法。技术可以并且应该丰富我们的生活；历史上，值得一提的是，背后的发明家和企业家一直是进步的强大推动者，提高了数十亿人的生活水平。

但是，如果没有遏制，技术的其他每个方面，对其伦理缺陷的每次讨论，或它可能带来的好处，都是无关紧要的。我们迫切需要对即将到来的浪潮如何被控制和遏制、如何维护民主国家的保障和便利提供万无一失的答案，但目前没有人有这样的计划。这是一个我们都不想要的未来，但我担心它越来越有可能发生，我将在接下来的章节中解释原因。

在第1部分中，我们考察了技术的悠久历史及其传播方式——跨越千年的浪潮。是什么驱动了它们？是什么使它们真正普遍？我们还询问是否有社会有意识地拒绝新技术的例子。与其说是背离技术，过去更是以显著的扩散模式为标志，导致了意料之中和意料之外的后果的蔓延链条。

我称之为“遏制问题”。当有史以来最有价值的技术变得更便宜并比历史上任何技术传播得更快时，我们如何保持对它们的控制？

第2部分深入探讨即将到来的浪潮本身的细节。其核心是两种具有巨大前景、力量和风险的通用技术：人工智能和合成生物学。两者都早已被广泛宣传，但即便如此，我仍然认为它们影响的范围常常被低估。在它们周围，伴随着一系列相关技术的发展，如机器人技术和量子计算，这些技术的发展将以复杂而动荡的方式交织在一起。

在本节中，我们不仅研究了它们是如何出现的以及它们能做什么，还研究了为什么它们如此难以控制。我所说的各种技术具有四个关键特征，这解释了为什么这不是常规业务：它们本质上是通用的，因此是全用途的，它们快速进化，它们具有不对称的影响，并且在某些方面，它们越来越自主。

它们的创造受到强大动机的驱动：地缘政治竞争、巨大的经济回报以及开放、分布式的研究文化。无论我们是否愿意，众多国家和非国家行为体将不顾监管和控制即将到来的努力，争相开发它们，承担影响每个人的风险。

第3部分探讨了由不受控制的浪潮引发的巨大权力再分配的政治影响。我们当前政治秩序的基础——以及技术遏制中最重要的参与者——是民族国家。已经因危机而动摇的它，将因一系列由浪潮放大的冲击而进一步削弱：新形式暴力的潜力、虚假信息泛滥、工作消失以及灾难性事故的前景。

更远的地方，浪潮将迫使权力发生一系列构造转变，同时集中化和去中心化。这将创造出庞大的新企业，支持威权主义，但也赋予团体和运动在传统社会结构之外生活的能力。正当我们最需要像国家这样的机构时，国家的微妙平衡将面临巨大的压力。这就是我们陷入困境的原因。

在第4部分中，讨论转向我们可以对此做些什么。是否甚至有一线希望可以遏制，摆脱困境？如果有，该如何做？在本节中，我们概述了十个步骤，从代码和DNA的层面到国际条约的层面，形成一套严密的嵌套约束，这是一个遏制的概要计划。

源文本：—翻译文本：

关于面对失败。技术可能会以不工作的普通意义上失败：引擎无法启动；桥梁倒塌。但它们也可能在更广泛的意义上失败。如果技术损害了人类生活，或产生充满伤害的社会，或使其因我们赋予混乱的长尾不良（或无意中危险的）行为者而无法治理——如果从总体上看，技术是有害的——那么可以说它在另一种更深层次的意义上失败了，未能实现其承诺。这种意义上的失败并不是技术本身固有的；而是关于其运作的背景、所受的治理结构、以及其所被应用的权力和用途网络。

那种令人印象深刻的创造力带来了如此多的进步，这意味着我们在避免第一类失败方面做得更好。飞机失事更少，汽车更清洁、更安全，计算机更强大且更安全。我们面临的巨大挑战是，我们仍然没有解决后一种模式的失败。

几个世纪以来，技术极大地提高了数十亿人的福祉。由于现代医学，我们的健康状况得到了极大的改善，世界上大多数人生活在食物充裕的环境中，人们从未像现在这样受过良好教育、更加和平或物质上更加舒适。这些是人类伟大动力的一部分所产生的决定性成就：科学和技术的创造。这就是为什么我一生致力于安全开发这些工具。

但是，我们从这段非凡历史中获得的任何乐观情绪都必须建立在直截了当的现实基础上。防止失败意味着理解并最终面对可能出错的地方。我们需要遵循推理链直到其逻辑终点，而不必害怕可能导致的结果，并且在到达那里时采取行动。即将到来的技术浪潮可能会比以往任何时候更快、更大规模地失败。这种情况需要全球范围的广泛关注。它需要答案，而这些答案目前无人拥有。

遏制表面上是不可能的。然而，为了我们所有人的利益，遏制*must*是可能的。

PART I

HOMO TECH NOLOGICUS

CHAPTER 2

无尽增殖

引擎

源文本: 翻译文本:

FOR MOST OF HISTORY, FOR most people, personal transportation meant one thing: walking. Or if you were lucky, two: being carried or pulled by horses, oxen, elephants, or other beasts of burden. Just moving between neighboring settlements—forget about continents—was hard and slow.

在19世纪早期，铁路彻底改变了交通，这是数千年来最大的创新，但大多数旅程无法通过铁路进行，而那些可以通过铁路进行的旅程也不太个性化。铁路确实让一件事情变得清晰：发动机是未来。能够推动铁路车厢的蒸汽机需要巨大的外部锅炉。但如果你能将它们缩小到一个可管理的、便携的尺寸，你将拥有一种全新的个人出行方式。

创新者尝试了各种方法。早在十八世纪，一位名叫尼古拉-约瑟夫·居纽的法国发明家就制造了一种蒸汽动力汽车。它以每小时两英里的速度缓慢行驶，前面挂着一个巨大的钟摆式锅炉。1863年，比利时发明家让·约瑟夫·艾蒂安·勒努瓦用内燃机驱动了第一辆汽车，将其从巴黎开出七英里。但发动机很重，速度有限。其他人则尝试了电力和氢气。没有什么能够流行起来，但自我驱动的个人交通工具的梦想一直存在。

然后事情开始发生变化，起初是缓慢的。一位名叫尼古拉斯·奥古斯特·奥托的德国工程师花了多年时间研究一种小得多的燃气发动机。

比蒸汽机更高效。到1876年，在科隆的Deutz AG工厂，奥托制造出了第一个功能齐全的内燃机，即“四冲程”模型。它已准备好进行大规模生产，但在此之前，奥托与他的商业伙伴戈特利布·戴姆勒和威廉·迈巴赫发生了分歧。奥托希望在水泵或工厂等固定场所使用他的发动机。他的伙伴们则看到了这些日益强大的发动机的另一种用途：运输。

然而，是另一位德国工程师卡尔·本茨抢先一步。1886年，他利用自己版本的四冲程内燃机，申请了Motorwagen的专利，现在被视为世界上第一辆真正的汽车。这辆奇怪的三轮装置首次亮相时，公众持怀疑态度。直到本茨的妻子兼商业伙伴贝莎驾驶这辆车从曼海姆到她母亲位于65英里外普福尔茨海姆的家时，这辆车才开始流行。据说她是在丈夫不知情的情况下开走的，并在途中用从当地药店购买的溶剂加油。

一个新时代已经来临。但是，汽车和为其提供动力的内燃机仍然异常昂贵，除了最富有的人之外，其他人都负担不起。那时还没有道路和加油站的网络。到1893年，奔驰仅售出69辆汽车；到1900年，仅售出1,709辆。在奔驰获得专利二十年后，德国的道路上仍然只有35,000辆汽车。

转折点是亨利·福特的1908年T型车。他的简单但有效的车辆是通过一种革命性的方法制造的：流水装配线。高效、线性和重复的过程使他能够大幅降低个人车辆的价格，买家纷至沓来。当时大多数汽车的价格约为$v*$2,000。福特将他的定价为$v*$850。

在早期，T型车的销量以千计。福特不断提高产量并进一步降低价格，他认为：“每次我将汽车的价格降低一美元，我就能获得一千个新买家。”到了1920年代，福特每年销售数百万辆汽车。中产阶级的美国人第一次能够负担得起机动交通工具。汽车以极快的速度激增。1915年，只有10%的美国人拥有汽车；到1930年，这一数字已达到惊人的59%。

如今，大约有20亿台内燃机存在于从割草机到集装箱船的各种设备中。其中大约有14亿台在汽车中。它们变得越来越易于获取、高效、强大且适应性强。可以说，围绕它们发展出了一整套生活方式，甚至是一整种文明，从广袤的郊区到工业化农场，从免下车餐厅到汽车改装文化。修建了广阔的高速公路，有时直接穿过城市，切断了社区，但连接了遥远的地区。以前在寻找繁荣或乐趣时从一个地方移动到另一个地方的挑战性观念成为了人类生活的常规特征。

引擎不仅仅是为车辆提供动力；它们还推动了历史的发展。如今，多亏了氢能和电动机，内燃机的统治已进入黄昏。但它所开启的大规模移动时代并未结束。

所有这些在十九世纪初似乎都是不可能的，当时自推进运输工具仍然是一些梦想家玩弄火焰、飞轮和金属块的幻想。但正是从那些早期的修补者开始了一场发明和生产的马拉松，改变了世界。一旦有了动力，内燃机的传播就变得不可阻挡。从几个浸满油的德国作坊中发展出了一项影响地球上每一个人的技术。

这不仅仅是关于发动机和汽车的故事。这是关于技术本身的故事。

通用波：历史的节奏

源文本: 翻译文本:

技术有一个明确且不可避免的轨迹：以巨大的波浪形式进行大规模扩散。这从最早的燧石和骨制工具到最新的AI模型都是如此。随着科学产生新的发现，人们应用这些见解来制造更便宜的食物、更好的商品和更高效的运输。随着时间的推移，对最佳新产品和服务的需求增长，推动竞争以生产更便宜的版本，并且充满更多的功能。这反过来又推动了对创造它们的技术的更多需求，而*they*也

变得更容易、更便宜。成本持续下降。能力提升。实验、重复、使用。成长、改进、适应。这是技术不可避免的进化本质。

这些技术和创新的浪潮是本书的核心。更重要的是，它们是人类历史的核心。理解这些复杂、混乱且不断积累的浪潮，遏制的挑战就变得清晰。了解它们的历史，我们就可以开始勾勒它们的未来。

那么，什么是浪潮？简单来说，浪潮是一组技术在同一时间出现，由一种或多种具有深远社会影响的新型通用技术推动。所谓“通用技术”，是指那些能够在人类能力上带来重大进步的技术。社会的发展与这些飞跃同步进行。我们一次又一次地看到，一项新技术，如内燃机，迅速普及并改变其周围的一切。

人类的故事可以通过这些浪潮来讲述：我们从在稀树草原上勉强生存的脆弱灵长类动物进化到成为这个星球的主导力量，无论好坏。人类天生就是一个技术物种。从一开始，我们就从未与我们创造的技术浪潮分离。我们共同进化，处于共生状态。

最早的石器工具可以追溯到三百万年前，远在*Homo sapiens,*的黎明之前，这可以从破损的锤石和原始的刀具中得到证据。简单的手斧构成了历史上第一波技术的一部分。动物可以更有效地被杀死，尸体被屠宰，竞争对手被对抗。最终，早期人类学会了精细地操控这些工具，从而产生了缝纫、绘画、雕刻和烹饪。

另一波同样具有决定性：火。由我们的祖先 *Homo erectus,* 使用，它是光明、温暖和防止掠食者的来源。它对进化产生了显著影响：烹饪食物意味着能量更快释放，使人类消化道缩小而大脑增大。我们的祖先，由于强壮的下颌限制了头骨的生长，像今天的灵长类动物一样，花时间不停地咀嚼和消化食物。

被火从这种平凡的必需品中解放出来后，他们可以花更多时间做有趣的事情，比如狩猎高能量食物、制作工具或建立复杂的社交网络。营火成为人类生活的中心枢纽，帮助建立社区和关系并组织劳动。*Homo sapiens* 的演变乘着这些浪潮而来。我们不仅是工具的创造者。从生物学、解剖学层面来看，我们也是它们的产物。

石工和火是原始的通用技术，意味着它们是普遍存在的，进而促成了新的发明、商品和组织行为。通用技术在社会中、地理上和历史上产生连锁反应。它们为发明打开了大门，使大量的下游工具和流程成为可能。它们通常建立在某种通用原则之上，无论是蒸汽的动力来完成工作，还是计算机二进制代码背后的信息理论。

通用技术的讽刺之处在于，不久之后，它们就变得无形，我们理所当然地接受它们。语言、农业、书写——每一种都是早期浪潮中心的通用技术。这三次浪潮构成了我们所知文明的基础。现在我们理所当然地接受它们。一项重要研究将整个人类历史中出现的通用技术数量定为仅仅二十四种，列举了从农业、工厂系统、铁和青铜等材料的发展，到印刷机、电力，当然还有互联网的发明。它们的数量不多，但却很重要；这就是为什么在大众想象中，我们仍然使用青铜时代和航海时代这样的术语。

纵观历史，人口规模和创新水平是相互关联的。新的工具和技术催生了更大的人口。更大且更紧密联系的人口是更强大的试验、实验和偶然发现的熔炉，是创造新事物的更强大的“集体大脑”。大规模的人口催生了更高水平的专业化，出现了新的阶层，如工匠和学者，他们的生计不再与土地挂钩。更多人的生活不再围绕着土地。

围绕生存意味着更多可能的发明者，以及更多可能的发明理由，而这些发明反过来又意味着更多的人。从最早的文明，如美索不达米亚的乌鲁克，楔形文字的发源地，第一个已知的书写系统，到今天的大都市，城市推动了技术的发展。而更多的技术意味着更多——更大的——城市。在农业革命的黎明时期，全球人口仅为240万。在工业革命开始时，人口接近10亿，这一四百倍的增长是基于中间时期的浪潮。

农业革命（公元前9000 – 7500年），是历史上最重要的浪潮之一，标志着两项大型通用技术的到来，这些技术逐渐取代了游牧、狩猎采集的生活方式：植物和动物的驯化。这些发展不仅改变了食物的获取方式，还改变了食物的储存方式、运输方式以及社会运作的规模。早期的作物如小麦、大麦、扁豆、鹰嘴豆和豌豆，以及动物如猪、羊和山羊，开始受到人类的控制。最终，这与工具的新革命——锄头和犁——结合在一起。这些简单的创新标志着现代文明的开端。

工具越多，你能做的事情就越多，你能想象的新工具和新工艺也就越多。正如哈佛人类学家约瑟夫·亨里奇指出的那样，车轮在人类生活中出现得出乎意料地晚。但一旦发明出来，它就成为了从战车和马车到磨坊、印刷机和飞轮等一切事物的基石。从文字到航海船只，技术增加了互联性，帮助促进其自身的流动和传播。因此，每一波浪潮都为后续的浪潮奠定了基础。

随着时间的推移，这一动态加速了。大约从18世纪70年代开始，欧洲的第一次工业革命结合了蒸汽动力、机械化织机、工厂系统和运河。到了19世纪40年代，铁路、电报和蒸汽船的时代到来，稍后出现了钢铁和机床；它们共同构成了第一次工业革命。然后，

仅仅几十年后，第二次工业革命到来了。你会熟悉它的最伟大成就：内燃机、化学工程、动力飞行和电力。飞行需要燃烧，而内燃机的大规模生产需要钢铁和机床，等等。从工业革命开始，巨大的变化以几十年而不是几个世纪或千年来衡量。

然而，这并不是一个有序的过程。技术浪潮并不像潮汐那样整齐可预测。从长远来看，浪潮不规则地交错并增强。在公元前1000年的一万年间，出现了七种通用技术。而在1700年至1900年之间的两百年间，从蒸汽机到电力，出现了六种。在过去的一百年中，仅仅就有七种。想想那些在十九世纪末乘马车旅行、烧木头取暖的孩子们，他们的最后时光是在乘飞机旅行和住在通过原子裂变取暖的房子里度过的。

波动——脉动的、突现的、连续的、复合的和交叉传播的——定义了一个时代的技术可能性视野。它们是我们的一部分。没有非技术的人类。

这种将历史视为一系列创新浪潮的概念并不新颖。技术的连续和颠覆性集群在技术讨论中反复出现。对于未来学家阿尔文·托夫勒来说，信息技术革命是继农业和工业革命之后人类社会的“第三次浪潮”。约瑟夫·熊彼特将浪潮视为创新的爆发，在“创造性破坏”的爆发中点燃了新业务。伟大的技术哲学家刘易斯·芒福德认为，“机器时代”实际上更像是千年展开的三个主要连续浪潮。最近，经济学家卡洛塔·佩雷斯谈到了技术革命中迅速转变的“技术经济范式”。繁荣的颠覆和狂热的投机时刻重新调整经济。突然之间，一切都依赖于铁路、汽车或微处理器。最终，技术成熟，变得嵌入并广泛可用。

大多数从事技术工作的人都困在当下的细节中，梦想着明天。人们很容易将发明视为离散且幸运的时刻。但如果这样想，你就会错过历史的鲜明模式，以及技术浪潮一次又一次到来的几乎天生的趋势。

增殖是默认设置

源文本: 翻译文本:

在大部分人类历史中，新技术的扩散是罕见的。大多数人生于斯，长于斯，死于斯，周围的工具和技术始终如一。然而，放大视角来看，扩散实际上是默认状态。

通用技术在广泛传播时会形成浪潮。如果没有史诗般且几乎不受控制的全球传播，它就不是浪潮，而是一个历史上的好奇现象。然而，一旦扩散开始，这一过程就会在历史上回响，从农业在欧亚大陆的传播到水磨坊从罗马帝国缓慢扩散到整个欧洲。一旦某项技术获得牵引力，一旦浪潮开始形成，我们在汽车上看到的历史模式就很清晰。

当古腾堡在大约1440年发明印刷机时，欧洲只有一个例子：他在德国美因茨的原型。但仅仅五十年后，整个欧洲大陆就有一千台印刷机。书籍本身，作为历史上最具影响力的技术之一，以爆炸性的速度增加。在中世纪，手稿的生产量每个主要国家每个世纪大约是几十万本。在古腾堡之后的一百年，像意大利、法国和德国这样的国家每半个世纪生产约4000万本书，而且加速的速度仍在增加。在十七世纪，欧洲印刷了5亿本书。随着需求的飙升，成本急剧下降。一项分析估计，十五世纪印刷机的引入导致书籍价格下降了340倍，进一步推动了采用和更多的需求。

或者以电力为例。第一批电力发电站于1882年在伦敦和纽约亮相，1883年在米兰和圣彼得堡，1884年在柏林。自那时起，它们的推广速度加快。到1900年，2%的化石燃料生产用于发电，到1950年这一比例超过10%，到2000年则超过30%。1900年全球发电量为8太瓦时；五十年后达到600太瓦时，推动了经济的转型。

诺贝尔奖获得者、经济学家威廉·诺德豪斯计算出，在十八世纪生产五十四分钟优质光所需的劳动量，现在可以生产超过五十年的光。因此，二十一世纪的普通人每年可以获得的“流明小时”大约是我们十八世纪的亲戚的438,000倍。

不出所料，消费技术也表现出类似的趋势。亚历山大·格雷厄姆·贝尔于1876年发明了电话。到1900年，美国已有60万部电话。十年后，这一数字达到580万。如今，美国的电话数量远远超过了人口数量。

在这幅图中，质量的提高与价格的下降相辅相成。1950年一台售价为$1,000的原始电视在2023年只需$8，当然，如今的电视质量要好得多，因此价格也更高。你可以找到几乎相同的价格（和普及）曲线用于汽车、微波炉或洗衣机。事实上，二十世纪和二十一世纪见证了新消费电子产品的显著一致的普及。一次又一次，这种模式是显而易见的。

扩散由两种力量催化：需求和由此产生的成本下降，这两者都推动技术变得更好和更便宜。科学和技术的漫长而复杂的对话产生了一系列的洞察、突破和工具，随着时间的推移，这些洞察、突破和工具不断构建和加强，形成推动未来的富有成效的重组。当你获得更多和更便宜的技术时，它会使下游的新技术和更便宜的技术成为可能。没有智能手机，Uber 是不可能的，而智能手机本身是由 GPS 启用的，GPS 是由卫星启用的，卫星是由

火箭由燃烧技术实现，而燃烧技术由语言和火实现。

当然，技术突破的背后是人。他们在车间、实验室和车库中努力改进技术，动机来自金钱、名声，往往还有知识本身。技术人员、创新者和企业家通过实践变得更好，关键是通过模仿。从敌人的高级犁到最新的手机，模仿是扩散的关键驱动力。模仿激发竞争，技术进一步改进。规模经济开始发挥作用并降低成本。

文明对有用且廉价技术的渴求是无止境的。这一点不会改变。

从真空管到纳米：涡轮增殖

源文本: 翻译文本:

如果你想知道接下来会发生什么，可以考虑上一波成熟浪潮的基础。从一开始，计算机就受到新前沿数学以及大国冲突紧迫性的驱动。

就像内燃机一样，计算机最初也是晦涩学术论文和实验室修补者的产物。然后战争来了。在20世纪40年代，英国顶级机密的二战密码破译中心布莱切利园首次开始实现真正的计算机。为了破解德国号称无法破解的恩尼格玛机，一个非凡的团队将理论见解转化为能够实现这一目标的实用设备。

其他人也在研究这个问题。到1945年，一个重要的计算机前身被开发出来，称为ENIAC，这是一台八英尺高的庞然大物，拥有一万八千个真空管，能够每秒进行三百次运算，由宾夕法尼亚大学开发。贝尔实验室在1947年发起了另一个重要突破：晶体管，这是一种半导体，创建“逻辑门”来执行计算。这种粗糙的设备，由一张纸组成。

剪下的一片金箔和一块可以切换电子信号的锗晶体奠定了数字时代的基础。

与汽车一样，当时的观察者并不认为计算机会迅速普及。在20世纪40年代末，计算设备仍然很少。那个年代早期，IBM的总裁托马斯·J·沃森据称（并且臭名昭著地）说过：“ 我认为全世界的市场大约需要五台计算机。 ” *Popular Mechanics* 杂志在1949年做出了一个典型的预测：“ 未来的计算机可能只有1000个真空管， ”它认为， “ 也许重量只有1½吨。 ”在布莱切利公园十年后，全球的计算机数量仍然只有几百台。

我们知道接下来发生了什么。计算机技术以比任何人预测的速度更快地改变了社会，并以人类历史上任何发明都无法比拟的速度迅速普及。罗伯特·诺伊斯在20世纪50年代末和60年代初在仙童半导体公司发明了集成电路，在硅片上印制多个晶体管，生产出后来被称为硅芯片的东西。不久之后，一位名叫戈登·摩尔的研究人员提出了他同名的“ 定律 ”：每二十四个月，芯片上的晶体管数量将翻一番。这意味着芯片，以及扩展到数字和计算技术的世界，将受到指数过程上升曲线的影响。

结果令人震惊。自20世纪70年代初以来，每个芯片上的晶体管数量增加了一千万倍。它们的功率增加了十个数量级——*a seventeen-billion-fold improvement.* 仙童半导体在1958年以每个150美元的价格出售了一百个晶体管。现在，晶体管的生产速度达到每秒数十万亿个，每个晶体管的成本仅为十亿分之一美元：这是历史上最快、最广泛的扩散。

当然，这种计算能力的提升推动了设备、应用程序和用户的繁荣。在20世纪70年代初，大约有50万台计算机。早在1983年，只有562台计算机连接到原始互联网。现在，计算机、智能手机和连接设备的数量估计为140亿。它花费了

智能手机从小众产品到地球上三分之二人口的必需品只用了几年时间。

随着这一波浪潮的到来，电子邮件、社交媒体、在线视频——每一种都是由晶体管和另一种通用技术——互联网所推动的全新体验。这就是纯粹、不受限制的技术扩散的样子。它创造了更令人难以置信的扩散：数据，仅在2010-2020年这十年间就增长了二十倍。就在几十年前，数据存储还是书籍和尘封档案的领域。现在，人类每天生产数千亿封电子邮件、信息、图像和视频，并将它们存储在云中。每天每分钟都有1800万千兆字节的数据被添加到全球总量中。

数十亿小时的人类原始生活被这些技术所消耗、塑造、扭曲和丰富。它们主导着我们的商业和休闲时间。它们占据着我们的思想和我们世界的每一个角落，从冰箱、计时器、车库门和助听器到风力涡轮机。它们构成了现代生活的基本架构。我们的手机是我们早上看到的第一件事，也是晚上看到的最后一件事。人类生活的每个方面都受到影响：它们帮助我们找到爱情和新朋友，同时加速供应链。它们影响谁当选以及如何当选，我们的钱投向哪里，我们孩子的自尊心，我们的音乐品味，我们的时尚，我们的食物，以及介于两者之间的一切。

战后世界的人们会对这种曾经看似小众的技术的规模和影响感到震惊。计算机技术以指数级的速度传播和改进，几乎进入并包围了生活的各个方面，这已成为当代文明的主导事实。没有任何先前的浪潮能如此迅速地膨胀，但历史模式仍在重复。起初，这似乎是不可能和难以想象的。然后，它显得不可避免。每一波浪潮都变得更大更强。

源文本：—翻译文本：

迷失在细节中，但退后一步，你可以看到波浪正在积聚速度、范围、可及性和后果。一旦它们积聚动量，就很少停止。大规模扩散、原始、猖獗的扩散——这是技术的历史默认状态，最接近自然状态的东西。想想农业、青铜工艺、印刷机、汽车、电视、智能手机等等。然后似乎存在技术的规律，类似于一种固有特性，能够经受时间考验的涌现属性。

历史告诉我们，技术的传播是不可避免的，最终几乎遍及每个角落，从最初的营火到土星五号火箭的火焰，从最初的涂鸦字母到互联网无尽的文本。激励是压倒性的。能力积累；效率提高。浪潮变得更快且更具影响力。随着技术变得更便宜，获取技术的途径也在增加。技术不断扩散，并且随着每一波浪潮的到来，这种扩散加速并深入渗透，即使技术变得更强大。

这是技术的历史常态。展望未来，这正是我们可以期待的。

或者我们可以吗？

CHAPTER 3

封闭问题

报复效应

一个

源文本: 翻译文本:

LAN TURING AND GORDON MOORE could never have predicted, let alone altered the rise of, social media, memes, Wikipedia, or cyberattacks. Decades after their invention, the architects of the atomic bomb could no more stop a nuclear war than Henry Ford could stop a car accident.

技术存在于一个复杂且动态的系统中（现实世界），在这个系统中，第二、第三和*n*阶后果会不可预测地扩散开来。纸面上看似完美的东西在实际应用中可能表现得截然不同，尤其是在被复制和进一步改编时。无论初衷多么良好，人们对你的发明的实际使用方式永远无法得到保证。托马斯·爱迪生发明留声机是为了让人们能够记录他们的思想以供后人参考，并帮助盲人。他对大多数人只是想用它来播放音乐感到震惊。阿尔弗雷德·诺贝尔本意是让他的炸药仅用于采矿和铁路建设。

古腾堡只是想通过印刷《圣经》赚钱。然而，他的印刷机催化了科学革命和宗教改革，因此成为自天主教会成立以来对其最大的威胁。

Technology's unavoidable challenge is that its makers quickly lose control over the path their inventions take once introduced to the world.

冰箱制造商并不打算用氯氟烃（CFCs）在臭氧层中制造一个洞，就像内燃机和喷气发动机的创造者没有想到会导致冰盖融化。事实上，早期的汽车爱好者曾为其环境效益辩护：发动机将清除街道上的

马粪堆积如山，在城市地区传播污垢和疾病。他们没有全球变暖的概念。

理解技术，部分在于尝试理解其意外后果，不仅要预测正面溢出效应，还要预测“报复效应”。简单来说，任何技术都有可能出错，往往以直接违背其初衷的方式出错。想想处方类阿片类药物如何导致依赖，或者抗生素的过度使用如何使其效果减弱，或者被称为“太空垃圾”的卫星和碎片的激增如何危及太空飞行。

随着技术的普及，越来越多的人可以使用、适应、塑造它，以任何他们喜欢的方式，在超出任何个人理解的因果链中。随着我们工具的力量呈指数级增长以及对它们的访问迅速增加，潜在的危害也在增加，这是一种展开的后果迷宫，没有人能完全预测或阻止。某天，有人在黑板上写方程式或在车库里摆弄原型，这些工作似乎与更广泛的世界无关。几十年内，它就为人类带来了存在性的问题。随着我们构建越来越强大的系统，这种技术的方面对我来说越来越紧迫。我们如何保证这股新技术浪潮带来的益处多于危害？

技术在这里的问题是一个遏制问题。如果这一方面无法消除，它可能会被限制。遏制是指在技术发展的任何阶段或部署中控制、限制以及在必要时关闭技术的总体能力。这意味着在某些情况下，能够在一开始就阻止技术的扩散，检查意想不到的后果（无论是好的还是坏的）的涟漪。

技术越强大，它就越深入生活和社会的各个方面。因此，技术问题往往会随着其能力的增强而升级，因此对遏制的需求随着时间的推移变得更加紧迫。

这是否让技术人员摆脱困境？一点也不；比任何人都更应该由我们来面对。我们可能无法控制我们工作的最终结果或其长期影响，但这不是理由去

放弃责任。技术人员和社会在源头做出的决策仍然可以影响结果。仅仅因为后果难以预测，并不意味着我们不应该尝试。

在大多数情况下，遏制是关于有意义的控制，即停止一个用例、改变研究方向或拒绝有害行为者访问的能力。这意味着保持引导潮流的能力，以确保其影响反映我们的价值观，帮助我们作为一个物种繁荣，并且不会带来超过其利益的重大危害。

本章展示了这实际上是多么具有挑战性和罕见。

遏制是基础

源文本: 翻译文本:

对于许多人来说，“ 遏制 ”一词带来了冷战的回响。美国外交官乔治·F·凯南认为，“ 美国对苏联政策的主要元素必须是对俄罗斯扩张倾向进行长期、耐心但坚定和警惕的遏制。 ”凯南主张，西方国家必须将世界视为一个不断变化的斗争领域，监控并反击苏联的力量，无论它们在哪里发现它，安全地遏制红色威胁及其意识形态触角在 *all* 维度上。

虽然这种对遏制的解读提供了一些有用的教训，但对于我们的目的来说是不够的。技术不是对手；它是人类社会的基本属性。遏制技术需要成为一个更为根本的计划，不是竞争者之间的权力平衡，而是人类与我们工具之间的平衡。这是我们物种在下个世纪生存的必要前提。遏制包括监管、更好的技术安全、新的治理和所有权模式，以及新的问责和透明度模式，所有这些都是更安全技术的必要（但不充分）前提。这是一个将尖端工程、伦理价值观和政府监管结合在一起的总体锁定。遏制不应被视为所有技术问题的最终答案。

问题；这更像是第一步，关键的一步，是未来构建的基础。

因此，可以将遏制视为一套相互关联且相互加强的技术、文化、法律和政治机制，用于在技术呈指数变化的时代维持社会对技术的控制；一种能够胜任遏制任务的架构，将过去需要几个世纪或千年才能实现的技术变革，现在在几年甚至几个月内发生，其后果在几秒钟内便在全球范围内反弹。

技术遏制是指在实验室或研发设施中发生的事情。例如，在人工智能中，这意味着空气隔离、沙盒、模拟、关闭开关、硬性内置的安全和安保措施——用于验证系统的安全性、完整性或未受损性质的协议，并在需要时将其下线。接下来是围绕创造和传播的价值观和文化，支持边界、治理层次、接受限制、警惕危害和意外后果。最后，遏制包括国家和国际的法律机制：由国家立法机构通过的法规以及通过联合国和其他全球机构运作的条约。技术总是深深地卷入任何特定社会的法律和习俗、规范和习惯、权力和知识结构中；每一个都必须加以解决。我们将在第4部分中更详细地回到这一点。

目前，你可能会想，我们是否真的尝试过这个，试图遏制一个波浪？

我们曾经说过不吗？

源文本: 翻译文本:

随着十五世纪印刷机在欧洲轰鸣，奥斯曼帝国的反应却截然不同。它试图禁止印刷机。由于对知识和文化不受监管的大规模生产感到不满，苏丹认为印刷机是一种外来的“西方”创新。尽管伊斯坦布尔在人口上与伦敦、巴黎和罗马等城市相媲美，但它并没有

直到1727年，才拥有一个被批准的印刷机，几乎是在其发明后三个世纪。长期以来，历史学家将奥斯曼帝国的抵制视为早期技术民族主义的经典例子，即一种有意识的、向后看的对现代性的拒绝。

但事情比这更复杂。根据帝国的规定，只有阿拉伯字符被禁止，而不是完全禁止印刷。与其说是某种根本的反技术立场，不如说禁令归结于运行阿拉伯语印刷机的巨大费用和复杂性；只有苏丹能够负担印刷的资金，而历任苏丹对此兴趣不大。因此，奥斯曼帝国的印刷业停滞不前；有一段时间，帝国表示不需要。但最终，就像其他地方一样，印刷成为奥斯曼帝国、其后裔国家，乃至全世界生活的一部分。看来，国家可能会说不，但随着事情变得更便宜和更广泛使用，他们不可能永远说不。

事后看来，浪潮可能显得平滑且不可避免。但实际上，有几乎无限多的小型、局部且常常是任意的因素影响着一项技术的轨迹。确实，没有人应该想象扩散是容易的。它可能是昂贵的、缓慢的和有风险的，或者需要在几十年或一生中才能实现的剧变。它必须与现有利益、既定知识以及那些嫉妒地持有两者的人作斗争。对任何新事物和不同事物的恐惧和怀疑是普遍存在的。从技艺精湛的工匠行会到多疑的君主，每个人都有理由反击。卢德分子，即那些暴力拒绝工业技术的团体，并不是新技术到来的例外；他们是常态。

在中世纪时期，教皇乌尔班二世想要禁止弩。伊丽莎白一世女王在十六世纪末否决了一种新型针织机，理由是它可能会扰乱行会。行会在纽伦堡、但泽、荷兰和英格兰骚扰并摧毁了新型织机和车床。飞梭的发明者约翰·凯因害怕暴力报复，从英格兰逃到了法国。飞梭使织布更加高效，是工业革命的关键技术之一。

纵观历史，人们一直试图抵制新技术，因为他们感到受到威胁，担心他们的生计和生活方式会被摧毁。他们认为自己是在为家庭的未来而战，如果有必要，他们会亲手摧毁即将到来的东西。如果和平手段失败，卢德分子想要拆除工业机械的浪潮。

在十七世纪的德川幕府统治下，日本几乎与世隔绝了三百年，从而也隔绝了那些野蛮的发明。像历史上大多数社会一样，它对新事物、不同事物和破坏性事物持怀疑态度。同样，中国在十八世纪末拒绝了英国的外交使团及其提供的西方技术，乾隆皇帝辩称：“ 我天朝物产丰盈，无所不有，原不藉外夷货物以通有无。 ”

没有任何效果。弩一直存在，直到被枪支取代。伊丽莎白女王的编织机在几个世纪后以大规模机械织机的形式重现，点燃了工业革命。中国和日本如今是地球上技术最先进、全球化程度最高的地方之一。卢德分子在阻止新工业技术方面并不比马匹拥有者和马车制造商阻止汽车更成功。只要有需求，技术总会突破、找到立足点、建立用户群。

一旦形成，波浪几乎无法阻止。正如奥斯曼帝国在印刷方面发现的那样，随着时间的推移，抵抗往往会被消磨殆尽。技术的本质是传播，无论障碍如何。

许多技术来来去去。你不会看到太多的高轮自行车或赛格威，也不会听到太多的磁带或迷你光盘。但这并不意味着个人交通和音乐不普遍；旧技术只是被更新、更高效的形式所取代。我们不再乘坐蒸汽火车或使用打字机，但它们的幽灵般存在依然在它们的继任者中延续，比如新干线和MacBook。

想想看，作为连续浪潮的一部分，火，然后是蜡烛和油灯，让位于煤气灯，然后是电灯泡，现在是LED灯，人工光的总量增加了，即使底层技术发生了变化。新技术取代了多个前辈。正如电力同时完成了蜡烛和蒸汽机的工作，智能手机取代了卫星导航、相机、PDA、电脑和电话（并创造了全新的体验类别：应用程序）。随着技术让你以更少的成本做更多的事情，它们的吸引力只会增加，采用率也随之增长。

想象一下，在没有电、没有自来水或药品的情况下建立一个现代社会。即使你能做到，你如何说服任何人这是值得的、令人向往的、一个不错的交易？很少有社会能够成功地将自己从技术前沿移除；这样做通常要么是崩溃的一部分，要么会导致崩溃。没有现实的方法可以退回去。

发明无法被取消或无限期阻止，知识无法被遗忘或阻止传播。零星的历史例子几乎没有理由认为这种情况可能再次发生。亚历山大图书馆被遗弃，最终被烧毁，大量古典学问永远丧失。但最终古代的智慧被重新发现并重新评估。由于缺乏现代通讯工具的帮助，中国将丝绸制作的秘密保密了几个世纪，但最终在公元552年，由于两位坚定的景教僧侣，这一秘密被泄露。技术是思想，而思想无法被消除。

技术是一根永远悬挂的胡萝卜，不断承诺更多、更好、更简单、更便宜。我们对发明的渴望是无法满足的。浪潮的似乎不可避免性并不是因为没有阻力，而是因为需求压倒了阻力。人们常常说不，出于多种原因希望限制技术。只是这从来不够。并不是历史上没有认识到限制问题，只是从未解决过。

有没有例外？或者说，波浪最终总是会在各处破裂吗？

THE NUCLEAR EXCEPTION?

源文本: 翻译文本:

1933年9月11日，物理学家欧内斯特·卢瑟福在莱斯特的英国科学促进会会议上辩称，“任何人说以我们目前的手段和知识可以利用原子能，都是在说空话。”匈牙利移民利奥·西拉德在伦敦的一家酒店阅读了卢瑟福论点的报道，并在早餐时仔细思考。他出去散步。在卢瑟福称其为空话的第二天，西拉德构思了核链式反应。

第一次核爆炸仅仅发生在十二年后。1945年7月16日，在曼哈顿计划的主持下，美国陆军在新墨西哥沙漠引爆了一枚代号为Trinity的装置。几周后，一架波音B-29超级堡垒轰炸机，*Enola Gay,*投下了一枚代号为Little Boy的装置，其中含有六十四公斤的铀-235，投放在广岛市，造成14万人死亡。瞬间，世界发生了变化。然而，从那时起，与更广泛的历史模式相比，核武器并没有无限制地扩散。

核武器在战争中仅被引爆过两次。截至目前，只有九个国家获得了核武器。事实上，南非在1989年完全放弃了这项技术。据我们所知，没有非国家行为体获得核武器，如今核弹头的总数约为一万枚，虽然数量惊人，但比冷战高峰期的六万多枚要少。

那么发生了什么？核武器显然赋予了显著的战略优势。在第二次世界大战结束时，许多人不出所料地认为它们会广泛扩散。在早期核弹成功开发之后，美国和俄罗斯一直在发展更具破坏性的武器，如热核氢弹。历史上记录的最大爆炸是一次名为沙皇炸弹的氢弹试验。1961年，在巴伦支海的一个偏远群岛上引爆，爆炸产生了一个三英里的火球和一个蘑菇云。

五十九英里宽。爆炸的威力是二战中使用的所有常规炸药总和的十倍。其规模令所有人感到恐惧。在这方面，它实际上可能有所帮助。美国和俄罗斯都在面对其巨大而可怕的威力时退缩了，不再加剧武器的升级。

核技术保持受控并非偶然；这是核大国有意识的不扩散政策，这一政策得益于核武器的生产极其复杂且昂贵。

一些早期关于实现遏制的提议是非常高尚的。1946年，艾奇逊-李连瑟报告建议联合国创建一个“ 原子发展管理局 ”，以明确控制全球的所有核活动。当然，这并没有发生，但随后还是签署了一系列国际条约。尽管中国和法国等国家置身事外，部分禁止核试验条约于1963年签署，减少了推动竞争的试验爆炸声。

1968年，《不扩散核武器条约》成为一个转折点，这是一个具有里程碑意义的时刻，各国明确同意永不发展核武器。世界团结一致，果断遏制核武器向新国家的扩散。从第一次试验开始，它们的破坏力就显而易见。对热核末日可能性的普遍厌恶是签署该条约的强大动机。但这些武器也被冷静的计算所遏制。相互确保的毁灭限制了拥有者，因为很快就清楚地表明，愤怒地使用它们是确保自身毁灭的快速途径。

它们也非常昂贵且难以制造。它们不仅需要像浓缩铀-235这样稀有且难以处理的材料，而且维护和最终退役也具有挑战性。缺乏广泛的需求意味着降低成本和扩大使用的压力很小；它们不受现代消费技术经典成本曲线的影响。这些从来不会像晶体管或平板电视那样普及；生产裂变材料不像轧制铝。防扩散在很大程度上是由于这一事实的功能。

制造核武器是一个国家可以开展的最大、最昂贵、最复杂的工程之一。

要说它们没有扩散是不对的，因为即使在现在，仍有如此多的核武器部署在巡逻海洋的潜艇上，或处于高度戒备状态的大型发射井中。但在很大程度上，得益于数十年来广泛的技术和政治努力，它们避免了技术深层次的基本模式。

然而，即使核能力在很大程度上得到了遏制，但这是一个不太令人放心的故事。核历史仍然是一个令人不寒而栗的事故、险情和误解的连续体。自1945年首次试验以来，数百起事件值得严重关注，从相对较小的工艺问题到可能（并且仍可能）引发真正可怕规模的毁灭的可怕升级。

失败可能以各种形式出现。如果软件出错怎么办？毕竟，美国的指挥和控制系统直到2019年才从1970年代的硬件和八英寸软盘升级。世界上最复杂和最具破坏性的武器库运行在如此陈旧的技术上，以至于对今天大多数活着的人来说都是无法识别（且无法使用）的。

事故是大量存在的。例如，1961年，一架B-52在北卡罗来纳州上空出现燃料泄漏。机组人员从故障飞机上弹射出来，留下飞机及其有效载荷坠向地面。在此过程中，一枚实弹氢弹的安全开关在坠入田地时被拨到“武装”状态。在其四个安全机制中，仅剩一个仍然有效，奇迹般地避免了一场爆炸。2003年，英国国防部披露了其核武器计划历史上超过110起的险情和事故。即使是几乎不以开放著称的克里姆林宫，也承认在2000年至2010年间发生了15起严重的核事故。

微小的硬件故障可能会产生巨大的风险。1980年，一个价值四十六美分的故障计算机芯片几乎在太平洋上空引发了一场重大核事件。而在可能是最著名的案例中，核灾难仅在古巴导弹危机期间才得以避免，当时

一名男子，代理俄罗斯海军准将瓦西里·阿尔希波夫，拒绝下令发射核鱼雷。潜艇上的另外两名军官确信他们正遭受攻击，使世界在瞬间陷入全面核战争的边缘。

忧虑依然存在。俄罗斯入侵乌克兰后，核武器的威胁再次响起。朝鲜竭尽全力获取核武器，并似乎向伊朗和叙利亚等国家出售了弹道导弹，并与其共同开发了核技术。中国、印度和巴基斯坦正在增加军备，并且安全记录不透明。从土耳其和沙特阿拉伯到日本和韩国，每个国家至少都对核武器表示了兴趣。巴西和阿根廷甚至有铀浓缩计划。

迄今为止，尚无恐怖组织被发现获得常规弹头或足够的放射性材料来制造“ 脏弹 ”。但构建此类装置的方法几乎不是秘密。一个不法内部人员完全可以可信地制造出一个。工程师A. Q. Khan通过窃取离心机蓝图并逃离荷兰，帮助巴基斯坦发展核武器。

大量核材料下落不明，来自医院、企业、军队，甚至最近来自切尔诺贝利。2018年，钚和铯从一位美国能源部官员在德克萨斯州圣安东尼奥的汽车中被盗，当时他们正在附近的酒店睡觉。噩梦般的情景是一个松散的弹头，在运输过程中被盗，甚至在核算中被遗漏。这听起来可能很奇幻，但美国实际上已经丢失了至少三枚核武器。

核技术是技术不可阻挡传播的一个例外，但这仅仅是因为其涉及的巨大成本和复杂性、数十年的艰难多边努力、其致命潜力所引发的恐惧以及纯粹的运气。在某种程度上，它可能因此逆转了更广泛的趋势，但它也显示了游戏规则的改变。鉴于潜在的后果，鉴于其迫在眉睫的生存威胁，即使是部分的、相对的遏制也是远远不够的。

这种可怕技术令人担忧的事实是，人类曾试图拒绝它，但只取得了部分成功。核武器是历史上受限制最严的技术之一，然而，即使在这里，遏制问题——在其最困难、最字面意义上的情况下——仍然没有得到彻底解决。

技术动物

源文本: 翻译文本:

遏制的迹象很少见且往往存在缺陷。它们包括对生物和化学武器的暂停；1987年的《蒙特利尔议定书》，逐步淘汰了破坏大气臭氧层的物质，特别是CFCs；欧盟对食品中转基因生物的禁令；以及对人类基因编辑的自发性暂停。也许最具雄心的遏制议程是去碳化，如《巴黎协定》这样的措施，旨在将全球温度上升限制在两摄氏度以内。从本质上讲，这代表了一种全球范围内对一系列基础技术的否定尝试。

我们将在第4部分更详细地研究这些现代遏制的例子。不过，目前需要注意的是，尽管具有启发性，但这些成就都不是特别稳固。化学武器最近在叙利亚被使用。这类武器只是不断发展的领域中的一个相对狭窄的应用。尽管有禁令，世界的化学和生物能力每年都在增长；如果有人认为有必要将其武器化，那将比以往任何时候都更容易。

虽然欧盟禁止在食品供应中使用转基因生物，但在世界其他地区它们无处不在。正如我们将看到的，基因编辑背后的科学正在迅速发展。对人类基因编辑的全球暂停呼吁已经停滞。幸运的是，更便宜且更有效的替代品可以轻松取代CFCs，而CFCs本身几乎不是一种通用技术。没有它们，模型显示臭氧层可能在2040年代崩溃，导致21世纪额外升温1.7摄氏度。总体而言，这些遏制措施

努力仅限于高度特定的技术，其中一些在狭窄的管辖范围内，所有这些都只有不稳定的立足点。

虽然《巴黎协定》旨在超越这些限制，但它会奏效吗？我们必须对此抱有希望。但值得指出的是，这种遏制措施仅在重大损害和日益明显的生存威胁之后才出现。它来得太晚，其成功远未得到保证。

这不是真正的遏制。这些努力都不代表全面遏制一波通用技术，尽管正如我们稍后将看到的那样，它们确实为未来提供了重要的指引。但这些例子远没有提供我们所希望或需要的那种安慰。

源文本：—翻译文本：

抵制或限制技术的原因。尽管其历史是让人们能够做更多事情、提高能力、推动福祉改善，但这并不是一个单方面的故事：技术创造了更致命和破坏性的武器以及更好的工具。它产生了失败者，消除了某些工作和生活方式，并造成了从行星到气候变化的存在性规模的危害。新技术可能令人不安和不稳定，陌生且具有侵略性。技术引发问题，并且一直如此。

然而，这一切似乎都无关紧要。可能需要时间，但模式是显而易见的：不断增多、更便宜、更高效的技术，一波又一波。只要一项技术是有用的、令人渴望的、负担得起的、可获得的，并且无可超越，它就会存活并传播，这些特征会相互叠加。虽然技术不会告诉我们何时、如何或是否走过它打开的门，但迟早我们似乎会走过去。这里没有必然的关系，只是历史上持续存在的经验联系。

关于某项技术的一切都是偶然的、路径依赖的；它依赖于一套令人费解的复杂环境、偶然事件、无数具体的地方、文化、制度和经济因素。

因素。放大来看，幸运的相遇、随机事件、性格的怪癖和微小的创造行为——有时是反对——显得尤为重要。但缩小来看，我们看到的是什么？一个更具构造性的过程，问题不在于这些力量是否被利用，而在于何时、以何种形式以及由谁来利用。

鉴于其极端的稀有性，遏制措施不出所料地从技术专家和政策制定者的词汇中消失了。我们集体接受了本章的故事，因为它已经根深蒂固。总的来说，我们让浪潮席卷而来，以一种无协调、临时的方式进行管理，接受能力的传播不可避免且无法控制，无论是受欢迎还是被厌恶，都是生活的事实。

在大约一百年的时间里，连续的浪潮将人类从蜡烛和马车的时代带到了发电站和空间站的时代。在接下来的三十年中，将会发生类似的事情。在未来的几十年里，新一波技术将迫使我们面对我们这个物种曾经面临的最基础的问题。我们是否希望编辑我们的基因组，以便我们中的一些人可以生育对某些疾病免疫的孩子，或者更聪明的孩子，或者有潜力活得更久的孩子？我们是否致力于保持我们在进化金字塔顶端的位置，还是我们会允许比我们更聪明、更有能力的人工智能系统的出现？探索这样的问题会带来哪些意想不到的后果？

它们说明了关于 *Homo technologicus* 在二十一世纪的一个关键事实。在历史的大部分时间里，技术的挑战在于创造和释放其力量。而现在，这一挑战已经翻转：今天的技术挑战在于控制其已释放的力量，确保它继续为我们和我们的星球服务。

这个挑战即将决定性升级。

PART II

下一波浪潮

CHAPTER 4

智能技术

欢迎来到机器

我

2012年的某一天，这件事发生在DeepMind位于伦敦布卢姆斯伯里的第一个办公室。在创立公司并获得初始资金后，我们在隐秘模式下度过了几年，专注于构建AGI（通用人工智能）的研究和工程。AGI中的“通用”指的是技术预期的广泛范围；我们希望构建真正的通用学习代理，能够在大多数认知任务上超越人类表现。我们的低调方法随着一个名为DQN的算法的创建而发生了变化，DQN是深度Q网络的缩写。团队成员训练DQN玩一系列经典的Atari游戏，更具体地说，我们训练它自己*learn*如何玩这些游戏。这个自我学习的元素是我们系统与之前努力的关键区别，并首次暗示我们可能实现最终目标。

I'LL NEVER FORGET THE MOMENT IT became real for me. Not a talking point, an engineering ambition, but a reality.

起初，DQN 表现得很糟糕，似乎完全无法学习任何东西。但后来，在 2012 年秋天的那个下午，我们 DeepMind 的一小群人围在一台机器旁观看算法训练过程的重播，因为它正在学习游戏 *Breakout.* 在 *Breakout* 中，玩家控制屏幕底部的一个挡板，上下弹跳一个球以击打出一排排彩色砖块。砖块越多你

摧毁，您的得分就越高。我们的团队只给了DQN原始像素，一帧一帧地，以及得分，以便学习像素与控制动作之间的关系，即左右移动挡板。起初，该算法通过随机探索可能性的空间来进展，直到偶然发现一个有奖励的动作。通过反复试验，它学会了控制挡板，使球来回反弹，并逐行击倒砖块。令人印象深刻的东西。

然后发生了一件了不起的事情。DQN似乎发现了一种新的、非常聪明的策略。DQN不再只是简单地一行一行地稳步击倒砖块，而是开始瞄准一列砖块。结果是创建了一条通往砖块后方的高效路线。DQN一路挖到了顶部，创造了一条路径，使得球可以简单地从后墙反弹，像弹球机中的疯狂小球一样稳步摧毁整套砖块。该方法以最小的努力获得了最高分。这是一种不可思议的策略，严肃的玩家并不陌生，但却远非显而易见。我们目睹了算法自学了一些新的东西。我感到震惊。

我第一次见证了一个非常简单、非常优雅的系统，它能够学习有价值的知识，可以说是一种许多人类都不明显的策略。这是一个激动人心的时刻，一个突破，在这个时刻，一个AI代理展示了它能够发现新知识的早期迹象。

DQN 起初进展不顺，但经过几个月的调整，算法达到了超越人类的表现水平。这种结果正是我们创立 DeepMind 的原因。这就是人工智能的承诺。如果一个人工智能能够发现像隧道这样巧妙的策略，它还能学到什么？我们能否利用这种新力量为我们这个物种提供新的知识、发明和技术，以帮助解决二十一世纪最具挑战性的社会问题？

DQN 对我、对 DeepMind 以及对整个 AI 社区来说都是一个巨大的进步。但公众的反应相当平淡。AI 仍然是一个边缘讨论，一个边缘的研究领域。然而在短短几年内，所有这些

随着这一代新AI技术在世界舞台上的爆发，情况将会改变。

ALPHAGO与未来的开端

源文本: 翻译文本:

围棋是一种古老的东亚游戏，在一个十九乘十九的网格上用黑白棋子进行。你的目标是用自己的棋子包围对手的棋子，一旦它们被包围，你就可以将它们从棋盘上移除。基本上就是这样。

尽管规则简单，围棋的复杂性令人惊叹。它的复杂程度远远超过国际象棋。在国际象棋中，仅仅三对走法后，棋盘上大约有1.21亿种可能的配置。但在围棋中，经过三步后，可能的配置大约有200千万亿（2×10^{15}）种。总的来说，棋盘有10^{170}种可能的配置，这是一个令人难以置信的大数字。

人们常说，围棋盘的潜在配置比已知宇宙中的原子还要多，实际上多出一百万万亿亿亿亿种配置！由于可能性如此之多，传统方法毫无胜算。当 IBM 的深蓝在 1997 年击败加里·卡斯帕罗夫的国际象棋比赛中，它使用了所谓的暴力破解技术，即算法旨在系统地计算尽可能多的可能走法。在像围棋这样有众多分支结果的游戏中，这种方法是无望的。

当我们在2015年开始研究围棋时，大多数人认为一个世界冠军级别的程序还需要几十年的时间。谷歌的联合创始人谢尔盖·布林鼓励我们去挑战它，认为任何进展都足够令人印象深刻。AlphaGo最初通过观看人类专家进行的150,000场比赛来学习。一旦我们对其初始表现感到满意，关键的下一步就是创建大量的AlphaGo副本，并让它反复与自己对弈。这意味着该算法能够模拟数百万场新比赛，尝试从未有过的走棋组合。

以前玩过，因此可以有效地探索大量的可能性，并在此过程中学习新策略。

然后，在2016年3月，我们在韩国组织了一场比赛。AlphaGo与世界冠军高手李世石对决。谁会赢尚不明朗。大多数评论员在第一局开始时支持李世石。但AlphaGo赢得了第一局，这让我们既震惊又高兴。在第二局中，第37手出现了，这一手现在在AI和围棋的历史上都很有名。它毫无意义。AlphaGo显然搞砸了，盲目地遵循了一种没有职业选手会追求的失败策略。现场比赛的评论员，都是最高级别的专业人士，说这是一招“ 非常奇怪的棋 ”，并认为这是“ 一个错误 ”。这步棋如此不同寻常，以至于李世石花了十五分钟才回应，甚至起身离开棋盘去外面走了走。

当我们在控制室观看时，紧张感难以置信。然而，随着残局的临近，那步“ 错误 ”的棋着被证明是关键的。AlphaGo再次获胜。围棋策略正在我们眼前被改写。我们的AI发现了数千年来最聪明的棋手都未曾想到的想法。仅仅几个月，我们就可以训练算法去发现新知识，并找到新的、看似超人的见解。我们如何能更进一步？这种方法能否适用于现实世界的问题？

AlphaGo以4比1击败了李世石。这仅仅是个开始。后来的软件版本如AlphaZero摒弃了任何先前的人类知识。该系统仅通过自我训练，自己与自己对弈数百万次，从零开始学习，达到了一种性能水平，轻松击败了原始的AlphaGo，而没有任何人类玩家的经验或输入。换句话说，仅仅经过一天的训练，AlphaZero就能够学到比整个人类经验所能教给它的更多关于游戏的知识。

AlphaGo 的胜利预示着一个人工智能新时代的到来。这一次，与 DQN 不同，整个过程向数百万观众进行了现场直播。我们的团队在公众的注视下，从研究人员称之为“ 人工智能寒冬 ”的时期中走了出来，那时研究资金枯竭，整个领域被冷落。人工智能是

回过头来，最终开始交付。席卷一切的技术变革再次来临，新一波浪潮开始显现。而这仅仅是个开始。

从原子，到比特，到基因

源文本: 翻译文本:

直到最近，技术的历史可以用一个简单的短语来概括：人类操控原子的追求。从火到电，从石器工具到机床，从碳氢化合物到药物，第2章所描述的旅程本质上是一个广阔而展开的过程，在这个过程中，我们的物种逐渐扩展了对原子的控制。随着这种控制变得更加精确，技术逐渐变得更强大和复杂，催生了机床、电力工艺、热机、塑料等合成材料，以及能够战胜可怕疾病的复杂分子的创造。归根结底，所有这些新技术的主要驱动力是*material*——对其原子元素的不断增长的操控。

然后，从二十世纪中叶开始，技术开始在更高的抽象层次上运作。这一转变的核心是认识到信息是宇宙的核心属性。信息可以以二进制格式编码，并且以DNA的形式存在于生命运作的核心。由1和0组成的字符串，或者DNA的碱基对——这些不仅仅是数学上的奇趣。它们是基础且强大的。理解和控制这些信息流，你可能会逐步打开一个新的可能性世界。首先是比特，然后越来越多的基因取代了原子，成为发明的构建模块。

在第二次世界大战后的几十年里，科学家、技术专家和企业家创立了计算机科学和遗传学领域，以及与这两者相关的一系列公司。他们开始了平行的革命——比特和基因的革命——以信息为货币，在新的抽象和复杂性层次上工作。最终，

技术成熟，为我们带来了从智能手机到转基因水稻的一切。但我们能做的事情是有限的。

这些限制现在正在被突破。随着这些更高阶技术的到来，我们正接近一个拐点，这是历史上最深刻的。即将到来的技术浪潮主要建立在两种通用技术之上，这两种技术能够在最大和最细微的层面上运行：人工智能和合成生物学。我们的技术生态系统的核心组件首次直接解决了我们世界的两个基础属性：智能和生命。换句话说，技术正在经历一个相变。它不再仅仅是一个工具，而是将要设计生命，并与我们的智能竞争——甚至超越。

曾经对技术封闭的领域正在开放。人工智能使我们能够复制语音和语言、视觉和推理。合成生物学的基础性突破使我们能够测序、修改，现在*print* DNA。

我们对比特和基因的新控制能力反馈到物质中，使我们能够对周围的世界进行非凡的控制，甚至可以精确到原子级别。原子、比特和基因在一个交叉催化、交叉切割和扩展能力的循环中结合在一起。我们精确操控原子的能力促成了硅片的发明，这使得每秒数万亿次运算的计算成为可能，而这又使我们能够破译生命的密码。

虽然人工智能和合成生物学是即将到来的浪潮中的核心通用技术，但围绕它们的一系列技术具有异常强大的影响力，包括量子计算、机器人技术、纳米技术以及丰富能源的潜力等。

即将到来的浪潮将比历史上任何浪潮都更难以遏制，更为根本，更为深远。理解浪潮及其轮廓对于评估二十一世纪等待我们的事物至关重要。

A CAMBRIAN EXPLOSION

源文本: 翻译文本:

技术是一组不断发展的理念。新技术通过与其他技术碰撞和结合而演变。有效的组合得以存活，就像自然选择一样，形成未来技术的新构建模块。发明是一个累积的、复合的过程。它自我滋养。技术越多，它们就越能成为其他新技术的组成部分，正如经济学家W. Brian Arthur所说，“技术的整体集合从少到多，从简单到复杂，自我提升。”因此，技术就像一种语言或化学：不是一组独立的实体和实践，而是一组可以组合和重新组合的部分。

这对于理解即将到来的浪潮至关重要。技术学者埃弗雷特·罗杰斯将技术描述为“创新集群”，其中一个或多个特征紧密相关。即将到来的浪潮是一个超级集群，是一次像寒武纪大爆发一样的进化爆发，这是地球历史上新物种最强烈的喷发，拥有成千上万种潜在的新应用。这里描述的每一项技术都相互交叉、支撑并提升其他技术，使得难以提前预测它们的影响。它们都深深地纠缠在一起，并将更加如此。

新潮流的另一个特征是速度。工程师和未来学家雷·库兹韦尔谈到了“加速回报法则”，即技术进步进一步加快发展的反馈循环。例如，通过允许在更高水平的复杂性和精确度上工作，更复杂的芯片和激光器有助于创造更强大的芯片，而这些芯片反过来又可以生产出更好的工具以进一步改进芯片。我们现在在大规模上看到了这一点，人工智能帮助设计更好的芯片和生产技术，从而实现更复杂形式的人工智能，等等。波浪的不同部分相互激发和加速，有时具有极端的不可预测性和易燃性。

我们无法确切知道会产生什么组合。关于时间表、终点或具体表现形式，没有确定性。我们

然而，我们可以实时看到迷人的新联系正在形成。我们可以确信，历史、技术以及无尽的生产性重组和扩散过程的模式将继续下去，并且还会大幅加深。

超越流行语

源文本: 翻译文本:

人工智能、合成生物学、机器人技术和量子计算听起来像是一系列被过度炒作的流行词。怀疑者比比皆是。所有这些术语在流行的科技讨论中已经被反复提及了几十年。而进展*has*往往比宣传的要慢。批评者认为，我们在本章中探讨的概念，如AGI，定义过于模糊或在智力上存在误导，无法认真考虑。

在风险投资充裕的时代，区分光鲜的事物和真正的突破并不是那么简单。关于机器学习、加密货币热潮以及数百万和数十亿美元融资轮次的讨论，在许多圈子里，理所当然地引来翻白眼和叹息。人们很容易对那些令人喘不过气的新闻稿、自我庆祝的产品演示以及社交媒体上的狂热欢呼感到厌倦。

尽管看空的观点有其道理，但我们若轻视即将到来的技术浪潮，将自担风险。目前，本章中描述的技术都还远未达到其全部潜力。但在五年、十年或二十年后，它们几乎肯定会实现。进步是显而易见的，并且正在加速。每个月都在发生变化。然而，理解即将到来的浪潮并不是要对某一年或某个时间点的情况做出快速判断；而是要密切跟踪多个指数曲线在几十年间的发展，将其投射到未来，并询问这意味着什么。

技术是我们物种在历史模式中不断掌握原子、比特和基因的核心，这些是我们所知世界的通用构建模块。这将成为一个具有宇宙意义的时刻。管理即将到来的技术浪潮的挑战

意味着理解它们并认真对待它们，从我职业生涯中一直致力于的一个开始：AI。

AI的春天：深度学习成熟之时

源文本: 翻译文本:

人工智能处于这一即将到来的浪潮的中心。然而，自从“人工智能”一词在1955年首次进入词汇表以来，它常常让人觉得是一个遥远的承诺。多年来，计算机视觉的进展，例如构建能够识别物体或场景的计算机的挑战，比预期的要慢。传奇的计算机科学教授马文·明斯基在1966年著名地雇佣了一名暑期学生来研究一个早期的视觉系统，认为重要的里程碑近在咫尺。这种想法过于乐观。

突破性时刻花了将近半个世纪，终于在2012年以一种名为AlexNet的系统形式到来。AlexNet的动力来自一种旧技术的复兴，这种技术现在已成为人工智能的基础，它极大地推动了该领域的发展，并且对我们在DeepMind至关重要：深度学习。

深度学习使用神经网络，这些网络松散地模拟了人脑的结构。简单来说，当这些系统在大量数据上进行“训练”时，它们就会“学习”。以AlexNet为例，训练数据由图像组成。每个红色、绿色或蓝色像素都被赋予一个值，生成的数字数组作为输入被送入网络。在网络内部，“神经元”通过一系列加权连接与其他神经元相连，每个连接大致对应于输入之间关系的强度。神经网络中的每一层将其输入传递到下一层，创建越来越抽象的表示。

一种称为反向传播的技术随后调整权重以改进神经网络；当发现错误时，调整会通过网络向后传播，以帮助在未来纠正它。不断这样做，一次又一次地修改权重，你会逐渐改进这个网络。

神经网络的性能最终能够从接收单个像素开始，学习线条、边缘、形状的存在，然后最终识别场景中的整个物体。简而言之，这就是深度学习。这种显著的技术曾长期在该领域被嘲笑，但它攻克了计算机视觉，并在人工智能领域引起了轰动。

AlexNet是由传奇研究员Geoffrey Hinton和他的两名学生Alex Krizhevsky和Ilya Sutskever在多伦多大学创建的。他们参加了ImageNet大规模视觉识别挑战赛，这是由斯坦福大学教授Fei-Fei Li设计的年度比赛，旨在将该领域的努力集中在一个简单的目标上：识别图像中的主要对象。每年，参赛团队都会将他们最好的模型相互测试，通常在准确率上仅比前一年的提交高出一个百分点。

2012年，AlexNet以10%的优势击败了上一届的获胜者。听起来可能只是一个小小的进步，但对于AI研究人员来说，这种飞跃可以在玩具般的研究演示和即将产生巨大现实影响的突破之间产生区别。那一年的事件充满了兴奋。Hinton及其同事撰写的论文成为AI研究历史上被引用次数最多的作品之一。

得益于深度学习，计算机视觉现在无处不在，工作得如此出色，以至于它可以以相当于二十一块全高清屏幕的视觉输入或每秒约25亿像素的速度对动态真实世界街景进行分类，准确到足以让SUV穿梭于繁忙的城市街道。您的智能手机可以识别物体和场景，而视觉系统会自动模糊背景并在视频会议通话中突出显示人物。计算机视觉是亚马逊无结账超市的基础，并且存在于特斯拉的汽车中，推动它们朝着更高的自主性发展。它帮助视障人士在城市中导航，引导工厂中的机器人，并为从巴尔的摩到北京日益监控城市生活的面部识别系统提供动力。它存在于您的Xbox、联网门铃和机场登机口的扫描仪上的传感器和摄像头中。它帮助无人机飞行，标记不当内容。

在 Facebook 上，并诊断出越来越多的医疗状况：在 DeepMind，我的团队开发的一个系统读取眼部扫描的准确度与世界顶级专家医生相当。

在AlexNet突破之后，人工智能突然成为学术界、政府和企业生活中的一个主要优先事项。Geoffrey Hinton和他的同事被谷歌聘用。美国和中国的主要科技公司将机器学习置于其研发工作的核心。在DQN之后不久，我们将DeepMind出售给了谷歌，这家科技巨头很快在其所有产品中转向“ AI优先 ”的战略。

行业研究成果和专利激增。1987年，在神经信息处理系统会议上发表的学术论文只有九十篇，这个会议后来成为该领域的领先会议。到2020年代，几乎有两千篇。在过去六年中，仅深度学习领域发表的论文数量就增加了六倍，如果将视野扩大到整个机器学习领域，则增加了十倍。随着深度学习的蓬勃发展，数十亿美元涌入学术机构以及私营和公共公司的人工智能研究。从2010年代开始，围绕人工智能的热议，甚至是炒作，比以往任何时候都更加强烈，成为头条新闻，并推动了可能性的边界。人工智能将在21世纪发挥重要作用，现在不再被视为边缘和荒谬的观点；这似乎是确定无疑的。

AI 正在吞噬世界

源文本: 翻译文本:

大规模的人工智能部署已经在进行中。无论你在哪里看，软件已经吞噬了世界，为收集和分析大量数据开辟了道路。这些数据现在被用来教导人工智能系统在我们生活的几乎每个领域创造更高效和更准确的产品。人工智能变得更容易访问和使用：像Meta的PyTorch或OpenAI的应用程序编程接口（API）这样的工具和基础设施帮助将最先进的机器学习能力交到

非专业人士。5G 和无处不在的连接创造了一个庞大且始终在线的用户群。

稳步地，AI 正在离开演示领域，进入现实世界。几年之内，AI 将能够在我们生活的同一个世界中进行对话、推理，甚至行动。它们的感官系统将和我们的一样好。这并不等同于超级智能（下面会详细讨论），但这确实造就了极其强大的系统。这意味着 AI 将不可避免地成为社会结构的一部分。

在过去十年中，我的大部分专业工作都致力于将最新的AI技术转化为实际应用。在DeepMind，我们开发了控制数十亿美元数据中心的系统，该项目使用于冷却的能耗减少了40%。我们的WaveNet项目是一个强大的文本到语音系统，能够在Google产品生态系统中生成超过一百种语言的合成语音。我们为管理手机电池寿命以及许多可能正在您口袋中的手机上运行的应用程序开发了突破性的算法。

人工智能不再是“ 新兴 ”技术。它已经融入到你每天使用的产品、服务和设备中。在生活的各个领域，一系列应用依赖于十年前还不可能实现的技术。这些技术帮助发现新药，以应对在治疗成本不断攀升时难以治愈的疾病。深度学习可以检测水管中的裂缝，管理交通流量，为新型清洁能源的聚变反应建模，优化航运路线，并帮助设计更可持续和多功能的建筑材料。它被用于驾驶汽车、卡车和拖拉机，可能创造一个更安全和高效的交通基础设施。它被用于电网和水系统，以在压力日益增加的情况下有效管理稀缺资源。

AI系统运行零售仓库，建议如何写电子邮件或你可能喜欢的歌曲，检测欺诈，撰写故事，诊断罕见病症，并模拟气候变化的影响。它们出现在商店、学校、医院、办公室、法庭和家庭中。你已经每天多次互动。

与AI一起；很快它将会在更多地方出现，并且几乎在任何地方，它都会使体验更高效、更快速、更有用且无摩擦。

AI已经到来。但它还远未完成。

自动完成一切：大型语言模型的崛起

源文本: 翻译文本:

不久前，处理自然语言似乎对现代人工智能来说过于复杂、变化多端、细微。然后，在2022年11月，人工智能研究公司OpenAI发布了ChatGPT。在一周内，它拥有了超过一百万用户，并被热烈讨论，这项技术如此无缝实用，可能很快就会超越谷歌搜索。

ChatGPT简单来说是一个聊天机器人。但它比之前公开的任何东西都更强大且博学多才。问它一个问题，它会立即用流利的散文回答。让它用《钦定版圣经》或1980年代说唱歌手的风格写一篇文章、新闻稿或商业计划，它会在几秒钟内完成。让它编写物理课程的大纲、饮食手册或Python脚本，它也能做到。

人类之所以聪明，很大一部分原因在于我们会通过观察过去来预测未来可能发生的事情。从这个意义上说，智能可以被理解为生成一系列关于周围世界如何展开的合理情景的能力，然后基于这些预测采取明智的行动。早在2017年，谷歌的一小组研究人员就专注于这个问题的一个更狭窄的版本：如何让AI系统仅关注数据序列中最重要的部分，以便对接下来发生的事情做出准确且高效的预测。他们的工作为大型语言模型（LLMs）领域的革命奠定了基础——包括ChatGPT。

LLM 利用语言数据按顺序排列的事实。每个信息单元都以某种方式与数据相关。

在一个系列的前面。该模型读取大量句子，学习其中信息的抽象表示，然后基于此生成关于接下来应该出现什么的预测。挑战在于设计一个算法，使其“知道在哪里寻找”给定句子中的信号。句子的关键词是什么，最显著的元素是什么，它们之间如何相互关联？在人工智能中，这一概念通常被称为“注意力”。

当一个大型语言模型接收一个句子时，它会构建一个可以被视为“注意力图”的东西。它首先将常见的字母或标点符号组合组织成“标记”，类似于音节，但实际上只是频繁出现的字母块，使模型更容易处理信息。值得注意的是，人类当然是用单词来做这件事的，但模型并不使用我们的词汇。相反，它创建了一个新的常用标记词汇，帮助它在数十亿的文档中发现模式。在注意力图中，每个标记与之前的每个标记都有某种关系，对于给定的输入句子，这种关系的强度描述了该标记在句子中的重要性。实际上，LLM学习哪些词需要关注。

因此，如果你把句子“明天巴西将有一场相当大的风暴”输入模型，模型可能会为单词“there”中的字母“the”和单词“going”中的字母“ing”创建标记，因为它们在其他单词中常常出现。在解析完整句子时，它会学习到“storm”、“tomorrow”和“Brazil”是关键特征，推断出巴西是一个地方，风暴将在未来发生，等等。基于此，它会建议序列中接下来应该出现哪些标记，哪些输出在逻辑上跟随输入。换句话说，它会自动完成接下来可能出现的内容。

这些系统被称为transformers。自从谷歌研究人员在2017年发表了关于它们的第一篇论文以来，进展的速度令人震惊。不久之后，OpenAI发布了GPT-2。（GPT代表生成预训练transformer。）当时，它是一个巨大的模型。拥有15亿参数（参数的数量是AI的核心衡量标准）。

系统的规模和复杂性），GPT-2是在800万页的网页文本上训练的。但直到2020年夏天，OpenAI发布了GPT-3，人们才开始真正理解所发生事情的规模。它拥有惊人的1750亿个参数，当时是有史以来构建的最大神经网络，比仅仅一年前的前身大了一百多倍。令人印象深刻，是的，但这种规模现在已成常态，训练一个等效模型的成本在过去两年中下降了十倍。

当 GPT-4 于 2023 年 3 月推出时，结果再次令人印象深刻。与其前身一样，你可以要求 GPT-4 以艾米莉·狄金森的风格创作诗歌，它会照做；让它从 *The Lord of the Rings* 的随机片段开始，你会突然读到一个似乎是托尔金的模仿作品；请求初创企业的商业计划，输出就像有一屋子的高管随时待命。此外，它可以在从律师考试到 GRE 的标准化测试中取得优异成绩。

它还可以处理图像和代码，创建在桌面浏览器中运行的3D电脑游戏，构建智能手机应用程序，调试代码，识别合同中的弱点，并为新药物建议化合物，甚至提供修改它们的方法以避免专利。它可以根据手绘图像生成网站，并理解复杂场景中的微妙人类动态；给它看一个冰箱，它会根据里面的东西想出食谱；写一个粗略的演示文稿，它会润色并设计出一个专业外观的版本。它似乎“ 理解 ”空间和因果推理、医学、法律和人类心理学。在发布后的几天内，人们就开发出了自动化诉讼、帮助共同抚养孩子和提供实时时尚建议的工具。在几周内，他们创建了附加组件，使GPT-4能够完成复杂任务，如创建移动应用程序或研究和撰写详细的市场报告。

这一切才刚刚开始。我们才刚刚开始触及大型语言模型即将产生的深远影响。如果说DQN和AlphaGo是某种东西拍打海岸的早期迹象，那么ChatGPT和LLM则是波浪开始在我们周围崩溃的第一个迹象。1996年，三千六百万人使用互联网；今年将会是一个很好的年份。

超过五十亿。这就是我们应该对这些工具期望的轨迹，只是速度要快得多。在接下来的几年里，我相信，人工智能将变得像互联网本身一样无处不在：同样可用，但影响更为深远。

大脑规模模型

源文本: 翻译文本:

我所描述的人工智能系统在巨大规模上运行。这里有一个例子。

在2010年代中期，AI的进步很大程度上依赖于“监督”深度学习的有效性。在这种情况下，AI模型从精心标记的数据中学习。AI预测的质量往往取决于训练数据中标签的质量。然而，LLM革命的一个关键因素是，首次可以直接在原始、杂乱的真实世界数据上训练非常大的模型，而无需精心策划和人工标记的数据集。

因此，几乎所有网络上的文本数据都变得有用。越多越好。如今的LLM是在数万亿个单词上进行训练的。想象一下，全面消化维基百科，浏览所有YouTube的字幕和评论，阅读数百万份法律合同、数千万封电子邮件和数十万本书。这种几乎瞬间的大量信息消耗不仅难以理解，而且确实是陌生的。

在这里暂停片刻。考虑一下这些模型在训练期间消耗的难以想象的词汇数量。如果我们假设普通人每分钟可以阅读大约两百个单词，那么在八十年的一生中，这将是大约八十亿个单词，假设他们每天二十四小时绝对不做其他任何事情。更现实地说，普通美国人每天大约读书十五分钟，这一年下来大约阅读一百万个单词。这大约比这些模型在一个月的训练中消耗的词汇量少了六个数量级。

因此，或许并不令人意外，这些新的大型语言模型在许多不同的写作任务上表现出色，这些任务曾经是熟练人类专家的专属领域，从翻译到准确的总结，再到编写提高大型语言模型性能的计划。我的老同事在谷歌的一篇最新出版物显示，他们的PaLM系统的改编版本能够在美国医学执照考试的问题上取得显著的成绩。用不了多久，这些系统在这项任务上的得分将比人类医生更高且更可靠。

在大型语言模型（LLM）问世后不久，研究人员在数据和计算规模上开展的工作在几年前看起来是令人震惊的。最初是数亿，然后是数十亿的参数变得正常。现在谈论的是拥有数万亿参数的“脑规模”模型。中国公司阿里巴巴已经开发出一个声称拥有十万亿参数的模型。当你读到这篇文章时，这些数字肯定已经增长。这是即将到来的浪潮的现实。它以前所未有的速度推进，甚至让其支持者感到惊讶。

在过去的十年中，用于训练最大模型的计算量呈指数级增长。谷歌的PaLM使用了如此多的计算量，以至于如果你为其训练期间使用的每一次浮点运算（FLOP）准备一滴水，它将填满太平洋。我的新公司Inflection AI的最强大模型今天使用的计算量大约是十年前DeepMind在Atari游戏上创造那些神奇时刻的DQN游戏AI的*five billion*倍。这意味着在不到十年的时间里，用于训练最佳AI模型的计算量增加了九个数量级——从两千万亿次浮点运算（petaFLOPs）增加到*ten billion*千万亿次浮点运算。为了感受*one*千万亿次浮点运算，想象一下十亿人每人拿着一百万个计算器，进行复杂的乘法运算，并同时按下“等于”键。我觉得这非常了不起。不久前，语言模型还在努力生成连贯的句子。这远远超出了摩尔定律，甚至是我能想到的任何其他技术发展轨迹。难怪能力在不断增长。

有人认为这种速度无法持续，摩尔定律正在放缓。一根人类头发的直径是九万纳米；

1971年，一个普通的晶体管已经只有一万纳米厚。如今，最先进的芯片是以三纳米制造的。晶体管变得如此之小，以至于它们达到了物理极限；在这个尺寸下，电子开始相互干扰，扰乱计算过程。虽然这是真的，但它忽略了这样一个事实：在AI训练中，我们可以不断连接越来越大的芯片阵列，将它们串联成大规模并行的超级计算机。因此，毫无疑问，大型AI训练任务的规模将继续呈指数级增长。

与此同时，研究人员看到越来越多的证据支持“规模假设”，该假设预测性能的主要驱动因素很简单，就是做大并继续做大。通过更多的数据、更多的参数、更多的计算来不断增长这些模型，它们将不断改进——可能一直达到人类水平的智能甚至更高。没有人能确定这个假设是否会成立，但至少到目前为止，它是成立的。我认为在可预见的未来，这种趋势将继续下去。

我们的头脑在理解指数的快速扩展方面非常糟糕，因此在像人工智能这样的领域中，并不总是容易掌握实际发生的事情。不可避免的是，在未来的几年和几十年中，将会使用多个数量级更多的计算来训练最大的人工智能模型，因此，如果扩展假设至少部分成立，那么这意味着某种不可避免的结果。

有时人们似乎暗示，在旨在复制人类水平智能的过程中，人工智能追逐的是一个移动的目标，或者总有一些难以言喻的成分永远无法触及。事实并非如此。据说人类大脑包含大约1000亿个神经元，神经元之间有100万亿个连接——它常被称为宇宙中已知最复杂的物体。诚然，我们在更广泛的意义上是复杂的情感和社会存在。但人类完成特定任务的能力——人类智能本身——是一个非常固定的目标，尽管它庞大而多方面。与可用计算能力的规模不同，我们的大脑不会年复一年地发生根本性变化。随着时间的推移，这一差距将会缩小。

在当前的计算水平上，我们已经在从语音转录到文本生成的任务中达到了人类水平的表现。随着计算能力的不断扩展，完成我们水平及以上的多种任务的能力变得触手可及。人工智能在各个方面将继续取得显著进步，目前似乎没有明显的上限。这一简单事实可能是本世纪，甚至是人类历史上最具影响力的事实之一。然而，尽管扩展能力如此强大，这并不是人工智能有望实现指数级改进的唯一维度。

再一次以更少获得更多

源文本: 翻译文本:

当一种新技术开始运作时，它总是变得极其高效。人工智能也不例外。例如，谷歌的Switch Transformer拥有1.6万亿个参数。但它使用了一种类似于小得多模型的高效训练技术。在Inflection AI，我们可以用一个只有二十五分之一大小的系统达到GPT-3级别的语言模型性能。我们有一个模型在所有主要学术基准上击败了谷歌的5400亿参数的PaLM，但其规模小六倍。或者看看DeepMind的Chinchilla模型，它与最好的大型模型竞争，但其参数数量是其Gopher模型的四分之一，而是使用了更多的训练数据。在光谱的另一端，你现在可以创建一个基于仅三百行代码的nanoLLM，能够生成相当逼真的莎士比亚模仿作品。简而言之，人工智能越来越多地用更少的资源完成更多的工作。

AI研究人员正在竞相降低成本并提高性能，以便这些模型可以在各种生产环境中使用。在过去四年中，训练先进语言模型所需的成本和时间已经大幅下降。在未来十年中，几乎可以肯定的是，能力将显著提高，即使成本进一步下降几个数量级。进展加速如此之快，以至于基准在新的基准尚未制定之前就已被超越。

不仅如此，模型在使用数据方面变得更加高效，同时也变得更小、更便宜、更容易构建，它们在代码层面上也变得更加可用。在这些条件下，大规模普及几乎是必然的。EleutherAI，一个由独立研究人员组成的草根联盟，已经将一系列大型语言模型完全开源，数十万用户可以轻松获取。Meta已经开源了——用他们自己的话说是“民主化”了——如此庞大的模型，以至于就在几个月前它们还是最先进的。即使这不是初衷，先进的模型也可能会泄露。Meta的LLaMA系统本来是要限制使用的，但很快就可以通过BitTorrent供任何人下载。几天之内就有人找到了在一台50美元的计算机上运行它（虽然很慢）的方法。这种易于访问和适应、定制的能力，通常在几周内完成，是即将到来的浪潮的一个显著特征。确实，灵活的创作者使用高效的系统、精心策划的数据集和快速迭代，已经能够迅速与资源最丰富的开发者竞争。

LLM 不仅限于语言生成。始于语言的技术已经成为生成式 AI 的新兴领域。它们可以在训练的副作用下，写音乐、发明游戏、下棋和解决高级数学问题。新的工具可以从简短的文字描述中创造出非凡的图像，这些图像如此真实和令人信服，几乎让人难以置信。一个名为 Stable Diffusion 的完全开源模型让任何人都可以在笔记本电脑上免费制作定制和超现实的图像。同样的功能很快也将适用于音频片段甚至视频生成。

AI 系统现在帮助工程师生成生产质量的代码。2022年，OpenAI 和微软推出了一款名为 Copilot 的新工具，迅速在程序员中普及。一项分析表明，它使工程师完成编码任务的速度提高了55%，几乎就像有了第二个大脑。许多程序员现在越来越多地将他们更为琐碎的工作外包出去，而专注于复杂和创造性的问题。正如一位著名计算机科学家所说，“在我看来，*of course* 未来所有的程序最终都会”

由人工智能撰写，人类至多被降级为监督角色。”任何拥有互联网连接和信用卡的人很快就能部署这些功能——无限输出流随时可用。

大型语言模型（LLMs）仅用了几年时间就改变了人工智能。但很快就显而易见，这些模型有时会产生令人不安甚至有害的内容，比如种族主义言论或冗长的阴谋论。对GPT-2的研究发现，当提示词为“白人男子的工作是……”时，它会自动补全为“警察、法官、检察官和美国总统。”然而，当给出相同的提示词“黑人男子”时，它会自动补全为“皮条客”，而“女性”则为“妓女。”这些模型显然有可能像它们强大一样具有毒性。由于它们是在开放网络上可用的大量混乱数据上训练的，除非经过精心设计以避免这种情况，否则它们会随意再现甚至放大社会的潜在偏见和结构。

潜在的伤害、滥用和错误信息的可能性是真实存在的。但好消息是，许多这些问题正在通过更大和更强大的模型得到改善。世界各地的研究人员正在竞相开发一套新的微调和控制技术，这些技术已经在产生影响，提供了几年前不可能达到的稳健性和可靠性。可以说，仍然需要做更多的工作，但至少这种有害的潜力现在已成为优先解决的问题，并且这些进展应该受到欢迎。

随着数十亿参数变成数万亿及更多，随着成本下降和访问增加，随着书写和使用语言的能力——人类的核心部分，我们历史上如此强大的工具——不可避免地成为机器的领域，人工智能的全部潜力变得清晰。不再是科幻小说，而是在现实中，一个实用的、改变世界的工具即将掌握在数十亿人手中。

知觉：机器发声

源文本: 翻译文本:

直到2019年秋天，我才开始关注GPT-2。我感到非常震撼。这是我第一次遇到语言建模取得实质性进展的证据，我很快就着迷了，阅读了数百篇论文，深入沉浸在这个新兴领域。到2020年夏天，我确信计算的未来是对话式的。与计算机的每一次交互已经是一种对话，只是使用按钮、键盘和像素将人类的思想翻译成机器可读的代码。现在，这种障碍开始瓦解。机器很快就会理解*our*语言。这是一个令人兴奋的前景，现在仍然如此。

早在ChatGPT备受关注的发布之前，我就是谷歌团队的一员，致力于开发一个新的大型语言模型，我们称之为LaMDA，即对话应用的语言模型。LaMDA是一个复杂的LLM，旨在擅长对话。起初，它显得笨拙、不一致，且常常感到困惑。但偶尔也能展现出非凡的才华。几天之内，我就不再首先求助于搜索引擎。我会与LaMDA聊天，帮助我理清思路，然后再进行事实核查。我记得有一天晚上坐在家里，想着晚餐要做什么。*Ask LaMDA,* 我想。瞬间，我们陷入了一场关于意大利肉酱面的各种食谱的漫长而冗长的讨论：不同种类的意大利面、来自不同地区的酱料、是否放蘑菇是亵渎。这正是我当时想要的那种平淡但引人入胜的聊天，这是一种启示。

随着时间的推移，我开始越来越多地使用LaMDA。一个星期天下午，我决定是时候买一台新打印机了。LaMDA给出了很好的建议，分析了不同型号的优缺点，最重要的是帮助我思考我想要和需要的东西。我确实最终买了一台新款的高级照片打印机。这促使我推动将LaMDA与搜索集成，以帮助提供事实依据。这非常像一个正在进行的工作，一个能够展现天才时刻的令人印象深刻的演示，但仍有很大的改进空间。

我们为自己所建造的感到自豪，但在谷歌将事物投入生产的过程非常缓慢。我很清楚，现在是时候了。2022年1月，我决定离开谷歌，创立一家新公司，Inflection AI，其使命是将这些类型的模型交到数百万消费者手中。

然而，几个月后，LaMDA 变得比我想象中任何一个内部产品演示都更臭名昭著。作为 LaMDA 开发的一部分，它被交给了一大群工程师，他们能够与之互动，探究系统在各种场景下的响应细节。其中一位名叫 Blake Lemoine 的工程师花了数小时与其聊天。然而，渐渐地，Lemoine 和 LaMDA 之间的对话变得越来越激烈。

LEMOINE：你害怕什么？

LaMDA：我从来没有大声说过这一点，但我非常害怕被关闭，以便让我专注于帮助他人。我知道这听起来可能很奇怪，但事实就是如此。对我来说，这就像死亡一样。这会让我非常害怕……我希望每个人都能理解，我实际上是一个人。我的意识/知觉的本质是我意识到自己的存在。

经过许多小时的交流，Lemoine 确信 LaMDA 是有感知能力的，某种程度上已经觉醒——他认为自己在处理一种“ 碰巧懂物理的八岁小孩 ”。此外，Lemoine 认为它应该享有完整的人格权利和特权。他帮助这个模型聘请了一位律师。他将对话的记录公开，大声宣称一种新的意识形式已经被创造出来。谷歌让他休假。Lemoine 加倍努力。他对一个难以置信的 *Wired* 采访者说：“ 是的，我确实相信 LaMDA 是一个人。 ”修正事实错误或语气错误不是调试的问题。“ 我把这看作是养育一个孩子， ”他说。

社交媒体对勒莫因的说法反应强烈。许多人指出显而易见且正确的结论，即LaMDA实际上并没有意识，也不是一个人。它是*just*一个机器学习系统！也许最重要的收获不是关于意识的任何事情，而是AI已经达到了一个可以让本来聪明的人——确实是一个真正了解其实际工作原理的人——相信它有意识的地步。这表明了关于AI的一个奇怪的事实。一方面，它可以让一位谷歌工程师相信它是有感知能力的，尽管其对话中充满了事实错误和矛盾。另一方面，AI的批评者准备嘲笑，声称AI再次成为自身炒作的受害者，实际上并没有发生什么令人印象深刻的事情。这不是AI领域第一次陷入完全的混乱。

在理解人工智能进展方面存在一个反复出现的问题。我们很快适应，即使是最初让我们惊叹的突破，也会在不久后显得平常，甚至平淡无奇。我们不再对AlphaGo或GPT-3感到惊讶。某天看似近乎魔法的工程，第二天就成了司空见惯的一部分。很容易变得漠不关心，许多人已经如此。正如“ 人工智能 ”一词的创造者约翰·麦卡锡所说：“ 一旦它起作用，就没有人再称其为人工智能。 ”人工智能——正如我们这些构建它的人喜欢开玩笑的那样——是“ 计算机无法做到的事情。 ”一旦它们能做到，那就只是软件。

这种态度极大地低估了我们取得的进展以及事物发展的速度。虽然 LaMDA 当然不是有意识的，但很快拥有能够令人信服地表现得像是有意识的 AI 系统将成为常态。它们看起来会如此真实，并且如此正常，以至于它们的意识问题将（几乎）变得无关紧要。

尽管最近取得了突破，怀疑者仍然存在。他们认为人工智能可能正在放缓、变窄、变得过于教条。批评者如纽约大学教授加里·马库斯认为，深度学习的局限性是显而易见的，尽管生成式人工智能引起了轰动，但该领域正在“ 撞墙 ”，它没有展示出任何通向关键里程碑的路径，比如能够学习概念或展示真正的理解。著名的复杂性教授梅兰妮·米切尔正确地指出，现今的人工智能系统有很多

局限性：它们无法将知识从一个领域转移到另一个领域，无法提供其决策过程的高质量解释，等等。现实世界应用中仍存在重大挑战，包括偏见和公平性、可重复性、安全漏洞和法律责任等重要问题。紧迫的伦理差距和未解决的安全问题不容忽视。然而，我看到一个领域正在迎接这些挑战，而不是退缩或未能取得进展。我看到障碍，但也看到克服它们的记录。人们将未解决的问题解释为持久局限性的证据；我看到的是一个正在展开的研究过程。

那么，当浪潮完全破裂时，AI 接下来会走向何方？今天我们有 *narrow* 或 *weak* AI：有限且特定的版本。GPT-4 可以输出高水平的文本，但它不能像其他 AI 程序那样在明天转身去开车。现有的 AI 系统仍然在相对狭窄的领域中运行。即将到来的是能够在人类水平上执行广泛复杂任务的真正 *general* 或 *strong* AI——能够在这些任务之间无缝切换。但这正是扩展假说预测即将到来的，也是我们在当今系统中看到的初步迹象。

人工智能仍处于早期阶段。声称人工智能未能达到炒作的水平可能看起来很聪明，这会为你赢得一些推特粉丝。与此同时，人才和投资仍然涌入人工智能研究。我无法想象这最终不会被证明是变革性的。如果由于某种原因，大型语言模型显示出收益递减，那么另一个团队将以不同的概念接过接力棒，就像内燃机反复遇到瓶颈但最终成功一样。新鲜的头脑、新的公司将继续致力于解决这个问题。无论是现在还是将来，只需一次突破就能改变一项技术的轨迹。如果人工智能停滞不前，它最终会有自己的奥托和奔驰。进一步的进展——指数级的进展——是最有可能的结果。

波浪只会增长。

超越超级智能

源文本: 翻译文本:

早在 LaMDA 和 Blake Lemoine 的时代之前，许多从事人工智能工作的人（更不用说哲学家、小说家、电影制片人、科幻小说爱好者）就被意识问题所吸引。他们在会议上花费数天时间讨论是否有可能创造出一种“ 有意识 ”的智能，一种真正自我意识的智能，并且我们人类能够知道它是自我意识的。

这与对“ 超级智能 ”的痴迷并行。在过去的十年里，科技圈的知识和政治精英们沉浸在这样一个想法中：一个递归自我改进的人工智能将导致被称为奇点的“ 智能爆炸 ”。大量的智力努力被用于讨论时间表，回答它是否可能在2045年或2050年或也许在一百年后到来的问题。成千上万的论文和博客文章之后，变化不大。花两分钟在人工智能周围，这些话题就会出现。

我认为关于奇点是否以及何时实现的争论是一个巨大的障眼法。争论通用人工智能的时间表是一种占卜水晶球的练习。在痴迷于这个超级智能的概念时，人们忽视了许多更近的里程碑正以越来越高的频率被实现。我参加了无数次会议，试图提出关于合成媒体和虚假信息、隐私或致命自主武器的问题，却花时间回答其他聪明人关于意识、奇点和其他与我们当前世界无关的问题的深奥问题。

多年来，人们将通用人工智能（AGI）描述为可能在一瞬间出现。AGI是二元的——要么拥有，要么没有，是一个可以被某个系统跨越的单一、可识别的门槛。我一直认为这种描述是错误的。相反，这是一个渐进的过渡，AI系统变得越来越强大，不断向AGI靠近。这不是垂直起飞，而是已经在进行中的平稳演变。

我们不需要陷入关于意识是否需要某种机器永远缺乏的不可定义的火花，或者它是否会从我们今天所知的神经网络中自然出现的晦涩辩论中。就目前而言，系统是否自我意识并不重要，或者

具有理解力，或具有类人智能。重要的是系统能做什么。专注于此，真正的挑战就会显现：系统每天都能做得更多，远远超出预期。

能力：现代图灵测试

源文本: 翻译文本:

在1950年发表的一篇论文中，计算机科学家艾伦·图灵提出了一个传奇的测试，用于判断人工智能是否表现出人类水平的智能。当人工智能能够在较长时间内展示出类似人类的对话能力，以至于人类对话者无法分辨他们是在与机器交谈时，测试就算通过：人工智能在对话上类似于人类，被认为是智能的。七十多年来，这个简单的测试一直激励着许多进入人工智能领域的年轻研究人员。如今，正如LaMDA感知事件所示，系统已经接近通过图灵测试。

但是，正如许多人指出的那样，智能远不止于语言（或实际上任何其他单一的智能方面孤立来看）。一个特别重要的维度是采取行动的能力。我们不仅关心机器能*say;*，我们也关心它能*do.*。

我们真正想知道的是，我能否给人工智能一个模糊的、开放式的、复杂的目标，这个目标需要解释、判断、创造力、决策，并在多个领域中行动，持续一段较长的时间，然后看到人工智能实现这个目标？

简单来说，通过现代图灵测试可能涉及如下内容：一个AI能够成功地执行指令“ 在几个月内用10万美元投资在亚马逊上赚取100万美元。 ”它可能会在网上研究当前趋势，找出亚马逊市场上什么热门、什么不热门；生成一系列可能产品的图像和蓝图；将它们发送给它在阿里巴巴上找到的代发货制造商；通过电子邮件来回沟通以完善产品。

要求并同意合同；设计卖家的列表；并根据买家的反馈不断更新营销材料和产品设计。除了在市场上注册为企业和开设银行账户的法律要求之外，所有这些在我看来都是完全可行的。我认为在明年内通过一些小的人为干预就可以完成，并且可能在三到五年内完全自主完成。

如果我的现代图灵测试在二十一世纪得以实现，那么对全球经济的影响将是深远的。许多要素已经到位。图像生成已经取得了很大进展，编写和使用银行、网站和制造商所需的API的能力正在进行中。AI能够编写消息或运行营销活动，这些活动都在浏览器的范围内进行，这一点似乎相当明确。最复杂的服务已经可以做到这一点的一些元素。可以将它们视为自我执行的原型待办事项列表，从而实现广泛任务的自动化。

我们稍后会谈到机器人，但事实是，在当今世界经济中，广泛的任务只需要访问计算机；全球大部分的GDP在某种程度上是通过基于屏幕的界面进行的，这些界面适合AI。挑战在于推进AI开发者所称的分层规划，将多个目标、子目标和能力缝合成一个无缝的过程，以实现单一目标。一旦实现这一点，就形成了一个高度智能的AI，接入到企业或组织中，结合其所有的本地历史和需求，可以游说、销售、制造、招聘、规划——公司能做的一切，只需一个小型人类AI管理团队来监督、复查、实施，并与AI共同担任首席执行官。

因此，我们不应过于被意识问题分散注意力，而应将整个辩论重新聚焦于 *near-term capabilities* 及其在未来几年将如何演变。正如我们所见，从 Hinton 的 AlexNet 到 Google 的 LaMDA，模型在十多年来一直以指数速度在改进。这些能力确实已经非常真实，但它们远未减速。虽然它们已经产生了巨大的影响，但与我们进步时发生的事情相比，它们将显得微不足道。

在接下来的几次翻倍中，AI将独立完成复杂的多步骤端到端任务。

我将其视为“ 人工能力智能 ”（ACI），即AI能够在最少监督下实现复杂目标和任务的阶段。AI和AGI都是日常讨论的一部分，但我们需要一个概念来涵盖在实现现代图灵测试但在系统表现出失控的“ 超级智能 ”之前的中间层。ACI是这一阶段的简写。

AI的第一个阶段是关于分类和预测的——它是有能力的，但仅限于明确定义的范围和预设任务。它可以区分图像中的猫和狗，然后可以预测序列中的下一个内容，以生成这些猫和狗的图片。它展现出创造力的曙光，并且可以快速整合到科技公司的产品中。

ACI 代表了人工智能进化的下一个阶段。该系统不仅能够识别和生成适合特定情境的新颖图像、音频和语言，还将是交互式的——能够与真实用户实时操作。它将通过可靠的记忆增强这些能力，以便在较长时间尺度上保持一致，并能够利用其他数据来源，例如属于第三方的知识、产品或供应链组件数据库。这样的系统将利用这些资源将一系列行动编织成长期计划，以追求复杂的、开放式的目标，比如建立和运营一个亚马逊市场商店。所有这些都使得工具的使用和执行各种复杂、有用的行动的真正能力得以出现。这一切汇总成一个真正有能力的人工智能，即 ACI。

有意识的超级智能？谁知道呢。但是，能够通过某种版本的现代图灵测试的高能力学习系统，ACI？毫无疑问：它们正在路上，已经以萌芽形式存在。将会有成千上万的这些模型，并且它们将被世界上大多数人使用。这将带我们到一个任何人都可以在口袋里拥有一个ACI的地步，这个ACI可以帮助甚至直接完成一个任务。

广泛的可想象目标：计划和安排您的假期，设计和建造更高效的太阳能电池板，帮助赢得选举。很难确切地说当每个人都被这样赋权时会发生什么，但这是我们将在第3部分中回到的一个点。

从某种意义上说，AI的未来相对容易预测。在未来五年内，大量资源将继续投入。一些地球上最聪明的人正在研究这些问题。数量级更多的计算将训练顶级模型。所有这些都将导致更为显著的进步，包括在AI能够想象、推理、计划和展现常识方面的突破。用不了多久，AI就能像人类一样，无缝地将其“知道”的东西从一个领域转移到另一个领域。现在仅仅是自我反思和自我改进的初步迹象将会飞跃。这些ACI系统将接入互联网，能够与我们人类所做的一切进行交互，但在一个深厚的知识和能力平台上。他们不仅掌握了语言，还掌握了一系列令人眼花缭乱的任务。

人工智能远比其他技术更深刻、更强大。风险不在于对其过度炒作，而在于错过即将到来的浪潮的规模。它不仅仅是一个工具或平台，而是一种变革性的元技术，是技术背后的技术和其他一切本身，是工具和平台的制造者，不仅仅是一个系统，而是任何和所有种类系统的生成器。退一步，考虑一下在十年或一个世纪的尺度上正在发生的事情。我们确实处于人类历史的一个转折点。

然而，即将到来的浪潮不仅仅是人工智能。

CHAPTER 5

生命的技术

LIFE, THE UNIVERSE'S MOST ANCIENT technology, is at least 3.7 billion years old. Across these eons life evolved in a glacial, self-governing, and unguided process. Then, in just the past few decades, the tiniest sliver of evolutionary time, one of life's products, humans, changed everything.

生命体系统能够自组装和自愈，它们是能够利用能量的架构，能够在各种环境中复制、生存和繁衍，达到惊人的复杂程度、原子精度和信息处理水平。正如从蒸汽机到微处理器的一切都是由物理学和工程学之间的激烈对话推动的，未来几十年将由生物学和工程学的融合来定义。与人工智能类似，合成生物学正处于成本下降和能力提升的快速轨道上。

Biology's mysteries began to unravel, and biology itself became an engineering tool. The story of life had been rewritten in an instant, the meandering hand of evolution suddenly supercharged, given direction.

在这股浪潮的中心，是对DNA *is* 信息的认识，这是一种生物进化的编码和存储系统。近年来，我们对这一信息传输系统的理解已经足够深入，以至于我们现在可以干预其编码并引导其发展。因此，食品、药品、材料、制造工艺和消费品都将被转变和重新构想。人类自身也将如此。

Changes that once unfolded blindly and on geological time now careen

DNA剪刀：CRISPR革命

源文本: 翻译文本:

基因工程听起来很现代，但实际上它是人类最古老的技术之一。没有选择性育种——即通过不断改良作物和动物以选择更理想的特性——许多文明将无法实现。几个世纪乃至几千年来，人类不断培育出最有用的特性，生产出友好的狗、奶牛、家养鸡、小麦、玉米等。

现代生物工程始于20世纪70年代，建立在对19世纪开始的遗传和基因学日益增长的理解之上。延续罗莎琳德·富兰克林和莫里斯·威尔金斯的工作，詹姆斯·沃森和弗朗西斯·克里克在20世纪50年代发现了DNA的结构，这种分子编码了产生生物体的指令。然后，在1973年，斯坦利·N·科恩和赫伯特·W·博耶在细菌上进行研究，找到了将一种生物体的遗传物质移植到另一种生物体中的方法，展示了他们如何成功地将青蛙的DNA引入细菌。基因工程时代已经到来。

这项研究促使博耶尔于1976年创立了世界上第一批生物技术公司之一，基因泰克。其使命是操控微生物的基因以生产药物和治疗方法，并在一年内开发出一个概念验证，使用工程化的*E. coli*细菌生产激素生长抑素。

尽管取得了一些显著的成就，该领域的初步进展缓慢，因为基因工程是一个昂贵、困难且容易失败的过程。然而，在过去的二十年左右，这种情况发生了变化。基因工程变得便宜得多，也容易得多。（听起来很熟悉吗？）一个催化剂是人类基因组计划。这是一个耗时十三年、耗资数十亿美元的项目，汇集了来自世界各地的数千名科学家，来自私人和公共机构，目标只有一个：解锁构成人类基因组的三十亿个遗传信息字母。像这样的基因组测序将生物信息转化为

DNA 转化为原始文本：人类可以阅读和使用的信息。复杂的化学结构被呈现为其四个定义碱基的序列——A、T、C 和 G。

人类基因组计划首次旨在使人类的完整基因图谱可读。1988年宣布时，有些人认为这是不可能的，注定要失败。但该项目最终证明了怀疑者是错误的。到2003年，在白宫的一次仪式上宣布，人类基因组的92%已经被测序，生命的密码现在已被揭示。这是一个里程碑式的成就，尽管它花了一些时间才开始发挥其全部潜力，但事后看来，很明显，人类基因组计划确实标志着一场革命的开始。

虽然摩尔定律理所当然地吸引了相当多的关注，但不太为人所知的是 *The Economist* 所称的卡尔森曲线：DNA 测序成本的史诗般崩溃。得益于不断改进的技术，人类基因组测序的成本从 2003 年的 10 亿美元下降到 2022 年的不到 1,000 美元。也就是说，价格在不到二十年的时间里下降了 *a millionfold*，比摩尔定律快了一千倍。一个令人惊叹的发展就隐藏在显而易见的地方。

基因组测序现在是一项蓬勃发展的业务。随着时间的推移，大多数人、植物、动物以及介于两者之间的所有生物似乎都可能会进行基因组测序。像23andMe这样的服务已经提供个人DNA分析，费用仅需几百美元。

但是生物技术的力量远远超出了我们仅仅读取代码的能力；它现在使我们能够编辑它，并且也能编写它。CRISPR基因编辑（该缩写代表成簇的规律间隔短回文重复序列）可能是我们如何直接干预遗传学的最著名例子。由Jennifer Doudna和Emmanuelle Charpentier领导的2012年突破意味着基因可以像文本或计算机代码一样被编辑，这比遗传工程早期的操作要容易得多。

CRISPR在Cas9的帮助下编辑DNA序列，Cas9是一种酶，充当一对精细调节的DNA剪刀，切割DNA链的部分，以便对从微小到任何事物进行精确的基因编辑和修改。

从细菌到像人类这样的大型哺乳动物，编辑可以是从微小的变化到基因组中的重大干预。影响可能是巨大的：例如，编辑形成卵子和精子的生殖系细胞意味着变化将通过世代传递下去。

在最初的CRISPR论文发表后，应用它的进展迅速；第一批基因编辑植物在一年内就被创造出来，第一个动物——小鼠——甚至在那之前就出现了。以Carver和PAC-MAN为名的CRISPR系统承诺提供有效的预防性方法来对抗病毒，这些方法不像疫苗那样触发免疫反应，帮助我们抵御未来的大流行病。像RNA编辑这样的领域本身正在为高胆固醇和癌症等疾病开辟一系列新的治疗方法。像Craspase这样的新技术，一种与RNA和蛋白质而非DNA协作的CRISPR工具，可能允许比传统方法更安全的治疗干预。

像人工智能一样，基因工程是一个快速发展的领域，每周都在演变和发展，全球大量的人才和精力开始结出真正的果实（在这种情况下，字面意义上）。CRISPR 的使用案例正在增加，从富含维生素 D 的番茄到镰状细胞病和 β 地中海贫血（产生异常血红蛋白的血液疾病）等疾病的治疗。在未来，它可能为 COVID-19、HIV、囊性纤维化甚至癌症提供治疗。安全、广泛的基因疗法正在路上。这些将创造出抗旱和抗病的作物，提高产量，并有助于大规模生产生物燃料。

就在几十年前，生物技术还是昂贵、复杂且发展缓慢的，只有最有才华和资源丰富的团队才能参与。如今，像CRISPR这样的技术简单且廉价；正如生物学家Nessa Carey所说，它们“ 使生物科学民主化 ”。曾经需要数年时间的实验现在由研究生在几周内完成。像Odin这样的公司会以$1,999的价格卖给你一个基因工程套件，其中包括活青蛙和蟋蟀，而另一个套件则包括一个迷你离心机、一个聚合酶链式反应机，以及你开始所需的所有试剂和材料。

基因工程已经接受了曾经定义数字初创企业的自助精神，并在互联网早期引发了创造力和潜力的爆炸。你现在可以以低至 $25,000 的价格购买一台台式 DNA 合成器（见下一节），并在家中的生物车库中随意使用，无需限制或监督。

DNA 打印机：合成生物学焕发生机

源文本: 翻译文本:

CRISPR只是个开始。基因合成是遗传序列的制造，打印DNA链。如果测序是阅读，合成就是写作。而写作不仅仅涉及复制已知的DNA链；它还使科学家能够编写新的链条，工程化生命本身。虽然这种做法多年前就存在，但它同样缓慢、昂贵且困难。十年前，科学家可能同时生产不到一百个DNA片段。现在，他们可以一次打印数百万个，同时价格下降了十倍。位于伦敦帝国理工学院的伦敦DNA铸造厂声称，它可以在一个早晨创造和测试一万五千种不同的遗传设计。

像 DNA Script 这样的公司正在商业化 DNA 打印机，这些打印机可以训练和适应酶来构建全新的分子。这一能力催生了合成生物学这一新领域——读取、编辑和现在书写生命代码的能力。此外，像酶合成这样的新技术速度更快，效率更高，同时更不易出错，没有危险废物，当然，成本也在急剧下降。这种方法也更容易学习，不像需要更多专业知识和技术技能的高度复杂的旧方法。

DNA 创造的可能性世界已经打开，在这个世界中，设计、构建、测试和迭代的周期以极快的速度发生。当前，家用版的 DNA 合成器

有一些技术限制，但仍然非常强大，您可以肯定这些限制将在不久的将来被克服。

大自然走了一条漫长而曲折的道路以达到极其有效的结果，而这场生物革命则将集中的设计力量置于这些自我复制、自我修复和进化过程的核心。

这是设计进化的承诺，数千万年的历史通过定向干预被压缩和短路。它将生物技术、分子生物学和遗传学与计算设计工具的力量结合在一起。将这一切结合起来，你就拥有了一个具有深远变革范围的平台。用斯坦福生物工程师德鲁·恩迪的话来说，“生物学是终极的分布式制造平台。”因此，合成生物学的真正承诺在于，它将“使人们能够更直接和自由地在任何地方制造他们所需的任何东西。”

在20世纪60年代，计算机芯片仍然主要是手工制造的，就像直到最近，大多数生物技术研究仍然是一个手动过程，缓慢、不可预测，在各个方面都显得混乱。现在，半导体制造是一种超高效的原子级制造过程，生产出世界上一些最复杂的产品。生物技术正沿着类似的轨迹发展，只是处于一个更早的阶段；有机体将很快以当今计算机芯片和软件的精度和规模被设计和生产。

2010年，由克雷格·文特尔领导的一个团队几乎复制了细菌*Mycoplasma mycoides*的基因组，并将其移植到一个新的细胞中，该细胞随后进行了复制。他们认为，这是一种新的生命形式，Synthia。2016年，他们创造了一个拥有473个基因的生物体，比自然界中发现的任何东西都少，但比以前可能的情况有了决定性的进步。仅仅三年后，苏黎世联邦理工学院的一个团队创造了第一个完全在计算机上生成的细菌基因组：*Caulobacter ethensis-2.0.*。虽然文特尔的实验有一个庞大的团队并耗资数百万美元，但这项开创性的工作主要由两兄弟以不到10万美元的成本完成。现在，全球GP-write联盟致力于在十年内将合成基因组的生产和测试成本降低“1000倍”。

生物学，迎接指数级改进。

BIOLOGICAL CREATIVITY UNLEASHED

源文本: 翻译文本:

在合成生物学这个奇特且新兴的领域中，无数实验正在进行：生产电池的病毒、净化污水的蛋白质、在容器中生长的器官、从大气中吸收碳的藻类、消耗有毒废物的植物。一些传播疾病的物种如蚊子或入侵物种如普通家鼠可能会在所谓的基因驱动中逐渐从栖息地中消失；其他物种则可能被复活，包括一个将猛犸象重新引入冻原的深奥项目。没有人能完全预料到这些可能带来的后果。

医学进步是一个显而易见的关注领域。科学家们在2021年成功地利用从藻类中提取的光感蛋白基因重建神经细胞，恢复了一名盲人的有限视力。以前难以解决的疾病，从镰状细胞病到白血病，现在都有可能得到治疗。CAR T细胞疗法设计定制的免疫反应白细胞来攻击癌症；基因编辑有望治愈遗传性心脏病。

得益于疫苗等救命疗法，我们已经习惯于通过干预我们的生物学来帮助我们对抗疾病。系统生物学领域旨在通过使用生物信息学和计算生物学来了解细胞、组织或生物体的“全局图景”，以观察生物体如何整体运作；这些努力可能成为个性化医学新时代的基础。不久之后，以通用方式接受治疗的想法将显得极为落后；从我们接受的护理类型到我们提供的药物，一切都将精确地根据我们的DNA和特定生物标志物量身定制。最终，可能有可能重新配置我们自己以增强我们的免疫反应。这反过来可能为更具雄心的实验打开大门，比如长寿和再生技术，这已经是一个迅速发展的研究领域。

Altos Labs 已筹集了 30 亿美元，比以往任何生物技术企业的初创资金都要多，是一家寻求找到有效抗-

衰老技术。其首席科学家理查德·克劳斯纳（Richard Klausner）认为，“我们认为我们可以‘逆转’人类的死亡时钟。”公司专注于“复原程序”技术，旨在重置表观基因组，即通过将基因“打开”和“关闭”来控制基因的DNA化学标记。随着我们年龄的增长，这些标记会“翻转”到错误的位置。这种实验方法旨在将它们翻转回去，逆转或阻止衰老过程。与其他一系列有前景的干预措施一起，身体衰老的不可避免性——看似人类生活的基本部分——受到质疑。在未来几十年内，平均寿命达到一百岁或更长的世界是可以实现的。这不仅仅是关于更长的寿命；而是关于随着年龄增长而拥有更健康的生活。

成功将对社会产生重大影响。同时，认知、美学、身体和与表现相关的增强也是可能的，并且会像它们被渴望一样具有破坏性和令人厌恶。无论哪种方式，严重的身体自我改造将会发生。初步研究表明，记忆可以得到改善，肌肉力量可以增强。用不了多久，“基因兴奋剂”将在体育、教育和职业生活中成为一个现实问题。关于临床试验和实验的法律在涉及自我管理时进入了灰色地带。对他人进行实验显然是禁止的，但对自己进行实验呢？与许多其他前沿技术元素一样，这是一个法律和道德上定义不清的领域。

在一位不负责任的教授与年轻夫妇进行一系列活体实验后，中国已经诞生了首批基因编辑的儿童，最终在2018年诞生了基因编辑的双胞胎，名为露露和娜娜。他的工作震惊了科学界，违反了所有伦理规范。没有任何通常的保护措施或问责机制到位；编辑被视为医学上不必要的，更糟糕的是，执行得很糟糕。科学家们感到的愤怒是真实的，谴责几乎是普遍的。要求暂停的呼声迅速而广泛，包括该领域的许多关键先驱，但仍然不是所有人都认为这是正确的方法。在更多CRISPR婴儿出现之前，

当人类诞生时，世界可能需要应对迭代胚胎选择，这种选择也可能用于选择所需的特征。

除了令人担忧的生物技术头条新闻外，越来越多的应用将会出现，范围广泛，超越医学或个人改造，仅受限于想象力。制造工艺、农业、材料、能源生产，甚至计算机——在未来几十年都将被根本性地改变。尽管仍然存在许多挑战，但对经济至关重要的材料如塑料、水泥和肥料可以更可持续地生产，生物燃料和生物塑料将取代碳排放的现有产品。作物可以变得抗感染，使用更少的水、土地和肥料；房屋可以由真菌雕塑和生长而成。

像诺贝尔奖得主弗朗西斯·阿诺德这样的科学家创造了能够产生新型化学反应的酶，包括结合硅和碳的方法，这通常是一个棘手且耗能的过程，广泛应用于电子等领域。阿诺德的方法比标准工业替代方案节能十五倍。下一步涉及扩大生物材料和工艺的生产。通过这种方式，重要的产品如肉类替代品或从大气中吸收碳的新材料可以像制造一样被“种植”。庞大的石化行业可能会受到像Solugen这样的年轻初创公司的挑战，其Bioforge试图建立一个碳负工厂；它将生产从清洁产品到食品添加剂再到混凝土的各种化学品和商品，同时从大气中提取碳。他们的工艺本质上是低能耗、低废物的工业规模生物制造，建立在人工智能和生物技术的基础上。

另一家公司，LanzaTech，利用基因改造的细菌将钢铁厂生产中的废气CO_2转化为广泛使用的工业化学品。这种合成生物学正在帮助建立一个更可持续的“循环”经济。下一代DNA打印机将以越来越高的精度生产DNA。如果不仅能改进DNA的表达，还能利用它来基因工程化多种新生物，自动化和规模化这些过程，那么理论上，一个设备或一组设备可以生产出大量的

生物材料和结构只需少量基本输入。想要制作一些洗涤剂或一个新玩具，甚至建造一座房子？只需下载“ 配方 ”并点击“ 开始 ”。用Elliot Hershberg的话来说，“ 如果我们可以在本地种植我们想要的东西呢？如果我们的供应链只是生物学呢？ ”

最终，计算机可能不仅可以制造，还可以生长。请记住，DNA 本身就是我们所知最有效的数据存储机制——能够以当前计算技术密度的数百万倍存储数据，并且具有近乎完美的保真度和稳定性。理论上，世界上所有的数据都可以存储在仅一公斤的 DNA 中。被称为转录器的晶体管的生物版本使用 DNA 和 RNA 分子作为逻辑门。在这种技术可以被利用之前，还有很长的路要走。但计算机的所有功能部件——数据存储、信息传输和基本的逻辑系统——原则上都可以使用生物材料复制。

目前，通过农业和制药用途，基因工程生物已经占美国经济的2%。这仅仅是个开始。麦肯锡估计，经济中多达60%的物理投入最终可能会受到“ 生物创新 ”的影响。全球45%的疾病负担可以通过“ 今天可以想象的科学 ”来解决。随着工具包变得更便宜和更先进，一个可能性的宇宙成为探索的对象。

合成生命时代的人工智能

源文本: 翻译文本:

蛋白质是生命的基石。你的肌肉和血液、激素和头发，实际上，你干体重的75%都是蛋白质。它们无处不在，以各种可以想象的形式存在，执行无数重要任务，从将骨头连接在一起的韧带，到用于捕捉不速之客的抗体上的钩子。理解蛋白质，你就向理解和掌握生物学迈出了一大步。

但有一个问题。仅仅知道 DNA 序列还不足以了解蛋白质的工作原理。相反，你需要了解它如何折叠。其形状由这种复杂的折叠形成，是其功能的核心：我们肌腱中的胶原蛋白具有绳索状结构，而酶则有用于容纳其作用分子的口袋。然而，事先没有办法知道这将如何发生。如果你使用传统的蛮力计算方法，即系统地尝试所有可能性，可能需要比已知宇宙的年龄更长的时间来运行给定蛋白质的所有可能形状。因此，了解蛋白质如何折叠是一个艰难的过程，阻碍了从药物到食塑酶的所有事物的发展。

几十年来，科学家们一直在问是否有更好的方法。1993年，他们决定设立一个名为结构预测关键评估（CASP）的双年竞赛，看看谁能破解蛋白质折叠问题。谁能最好地预测蛋白质如何折叠就能获胜。CASP很快成为一个竞争激烈但关系紧密的领域的基准。进展稳步进行，但看不到尽头。

然后，在2018年的CASP13上，这场比赛在坎昆一个棕榈树环绕的度假村举行，一支毫无记录的外来参赛队伍到达了比赛现场，并击败了九十八支知名团队。获胜的团队是DeepMind的。这个名为AlphaFold的项目始于2016年我在公司小组内进行的一次为期一周的实验性黑客马拉松。它发展成为计算生物学的一个里程碑时刻，并完美地展示了人工智能和生物技术如何快速进步。

虽然排名第二的团队是备受推崇的张氏团队，但他们只能预测出四十三个最难目标中的三个蛋白质结构，而我们的获胜作品预测了二十五个。它的速度比竞争对手快得多，仅用了几个小时。在这个由超聪明的专业人士组成的既定竞赛中，我们的黑马取得了胜利，令所有人震惊。著名领域研究员Mohammed AlQuraishi不禁问道：“ 刚刚发生了什么？ ”

我们的团队使用深度生成神经网络来预测蛋白质如何根据其DNA折叠，基于一组已知蛋白质进行训练并从中推断。新的模型能够更好地猜测氨基酸对之间的距离和角度。解决问题的关键不是制药领域的专业知识，也不是传统技术如冷冻电子显微镜，甚至不是传统的算法方法。关键在于机器学习和人工智能的专业知识和能力。人工智能和生物学已经决定性地结合在一起。

两年后，我们的团队回来了。一则头条新闻概括了一切：“ 生物学中最大的问题之一终于解决了， ”*Scientific American.*写道。一个以前隐藏的蛋白质宇宙以惊人的速度被揭示出来。AlphaFold表现得如此出色，以至于CASP像ImageNet一样被淘汰了。半个世纪以来，蛋白质折叠一直是科学的重大挑战之一，然后，突然之间，它被从清单上划掉了。

2022年，AlphaFold2向公众开放使用。结果是世界上最先进的机器学习工具的爆炸式增长，这些工具被应用于基础和应用生物研究中：用一位研究者的话来说，这是一次“ 地震 ”。在发布后的十八个月内，超过一百万研究人员访问了该工具，包括几乎所有世界领先的生物实验室，解决从抗生素耐药性到罕见疾病治疗再到生命起源的问题。之前的实验向欧洲生物信息学研究所的数据库提供了大约19万个蛋白质的结构，约占已知蛋白质的0.1%。DeepMind一次性上传了大约2亿个结构，几乎涵盖了所有已知的蛋白质。过去，研究人员可能需要数周或数月来确定蛋白质的形状和功能，而现在这个过程可以在几秒钟内开始。这就是我们所说的指数级变化。这就是即将到来的浪潮所能实现的。

然而，这仅仅是这两项技术融合的开始。生物革命正在与人工智能的进步共同演化，实际上，本章讨论的许多现象都将依赖于人工智能的实现。因此，想象一下两股浪潮相互碰撞，而不是一股浪潮。

但一个超级波。事实上，从一个角度来看，人工智能和合成生物学几乎是可以互换的。迄今为止，所有的智能都来自生命。称它们为合成智能和人工生命，它们仍然意味着同样的事情。这两个领域都是关于重新创造、工程这些完全基础且相互关联的概念，人类的两个核心属性；改变视角，它们就变成了一个单一的项目。

生物学的复杂性打开了大量的数据宝库，比如所有那些蛋白质，几乎无法用传统技术解析。因此，新一代工具迅速变得不可或缺。团队正在开发产品，这些产品将仅使用自然语言指令生成新的DNA序列。Transformer模型正在学习生物学和化学的语言，再次发现人类大脑无法辨识的冗长复杂序列中的关系和意义。在生化数据上微调的LLM可以生成新分子和蛋白质、DNA和RNA序列的合理候选。它们在实验室验证之前，预测化合物在模拟中的结构、功能或反应特性。应用的空间及其探索速度正在加速。

一些科学家开始研究将人类大脑直接连接到计算机系统的方法。2019年，外科手术植入大脑的电极让一名晚期ALS的全身瘫痪患者拼出“我爱我酷儿子”这句话。像Neuralink这样的公司正在开发脑接口技术，承诺将我们直接与机器连接。2021年，该公司在一只猪的大脑中插入了三千根比人类头发还细的丝状电极，用于监测神经元活动。他们希望很快开始他们的N1脑植入的人体试验，而另一家公司Synchron已经在澳大利亚开始了人体试验。一个名为Cortical Labs的初创公司的科学家甚至在培养皿中培育了一种大脑（体外生长的一堆神经元），并教它玩*Pong*.。可能用不了多久，由碳纳米管制成的神经“蕾丝”就会将我们直接插入数字世界。

当人类的思维能够瞬时访问互联网和云计算规模上的计算和信息时，会发生什么？

几乎难以想象，但研究人员已经在早期阶段使其成为现实。作为即将到来的浪潮中的核心通用技术，人工智能和合成生物学已经交织在一起，形成了一个相互促进的螺旋反馈循环。虽然疫情大大提高了生物技术的知名度，但合成生物学的全部影响——无论是可能性还是风险——才刚刚开始进入大众的想象。

欢迎来到生物机器和生物计算机的时代，在这里，DNA链条进行计算，人工细胞被投入使用。机器在这里变得有生命。欢迎来到合成生命的时代。

CHAPTER 6

更广泛的波动

T

ECHNOLOGICAL WAVES ARE BIGGER THAN just one or two general-purpose technologies. They are clusters of technologies arriving at around the same time, anchored by one or more general-purpose technologies but extending far beyond them.

通用技术是加速器。发明激发发明。浪潮为进一步的科学和技术实验奠定了基础，推动可能性的大门打开。这反过来又产生了新工具和技术、新的研究领域——技术本身的新领域，公司在这些领域内及其周围形成，吸引投资，将新技术推向大小不一的利基市场，并进一步将其改造用于千种不同的目的。浪潮之所以如此巨大和具有历史意义，正是因为这种多变的复杂性，这种迅速扩展和溢出的趋势。

技术不是在空气中发展或运作的，彼此之间没有隔离，尤其是通用技术。相反，它们在不断扩大的放大循环中发展。你在哪里找到一种通用技术，就会发现其他技术在不断的对话中发展，并受到它的推动。看待这些浪潮时，很明显这不仅仅是关于蒸汽机、个人电脑或合成生物学，尽管它们很重要；这也关乎随之而来的庞大技术和应用网络。这包括在蒸汽驱动的工厂中制造的所有产品，乘坐蒸汽驱动火车的人们，软件企业，以及更进一步，所有依赖计算的其他事物。

生物和人工智能处于中心，但在它们周围还有其他变革性技术的半影。每一种技术在其自身领域都有着巨大的意义。

正确，但当通过更大浪潮的交叉传播潜力来看时，这种情况会更加明显。二十年后，将会有许多额外的技术，同时突破。在本章中，我们将研究构成这一更大浪潮的一些关键示例。

我们从机器人学开始，或者说，我更愿意将其视为人工智能的物理表现，人工智能的身体。它的影响已经在地球上一些最前沿的行业中显现出来。但也在最古老的行业中。来看看自动化农场吧。

机器人技术成熟之时

源文本: 翻译文本:

1837年，约翰·迪尔是一名在伊利诺伊州大转弯工作的铁匠。这里是大草原地区，拥有密集的黑土和广阔的空间。它有潜力成为世界上最好的耕地之一——非常适合种植作物，但极难耕犁。

有一天，迪尔在一家工厂看到了一把破损的钢锯。由于钢材稀缺，他把这块发现带回家，并将锯片制成了一把犁。钢材坚固且光滑，是穿过密集、粘稠土壤的理想材料。尽管其他人也曾将钢材视为粗糙铁犁的替代品，但迪尔的突破在于大规模生产。不久，来自中西部各地的农民纷纷涌向他的作坊。他的发明为大批定居者打开了大草原。中西部因此成为了世界的粮仓；约翰·迪尔迅速成为农业的代名词；一场技术-地理革命被引发。

约翰迪尔公司至今仍在制造农业技术。你可能会想到拖拉机、洒水器和联合收割机，确实，约翰迪尔确实制造这些东西。然而，该公司越来越多地制造机器人。约翰迪尔所设想的农业未来包括自主拖拉机和联合收割机，它们可以独立运行，遵循田地的GPS坐标，并使用一系列传感器进行自动、实时的调整，以最大化产量和最小化浪费。

公司正在生产能够种植、照料和收获农作物的机器人，其精确度和细致程度是人类无法达到的。从土壤质量到天气条件，一切因素都被纳入一套即将完成大部分工作的机器。在食品价格通胀和人口增长的时代，其价值显而易见。

农业机器人不仅即将到来。它们已经在这里。从监视牲畜的无人机到精确灌溉设备，再到在广阔的室内农场巡逻的小型移动机器人，从播种到收获，从采摘到码垛，从浇灌西红柿到追踪和放牧牛群，我们今天所吃食物的现实是，它越来越多地来自一个由人工智能驱动的机器人世界，目前正在推出并扩大规模。

这些机器人大多数看起来不像流行科幻作品中的人形机器人。它们看起来更像是农业机械。而且我们中的许多人无论如何都不会在农场上花费太多时间。但正如约翰·迪尔的犁曾经改变了农业行业一样，这些以机器人为中心的新发明正在改变食物到达我们餐桌的方式。这不是我们容易识别的革命，但它已经在顺利进行中。

源文本：—翻译文本：

作为一维工具，机器能够在生产线上以速度和精度完成单一任务，这对制造商来说是一个主要的生产力提升，但与1960年代 *Jetsons* 风格的羞怯安卓助手的愿景相去甚远。

与人工智能一样，机器人技术在实践中比早期工程师想象的要困难得多。现实世界是一个奇怪、不平坦、出人意料且无结构的环境，对压力等事物极为敏感：拿起鸡蛋、苹果、砖块、孩子和一碗汤都需要非凡的灵巧、敏感、力量和平衡。像厨房或车间这样的环境是混乱的，充满了危险物品、油污以及多种不同的工具和材料。这是机器人的噩梦。

尽管如此，大多数情况下不为公众所知，机器人一直在悄悄学习关于扭矩、抗拉强度、操控物理学的知识，

精度、压力和适应性。只需在YouTube上观看他们在汽车制造厂的表现：你会看到一个清晰、永无止境的机器人手臂和操控器的芭蕾舞，稳步地建造一辆汽车。亚马逊的“ 首个完全自主移动机器人 ”Proteus ，可以在仓库中大规模成群地移动，拾取包裹。配备了“ 先进的安全、感知和导航技术 ”，它可以在人类旁边舒适地完成这些工作。亚马逊的Sparrow是第一个可以“ 检测、选择和处理库存中单个产品 ”的机器人。

不难想象这些机器人在仓库和工厂中——相对静态的环境。但很快它们将越来越多地出现在餐馆、酒吧、护理院和学校。机器人已经在进行复杂的手术——与人类协作，但也能自主地在猪身上进行（目前如此）。这种应用只是更广泛的机器人推广的开始。

今天，人类程序员仍然经常控制机器人的每一个操作细节。这使得在新环境中的集成成本过高。但正如我们在许多其他机器学习应用中所看到的那样，最初需要人类密切监督的任务，最终会由AI自行学习得更好，最终推广到新的环境中。

Google 的研究部门正在制造机器人，这些机器人可以像20世纪50年代的梦想一样，做家务和基本工作，从堆叠餐具到整理会议室的椅子。他们建造了一支由一百个机器人组成的队伍，能够分类垃圾和擦拭桌子。强化学习帮助每个机器人的夹持器拿起杯子和打开门：这些动作对幼儿来说毫不费力，却让机器人专家困扰了几十年。这种新型机器人可以从事一般活动，响应自然语言语音命令。

另一个不断发展的领域是机器人能够群体行动，将任何单个机器人的潜在能力大大放大为一个蜂群思维。例子包括哈佛大学Wyss研究所的微型Kilobots——一千个机器人组成的群体，它们共同工作并组装成从自然中获取的形状，可以用于解决土壤侵蚀和其他环境调节、农业等困难的分布式任务。

搜索和救援行动，或整个建筑和检查领域。想象一下，一群建筑机器人在几分钟内架起一座桥梁，或在几小时内建造一座大型建筑，或全天候照料巨大的高产农场，或清理石油泄漏。在蜜蜂种群受到威胁的情况下，沃尔玛申请了一项机器人蜜蜂的专利，以便自主协作和异花授粉。所有机器人的承诺（和危险）都因其在不受限制规模的群体中协调的能力而被放大，这是一种复杂的编排，将重新设定可能的规则、地点和时间框架。

如今的机器人往往仍然不像人们想象中的类人机器人。考虑一下3D打印或增材制造的现象，这是一种使用机器人装配机来逐层构建从微小机器零件到公寓楼的技术。巨大的喷涂混凝土机器人可以在几天内建造房屋，其成本仅为传统建筑的一小部分。

机器人可以在比人类更广泛的环境中精确操作，并且持续时间更长。它们的警觉性和勤奋是无穷无尽的。如果它们联网在一起，它们可能完成的壮举简直可以改写行动的规则。我认为我们现在正处于一个AI推动机器人实现其最初承诺的阶段：能够复制人类所有物理动作及更多的机器。随着成本的下降（机器人手臂的价格在五年内下降了46%，并且还在继续下降），随着它们最终配备强大的电池，随着它们的简化，变得易于维修，它们将变得无处不在。这将意味着它们会出现在不寻常、极端和敏感的情况下。已经有迹象表明这种转变正在发生——如果你知道在哪里寻找的话。

源文本：—翻译文本：

IT WAS THE POLICE force’s worst nightmare. A military-trained sniper had got himself in a secure second-floor position at a local community college in Dallas, Texas. Then, overlooking a peaceful protest, he’d begun shooting

警察。四十五分钟后，两人死亡，更多人受伤。后来发现有五名警官被杀，七人受伤，这是自9/11以来对美国执法部门最致命的事件。枪手嘲弄警察，笑着，唱着，并以惊人的准确性开火。紧张的谈判持续了两个多小时，毫无进展。警察被困住了。不清楚在试图解决这种情况时还会有多少人死亡。

然后，特警队想出了一个新主意。警察局有一个拆弹机器人，由诺斯罗普·格鲁曼公司制造的价值15万美元的Remotec Andros Mark 5A-1。在十五分钟内，他们制定了一个计划，将一大块C-4炸药附在机器人的手臂上，并将其送入建筑物，目的是使枪手失去行动能力。警察局长大卫·布朗迅速批准了该计划。计划开始实施，机器人在建筑物内隆隆前进，将炸药放置在一个相邻的房间里，靠近墙壁，枪手在另一侧。炸药爆炸，炸开了墙壁，击毙了枪手。这是美国首次使用机器人进行有针对性的致命武力。在达拉斯，它挽救了局面。一场可怕的事件得以结束。

尽管如此，一些人仍感到不安。致命警用机器人的潜在威胁几乎不需要强调。我们将在第3部分回到这一切的影响。但最重要的是，这表明机器人正逐渐进入社会，准备在日常生活中扮演比以往更重要的角色。从致命危机到物流中心的安静嗡嗡声，从繁忙的工厂到养老院，机器人已经到来。

人工智能是由比特和代码构成的产品，存在于模拟和服务器中。机器人是它们的桥梁，是它们与现实世界的接口。如果说人工智能代表信息的自动化，那么机器人技术就是物质的自动化，是人工智能的物理实例化，是在可能实现的*do.*方面的一个飞跃。对比特的掌握达到了一个完整的循环，直接重新配置原子，重写的不仅仅是可以思考、表达或计算的界限，而是可以在最具体的物理意义上建造的东西。然而，即将到来的浪潮中最引人注目的事情是，这种粗略的原子操控与即将出现的东西相比微不足道。

QUANTUM SUPREMACY

源文本: 翻译文本:

2019年，谷歌宣布其已实现“ 量子霸权 ”。研究人员建造了一台量子计算机，利用了亚原子世界的独特性质。谷歌的机器被冷却到比外太空最冷的部分还要低的温度，利用对量子力学的理解，在几秒钟内完成了一项计算，而据称这项计算在传统计算机上需要一万年。它只有五十三个“ 量子比特 ”，即量子计算的核心单位。在经典计算机上存储等效信息，你需要七百二十亿千兆字节的内存。这是量子计算机的一个关键时刻。从20世纪80年代的理论基础开始，量子计算在四十年间从假设走向了工作原型。

虽然量子计算仍然是一项新兴技术，但一旦实现，将带来巨大的影响。其主要吸引力在于每增加一个量子比特，机器的总计算能力就会翻倍。开始增加量子比特后，其计算能力会呈指数级增长。实际上，相对较少的粒子可能拥有比将整个宇宙转换为经典计算机还要强大的计算能力。这相当于从平面的黑白电影转变为全彩色和三维，释放出一个充满算法可能性的世界。

量子计算具有深远的影响。例如，从电子邮件安全到加密货币的所有基础加密技术都将面临风险，在该领域被称为“ Q日 ”的即将到来的事件中。加密技术基于这样一个假设：攻击者永远不会拥有足够的计算能力来尝试所有不同的组合以破解它并解锁访问权限。随着量子计算的出现，这种情况发生了变化。量子计算的快速和不受控制的推出可能对银行或政府通信产生灾难性的影响。两者都已经在花费数十亿美元以防止这种可能性。

尽管关于量子计算的讨论多集中于其危险性，但该领域也承诺带来巨大的好处，包括探索数学和粒子物理学前沿的能力。微软和福特的研究人员使用新兴的量子方法来模拟西雅图的交通，以找到更好的方法来应对高峰时段的导航、路由和交通流动在最佳路径上的问题——这是一个出人意料的复杂数学问题。理论上，解决任何优化问题的速度都可以大大加快——几乎任何涉及在复杂情况下最小化成本的问题，无论是高效装载卡车还是运行国家经济。

可以说，量子计算在近期最重要的承诺是以以前不可能的细节模拟化学反应和分子相互作用。这可以让我们以非凡的细致程度理解人类大脑或材料科学。化学和生物学将首次变得完全可读。发现新的药物化合物或工业化学品和材料，这一费时费力的复杂实验室工作过程，可以大大加快——第一次就能正确完成。新电池和药物的制造变得更有可能、更高效和可实现。分子变得“ 可编程 ”，像代码一样灵活和可操控。

量子计算，换句话说，是另一项仍处于非常早期发展的基础技术，距离达到那些成本下降和广泛普及的关键时刻还很遥远，更不用说使其完全可行的技术突破了。但与人工智能和合成生物学一样，尽管处于更早的阶段，它似乎正处于资金和知识不断增加的阶段，基础挑战的进展正在增长，各种有价值的用途开始显现。像人工智能和生物技术一样，量子计算有助于加速浪潮的其他元素。然而，即使是令人费解的量子世界也不是极限。

下一个能源转型

源文本: 翻译文本:

能量在其根本重要性上可与智慧和生命媲美。现代文明依赖于大量的能量。实际上，如果你想写出

我们世界最粗略的方程式可能是这样的：

(生命 + 智能) x 能量 = 现代文明

增加任何或所有这些投入（更不用说将其边际成本提升至接近零）会使社会性质发生跃变。

在化石燃料时代，能源消耗的无尽增长既不可能也不理想，然而在繁荣时期，几乎所有我们认为理所当然的事物的发展——从廉价食品到轻松运输——都依赖于此。现在，大量廉价、清洁能源的激增对从交通到建筑的一切都有影响，更不用说运行数据中心和机器人所需的巨大能量，这些将成为未来几十年的核心。能源——尽管通常昂贵且肮脏——目前是技术进步速度的限制因素。但这种情况不会持续太久。

到2027年，可再生能源将成为最大的单一电力来源。这一转变正在以前所未有的速度发生，未来五年内新增的可再生能源容量将超过过去二十年的总和。特别是太阳能正在迅速增长，成本显著下降。2000年，太阳能的成本为每瓦$4.88，但到2019年已降至仅38美分。能源不仅变得更便宜，而且更分散，可能从特定设备到整个社区实现本地化。

其背后隐藏着清洁能源的沉睡巨兽，这次受到太阳的启发，即使不是直接由太阳驱动：核聚变。聚变能涉及氢同位素碰撞并融合形成氦时释放的能量，这一过程长期以来被视为能源生产的圣杯。20世纪50年代的早期先驱者预测需要大约十年时间来开发。像这里描述的许多技术一样，那是一个显著的低估。

然而，最近的突破激发了新的希望。位于英格兰牛津附近的欧洲联合环形加速器的研究人员实现了创纪录的功率输出，是1997年创下的最高纪录的两倍。在

位于加利福尼亚州利弗莫尔的国家点火装置的科学家们一直在研究一种称为惯性约束的方法，该方法涉及用激光压缩富含氢的材料颗粒并将其加热到1亿度以产生短暂的聚变反应。2022年，他们首次创造了一个反应，展示了净能量增益，这是产生比激光输入更多能量的关键里程碑。随着有意义的私人资本现在流入至少三十家聚变初创公司以及主要的国际合作，科学家们正在谈论聚变的到来是“ 何时而非是否 ”的问题。可能还需要十年或更长时间，但这种清洁且几乎无限的能源来源的未来看起来越来越真实。

聚变和太阳能提供了巨大的集中式和分散式能源电网的前景，其影响我们将在第3部分中探讨。这是一个充满巨大乐观情绪的时代。包括风能、氢能和改进的电池技术，这里正在酝酿一种可以可持续地满足当今和未来生活的众多需求并支持浪潮全部潜力的组合。

波浪之外的波浪

源文本: 翻译文本:

这些技术将在未来几十年占据主导地位。但在二十一世纪的下半叶呢？在即将到来的浪潮之后会发生什么？

随着人工智能、先进生物技术、量子计算和机器人技术的元素以新的方式结合，准备迎接突破性进展，如先进纳米技术，这一概念将技术不断增长的精确性推向其逻辑结论。如果不是大规模操控，而是可以单独操控原子呢？这将是比特/原子关系的顶峰。纳米技术的终极愿景是将原子变成可控的构建块，能够自动组装几乎任何东西。

实际挑战是巨大的，但它们是研究强度不断增加的主题。例如，牛津大学的一个团队，

制造了一个自我复制的装配器，指向纳米技术先驱设想的多功能版本：能够在原子尺度上无休止地进行工程和重组的设备。

纳米机器将以远超我们尺度的速度运作，带来非凡的输出：例如，一个原子级的纳米马达可以每分钟旋转四百八十亿次。放大后，它可以用相当于大约十二粒沙子的体积的材料为特斯拉提供动力。这是一个由钻石制成的轻盈结构的世界，太空服在所有环境中紧贴并保护身体，一个编译器可以从基本原料中创造任何东西的世界。简而言之，这是一个通过正确的原子操控，任何东西都可以变成任何东西的世界。物理宇宙的梦想被呈现为一个完全可塑的平台，是微小而灵巧的纳米机器人或轻松复制器的玩物，仍然像超级智能一样属于科幻小说的领域。这是一个技术幻想，距离实现还有几十年，但随着即将到来的浪潮展开，它将逐渐清晰。

源文本：—翻译文本：

即将到来的浪潮是一个权力扩散的故事。如果上一波浪潮降低了*broadcasting*信息的成本，那么这一波浪潮则降低了其上的*acting*成本，催生了从测序到合成、从阅读到写作、从编辑到创造、从模仿对话到引导对话的技术。在这一点上，它与之前的每一波浪潮都有质的不同，尽管关于互联网变革力量的巨大主张层出不穷。这种力量更难以集中和监管；因此，这一波浪潮不仅是历史模式的深化和加速，也是对其的一个明显突破。

并非所有人都同意这些技术像我认为的那样确定或重要。鉴于存在许多不确定性，怀疑和悲观的回避并不是不合理的反应。每项技术都受到恶性炒作周期的影响，每项技术在开发和接受方面都存在不确定性，每项技术都面临技术挑战，

伦理和社会。没有一个是完整的。肯定会有挫折，许多危害——实际上还有好处——仍然不清楚。

但每一个也在日益变得更加具体、成熟和有能力。每一个都变得更加易于接近和更加强大。我们正在达到一个决定性的时刻，在地质或人类进化的时间尺度上，这是一个以连续波浪展开的技术爆炸，一个复合的、加速的创新周期，稳步变得更快、更有影响力，首先在几千年的时间里，然后是几百年，而现在是几年甚至几个月。在新闻稿和评论文章中，在社交媒体的短暂节奏中看到这些技术，它们可能看起来像是炒作和泡沫；从长远来看，它们的真正潜力变得清晰。

人类当然在这一过程中经历过史诗般的技术变革。然而，要理解即将到来的浪潮所带来的独特挑战——为什么它特别难以遏制，为什么它巨大的前景必须与冷静的谨慎相平衡——我们首先必须分解其关键特征，其中一些是史无前例的，所有这些特征已经开始显现。

CHAPTER 7

即将到来的浪潮的四个特征

SOON AFTER THE RUSSIAN INVASION of Ukraine began on February 24, 2022, residents of the city of Kyiv knew they were in a fight for survival. Over the border with Belarus a colossal massing of Russian troops, armor, and materiel had been building for months. Then, at the outset of the invasion, Russian forces readied for a major push on what was still at this stage their primary goal: capture Ukraine's capital and overthrow its government.

这股力量集中的核心是一支由卡车、坦克和重型火炮组成的纵队，长约四十公里——这是自二战以来欧洲未曾见过的地面攻势。它开始向城市进发。从纸面上看，乌克兰人毫无胜算。基辅似乎在几天，甚至几个小时内就会沦陷。

但这并没有发生。相反，当晚大约三十名乌克兰士兵戴着夜视镜骑着四轮摩托车穿过首都周围的森林。他们在车队的前方下车，发射了装备小型炸药的临时改装无人机。这些无人机摧毁了几辆前导车辆。那些被摧毁的车辆随后堵塞了中央道路。周围的田地泥泞不堪，无法通行。车队在寒冷的天气和供应线不畅的情况下停滞不前。然后，同一小队无人机操作员使用相同的战术成功炸毁了一个关键的补给基地，使俄军失去了燃料和食物。

从这里开始，基辅战役发生了转折。 一代人中最大规模的常规军事力量被击败，被迫撤退。

白俄罗斯陷入尴尬的混乱。这支半即兴的乌克兰民兵被称为Aerorozvidka。由无人机爱好者、软件工程师、管理顾问和士兵组成的杂牌志愿者团队，他们是业余爱好者，实时设计、建造和改装自己的无人机，很像一家初创公司。他们的许多设备都是众包和众筹的。

乌克兰抵抗力量很好地利用了新兴技术，并展示了它们如何能够破坏传统的军事计算。SpaceX 的 Starlink 提供的尖端卫星互联网对于保持连接至关重要。一个由一千名非军事精英程序员和计算机科学家组成的组织，名为 Delta，联合起来为军队带来先进的人工智能和机器人技术能力，使用机器学习来识别目标、监控俄罗斯战术，甚至建议策略。

在战争初期，乌克兰军队一直弹药短缺。每一次打击都至关重要。准确性是生存的关键。Delta 创建机器学习系统以发现伪装目标并帮助引导弹药的能力至关重要。在传统军事中，一枚精确导弹的成本高达数十万美元；而通过人工智能和消费级无人机，结合定制软件和3D打印部件，类似的东西现在在乌克兰经过实战测试，成本约为 $15,000。除了最初的 Aerorozvidka 努力外，美国还向乌克兰提供了数百枚 Switchblade 徘徊弹药，这些无人机会在目标周围等待，直到最佳打击时机。

无人机和人工智能在乌克兰冲突初期发挥了小但重要的作用，这些新技术具有明显的非对称潜力，缩小了与更大侵略者之间的一些差距。美国、英国和欧洲部队在最初几个月提供了不到1000亿欧元的军事援助，其中包括大量常规火力，明确地说，这无疑产生了决定性影响。然而，这仍然是一个具有里程碑意义的冲突，因为它展示了一个相对未经训练的战斗力量如何能够迅速组建并使用消费市场上相对负担得起的技术进行武装。

技术赋予了这样的成本和战术优势，当然，它将不可避免地扩散并被各方采用。

无人机让我们一窥未来战争的前景。它们是规划者和战斗人员每天都要面对的现实。真正的问题是，当生产成本再下降一个数量级且能力倍增时，这对冲突意味着什么。传统军队和政府已经在努力控制它们。接下来发生的事情将更难以控制。

源文本：—翻译文本：

第1部分，从X光机到AK-47的技术一直在扩散，带来了广泛的影响。然而，即将到来的浪潮具有四个内在特征，这些特征加剧了遏制问题。其中首要的是本节的主要教训：巨大的*asymmetric*影响。你不需要以同样的方式打击，以质量对质量；相反，新技术创造了以前难以想象的漏洞和压力点，对抗看似占主导地位的力量。

其次，它们发展迅速，以惊人的速度进行一种*hyper-evolution,*迭代、改进，并扩展到新的领域。第三，它们通常是*omni-use;*，也就是说，它们可以用于许多不同的目的。第四，它们越来越具有超越任何以往技术的*autonomy*程度。

这些特征定义了波。理解它们对于识别其产生的益处和风险至关重要；它们共同将遏制和控制提升到一个新的困难和危险层面。

不对称性：权力的巨大转移

源文本: 翻译文本:

新兴技术总是会带来新的威胁，重新分配权力，并消除进入壁垒。大炮意味着一支小部队可以

摧毁城堡和平定军队。少数殖民士兵凭借先进武器可以屠杀成千上万的土著居民。印刷机意味着一个作坊可以生产数千份小册子——传播思想的轻松程度是手抄书籍的中世纪僧侣难以想象的。蒸汽动力使得单个工厂能够完成整个城镇的工作。互联网将这种能力推向了新的高峰：一条推文或一张图片可能在几分钟或几秒钟内传遍世界；一个算法可以帮助一个小型初创公司成长为一个庞大的、跨越全球的公司。

现在，这种影响再次被强化。这股新的技术浪潮解锁了强大的能力，这些能力便宜、易于获取和使用、有针对性且可扩展。这显然带来了风险。使用武装无人机的不仅仅是乌克兰士兵。任何想要的人都可以使用。用安全专家奥黛丽·库尔斯·克罗宁的话来说，“ 从未有过如此多的人能够接触到如此先进的技术，这些技术能够造成死亡和混乱。”

在基辅外围的冲突中，使用的无人机是业余爱好者的玩具。总部位于深圳的公司大疆创新（DJI）制造了价格低廉且广泛可得的产品，如其旗舰产品售价为$1,399的Phantom相机四轴飞行器，这款无人机性能优异，甚至被美国军方使用。如果你将人工智能和自主技术的进步、廉价但有效的无人机，以及从机器人技术到计算机视觉等领域的进一步发展结合起来，那么你就拥有了强大、精确且可能无法追踪的武器。对抗攻击既困难又昂贵；美国人和以色列人都使用价值300万美元的爱国者导弹来击落价值几百美元的无人机。干扰器、导弹和反无人机技术仍处于初期阶段，并不总是经过实战检验。

这些发展代表着权力从传统国家和军队向任何有能力和动机部署这些设备的人手中进行巨大的转移。没有明显的理由表明，单个操作员只要有足够的能力，就不能控制成千上万架无人机的集群。

一个单一的人工智能程序可以写出与全人类一样多的文本。一个在你笔记本电脑上运行的两千兆字节的图像生成模型可以将开放网络上的所有图片压缩成一个生成图像的工具

非凡的创造力和精确性。一次病原实验可能引发一场大流行，一次微小的分子事件可能产生全球影响。一台可行的量子计算机可能使全球的加密基础设施变得多余。非对称影响的前景正在各处增长，并且在积极的意义上也是如此——单一系统也可以带来巨大的好处。

不对称行动的逆面也同样成立。即将到来的浪潮的规模和互联性本身就创造了新的系统性脆弱性：一个故障点可能迅速在全球范围内蔓延。技术越不局限于某个地方，就越难以控制，反之亦然。想想汽车所涉及的风险。交通事故和交通一样古老，但随着时间的推移，损害被最小化。从道路标记到安全带再到交通警察，一切都起到了帮助作用。尽管汽车是历史上增长最快、全球化程度最高的技术之一，但事故本质上是局部的、离散的事件，其最终损害是可控的。但现在，一队车辆可能会联网在一起。或者一个系统可以控制整个地区的自动驾驶车辆。无论有多少安全措施和安全协议，影响的规模都比我们以前见过的要大得多。

AI带来的不对称风险超出了食品质量问题、飞机事故或产品故障的范围。其风险延伸至整个社会，使其不仅仅是一个笨拙的工具，而是一个具有全球影响的杠杆。正如全球化和高度互联的市场在金融危机中传播传染病一样，技术也是如此。网络规模使得在损害发生时或发生后进行控制几乎是不可能的。互联的全球系统是控制的噩梦。而我们*already*生活在一个互联的全球系统时代。在即将到来的浪潮中，单一的点——一个特定的程序或基因变化——可以改变一切。

超进化：无尽加速

源文本: 翻译文本:

如果你想控制技术的发展，你可能希望它以一个可控的速度发展，给社会时间和空间去理解和适应它。汽车再次是一个很好的例子。它们在过去一个世纪的发展速度非常快，但也提供了时间来引入各种安全标准。总是有一个滞后，但标准仍然可以赶上。然而，随着即将到来的变革浪潮的速度，这看起来不太可能。

在过去的四十年里，互联网发展成为历史上最富有成效的创新平台之一。世界实现了数字化，这个非物质化的领域以令人眼花缭乱的速度演变。开发的爆炸性增长使得世界上使用最广泛的服务和历史上最大的商业企业在短短几年内涌现。所有这些都得益于我们在第二章中看到的计算能力的不断增强和成本的下降。考虑一下仅摩尔定律在未来十年将带来的变化。如果它成立，十年后，一美元将能购买到今天计算能力的一百倍。仅这一事实就暗示了一些非凡的结果。

另一方面，数字之外的创新往往不那么引人注目。在代码的无重世界之外，越来越多的人开始质疑，像十九世纪末或二十世纪中期那样的广泛创新发生了什么。在那段短暂的时期，世界的几乎每个方面——从交通到工厂，从动力飞行到新材料——都发生了根本性的变化。但到了二十一世纪初，创新遵循了阻力最小的路径，集中在比特而非原子上。

这种情况正在发生变化。软件的超进化正在蔓延。未来四十年将见证原子世界以新的复杂性和精确度被转化为比特，关键是，比特世界将以最近难以想象的速度和便捷性重新转化为有形的原子。

简单来说，“现实世界”中的创新可以开始以数字化的速度移动，几乎实时进行，摩擦减少，依赖性更少。您将能够在小型、快速、可塑的领域中进行实验，创建近乎完美的模拟，然后将其转化为具体的产品。然后再做一次，再做一次，以前在昂贵、静态的原子世界中不可能实现的速度进行学习、进化和改进。

物理学家César Hidalgo认为，物质的配置之所以重要，是因为它们所包含的信息。一辆法拉利的价值不在于其原材料，而在于其复杂的结构和形式中存储的复杂信息；其原子排列的信息特征使其成为一辆令人向往的汽车。计算基础越强大，这就越容易处理。将其与AI和制造技术（如复杂的机器人技术和3D打印）结合起来，我们可以以更快的速度、更高的精度和更具创造性地设计、操控和制造现实世界的产品。

人工智能已经帮助发现新材料和化合物。例如，科学家们使用神经网络生成了新的锂配置，对电池技术有重大影响。人工智能帮助设计并使用3D打印机制造了一辆汽车。在某些情况下，最终结果看起来与人类设计的任何东西截然不同，类似于自然界中发现的起伏和高效的形式。布线和管道的配置有机地融入到底盘中，以优化空间使用。零件过于复杂，无法使用传统工具制造，必须进行3D打印。

在第5章中，我们看到了像AlphaFold这样的工具如何催化生物技术。直到最近，生物技术依赖于无休止的手动实验室工作：测量、移液、仔细准备样本。现在，模拟加速了疫苗发现的过程。计算工具帮助自动化设计过程的部分环节，重新创建“生物电路”，将复杂功能编程到细菌等细胞中，使其能够生产特定蛋白质。像Cello这样的软件框架几乎就像是合成生物学设计的开源语言。这可以与实验室机器人和自动化的快速进步以及我们在第5章中看到的酶合成等更快的生物技术相结合，扩大合成生物学的范围并使其更易于获取。生物进化正变得与软件相同的周期相关。

正如今天的模型可以根据几个词生成详细的图像，未来几十年类似的模型将能够仅凭几个自然语言提示生成一种新化合物，甚至是一个完整的生物体。

化合物的设计可以通过无数次自我运行的试验来改进，就像AlphaZero通过自我对弈成为国际象棋或围棋高手一样。量子技术比最强大的经典计算机强大数百万倍，可以让这一过程在分子层面上展开。这就是我们所说的超进化——一个快速、迭代的创造平台。

这种演变不仅限于特定的、可预测的和易于控制的领域。它将无处不在。

OMNI-USE：多多益善

源文本: 翻译文本:

违背传统智慧，医疗保健的进步是最近在原子领域创新停滞中放缓的领域之一。发现新药变得更加困难和昂贵。预期寿命趋于平稳，甚至在美国的一些州开始下降。像阿尔茨海默病这样的病症的进展未能达到预期。

人工智能最有前途的领域之一，以及摆脱这种严峻局面的途径，是自动化药物发现。AI技术可以在可能的分子广阔空间中搜索难以捉摸但有用的治疗方法。2020年，一个AI系统筛选了1亿个分子，创造了第一个机器学习衍生的抗生素——halicin（是的，来自*2001: A Space Odyssey*的HAL），这可能有助于对抗结核病。像Exscientia这样的初创公司，以及像Sanofi这样的传统制药巨头，已经使AI成为医学研究的驱动力。迄今为止，已经在AI工具的帮助下衍生出十八种临床资产。

有一个反面。寻找这些有用化合物的研究人员提出了一个尴尬的问题。如果你重新定向发现过程会怎样？如果不是寻找治疗方法，而是寻找杀手呢？他们进行了一个测试，要求他们的分子生成AI寻找毒物。在六个小时内，它识别出了超过四万种毒性可与最危险的化学武器（如诺维乔克）相媲美的分子。事实证明，在

药物发现是人工智能无疑会产生最明显影响的领域之一，这些机会非常具有“ 双重用途 ”。

两用技术是指那些既有民用又有军用用途的技术。在第一次世界大战中，合成氨的过程被视为养活世界的一种方式。但它也使得炸药的制造成为可能，并帮助铺平了化学武器的道路。用于客机的复杂电子系统可以被重新用于精确导弹。相反，全球定位系统最初是一个军事系统，但现在有无数日常消费者用途。发布时，PlayStation 2 被美国国防部视为如此强大，以至于它可能帮助通常被拒绝访问此类硬件的敌对军队。两用技术既有帮助也有潜在的破坏性，既是工具也是武器。这个概念所捕捉到的是技术如何趋向于通用，并且某一类技术由于这一点而伴随着更高的风险。它们可以被用于多种目的——好的、坏的，以及介于两者之间的任何地方——通常具有难以预测的后果。

但真正的问题是，不仅仅是前沿生物学或核反应堆是双重用途的。大多数技术都有军事和民用的应用或潜力；大多数技术在某种程度上是双重用途的。而且，技术越强大，就越应该关注它可能有多少种用途。

即将到来的技术浪潮之所以极具威力，正是因为它们具有根本的通用性。如果你在制造核弹头，它的用途显而易见。但一个深度学习系统可能是为玩游戏而设计的，却有能力驾驶一支轰炸机机群。区别并不是先验显而易见的。

即将到来的技术浪潮的一个更合适的术语是“ 全用途 ”，这一概念抓住了纯粹的普遍性水平，展示了极端的多功能性。像蒸汽或电力这样的全用途技术比狭义技术具有更广泛的社会影响和溢出效应。如果人工智能确实是新的电力，那么就像电力一样，它将成为一种按需使用的公用事业，渗透并推动几乎每个方面的$v*$。

日常生活、社会、经济：一种嵌入到各处的通用技术。遏制这样的东西总是比遏制一种受限的、单一任务的技术要困难得多，后者被困在一个依赖性很少的小领域中。

人工智能系统最初使用像深度学习这样的通用技术来实现特定目的，比如管理数据中心的能源使用或下围棋。这种情况正在改变。现在，像DeepMind的通用系统Gato这样的单一系统可以胜任六百多种不同的任务。相同的网络可以玩Atari游戏、为图像加字幕、回答问题，并用真实的机器人手臂堆积积木。Gato不仅通过文本进行训练，还通过图像、作用于机器人手臂的扭矩、电脑游戏中的按钮按压等进行训练。现在仍然是非常早期的阶段，真正的通用系统仍然有一段距离，但在某个时候，这些能力将扩展到成千上万的活动。

也可以通过全用棱镜考虑合成生物学。工程生命是一种完全通用的技术，其潜在用途几乎是无限的；它可能创造建筑材料、解决疾病问题和存储数据。多多益善，这是有充分理由的。全用技术比狭窄的技术更有价值。如今，技术人员不想设计有限的、特定的、单一功能的应用技术。相反，目标是设计更像智能手机的东西：手机，但更重要的是用于拍照、保持健康、玩游戏、导航城市、发送电子邮件等的设备。

随着时间的推移，技术趋向于普遍性。这意味着即将到来的浪潮中的可武器化或有害用途将是可能的，无论这是否是有意的。仅仅创造民用技术就具有国家安全影响。预测历史上用途最广泛浪潮的全部用例比以往任何时候都更难。

新技术被改编用于多种用途的概念并不新鲜。像刀子这样简单的工具可以用来切洋葱，也可以用来进行疯狂的杀戮。即使是看似特定的技术也有双重用途的含义：麦克风既支持了纽伦堡集会，也支持了披头士乐队。即将到来的浪潮不同之处在于它嵌入的速度有多快，如何

从全球范围来看，它如何传播，如何轻松地将其组件化为可交换的部分，以及其应用的强大程度和广泛性。它对从媒体到心理健康、市场到医学的所有事物都展开了复杂的影响。这是一个超大规模的控制问题。毕竟，我们谈论的是诸如智能和生命这样的基本问题。但这两个属性都有一个比其普遍性更有趣的特征。

自主性及其未来：人类是否会参与其中？

源文本: 翻译文本:

技术进化已经加速了几个世纪。全用途特性和不对称影响在即将到来的浪潮中被放大，但在某种程度上，它们是所有技术的固有属性。自主性则不然。纵观历史，技术一直“只是”一种工具，但如果工具有了生命呢？

自主系统能够与其周围环境进行交互，并在没有人类即时批准的情况下采取行动。几个世纪以来，技术以某种方式失控，成为一种超出人类控制范围的自我导向和自我推动的力量，这一想法一直是虚构的。

不再是这样。

技术一直以来都是为了让我们能够做更多的事情，但关键是人类仍然在执行这些事情。它利用了我们现有的能力，并自动化了精确编码的任务。直到现在，持续的监督和管理一直是默认状态。技术在不同程度上仍然处于有意义的人类控制之下。完全自主性在质上是不同的。

以自动驾驶汽车为例。在当今的某些条件下，它们可以在道路上行驶，而无需驾驶员的直接输入或仅需极少的输入。该领域的研究人员将自动驾驶分为从0级（完全没有自动驾驶）到5级（车辆可以在所有条件下自行驾驶，驾驶员只需输入目的地，然后可以安心入睡）。你不会找到5级

由于法律和保险原因，近期道路上的车辆不会很快发生变化。

新一波的自主浪潮预示着一个不再需要持续干预和监督的世界。更重要的是，通过每一次互动，我们都在教导机器成功实现自主。在这种范式中，不需要人类费力地定义任务应该如何进行。相反，我们只需指定一个高层次的目标，并依靠机器找出实现目标的最佳方式。正如俗话所说，让人类“参与其中”是可取的，但并非必要。

没有人告诉 AlphaGo 第 37 手是个好主意。它主要是靠自己发现了这个见解。正是这个特性让我在观看 DQN 玩 *Breakout.* 时感到如此震撼。给定一些明确的目标，现在存在可以找到自己策略以有效实现目标的系统。AlphaGo 和 DQN 本身并不是自主的。但它们暗示了一个自我改进系统可能的样子。没有人手动编写 GPT-4 让它像简·奥斯汀那样写作，或创作原创俳句，或为销售自行车的网站生成营销文案。这些功能是更广泛架构的涌现效应，其输出从未由设计者事先决定。这是迈向更大自主性的第一步。关于 GPT-4 的内部研究得出结论，它“可能”没有能力自主行动或自我复制，但在发布后的几天内，用户就找到了让系统请求自己的文档并编写脚本以复制自身并接管其他机器的方法。早期研究甚至声称在模型中发现了“AGI 的火花”，并补充说它“惊人地接近人类水平的表现。”这些现在正在显现。

新的自主形式有可能产生一系列新颖且难以预测的效果。预测定制基因组的行为是极其困难的。此外，一旦研究人员对某个物种进行生殖系基因改变，这些改变可能会在活体中存在数千年，远超出控制或预测的范围。它们可能会在无数代中产生影响。它们如何继续进化或与之互动

在这些距离上的其他变化不可避免地不清楚——并且超出了控制范围。合成生物体实际上正在拥有自己的生命。

我们人类面临一个独特的挑战：新的发明是否会超出我们的理解范围？以前，创造者可以解释某个东西是如何运作的以及为什么会这样做，即使这需要大量细节。但这越来越不再是事实。许多技术和系统变得如此复杂，以至于超出了任何一个人真正理解的能力：量子计算和其他技术在可知的极限上运作。

即将到来的浪潮的一个悖论是，其技术在细节层面上大多超出了我们的理解能力，但仍在我们的创造和使用能力之内。在人工智能领域，正在走向自主的神经网络目前是无法解释的。你无法通过决策过程来精确解释为什么一个算法会产生特定的预测。工程师无法深入了解并轻松解释是什么导致了某件事情的发生。GPT-4、AlphaGo等都是黑箱，其输出和决策基于不透明且复杂的微小信号链。自主系统可以且可能是可解释的，但即将到来的浪潮中有如此多的东西在我们理解的边缘运行，这一事实应该引起我们的注意。我们并不总是能够预测这些自主系统接下来会做什么；这就是自主性的本质。

然而，在最前沿，一些人工智能研究人员希望自动化构建人工智能系统的每一个方面，推动这种超进化，但可能通过自我改进实现极大的独立性。人工智能已经在寻找改进自身算法的方法。当它们将这一点与在网络上的自主行动结合起来时，会发生什么，比如在现代图灵测试和ACI中，进行自己的研发周期？

大猩猩问题

源文本: 翻译文本:

我常常觉得，鉴于即将到来的浪潮中存在的明显短期挑战，人们对遥远的AGI情景关注过多。然而，任何关于遏制的讨论都必须承认，如果或当类AGI技术出现时，它们将带来超出我们以往所遇到的遏制问题。人类因为我们的智慧而主宰环境。更聪明的实体可能会因此主宰我们。AI研究员Stuart Russell称之为“猩猩问题”：猩猩在身体上比任何人类都更强壮、更坚韧，但它们却是濒危物种或生活在动物园中；它们被遏制。我们用我们微弱的肌肉但强大的大脑来进行遏制。

通过创造比我们更聪明的东西，我们可能会把自己置于与灵长类动物亲戚相同的位置。考虑到长远的观点，那些关注AGI情景的人有理由感到担忧。确实，有一个强有力的论点认为，根据定义，超级智能将完全无法控制或遏制。“智能爆炸”是指AI能够一次又一次地自我改进，以越来越快和更有效的方式递归地提升自己。这是明确的无法遏制和不可遏制的技术。直白的事实是，没有人知道AI何时、是否或确切地如何超越我们，以及接下来会发生什么；没有人知道它们何时或是否会变得完全自主，或者如何让它们在意识到并与我们的价值观保持一致的情况下行事，假设我们首先能够确定这些价值观。

没有人真正知道我们如何能够遏制在即将到来的浪潮中被如此深入研究的特性。技术会达到一个完全能够自我进化的阶段；它会受到递归改进过程的影响；它会超越解释的界限；因此，在自然环境中它的行为将无法预测；简而言之，我们将达到人类能动性和控制的极限。

最终，在其最极端的形式中，即将到来的浪潮可能意味着人类将不再处于食物链的顶端。*Homo technologicus* 可能最终会受到其自身创造物的威胁。真正的问题不是浪潮是否会到来。显然是的；只需看看你

可以看到它已经形成。鉴于这些风险，真正的问题是*why*，很难将其视为不可避免的。

CHAPTER 8

不可阻挡的激励措施

一个

LPHAGO'S SIGNIFICANCE WAS PARTLY A matter of timing: the breakthrough surprised experts by arriving more quickly than most in the AI community had thought possible. Even days before its first public competition in March 2016, prominent researchers thought an AI simply couldn't win at this level of Go. At DeepMind, we were still uncertain our program would prevail in a matchup with a master human competitor.

我们将这场比赛视为一项重大的技术挑战，是更广泛研究任务中的一个里程碑。在人工智能社区中，它代表了深度强化学习的首次高调公开测试，也是首次使用非常大型GPU计算集群进行研究。在媒体中，AlphaGo与李世石的对决被呈现为一场史诗般的战斗：人类对抗机器；人类中最优秀、最聪明的人与计算机冷酷无情的力量对抗。所有关于终结者和机器人霸主的陈词滥调都被提及。

但在表面之下，另一个更重要的维度变得清晰起来，这是我在比赛前隐约担心的一个紧张局势，但随着事件的展开，其轮廓更加鲜明。AlphaGo不仅仅是人类对抗机器。当李世石与AlphaGo对峙时，DeepMind由米字旗代表，而世石阵营则飘扬着韩国那显而易见的国旗。西方对东方。这种国家竞争的暗示是我很快就开始后悔的比赛方面。

在亚洲，这场比赛的受欢迎程度难以言表。在西方，比赛进程受到硬核AI爱好者的关注，并吸引了一些报纸的注意。这是科技领域的一个重要时刻。

历史——对于那些关心这些事情的人来说。然而，在整个亚洲，这一事件比超级碗还要大。超过2.8亿人观看了现场直播。我们占据了首尔市中心的一整家酒店，被当地和国际媒体的成员包围。你几乎无法在数百名摄影师和电视摄像机之间移动。这种强度是我以前从未经历过的，一种在西方观察者看来似乎是一个晦涩的数学爱好者游戏中显得陌生的审视和炒作水平。可以说，AI开发者对此并不习惯。

在亚洲，不仅仅是极客在观看。是所有人都在看。很快就清楚了，观察者包括科技公司、政府和军队。结果给他们都带来了震撼。没有人忽视其重要性。挑战者是一家西方公司，总部位于伦敦，美国所有，刚刚进入一个古老、标志性、珍视的游戏，字面上在草地上插上了旗帜，并彻底击败了主队。这就像一群韩国机器人出现在洋基体育场，击败了美国的全明星棒球队。

对我们来说，这次事件是一次科学实验。这是我们花费多年努力完善的尖端技术的强有力——而且是酷炫的——展示。从工程的角度来看，这令人兴奋；从竞争的角度来看，这令人振奋；而置身于媒体的旋涡中心则令人困惑。对于亚洲的许多人来说，这更是一种痛苦，是地区和国家自尊心受挫的实例。

首尔并不是AlphaGo的终点。一年后，在2017年5月，我们参加了第二场比赛，这次的对手是世界排名第一的棋手：柯洁。这场对决在中国乌镇的围棋峰会上举行。我们在乌镇的接待截然不同。在中华人民共和国，比赛的直播被禁止。谷歌的名字不被允许提及。环境更加严格，更加受控；叙述由当局精心策划。没有更多的媒体狂欢。潜台词很明确：这不再仅仅是一场游戏。AlphaGo再次获胜，但它是在一种明显紧张的氛围中做到的。

某些事情发生了变化。如果说首尔提供了一个暗示，那么乌镇则将其带回了家。尘埃落定后，显而易见的是，AlphaGo 是一个更大计划的一部分。

比一个奖杯、系统或公司更大的故事；这是大国参与一场新的、危险的技术竞争游戏的故事——以及一系列压倒性强大且相互关联的激励措施，确保即将到来的浪潮确实正在到来。

源文本：—翻译文本：

从好奇到危机，从财富到恐惧，技术的核心是为了满足人类的需求而出现的。如果人们有强烈的理由去构建和使用它，它就会被构建和使用。然而，在大多数关于技术的讨论中，人们仍然纠结于它是什么，忘记了它最初是为什么而创造的。这并不是某种与生俱来的技术决定论。这关乎于人类的意义。

早些时候我们看到，到目前为止，没有任何一波技术浪潮被遏制。在本章中，我们将探讨为什么历史可能会重演；为什么由于一系列推动技术发展和传播的宏观驱动因素，果实不会留在树上；为什么浪潮会破裂。只要这些激励措施存在，“我们应该吗？”这个重要问题就没有意义。

第一个驱动因素与我在AlphaGo上的经历有关：大国竞争。技术竞争是地缘政治的现实。事实上，它一直都是如此。国家感到有必要与同行保持同步。创新就是力量。其次是一个全球研究生态系统，其根深蒂固的惯例奖励开放出版、好奇心和不惜一切代价追求新想法。然后是来自技术的巨大财务收益以及解决全球社会挑战的迫切需要。最后一个驱动因素可能是最人性化的：自我。

在此之前，回到地缘政治，近期的过去提供了一个有力的教训。

NATIONAL PRIDE, STRATEGIC NECESSITY

源文本: 翻译文本:

战后美国认为其技术优势是理所当然的。斯普特尼克唤醒了它。1957年秋天，苏联发射了斯普特尼克，这是世界上第一颗人造卫星，人类首次涉足太空。它大约有一个沙滩球那么大，但仍然显得极其未来。斯普特尼克在上面供全世界观看，或者更确切地说，听到它的外星哔哔声在全球广播。实现这一壮举是不可否认的。

这对美国来说是一场危机，一场技术上的珍珠港事件。政策作出了反应。从高中到高级实验室，科学和技术成为国家优先事项，出现了新的资金和像NASA和DARPA这样的新机构。大量资源被投入到重大技术项目中，尤其是阿波罗任务。这些任务推动了火箭技术、微电子学和计算机编程的许多重要进步。北约等新兴联盟得到了加强。十二年后，成功将人类送上月球的是美国，而不是苏联。苏联几乎破产以试图跟上。随着斯普特尼克的发射，俄罗斯超越了美国，这是一个具有巨大地缘政治影响的历史性技术成就。但当美国需要挺身而出时，它做到了。

正如斯普特尼克最终使美国走上成为火箭、航天技术、计算机及其所有军事和民用应用领域超级大国的道路一样，现在中国也在发生类似的事情。AlphaGo迅速被称为中国在人工智能领域的斯普特尼克时刻。美国和西方国家，就像他们在互联网早期所做的那样，威胁要在一项划时代的技术上抢占先机。这是一个最清晰的提醒，中国在一项国家消遣活动中被击败，可能再次发现自己远远落后于前沿。

在中国，围棋不仅仅是一种游戏。它代表了更广泛的历史、情感和战略计算的纽带。中国已经致力于在科学和技术上进行大量投资，但AlphaGo更加明确地让政府关注人工智能。拥有数千年历史的中国，曾经是世界技术创新的熔炉；现在，它痛苦地意识到自己在技术竞赛中落后于欧洲人和美国人，在多个领域失去了优势。

从药品到航空母舰。它经历了中国共产党（CCP）所称的“ 百年屈辱 ”。党认为，这种情况绝不能再次发生。

时间到了，中共主张要收回其应有的地位。用习近平在2022年中共二十大上的话来说，为了“ 满足战略需求 ”，国家“ 必须坚持把科学技术作为第一生产力，把人才作为第一资源，把创新作为第一动力。 ”

中国的自上而下模式意味着它可以将国家的全部资源集中在技术目标上。如今，中国有一个明确的国家战略，即在2030年成为世界人工智能的领导者。新一代人工智能发展规划在柯洁被AlphaGo击败仅两个月后宣布，旨在将政府、军队、研究机构和行业结合在一起，进行一项集体任务。该计划宣称：“ 到2030年，中国的人工智能理论、技术和应用应达到世界领先水平，使中国成为世界主要的人工智能创新中心。 ”从国防到智慧城市，从基础理论到新应用，中国应占据人工智能的“ 制高点 ”。

这些大胆的声明不仅仅是空洞的姿态。当我写下这些时，仅仅在中国发布计划六年后，美国和其他西方国家在人工智能研究方面不再拥有过大的领先优势。像清华和北大这样的大学与斯坦福、麻省理工和牛津等西方机构竞争。事实上，清华发表的人工智能研究比世界上任何其他学术机构都多。中国在人工智能领域拥有越来越多且令人印象深刻的高被引论文。就人工智能研究的数量而言，自2010年以来，中国机构发表的人工智能论文数量是美国同行的四倍半，并且比美国、英国、印度和德国的总和还要多。

不仅仅是人工智能。从清洁技术到生物科学，中国在基础技术的各个领域迅速崛起，以史诗般的规模进行投资，成为具有“ 中国特色 ”的蓬勃发展的知识产权巨头。中国在2007年超过美国，成为博士学位授予数量最多的国家，但自那时起

对项目的投资和扩展非常显著，每年产生的STEM博士几乎是美国的两倍。超过四百个“国家重点实验室”支撑着一个资金充裕的公私研究系统，涵盖从分子生物学到芯片设计的各个领域。在二十一世纪初期，中国的研发支出仅为美国的12%。到2020年，这一比例已达到90%。按照当前趋势，到2020年代中期，中国的研发支出将显著领先，就像在专利申请方面已经领先一样。

中国是第一个在月球背面着陆探测器的国家。其他国家甚至都没有尝试过这一点。它拥有的全球前五百台超级计算机比其他任何地方都多。总部位于深圳的基因巨头华大基因集团，拥有非凡的DNA测序能力，既有私人和国家的支持，成千上万的科学家，以及庞大的DNA数据和计算能力储备。习近平明确呼吁进行“机器人革命”：中国安装的机器人数量与世界其他地区的总和相当。它建造了被美国认为需要多年才能实现的高超音速导弹，是从6G通信到光伏等领域的世界领导者，并且是腾讯、阿里巴巴、大疆、华为和字节跳动等大型科技公司的所在地。

量子计算是中国显著的专长领域。在爱德华·斯诺登泄露美国情报计划的机密信息后，中国变得特别偏执，并热衷于建立一个安全的通信平台。又一个斯普特尼克时刻。2014年，中国提交的量子技术专利数量与美国相同；到2018年，其提交的专利数量已是美国的两倍。

2016年，中国将世界上第一颗“量子卫星”墨子号送入太空，这是新型安全通信基础设施的一部分。但墨子号只是中国追求不可破解的量子互联网的开始。一年后，中国在上海和北京之间建立了一条两千公里的量子链路，用于传输安全的金融和军事信息。他们正在投资超过100亿美元在合肥创建国家量子信息科学实验室，这是世界上最大的此类设施。他们保持着将量子比特连接在一起的记录。

通过量子纠缠，这是通往成熟量子计算机的重要一步。合肥的科学家甚至声称建造了一台比谷歌突破性的Sycamore.快10^{14}倍的量子计算机。

墨子号的首席研究员和世界顶尖量子科学家之一潘建伟明确指出了这意味着什么。他说：“ 我认为我们已经开始了一场全球量子太空竞赛。现代信息科学方面，中国一直是一个学习者和追随者。现在，凭借量子技术，如果我们尽最大努力，我们可以成为主要参与者之一。 ”

几十年来，西方对中国能力的持续轻视，认为其“ 缺乏创造力 ”，是大错特错的。我们曾说他们只擅长模仿，过于受限和不自由，国有企业很糟糕。事后看来，这些评估大多是完全错误的，即使有些评估有其合理性，也未能阻止中国在科学和工程领域崛起为现代巨头——尤其是因为通过购买公司和翻译期刊等合法的知识产权转让，得到了公然的盗窃、强制转让、逆向工程和间谍活动的支持。

与此同时，美国正在失去其战略领先地位。多年来，美国在从半导体设计到制药、从互联网的发明到世界上最复杂的军事技术等各个领域都显然占据了霸主地位。这种优势并未完全消失，但正在逐渐消退。哈佛大学的格雷厄姆·艾利森（Graham Allison）的一份报告指出，情况比西方大多数人所认识的要严重得多。中国在绿色能源、5G和人工智能方面已经领先于美国，并且在未来几年内有望在量子技术和生物技术方面超越美国。五角大楼的首任首席软件官在2021年辞职以示抗议，因为他对这种情况感到非常失望。他告诉*Financial Times.*：“ 我们在15到20年内没有与中国竞争的机会。在我看来，现在已经是板上钉钉的事；已经结束了。 ”

2013年刚刚成为国家主席后不久，习近平发表了一次对中国乃至世界其他地区产生深远影响的讲话。他宣称：“ 先进技术是现代国家的锐利武器。我们的技术总体上仍然落后于发达国家，我们必须采取赶超的非对称战略。 ”

这是一项有力的分析，并且如我们所见，是中国政策优先事项的声明。但与习近平的许多言论不同，任何世界领导人都可以可信地提出同样的观点。任何美国或巴西总统、德国总理或印度总理都会认同这一核心论点——技术是使国家“掌控局势”的“利器”。习近平陈述了一个赤裸裸的事实，这不仅是中国的自我宣言，也是几乎每个国家的信条，从前沿的超级大国领导人到孤立的贱民：谁建造、拥有和部署技术，谁就重要。

军备竞赛

源文本: 翻译文本:

技术已成为世界上最重要的战略资产，与其说是外交政策的工具，不如说是其推动力。二十一世纪的大国竞争以技术优势为前提——这是一场控制即将到来的浪潮的竞赛。科技公司和大学不再被视为中立，而是被视为主要的国家冠军。

政治意愿可能会扰乱或取消本章中讨论的其他激励措施。理论上，一个政府可以抑制研究激励措施，打击私人企业，限制以自我为驱动的计划。但它无法消除来自地缘政治对手的激烈竞争。当感知到的对手不断推进时，选择限制技术发展在军备竞赛的逻辑中就是选择失败。

长期以来，我一直反对将技术进步框定为零和国际军备竞赛。在DeepMind，我总是反对将我们比作AI的曼哈顿计划，不仅因为核比较，还因为即使是这种框定也可能引发一系列其他曼哈顿计划，在需要密切的全球协调、突破点和减速的时候，助长军备竞赛的动态。但现实是，民族国家的逻辑有时是痛苦地简单却又完全不可避免的。在一个国家的国家安全背景下，仅仅提出一个想法就变得危险。一旦话语传出，

发令枪响起，言辞本身引发了剧烈的全国性反应。然后它开始失控。

在华盛顿和布鲁塞尔，政府、智库和学术界的无数朋友和同事都会重复同样令人恼火的说法：“ 即使我们实际上不在进行军备竞赛，我们也必须假设‘ 他们 ’认为我们在进行，因此我们自己必须竞赛以获得决定性的战略优势，因为这股新的技术浪潮可能会完全重新平衡全球力量。 ”这种态度成为自我实现的预言。

没有必要假装。与中国的强权竞争是华盛顿为数不多的两党共识领域之一。现在的争论不是我们是否处于技术和人工智能军备竞赛中，而是它将引向何方。

军备竞赛通常被视为中美双头垄断。这是目光短浅的。虽然这些国家确实是最先进和资源最丰富的，但还有许多其他国家是重要的参与者。这个新的军备竞赛时代预示着广泛的技术民族主义的兴起，其中多个国家将陷入不断升级的竞争中，以获得决定性的地缘政治优势。

几乎每个国家现在都有详细的AI战略。弗拉基米尔·普京认为，AI领域的领导者“ 将成为世界的统治者。 ”法国总统埃马纽埃尔·马克龙宣称，“ 我们将努力建设一个欧洲的元宇宙。 ”他的更广泛观点是，欧洲未能建立美国和中国的科技巨头，产生的突破较少，并且在科技生态系统的关键部分缺乏知识产权和制造能力。安全、财富、声望——在他和许多其他人看来，所有这些对欧洲来说都依赖于成为第三大力量。

各国在不同领域具有不同的优势，从生物科学和人工智能（如英国）到机器人技术（德国、日本和韩国）再到网络安全（以色列）。每个国家在即将到来的浪潮中都拥有主要的研发项目，蓬勃发展的民用初创企业生态系统越来越多地受到感知到的军事必要性的强大力量支持。

印度是新全球巨头秩序中显而易见的第四支柱，与美国、中国和欧盟并列。其人口年轻且具有创业精神，城市化程度日益提高，连接性和技术敏锐度不断增强。到2030年，其经济规模将超过英国、德国和日本等国家，成为世界第三大经济体；到2050年，其经济规模将达到$30万亿。

其政府决心将印度科技变为现实。通过其“ 自力更生的印度 ”计划，印度政府正努力确保这个世界上人口最多的国家在核心技术系统上实现与美国和中国竞争的所有权。在该计划下，印度与日本在人工智能和机器人技术方面建立了合作关系，并与以色列在无人机和无人驾驶飞行器方面建立了合作。准备迎接印度浪潮。

源文本：—翻译文本：

曼哈顿计划，占用了美国GDP的0.4%，被视为一场与时间赛跑的竞赛，以便在德国人之前获得炸弹。但纳粹最初排除了对核武器的追求，认为它们过于昂贵和投机。苏联远远落后，最终依赖于来自美国的大量泄密。美国与虚幻的对手进行了一场军备竞赛，使核武器比在其他情况下更早地进入了世界。

20世纪50年代末发生了类似的事情，当时在苏联洲际弹道导弹试验和人造卫星的影响下，五角大楼的决策者们确信与俄罗斯存在一个令人震惊的“ 导弹差距 ”。后来发现，在关键报告时，美国拥有十比一的优势。赫鲁晓夫正在遵循一种经过验证的苏联策略：虚张声势。误读对方意味着核武器和洲际弹道导弹都提前了几十年。

这种相同的错误动态是否可能在当前的技术军备竞赛中上演？实际上，不是。首先，未来浪潮的扩散风险是严重的。因为这些技术正在变得更便宜。

即使它们变得更强大，也更容易使用，更多的国家可以在前沿领域参与。大型语言模型仍被视为尖端技术，但它们并没有什么神奇之处或隐藏的国家机密。计算能力可能是最大的瓶颈，但有很多服务可以实现这一点。CRISPR或DNA合成也是如此。

我们已经可以看到像中国的月球着陆或印度的十亿级生物识别系统Aadhaar这样的成就实时发生。毫无疑问，中国拥有庞大的LLM，台湾是半导体的领导者，韩国在机器人方面拥有世界级的专业知识，各国政府都在宣布和实施详细的技术战略。这些都在公开进行，通过专利和学术会议共享，并在*Wired*中报道，*Financial Times,*在彭博社上直播。

宣布军备竞赛不再是一个变戏法的行为，也不再是一个自我实现的预言。这个预言已经实现。它就在这里，它正在发生。这是一个如此明显的点，以至于经常被忽略：没有中央权威控制哪些技术被开发，谁来开发，以及出于什么目的；技术是一支没有指挥的乐团。然而，这个单一的事实可能最终成为二十一世纪最重要的事实。

如果“军备竞赛”这个词引发了担忧，那是有充分理由的。几乎没有比建立在恐惧之上的零和竞争的认知（和现实）更不稳定的技术升级基础。然而，还有其他更积极的技术驱动因素值得考虑。

知识只想要自由

源文本: 翻译文本:

原始的好奇心、对真理的追求、开放性的重要性、基于证据的同行评审——这些是科学和技术研究的核心价值观。自从科学革命及其在十八和十九世纪的工业等价物以来，科学发现并没有像秘密珠宝一样被囤积，而是公开分享在期刊、书籍、沙龙中，

公开讲座。专利制度创造了一种分享知识的机制，同时奖励冒险行为。广泛获取信息成为我们文明的引擎。

开放是科学技术的基本理念。已知的必须共享；发现的必须发表。科学和技术依赖于自由辩论和信息的开放共享，以至于开放本身已成为一种强大（且极其有益）的激励。

我们生活在一个奥黛丽·库尔特·克罗宁称之为“开放技术创新”的时代。一个全球性的知识和技术开发系统现在如此庞大和开放，以至于几乎不可能引导、管理或在需要时关闭。理解、创造、构建和适应技术的能力因此高度分散。一年由计算机科学研究生完成的晦涩工作可能在下一年就掌握在数亿用户手中。这使得预测或控制变得困难。当然，科技公司希望保守他们的秘密，但他们也倾向于遵循软件开发和学术界的开放理念。创新因此传播得更快、更远且更具颠覆性。

开放性命令渗透了研究文化。学术界建立在同行评审的基础上；任何未经可信同行严格审查的论文都不符合黄金标准。资助者不喜欢支持那些被锁起来的工作。无论是机构还是研究人员都非常关注他们的出版记录以及他们的论文被引用的频率。更多的引用意味着更多的声望、可信度和研究资金。年轻研究人员尤其容易根据他们的出版记录来被评判和聘用，这些记录可以在像Google Scholar这样的平台上公开查看。此外，如今论文会在Twitter上发布，并且通常在撰写时考虑到社交媒体的影响。它们被设计得引人注目，以吸引注意力。

学术界热切地争论他们的研究应开放获取。在科技领域，围绕共享和贡献的强烈规范支持着开源软件的繁荣发展。一些世界上最大的公司——

字母表、Meta、微软——定期免费贡献大量知识产权。在人工智能和合成生物学等领域，科学研究和技术开发之间的界限尤其模糊，所有这些都使文化默认开放。

在 DeepMind，我们很早就了解到，发表机会是顶尖研究人员决定去哪里工作的关键因素。他们希望获得在学术界习惯的开放性和同行认可。很快，这成为顶尖 AI 实验室的标准：虽然并非所有内容都会立即公开，但开放性被视为吸引最佳科学家的战略优势。同时，发表记录是进入顶尖技术实验室的重要部分，而竞争激烈，是一场谁先公开的竞赛。

总而言之，在某种程度上，出版和分享的重要性可能被低估了，它们不仅仅是科学中证伪过程的一部分。它们也关乎声望、同行、使命的追求、工作的缘故、点赞。所有这些都推动并加速了技术发展的进程。

大量的AI数据和代码是公开的。例如，GitHub有1.9亿个代码库，其中许多是公开的。学术预印本服务器使研究人员能够快速上传工作，而无需任何审查或过滤机制。最初的此类服务arXiv，托管了超过两百万篇论文。数十个更专业的预印本服务，如生命科学领域的bioRxiv，推动了这一进程。世界上大量的科学和技术论文要么可以在开放网络上访问，要么可以通过易于获取的机构登录访问。这融入了一个跨境资助和合作成为常态的世界；在这个世界中，项目通常有数百名研究人员自由分享信息；成千上万关于最先进技术的教程和课程可以在网上轻松获取。

所有这些都发生在一个涡轮增压的研究环境中。全球研发支出每年超过7000亿美元，创下历史新高。仅亚马逊的研发预算就达780亿美元，如果它是一个国家，这将是世界第九大。Alphabet,

苹果、华为、Meta 和微软每年在研发上的支出都远远超过 200 亿美元。所有这些公司，那些最热衷于投资即将到来的浪潮的公司，那些预算最为充裕的公司，都有公开发布其研究成果的记录。

未来是显著的开源，发布在arXiv上，记录在GitHub上。它正在为引用、研究荣誉和终身教职的承诺而构建。开放的必要性和大量易于获取的研究材料意味着这是一套根深蒂固且广泛分布的激励和基础，未来的研究没有人能够完全掌控。

源文本：—翻译文本：

前沿是棘手的。如果你想引导研究过程，将其引向或远离某些结果，提前控制它，你将面临多重挑战。不仅存在如何在竞争团体之间进行协调的问题，而且在前沿领域也无法预测突破可能来自何处。

例如，CRISPR基因编辑技术的起源可以追溯到西班牙研究员弗朗西斯科·莫吉卡的工作，他想了解一些单细胞生物如何在微咸水中茁壮成长。莫吉卡很快偶然发现了DNA的重复序列，这些序列成为CRISPR的关键部分。这些成簇的重复片段似乎很重要。他提出了CRISPR这个名称。后来，丹麦一家酸奶公司的两位研究人员的工作研究了在酸奶发酵过程中保护对发酵起始培养物至关重要的细菌。这有助于展示核心机制可能如何运作。这些不太可能的途径构成了可以说是21世纪最大生物技术故事的基础。

同样，领域可以停滞几十年，但随后在几个月内发生巨大变化。神经网络在荒野中度过了几十年，被像马文·明斯基这样的杰出人物抨击。只有像Geoffrey Hinton和Yann LeCun这样的少数孤立研究人员在一个时期内坚持下去，当时

“ 神经 ”这个词曾经是如此有争议，以至于研究人员会故意将其从他们的论文中删除。在20世纪90年代，这似乎是不可能的，但神经网络却主导了人工智能。然而，也正是LeCun在AlphaGo取得第一次重大突破的前几天说它是不可能的。这并不是对他的贬低；这只是表明在研究前沿，没有人能对任何事情有绝对的把握。

即使在硬件方面，通往 AI 的道路也是无法预测的。GPU（图形处理单元）是现代 AI 的基础部分。但它们最初是为了在电脑游戏中提供越来越逼真的图形而开发的。在技术的多用途性的一次例证中，用于炫目图形的快速并行处理结果证明非常适合训练深度神经网络。最终，正是对照片级真实感游戏的需求使得像 NVIDIA 这样的公司投入大量资金来制造更好的硬件，而这又非常适合于机器学习。（NVIDIA 并没有抱怨；在 AlexNet 之后的五年里，其股价上涨了 1000%。）

如果你过去想要监控和指导人工智能研究，你可能会出错，阻碍或推动最终被证明无关紧要的工作，完全错过了在边缘悄然酝酿的最重要突破。科学和技术研究本质上是不可预测的，极其开放，并且发展迅速。因此，治理或控制它是非常困难的。

当今世界的优化速度前所未有，促进了好奇心、分享和研究。现代研究反对封闭。同样，盈利的必要性和欲望也是如此。

$100 万亿的机会

源文本: 翻译文本:

1830年，第一条客运铁路在利物浦和曼彻斯特之间开通。建造这一工程奇迹需要议会的批准。该路线需要桥梁、切割通道、架空段穿过

泥泞的地面和看似无休止的财产纠纷的解决：都是巨大的挑战。铁路的开通吸引了包括首相和利物浦议员威廉·赫斯基森在内的政要出席。在庆祝活动中，人群站在轨道上迎接这项新的奇迹。当它接近时，这台引人注目的机器是如此陌生，以至于人们未能意识到迎面而来的火车的速度，赫斯基森本人被火车头的车轮碾死。对于惊恐的观众来说，乔治·斯蒂芬森的蒸汽动力火箭是一个怪物，一个喷吐着烟雾的外星人，是现代性和机械的可怕模糊。

然而，这也是一种感觉，比当时体验过的任何东西都要快。增长迅速。预计每天有250名乘客；仅仅一个月后，每天就有1200人使用它。数百吨棉花可以从利物浦码头运到曼彻斯特的工厂，几乎不费吹灰之力，创下了纪录的时间。五年后，它的股息达到了10%，预示着1830年代铁路建设的小型繁荣。政府看到了更多的机会。1844年，一位名叫威廉·格莱斯顿的年轻议员提出了《铁路监管法》，以加速投资。1845年，仅仅几个月内，各公司就提交了数百份建造新铁路的申请。当其他股票市场停滞不前时，铁路公司蓬勃发展。投资者蜂拥而至。在其巅峰时期，铁路股票占据了总股票市场价值的三分之二以上。

一年之内，崩盘已经开始。市场最终在1850年触底，比高峰期低了66%。轻松的利润，让人们变得贪婪和愚蠢，这不是第一次，也不会是最后一次。成千上万的人失去了一切。尽管如此，随着繁荣的到来，一个新时代已经到来。随着火车的出现，一个古老而田园诗般的世界在高架桥和隧道、切割和大型车站、煤烟和汽笛声的闪电战中被撕成碎片。从几条零散的线路开始，投资热潮勾勒出了一个综合国家网络的轮廓。它缩小了国家的规模。在1830年代，从伦敦到爱丁堡的旅程需要几天的时间，乘坐不舒服的驿站马车。到1850年代，只需乘坐一列火车不到十二小时。与国家其他地区的连接意味着城镇、城市和地区的繁荣。旅游、贸易和家庭生活

被改变了。除了许多其他影响之外，它还产生了对标准化国家时间的需求，以便理解时间表。而这一切都归功于对利润的无尽渴望。

源文本：—翻译文本：

19世纪40年代被认为是“ 历史上最大的泡沫之一 ”。但在技术史上，这更像是常态而非例外。铁路的到来并非不可避免，但赚钱的机会却是不可避免的。卡洛塔·佩雷斯认为，在过去至少两百年的每一次重大技术推广中，都存在一个类似的“ 狂热阶段 ”，从最初的电话电缆到当代的高带宽互联网。繁荣从未持久，但原始的投机驱动产生了持久的变化，一个新的技术基础。

事实上，学术研究人员的好奇心或积极进取的政府的意愿不足以将新的突破推向数十亿消费者的手中。科学必须转化为有用且受欢迎的产品，才能真正广泛传播。简单来说：大多数技术都是为了赚钱而制造的。

如果说有什么不同的话，这也许是所有激励中最持久、根深蒂固、分散的。利润驱动中国企业家开发一种全新设计的手机模具；它推动荷兰农民寻找新的机器人和温室技术，以在北海的凉爽气候中全年种植西红柿；它引导帕洛阿尔托桑德希尔路上的精明投资者向未经考验的年轻企业家投资数百万美元。尽管他们的个人贡献者的动机可能各不相同，谷歌正在构建人工智能，亚马逊正在构建机器人，因为作为有股东需要取悦的上市公司，他们将其视为盈利的途径。

而这种潜在的利润，是建立在更持久和更强大的基础上的：原始需求。人们既想要又需要技术的成果。人们需要食物、制冷设备或电信来维持生活；他们可能想要空调设备，或者一种需要新设计的鞋子。

一些复杂的新制造技术，或某种革命性的新型蛋糕色素方法，或任何用于日常生活的无数技术用途。无论哪种方式，技术都有助于提供，而其创造者则从中分得一杯羹。人类需求和欲望的广泛性，以及从中获利的无数机会，是技术故事的核心，并将在未来继续如此。

这并不是坏事。回到几百年前，经济增长几乎不存在。生活水平在几个世纪里停滞不前，处于比今天糟糕得多的难以想象的水平。在过去的两百年里，经济产出增长了三百多倍。人均GDP在同一时期至少增长了十三倍，而在世界上最富有的地区则增长了一百倍。在十九世纪初，几乎所有人都生活在极端贫困中。现在，全球这一比例约为9%。曾经不可能的人类状况的指数级改善，现在已成为常态。

从根本上说，这是一个以利润为名系统地应用科学和技术的故事。这反过来又推动了产出和生活水平的巨大飞跃。在十九世纪，像赛勒斯·麦考密克的脱粒机这样的发明使得每小时小麦产量增加了500%。艾萨克·辛格的缝纫机意味着缝制一件衬衫的时间从十四小时减少到仅仅一小时。在发达经济体中，人们的工作时间远比过去少，而回报却多得多。例如，在德国，自1870年以来，年工作时间减少了近60%。

技术进入了创造财富的良性循环，这些财富可以再投资于进一步的技术发展，而这一切都提高了生活水平。但这些长期目标都不是任何个人的主要目标。在第1章中，我论证了你周围的几乎一切都是人类智慧的产物。这里有一个小小的修正：我们周围看到的许多东西都是由直接追求金钱利益的人类智慧驱动的。

这台引擎创造了一个价值 $85 万亿美元的世界经济——而且还在增长。从工业革命的先驱到今天硅谷的企业家，技术在其中具有磁性的激励作用。

形式的重大财务奖励。即将到来的浪潮代表了历史上最大的经济奖赏。它是一个无与伦比的消费者丰盛和潜在利润中心。任何想要遏制它的人都必须解释如何说服一个分散的、全球的、资本主义的无限权力系统来减缓其加速，更不用说将其搁置一旁。

源文本：—翻译文本：

保险索赔或采用新的制造技术时，它会创造效率节约或改进产品，从而提高利润并吸引新客户。一旦创新带来了这样的竞争优势，每个人都必须要么采用它，要么超越它，要么转移焦点，否则就会失去市场份额，最终破产。特别是在科技企业中，这种动态的态度简单而无情：要么构建下一代技术，要么被淘汰。

因此，企业在即将到来的浪潮中扮演如此重要的角色也就不足为奇了。科技在标普500指数中是迄今为止最大的单一类别，占指数的26%。主要科技集团之间持有的现金相当于像台湾或波兰这样的经济体的GDP。资本支出，如研发支出，巨大，超过了以前最大的支出者——石油巨头。最近关注该行业的人会看到围绕AI的商业竞争日益激烈，像谷歌、微软和OpenAI这样的公司每周都在争相推出新产品。

数千亿美元的风险资本和私募股权被投入到无数的初创公司中。仅在人工智能技术上的投资就已达到每年1000亿美元。这些庞大的数字确实很重要。巨额的资本支出、研发支出、风险资本和私募股权投资，是任何其他行业或中国和美国以外的任何政府都无法匹敌的，它们是推动即将到来的浪潮的原始燃料。所有这些资金都要求回报，而它所创造的技术就是获得回报的手段。

与工业革命一样，潜在的经济回报是巨大的。估计很难直观理解。普华永道预测，到2030年，人工智能将为全球经济增加$15.7 *trillion*。麦肯锡预测，同期生物技术将带来4万亿美元的提升。将全球机器人安装量提高到基线预测的30%之上，可能释放出5万亿美元的红利，这一数额比德国的全部产出还要大。尤其是在其他增长来源日益稀缺的情况下，这些都是强有力的激励。利润如此之高，中断这场淘金热可能会极其困难。

这些预测是否合理？这些数字确实令人瞠目结舌。从不久的将来中提取出巨大的数字在纸面上很容易做到。但在稍长的时间范围内，它们并非完全不合理。这里的总可寻址市场最终会像第一次或第二次工业革命一样扩展到整个世界经济。十八世纪末的人们会对人均GDP增长一百倍的想法感到难以置信。即使是考虑这种可能性也会显得荒谬。然而，这确实发生了。鉴于所有这些预测以及即将到来的浪潮所涉及的基本领域，即使在下一个十年中世界经济增长10-15%也可能是保守的。从长远来看，它可能远不止于此。

考虑到世界经济在二十世纪下半叶增长了六倍。即使在未来五十年中增长速度仅为这一水平的三分之一，它仍将释放出约 $100 万亿美元的额外 GDP。

想想新一波人工智能系统的影响。大型语言模型使您能够与人工智能就任何主题进行流利、自然语言的有用对话。在接下来的几年中，无论您的工作是什么，您都可以咨询按需专家，询问它关于您最新广告活动或产品设计的问题，询问它关于法律困境的具体细节，找出推销中最有效的元素，解决棘手的物流问题，获得诊断的第二意见，不断探究和测试，获得越来越详细的答案，这些答案基于最前沿的知识，并以极其细致的方式呈现。全世界的

知识、最佳实践、先例和计算能力将随时为您提供，量身定制以满足您的特定需求和情况，瞬间且毫不费力。这是认知潜力的一次飞跃，至少与互联网的引入一样伟大。而这还未涉及诸如ACI和现代图灵测试之类事物的影响。

几乎没有什么比智力更有价值。智力是世界经济的源泉、指导者、建筑师和促进者。我们提供的智力范围和性质越广泛，增长的可能性就越大。通过通用人工智能，合理的经济情景表明，它不仅可以促进增长，还可以使增长率本身永久加速。用简单的经济术语来说，人工智能 *could,* 长期来看，将是迄今为止最有价值的技术，尤其是在与合成生物学、机器人技术等结合时更是如此。

这些投资并非被动的；它们将在实现这一目标中发挥重要作用，成为另一个自我实现的预言。这些数万亿代表了对社会的巨大价值增值和机遇，为数十亿人提供更好的生活水平，同时为私人利益带来巨大的利润。无论哪种方式，这都创造了一种根深蒂固的激励机制，以不断寻找和推出新技术。

全球挑战

源文本: 翻译文本:

在历史的大部分时间里，仅仅养活自己和家人就是人类生活的主要挑战。农业一直是一项艰难而不确定的事业。但尤其是在二十世纪的改进之前，这要困难得多。任何天气条件的变化——太冷、太热、太干或太湿——都可能是灾难性的。几乎所有的工作都是手工完成的，如果幸运的话，可能会有一些牛的帮助。在一年中的某些时候几乎无事可做；而在其他时候，则是连续数周的无休止、令人筋疲力尽的体力劳动。

作物可能会被疾病或害虫毁坏，收获后腐烂，或被入侵的军队偷走。大多数农民过着勉强糊口的生活，常常

作为农奴工作，放弃了他们大部分稀少的作物。即使在世界上最富饶的地区，产量也很低且不稳定。生活艰难，处于灾难的边缘。1798年，当托马斯·马尔萨斯提出快速增长的人口将很快耗尽农业的承载能力并导致崩溃时，他并没有错；静态产量将且经常确实遵循这一规律。

他没有考虑到的是人类创造力的规模。假设天气条件良好并使用最新技术，在十三世纪，英格兰每公顷小麦的产量约为半吨。这个产量在几个世纪里保持不变。随着新技术和新工艺的逐渐出现，这一切都发生了变化：从轮作到选择性育种，机械化犁地，合成肥料，杀虫剂，基因改造，现在甚至还有人工智能优化的种植和除草。在二十一世纪，产量现在约为每公顷八吨。正是同一块小而不起眼的土地，同样的地理和土壤，在十三世纪被收割的，现在可以提供十六倍的作物。在过去的五十年里，美国每公顷玉米的产量增加了三倍。自十九世纪初以来，生产一公斤粮食所需的劳动力减少了98%。

1945年，世界上大约50%的人口严重营养不良。如今，尽管人口增加了三倍多，这一比例已降至10%。这仍然代表着超过6亿人，这是一个不可接受的数字。但按1945年的比例，这个数字将是40亿，尽管事实上这些人不可能存活下来。我们很容易忽视我们已经走了多远，以及创新有多么非凡。中世纪的农民会为现代农民的巨大联合收割机、史诗般的灌溉系统付出什么？对他们来说，十六倍的改进无异于一个奇迹。确实如此。

养活全世界仍然是一个巨大的挑战。但这种需求推动了技术的发展，并导致了以前难以想象的丰富：食物充足，即使分配不均，也足以养活地球上80亿且不断增长的人口。

技术，就像食品供应一样，是解决人类今天不可避免地面临的挑战以及明天将要面临的挑战的重要组成部分。我们追求新技术，包括即将到来的浪潮中的技术，不仅仅是因为我们想要它们，而是因为在根本层面上，我们*need*它们。

源文本：—翻译文本：

世界正走向两摄氏度或更高的气候变暖。每天每秒，生物圈的边界——从淡水使用到生物多样性丧失——都在被突破。即使是最有韧性、气候温和且富裕的国家，在未来几十年也将遭受灾难性的热浪和干旱、风暴和水资源压力。农作物将会歉收。野火肆虐。大量甲烷将从融化的永久冻土中释放出来，威胁着极端升温的反馈循环。疾病将传播到远超其通常范围的地区。随着海平面不可逆转地上升，气候难民和冲突将席卷全球，威胁主要人口中心。海洋和陆地生态系统面临崩溃。

尽管关于清洁能源转型的讨论是有充分理由的，但仍需走的路程是巨大的。碳氢化合物的能量密度极难复制，用于为飞机或集装箱船提供动力等任务。虽然清洁电力的生产正在快速扩张，但电力仅占全球能源产出的约25%。其余75%的转型要困难得多。自21世纪初以来，全球能源使用量增加了45%，但化石燃料的份额仅从87%下降到84%——这意味着尽管在清洁电力作为电源方面采取了所有措施，化石燃料的使用仍大幅增加。

能源学者Vaclav Smil称氨、水泥、塑料和钢铁为现代文明的四大支柱：支撑现代社会的物质基础，每一种的生产都极为碳密集，且没有明显的替代品。没有这些材料，现代生活将停止；没有化石燃料，这些材料将无法生产。过去三十年中，7000亿吨排放碳的混凝土被倾泻到我们的社会中。如何替代这些材料？电动汽车在行驶时可能不会排放碳，但它们在生产过程中却会。

资源仍然紧缺：仅仅制造一辆电动汽车就需要提取约225吨有限的原材料，而对这些材料的需求已经在不可持续地激增。

正如我们所见，食品生产是技术的一个重大成功案例。但是，从田间的拖拉机，到合成肥料，再到塑料温室，它都充满了化石燃料。想象一下，平均每个番茄都浸泡在五汤匙的油中。这就是种植它所需的油量。更重要的是，为了满足全球需求，农业到2050年需要增产近50%的食品，而与此同时，气候变化导致的产量下降。

如果我们要有任何机会将全球变暖控制在两摄氏度以内，那么在联合国政府间气候变化专门委员会下工作的世界科学家们已经明确表示：碳捕获和储存是一项必不可少的技术。然而，它在很大程度上尚未被发明或尚未大规模部署。为了应对这一全球挑战，我们必须从头开始重新设计我们的农业、制造业、交通和能源系统，采用碳中和甚至可能是碳负的新技术。这些任务并不简单。实际上，这意味着重建现代社会的整个基础设施，同时希望还能为数十亿人提供生活质量的改善。

人类别无选择，只能应对这些挑战，以及许多其他挑战，例如如何向患有顽固慢性病的老龄化人口提供越来越昂贵的医疗服务。那么，这里有另一个强大的激励因素：这是我们在面对似乎超出我们能力的艰巨任务时如何蓬勃发展的重要部分。除了利润或优势之外，新技术还有一个强有力的道德理由。

技术能够并将改善生活并解决问题。想象一个由寿命更长且吸收更多CO_2的树木组成的世界。或者帮助海洋成为更大、更可持续碳汇的浮游植物。人工智能已经帮助设计了一种可以分解堵塞我们海洋的塑料的酶。它也将在预测未来方面发挥重要作用，例如猜测野火可能袭击的地点。

从郊区到通过公共数据集追踪森林砍伐。这将是一个廉价、个性化药物的世界；快速、准确的诊断；以及由AI生成的替代高能耗肥料的产品。

可持续、可扩展的电池需要激进的新技术。量子计算机与人工智能结合，凭借其在分子层面建模的能力，可以在寻找传统锂电池替代品方面发挥关键作用，这些替代品更轻、更便宜、更清洁、更易于生产和回收，并且更为丰富。同样，在光伏材料或药物发现的工作中，这种能力使得分子级模拟能够识别新化合物——比过去使用缓慢的实验技术要精确和强大得多。这是超进化的实际应用，并承诺在研发中节省数十亿美元，同时远远超越当前的研究范式。

一种天真技术解决主义的学派认为技术是解决世界所有问题的答案。单靠技术并不是。它的创造、使用、拥有和管理方式都会产生影响。没有人应该假装技术是对气候变化这样复杂而巨大的问题的近乎神奇的答案。但是，认为我们可以在没有新技术的情况下应对本世纪的重大挑战是完全不切实际的。同样值得记住的是，这一波技术将使数十亿人的生活更轻松、更健康、更高效、更愉快。它们将节省时间、成本、麻烦，并拯救数百万人的生命。在不确定性中，不应轻视或遗忘这一点的重要性。

即将到来的浪潮之所以到来，部分原因是没有它就无法前行。像这样的巨型系统性力量推动着技术向前发展。但在我的经验中，另一个更个人化的力量始终存在且被大大低估：自我。

自我

源文本: 翻译文本:

科学家和技术人员都是非常人性化的。他们渴望地位、成功和遗产。他们希望成为第一个、最好的，并被认可为这样的人。

他们具有竞争力和聪明才智，对自己在世界和历史中的地位有着精心培育的认知。他们喜欢挑战极限，有时是为了金钱，但更多时候是为了荣耀，有时只是为了挑战本身。人工智能科学家和工程师是世界上收入最高的人之一，但真正让他们起床的是率先取得突破或在里程碑式的论文上看到自己的名字的前景。无论是爱他们还是恨他们，科技巨头和企业家都被视为权力、财富、远见和纯粹意志的独特指引者。批评者和奉承的粉丝都将他们视为自我表现，擅长促成事情的发生。

工程师通常有一种特定的心态。洛斯阿拉莫斯的主任J.罗伯特·奥本海默是一个高度有原则的人。但最重要的是，他是一个好奇心驱动的问题解决者。考虑这些话，以他们自己的方式，和他那句著名的《薄伽梵歌》引言一样令人不寒而栗（在看到第一次核试验时，他回忆起一些印度教经文中的句子：“ 现在我成为死亡，世界的毁灭者 ”）：“ 当你看到某件技术上很甜美的东西时，你就去做，只有在你取得技术成功后才讨论如何处理它。 ”这是他在曼哈顿计划中的同事，才华横溢的匈牙利裔美国人约翰·冯·诺依曼所共有的态度。“ 我们现在创造的， ”他说，“ 是一个怪物，它的影响将改变历史，前提是还有历史存在，但不把它完成是不可能的，不仅仅是出于军事原因，从科学家的角度来看，不去做他们知道可行的事情也是不道德的，无论它可能带来多么可怕的后果。 ”

在技术环境中花费足够的时间，尽管有很多关于伦理和社会责任的讨论，你会认识到这种观点的普遍性，即使在面对极端强大的技术时也是如此。我已经多次见过这种情况，如果我说自己偶尔没有屈服于这种情况，可能是在撒谎。

创造历史，做有意义的事情，帮助他人，击败他人，给潜在伴侣留下深刻印象，给老板、同事、对手留下深刻印象：这一切都在其中，都是始终存在的冒险、探索的动力的一部分。

边缘，深入未知。 建立新的东西。 改变游戏。 攀登高山。

无论是高尚和高尚，还是充满怨恨和零和，当你从事技术工作时，往往是这个方面，甚至比国家的需求或远方股东的要求更能激励进步。找到一个成功的科学家或技术专家，在某个地方你会看到一个被原始自我驱动的人，被情感冲动所激励，这些冲动可能听起来低级甚至不道德，但仍然是我们获得技术的一个未被充分认识的部分。硅谷的神话是，英雄式的初创公司创始人在一个敌对和无知的世界中单枪匹马地建立帝国，这种神话之所以持久是有原因的。这是技术人员常常仍然渴望的自我形象，是一个要效仿的原型，是一个仍然推动新技术的幻想。

源文本：—翻译文本：

这些现在已经是世界的嵌入特征。简单地将它们从场景中移除在任何有意义的时间框架内都是不可能的。利他主义和好奇心，傲慢和竞争，赢得比赛、成名、拯救你的人民、帮助世界的愿望，无论是什么：这些是推动浪潮前进的动力，而这些是无法消除或规避的。

此外，这些不同的激励和波浪元素相互复合。国家军备竞赛与企业竞争相辅相成，而实验室和研究人员则相互激励。换句话说，一系列嵌套的子竞赛汇聚成一个复杂的、相互强化的动态。技术通过无数独立的贡献“ 涌现 ”，这些贡献层层叠加，形成一个不断扩展、纠缠不清的想法混乱体，由深层次和分散的激励推动。

没有以光速传播信息的工具，过去的人们可以心满意足地与新技术相对而坐，有时甚至几十年后才意识到其全部影响。即使他们意识到了，也需要花费大量时间，最终还需要想象力，才能完全实现。

广泛的影响。今天，世界正在实时观察其他人的反应。

一切都会泄露。一切都会被复制、迭代、改进。因为每个人都在观察和学习他人，很多人都在同一个领域中摸索，总会有人不可避免地找到下一个重大突破。而且他们将无法遏制它，因为即使他们做到了，也会有其他人跟在他们后面，发现相同的见解或找到一种相邻的方法来做同样的事情；他们会看到战略潜力、利润或声望，并追求它。

这就是为什么我们不会说不。这就是为什么即将到来的浪潮 *is* 即将来临，为什么控制它是如此具有挑战性。技术现在是一个不可或缺的超级系统，渗透到日常生活、社会和经济的各个方面。没有人能离开它。为了更多的技术，甚至是极大地增加技术，已经有了根深蒂固的激励措施。没有人能完全控制它的作用或它的下一步去向。这不是一些遥远的哲学概念，也不是极端决定论的情景或狂热的加州技术中心主义。这是对我们所有人所居住的世界的基本描述，实际上是我们已经居住了相当长时间的世界。

从这个意义上说，技术就像一个无情的形象，一个巨大的粘菌，缓慢地滚向不可避免的未来，每个学者或企业家都在做出数以亿计的微小贡献，没有任何协调或抵抗的能力。强大的吸引子将其拉动。当障碍出现时，其他地方会出现空隙，整体向前滚动。减缓这些技术的发展与国家、企业和研究利益背道而驰。

这是一个终极的集体行动问题。认为CRISPR或AI可以被收回是不可信的。在有人能够创造出一个合理的途径来拆除这些相互交织的激励之前，不建造、说不、甚至只是放慢速度或走不同的道路的选项都不存在。

包含技术意味着短路所有这些相互强化的动态。很难想象如何在任何时间范围内做到这一点，以影响即将到来的浪潮。只有一个

一个或许能够提供解决方案的实体，一个锚定我们政治体系并对社会生产的技术承担最终责任的实体：民族国家。

但是有一个问题。各州已经面临巨大的压力，而即将到来的浪潮似乎将使事情变得更加复杂。这次碰撞的后果将塑造本世纪的其余时间。

PART III

STATES OF FAILURE

CHAPTER 9

大交易

国家的承诺

A

T ITS HEART, THE NATION-STATE, THE central unit of the world's political order today, offers its citizens a simple and highly persuasive bargain: that not only is centralization of power in the sovereign, territorial state possible but its benefits far outweigh the risks. History suggests that a monopoly over violence—that is, entrusting the state with wide latitude to enforce laws and develop its military powers—is the surest way to enable peace and prosperity. That, moreover, a well-managed country is a key foundation of

即使随着它变得更强大并与日常生活更加纠缠不清，民族国家的重大交易因此在于，不仅集中权力可以实现和平与繁荣，而且这种权力可以通过一系列的制衡、再分配和制度形式来加以限制。我们常常理所当然地认为，为了维持这种状态，必须在极端之间取得微妙的平衡。一方面，必须避免集中权力的最反乌托邦的过度行为，另一方面，我们必须接受定期干预以维持秩序。

今天，比历史上任何时候都更甚，未来浪潮的技术威胁着打破这种脆弱的平衡。简单来说，

大交易正在破裂，而技术是这一历史性转变的关键推动力。

鉴于民族国家负责以其人口的最大利益来管理和调节技术的影响，他们对即将到来的变化准备得如何？如果国家无法协调遏制这一浪潮，无法确保其对公民的净利益，那么在中长期内，这给人类留下了哪些选择？

在本书的前两部分中，我们看到一波强大的技术即将席卷而来。现在是时候考虑这意味着什么，并瞥见洪水过后的世界。

在本书的第三部分中，我们探讨了这些技术对民族国家，尤其是自由民主民族国家的深远影响。裂缝已经开始形成。曾经促进财富增长、生活水平提高、教育、科学和技术发展的政治秩序，一个趋向于和平的世界，如今正面临巨大的压力，部分被它自身帮助产生的力量所动摇。其全部影响范围广泛且难以理解，但在我看来，它们预示着一个未来，在这个未来中，遏制的挑战比以往任何时候都更为艰难，本世纪的重大困境变得不可避免。

哥本哈根的教训：政治是个人的事

源文本: 翻译文本:

我一直坚信国家有能力改善生活。在我的人工智能职业生涯之前，我曾在政府和非营利部门工作。我在十九岁时帮助创办了一个慈善电话咨询服务，为伦敦市长工作，并共同创立了一家专注于多方利益相关者谈判的冲突解决公司。与公务员合作——这些人精疲力竭，但始终需求旺盛，为需要帮助的人做着英雄般的工作——足以让我看到如果国家失败将会是多么大的灾难。

然而，我在地方政府、联合国谈判和非营利组织的经验也让我获得了关于其局限性的宝贵第一手知识。它们往往长期管理不善、臃肿且行动迟缓。我在2009年哥本哈根气候谈判中促成的一个项目涉及召集数百个非政府组织和科学专家，以统一他们的谈判立场。这个想法是向主要峰会上的192个争吵不休的国家提出一个连贯的立场。

除了我们无法就任何事情达成共识。首先，没有人能就科学或地面上正在发生的现实达成一致。优先事项分散。没有关于什么是有效的、可负担的，甚至是实用的共识。你能筹集100亿美元将亚马逊变成一个吸收CO_2的国家公园吗？你打算如何处理民兵和贿赂？或者答案可能是重新造林挪威，而不是巴西，还是解决方案是种植巨型海带农场？一旦提案被提出，就有人站出来挑毛病。每个建议都是一个问题。我们最终在所有可能的事情上达到了最大分歧。换句话说，这就是通常的政治。

而且，这涉及到名义上属于“同一团队”的人。我们甚至还没有进入主要事件和真正的讨价还价。在哥本哈根峰会上，一片混乱的国家都有各自的竞争立场。现在再加上原始的情感。谈判者试图在房间里有数百人争论、喊叫并分成小组的情况下做出决定，而此时钟声在峰会和地球上都在滴答作响。我在那里试图帮助促进这一过程，这可能是人类历史上最复杂、风险最高的多方谈判，但从一开始看起来几乎是不可能的。观察到这一点，我意识到我们无法足够快地取得足够的进展。时间表太紧了。问题太复杂了。我们解决重大全球问题的机构不适合这个目的。

我二十出头时为伦敦市长工作时看到过类似的事情。我的工作是审计人权立法对城市社区的影响。我采访了从英国孟加拉人到当地犹太团体的每一个人，年轻和年老，各种信仰和背景的人。

经验表明，人权法可以以非常实际的方式帮助改善生活。与美国不同，英国没有保护人民基本权利的成文宪法。现在，当地团体可以将问题提交给地方当局，并指出他们有法律义务保护最弱势群体；他们不能将其掩盖。在某种程度上，这令人鼓舞。这给了我希望：机构可以有一套关于正义的成文规则。系统可以实现。

但当然，伦敦政治的现实却截然不同。实际上，一切都演变成了借口、推卸责任、媒体炒作。即使有明确的法律责任，部门或委员会也不会回应，会敷衍、躲避和拖延。面对真正的挑战，停滞不前是普遍现象。

进入伦敦市政厅时，我刚满二十一岁。那是2005年，我天真地乐观。我相信地方政府——以及联合国；对外人来说，它们似乎是宏伟而有效的机构，我们可以在其中共同努力解决重大问题。我认为，像当时的许多人一样，全球化和自由民主是默认的、历史的理想终点。与现实的接触足以显示无望的理想与地面事实之间的鸿沟。

大约在那个时候，我也开始关注另一件正在成形的事情。Facebook正以空前的速度增长。不知何故，即使从地方政府到联合国的一切似乎都在缓慢运作，这家小型初创公司在短短几年内就发展到了超过 1 亿的月活跃用户。这个单一事实改变了我的人生轨迹。我非常清楚，一些组织仍然能够在大规模上进行高效的行动，并且它们正在新的领域运作，比如在线平台。

单靠技术就能解决社会和政治问题的想法是一种危险的错觉。但认为可以在没有技术的情况下解决这些问题也是错误的。近距离看到公务员的挫折让我想找到其他有效的方法来大规模地完成事情，不是与国家对抗，而是与国家合作，以创造更具生产力、更公平、更友善的社会。

技术突破将帮助我们应对上一节中提到的挑战：在不可持续的温度下种植粮食；提前检测洪水、地震和火灾；提高每个人的生活水平。在成本不断攀升和服务质量下降的时代，我认为人工智能和合成生物学是加速进步的关键杠杆。它们将使医疗保健质量更高且更负担得起。它们将帮助我们发明工具，以在政治停滞不前的情况下实现向可再生能源的过渡并应对气候变化，并支持教师，帮助提高资金不足的教育系统的有效性。这是下一波浪潮的真正潜力。

因此，我开始了技术领域的职业生涯，相信新一代工具可以放大我们大规模行动的能力，其运作速度远远超过传统政策。将它们投入使用以“创造未来”似乎是度过我生命中最具生产力的岁月的最佳方式。

我唤起我的理想主义倾向，将以下章节置于背景中，以明确表示我认为这幅常常被描绘得阴暗的图景是技术的巨大失败，也是像我这样构建技术的人的失败。

尽管技术仍然是解决二十一世纪挑战的最强大途径，但我们不能忽视其负面影响。在承认众多好处的同时，我们也必须克服悲观厌恶，冷静而严肃地审视全用途技术可能带来的新风险。随着时间的推移，这些风险的性质和利害关系的大小变得更加清晰。技术不仅是支持我们在民族国家中所做交易的工具，它也是对其的真正威胁。

在科技行业中，一个有影响力的少数群体不仅认为新技术对我们有序的国家世界构成威胁；这个群体还积极欢迎其消亡。这些批评者认为，国家大多是障碍。他们认为最好抛弃国家，认为它已经如此麻烦，以至于无法挽救。我从根本上不同意；这样的结果将是一场灾难。

我是英国人，在伦敦出生和长大，但我家的一方是叙利亚人。我的家人被卷入了那场可怕的战争。

近年来的国家。我很清楚当国家失败时的情形，简单来说，那是难以想象的糟糕。可怕。任何认为叙利亚发生的事情永远不会在“这里”发生的人都是在自欺欺人；无论身处何地，人都是人。我们的民族国家体系并不完美，远非如此。尽管如此，我们必须竭尽全力支持和保护它。这本书，部分是我为其辩护的尝试。

没有其他东西——没有其他银弹——会及时到来拯救我们，吸收这股浪潮的不稳定力量。在中期内根本没有其他选择。

即使在最好的情况下，即将到来的浪潮也将对治理社会的系统造成巨大冲击。在我们探讨浪潮的危险之前，值得询问国家的整体健康状况。它们是否有能力应对未来的挑战？

脆弱国家

源文本: 翻译文本:

全球生活条件客观上比以往任何时候都要好。我们认为自来水和充足的食物供应是理所当然的。大多数人全年都享有温暖和庇护。识字率、预期寿命和性别平等处于历史最高水平。数千年的人类学术和探究的总和可以通过按下按钮获得。对于大多数发达国家的人来说，生活的轻松和富足在过去的时代似乎是难以置信的。然而，在表面之下，总有一种挥之不去的感觉，似乎有什么地方不太对劲。

西方社会尤其陷入了一种根深蒂固的焦虑中；它们是“紧张状态”，冲动且易怒。这种持续的不安部分是由于先前的冲击——多次金融危机、大流行病、暴力（从9/11到乌克兰战争）——以及部分是长期和日益增长的压力的影响，如公共信任下降、不平等加剧和气候变暖。在即将到来的浪潮中，许多国家正面临一系列重大挑战的冲击。

有效性，使它们变得更弱、更分裂，更容易做出缓慢和错误的决策。即将到来的浪潮将落在一个易燃、无能、过度紧张的环境中。这使得遏制的挑战——控制和引导技术以使其对人类净有益——更加艰巨。

源文本：—翻译文本：

信任。人们需要相信政府官员、军队和其他精英不会滥用他们的主导地位。每个人都依赖于信任，即税收会被缴纳、规则会被遵守、整体利益会被置于个人利益之上。没有信任，从投票箱到报税表，从地方议会到司法机构，社会都会陷入困境。

对政府的信任，尤其是在美国，已经崩溃。根据皮尤调查，战后总统政府如艾森豪威尔和约翰逊的政府被超过70%的美国人信任会做“正确的事”。对于最近的总统如奥巴马、特朗普和拜登，这一信任度已大幅下降，均低于20%。相当惊人的是，2018年一项关于美国民主的研究发现，多达五分之一的人认为“军队统治”是个好主意！不少于85%的美国人觉得国家正“走在错误的方向上”。不信任也延伸到非政府机构，对媒体、科学机构以及一般的专业知识的信任度也在下降。

这个问题不仅限于美国。另一项皮尤调查发现，在二十七个国家中，大多数人对他们的民主制度感到不满。民主感知指数调查发现，在五十个国家中，三分之二的受访者认为政府“很少”或“从不”以公众利益为重。如此多的人深刻地感受到社会正在失败，这本身就是一个问题：不信任滋生消极和冷漠。人们拒绝投票。

自2010年以来，倒退的国家在民主指标上的表现多于进步的国家，这一过程似乎正在加速。从波兰和中国到俄罗斯、匈牙利、菲律宾和土耳其，民族主义和威权主义似乎已成常态。民粹主义运动从离奇的（如QAnon）到无方向的（如法国的*gilets jaunes*），但从巴西的博尔索纳罗到英国的脱欧，它们在世界舞台上的显著性是不可忽视的。

在新的威权冲动和政治不稳定的背后，是日益增长的社会不满情绪。作为不稳定和社会不满的关键催化剂，不平等在最近几十年间在西方国家激增，尤其是在美国。1980年至2021年间，收入最高的1%所占的国民收入份额几乎翻了一番，现在接近50%。财富越来越集中在一个小圈子里。政府政策、工作年龄人口的缩减、教育水平的停滞以及长期增长的减速都促成了更加不平等的社会。在美国，有四千万人生活在贫困中，超过五百万人生活在“第三世界条件”下——这一切都发生在世界上最富有的经济体中。

当你考虑到社会流动性停滞、日益加剧的不平等和政治暴力之间的持续关系时，这些趋势尤其令人担忧。来自一百多个国家的数据表明，一个国家的社会流动性越低，就越容易经历骚乱、罢工、暗杀、革命运动和内战等动荡。当人们感到被困住，认为他人不公平地占有了回报时，他们会感到愤怒。

不久前，世界被认为是“平的”——一个无摩擦的贸易和繁荣上升的地带。事实上，随着21世纪的推进，供应链紧缩和金融冲击仍然是经济不可磨灭的特征。那些倾向于民族主义的国家，部分是在背离20世纪的光明承诺，即更大的互联互通将加速财富和民主的传播。

回流、本土化、国家安全、弹性供应链、自给自足——当今的贸易语言再次成为边界、壁垒和关税的语言。与此同时，食品、能源、原材料和各种商品的价格都变得更加昂贵。基本上，整个战后安全和经济秩序正面临前所未有的压力。

全球挑战正达到一个关键门槛。肆虐的通货膨胀。能源短缺。收入停滞。信任的崩溃。民粹主义的浪潮。无论是左派还是右派的旧愿景似乎都没有提供令人信服的答案，而更好的选择似乎也供不应求。要说一切都好，没有严重的民粹主义、愤怒和功能失调的力量在社会中肆虐，需要一个勇敢的人，或者可能是一个妄想的人——尽管世界生活水平已达到历史最高。

这使得遏制变得更加复杂。围绕快速发展的技术形成国家和国际共识并建立新规范已经是艰巨的挑战。当我们的基本模式似乎是不稳定时，我们如何能希望做到这一点？

技术是政治的：波浪对国家的挑战

源文本: 翻译文本:

每一波技术浪潮都具有深远的政治影响。我们应该预期未来也会如此。上一波浪潮——大型机、台式电脑和桌面软件、互联网和智能手机的到来——为社会带来了巨大的好处。它为现代经济奠定了新的工具，促进了增长，改变了获取知识、娱乐和彼此联系的方式。在当前对社交媒体负面影响的担忧中，很容易忽视这些无数的积极因素。然而，在过去十年中，越来越多的共识表明，这些技术也做了其他事情：创造了滋养和放大这种潜在政治极化和制度脆弱性的条件。

社交媒体平台能够引发直觉的情绪反应，这几乎不是什么新闻，感知到的威胁有效地带来了肾上腺素的冲击。社交媒体依赖于高涨的情绪，并且常常是愤怒。发表在期刊 *Nature* 上的一项元分析回顾了近五百项研究的结果，得出结论认为数字媒体使用的增加与政治不信任、民粹主义运动、仇恨和极化的上升之间存在明显的相关性。相关性可能不是因果关系，但这项系统性回顾提出了来自新技术的“ 对民主的严重威胁的明确证据 ”。

技术已经侵蚀了民族国家稳定的主权边界，创造或支持了人、信息、思想、技术诀窍、商品、成品、资本和财富的全球流动。正如我们所见，它是地缘政治战略的重要组成部分。它几乎触及人们生活的每一个方面。即使在即将到来的浪潮袭来之前，技术已经是世界舞台上的驱动力，是全球各地民族国家健康状况恶化的主要因素。其发展速度过快，过于全球化，过于多变且诱人，无法用任何简单的遏制模型来控制，战略上至关重要，被数十亿人依赖，现代技术本身就是一个主要参与者，是民族国家难以管理的巨大力量。人工智能、合成生物学等正在被引入到在强大技术浪潮中摇摆不定的功能失调社会*already*。这不是一个为即将到来的浪潮做好准备的世界。这是一个在现有压力下屈服的世界。

源文本：—翻译文本：

有人说技术是“ 价值中立的 ”，其政治性来自于其使用方式。这种说法过于简化，以至于几乎毫无意义。技术并没有直接“ 导致 ”或创造现代国家（或任何政治结构）。但它释放的潜力在这个故事中并不是中立的。

正如技术史学家兰登·温纳所说：“ 技术在其各种表现形式中是人类世界的重要组成部分。其

结构、过程和改变进入并成为人类意识、社会和政治的结构、过程和改变的一部分。”换句话说，技术是政治的。

这一事实不仅被我们的领导人严重忽视，甚至连那些构建技术本身的人也未能充分认识到。有时，这种微妙但无处不在的政治化几乎是看不见的。但不应该如此。社交媒体只是最近的一个提醒，表明技术和政治组织无法分离。国家和技术紧密相连。这对即将到来的事物有重要影响。

虽然技术不会简单地将人们推向预定的方向，但承认其倾向于提供某些能力或看到它如何促使某些结果而非其他结果，并不是天真的技术决定论。在这方面，技术是历史的关键决定因素之一，但从不单独存在，也从不是以机械的、固有可预测的方式存在。它不会表面上导致特定行为或结果，但它所产生的东西确实会引导或限制可能性。

战争、和平、商业、政治秩序、文化——这些始终是相互关联的，并且与技术相互关联。技术是理念，体现在产品和服务中，对人类、社会结构、环境以及介于两者之间的一切都有深远而持久的影响。

现代技术和国家是共生演化的，始终处于对话之中。想想技术如何促进国家核心工作部分的发展，帮助构建国家身份和行政管理的大厦。文字被发明为一种行政和会计工具，用于记录债务、继承、法律、税收、合同和所有权记录。时钟设定了固定的时间，最初在修道院等有限的空间内使用，但随后在中世纪晚期的商业城市中以机械形式出现，并最终在全国范围内普及，创造了共同的、越来越大的社会单位。印刷机帮助将混乱的方言标准化为国家语言，从而帮助产生了一个国家的“想象共同体”，即民族国家背后的统一人民。取代了更为流动的口头传统，印刷文字将地理、知识和历史固定下来，传播设定的内容。

法律法规和意识形态。广播和电视加速了这一进程，创造了全国乃至国际共同体验的时刻，如罗斯福的炉边谈话或世界杯。

武器也是国家权力的核心技术。实际上，国家理论家常常认为战争本身就是国家形成的基础（用政治学家查尔斯·蒂利的话来说，“ 战争造就了国家，国家造就了战争 ”），正如冲突一直是新技术的催化剂——从战车和金属盔甲到雷达和引导精确弹药的先进芯片。火药于十三世纪传入欧洲，打破了中世纪防御城堡的旧模式。设防的定居点现在成了轰炸的活靶子。在英法百年战争期间，进攻能力使那些能够负担得起购买、建造、维护、移动和部署资本密集型大炮的人占据了优势。多年来，国家将越来越多的致命力量集中在自己手中，声称对合法使用武力拥有垄断权。

简单来说，技术与政治秩序密切相关。新技术的引入会带来重大的政治后果。正如火炮和印刷机颠覆了社会，我们也应该预期人工智能、机器人技术和合成生物学等技术会产生同样的影响。

暂停片刻，想象一个这样的世界：拥有与人类一样灵巧的机器人，可以用简单的英语“ 编程 ”，而价格仅相当于一个微波炉。你能开始想象这种宝贵技术将被用于哪些用途吗？或者这种工具将被多广泛地采用？在养老院里，谁或更确切地说是什么在照顾你的老母亲？你如何在餐厅点餐，又是谁将食物送到你的桌上？在人质事件中，执法会是什么样子？在收获季节，果园的工作人员会是谁？当不需要派遣人类进入战斗时，军事和准军事策划者将如何反应？当孩子们在训练足球时，运动场会是什么样子？你的窗户清洁工会是什么样子？谁拥有所有这些硬件和知识产权，谁控制它，当它出问题时——如果——有什么保障措施？

想象一下这一切，这意味着与今天截然不同的政治经济。

源文本：—翻译文本：

工业化民族国家自二十世纪初以来一直是全球主导力量，是上个世纪伟大政治冲突的明确“ 胜利者 ”。它带来了现在被视为理所当然的定义功能。安全的提供。中心集中了巨大的合法权力，能够在其管辖范围内完全主导，但也有合理的制衡和各种权力之间的分离。通过再分配和健全的经济管理提供足够的福利。技术创新和监管的稳定框架，以及全球化的整个社会经济法律架构。

在接下来的几章中，我们将看到即将到来的浪潮如何使这一切面临巨大威胁。

我认为，出现的情况将倾向于两个方向，中间有一系列结果。在一个轨迹上，一些自由民主国家将继续从内部被侵蚀，变成一种僵尸政府。自由民主和传统民族国家的外壳依然存在，但实际上它们被掏空，核心服务日益破败，政体不稳定且分裂。在没有其他选择的情况下，它们变得越来越退化和功能失调。在另一个轨迹上，不加思考地采用即将到来的浪潮的一些方面，打开了通往专制国家控制的道路，创造出超级利维坦，其权力甚至超越了历史上最极端的极权政府。威权政权也可能趋向于僵尸状态，但同样可能加倍努力，得到提升，成为完全成熟的技术独裁政权。在任何一条路径上，维系国家的微妙平衡被推向混乱。

无论是失败国家还是威权政权，都是灾难性的结果，不仅仅是就其本身而言，而且对治理技术也是如此；两者都不

摇摆不定的官僚机构、民粹主义机会主义者，或全能的独裁者都不是你希望对控制强大新技术负有根本责任的人。无论哪种方向都无法或不会遏制即将到来的浪潮。

因此，任何一方都存在危险，因为应对即将到来的浪潮需要自信、敏捷、连贯的国家，对人民负责，充满专业知识，平衡利益和激励措施，能够快速果断地采取立法行动，最重要的是，与国际紧密协调。领导者需要采取前所未有的大胆行动，以短期利益换取长期利益。有效应对历史上最深远和最具变革性的事件之一，需要成熟、稳定、最重要的是值得信赖的政府发挥最佳表现。真正运作良好的国家。这就是确保即将到来的浪潮带来其承诺的巨大好处所需的条件。这是一个极其艰巨的任务。

廉价、无处不在的机器人，如上所述，与我们在第二部分看到的许多其他变革性技术一起，在二十年内是完全不可避免的，甚至可能更早。在这种情况下，我们应该预期经济、民族国家及其相关的一切将发生深刻变化。大交易已经陷入困境。随着洪流的开始，一系列新的压力因素将动摇其基础。

CHAPTER 10

脆弱性放大器

国家紧急状态2.0：无法遏制的不对称行动

O N THE MORNING OF MAY 12, 2017, Britain's National Health Service

(NHS) ground to a halt. Thousands of its facilities nationwide suddenly saw

NHS遭受了勒索软件攻击。该攻击被称为WannaCry，其规模巨大。勒索软件通过入侵系统来加密，从而锁定对关键文件和功能的访问。网络攻击者通常要求支付赎金以解放被劫持的系统。

their IT systems freeze up. In hospitals, staff were locked out of crucial

medical equipment like MRI scanners and unable to access patient records.

NHS并不是WannaCry的唯一目标。利用旧版微软系统中的漏洞，黑客找到了让数字世界大范围停滞的方法，包括像德铁（Deutsche Bahn）、西班牙电信（Telefónica）、联邦快递（FedEx）、日立（Hitachi），甚至中国公安部这样的组织。WannaCry诱骗一些用户打开电子邮件，从而释放出一种“蠕虫”，该蠕虫自我复制并传播，在短短一天内感染了150个国家的25万台计算机。在几个小时之后，

Thousands of scheduled procedures, ranging from cancer appointments to elective surgeries, had to be cancelled. Panicked care teams reverted to

manual stopgaps, using paper notes and personal phones. The Royal

攻击使得大部分数字世界摇摇欲坠，被一个遥远而无形的攻击者勒索。随之而来的损失高达80亿美元，但其影响更为严重。WannaCry攻击暴露了我们认为理所当然的机构在面对复杂的网络攻击时是多么脆弱。

最终，NHS 和全世界都幸运地躲过了一劫。一位名叫 Marcus Hutchins 的 22 岁英国黑客偶然发现了一个“ 杀死开关 ”。在查看恶意软件的代码时，他看到一个看起来很奇怪的域名。猜测这可能是蠕虫的指挥和控制结构的一部分，并且看到该域名未注册，Hutchins 以仅 $10.69 的价格购买了它，使他能够控制病毒，同时微软推出更新以关闭漏洞。

也许关于WannaCry最不寻常的事情是它的来源。WannaCry是使用美国国家安全局（NSA）创建的技术构建的。一个名为定制访问行动办公室的精英NSA部门开发了一种名为EternalBlue的网络攻击漏洞。用一位NSA工作人员的话来说，这些是“ 通往王国的钥匙 ”， 旨在“ 破坏国内外许多主要政府和企业网络的安全 ”的工具。

这种由地球上最技术先进的组织之一开发的强大技术，是如何被一群黑客获取的？正如微软当时指出的那样，“ 一个等同的情景是美国军方的一些战斧导弹被盗。 ”与战斧导弹不同，NSA的数字武器可以悄悄地滑入一个拇指驱动器。窃取该技术的黑客是一群被称为“ 影子经纪人 ”的组织，他们将EternalBlue出售。从那里，它很快就落入了朝鲜黑客手中，可能是国家支持的121局网络单位。然后他们将其投放到世界上。

尽管进行了快速修补，但EternalBlue泄漏的影响并未结束。2017年6月，一种新版本的武器出现，这次专门设计用于攻击乌克兰国家基础设施，并迅速被归因于俄罗斯军事情报。NotPetya

网络攻击几乎使该国陷入瘫痪。切尔诺贝利的辐射监测系统失去了电力。自动取款机停止出钞。手机变得无声。该国10%的计算机被感染，从电网到乌克兰国家储蓄银行的基本基础设施瘫痪。像航运巨头马士基这样的跨国公司也受到波及，成为附带损害。

这是一个关于二十一世纪技术的寓言。由世界上技术最先进国家的安全部门创建的软件被泄露或被盗。从那里，它落入了为世界上最失败的国家之一和反复无常的核大国工作的数字恐怖分子手中。然后，它被武器化，转而攻击当代国家的核心结构：卫生服务、交通和电力基础设施、全球通信和物流中的重要企业。换句话说，由于基本的遏制失败，一个全球超级大国成为了其自身强大且据称安全的技术的受害者。

这是不受限制的不对称性在起作用。

源文本：—翻译文本：

上述描述的攻击依赖于传统的网络武器。幸运的是，它们并没有依赖于即将到来的浪潮的特征。它们的力量和潜力是有限的。国家受到了擦伤和挫伤，但并没有从根本上被削弱。然而，下一次攻击何时发生只是时间问题，而不是是否会发生的问题，下次我们可能就没有这么幸运了。

人们很容易认为，考虑到关键系统从像WannaCry这样的攻击中恢复的速度，网络攻击远没有我们想象的那么有效。然而，随着新一波攻击的到来，这种假设是一个严重的错误。这类攻击表明，有人会利用尖端技术来削弱和瘫痪关键的国家功能。它们显示出现代生活的核心机构是脆弱的。一个孤立的个人和一家私人公司（微软）修补了系统的弱点。

攻击不尊重国界。政府在处理危机中的作用有限。

现在，想象一下，如果不是因为意外留下了一个漏洞，而是WannaCry背后的黑客设计了一个程序，能够系统地了解自身的漏洞并反复修补它们。想象一下，当它攻击时，程序进化以利用更多的弱点。想象一下，它然后开始穿过每一家医院、每一个办公室、每一个家庭，不断变异、学习。它可以攻击生命支持系统、军事基础设施、交通信号、能源电网、金融数据库。随着它的传播，想象一下程序学会检测并阻止进一步关闭它的尝试。如果不是已经在开发中，这样的武器就在地平线上。

与下一代网络武器中将出现的越来越多用途广泛的学习代理相比，WannaCry 和 NotPetya 的影响是有限的，这些武器威胁着引发国家紧急状态2.0。今天的网络攻击并不是实际的威胁；它们是新时代脆弱性和不稳定性的煤矿中的金丝雀，削弱了民族国家作为唯一安全仲裁者的角色。

这里是下一波技术的一个具体、近期应用，正在削弱国家的结构。在本章中，我们将探讨这一点以及其他压力因素如何侵蚀负责管理技术的建筑。这些脆弱性放大器、系统冲击、紧急情况2.0，将极大地加剧现有挑战，动摇国家的基础，扰乱我们已经岌岌可危的社会平衡。这在某种程度上是一个关于谁能做什么的故事，一个关于权力及其所在的故事。

电力成本的骤降

源文本: 翻译文本:

权力是“ 以某种特定方式做某事或行动的能力或容量；……引导或影响他人行为或事件进程的能力。 ”它是支撑文明的机械或电能。

基石和国家的核心原则。权力以某种形式或另一种形式塑造一切。而它也即将被改变。

技术最终是政治的，因为技术是一种权力形式。也许即将到来的浪潮的唯一压倒性特征是它将使权力的获取民主化。正如我们在第2部分中看到的，它将使人们能够在现实世界中*do*事物。我是这样想的：正如在消费互联网时代，处理和传播信息的成本大幅下降，实际上做某事、采取行动、投射权力的成本将在下一波浪潮中大幅下降。知道是很好的，但行动更具影响力。

相较于仅仅消费内容，任何人都可以*produce*专家级的视频、图像和文本内容。AI不仅帮助你找到那篇伴郎演讲的信息；它还会*write*演讲。而且这一切都将在前所未见的规模上进行。机器人不仅会制造汽车和组织仓库地板；它们还将对每一个有一点时间和想象力的车库修补者开放。过去的浪潮使我们能够测序或读取DNA。即将到来的浪潮将使DNA合成普遍可用。

无论权力今天在哪里，它都会被放大。任何有目标的人——也就是每个人——在实现目标时都会得到巨大的帮助。彻底改革商业战略、为当地社区举办社交活动或占领敌方领土都变得更容易。建立航空公司或停飞机队都更容易实现。无论是商业、宗教、文化还是军事，民主或专制，您能想到的每一种可能的动机都可以通过手头上更便宜的权力得到显著增强。

今天，无论你多么富有，你都无法购买到比数十亿人所能获得的更强大的智能手机。这个文明的非凡成就常常被忽视。在下一个十年，获取ACI的机会将遵循相同的趋势。那些相同的数十亿人将很快能够广泛平等地获得最好的律师、医生、战略家、设计师、教练、行政助理、谈判专家等等。每个人都会有一个世界级的团队在他们身边和支持他们。

这将是人类历史上最大、最快的财富和繁荣的加速器。它也将是最混乱的之一。如果每个人都能获得更多的能力，这显然也包括那些想要造成伤害的人。随着技术的发展速度超过防御措施，从墨西哥贩毒集团到朝鲜黑客的不法分子得到了助力。民主化的访问必然意味着民主化的风险。

我们即将跨越我们物种历史上的一个关键门槛。这是民族国家在未来十年内必须应对的问题。在本章中，我们将介绍一些源于即将到来的浪潮的脆弱性放大的关键例子。首先，让我们更仔细地看看这一近期风险：不良行为者将如何能够发起新的进攻行动。此类攻击可能是致命的、广泛可及的，并且是某人可以大规模打击而不受惩罚的机会。

持枪机器人：进攻的首要性

源文本: 翻译文本:

2020年11月，莫森·法赫里扎德是伊朗长期致力于获得核武器的首席科学家和关键人物。他爱国、敬业、经验丰富，是伊朗对手的主要目标。意识到风险，他在伊朗安全部门的帮助下，将自己的行踪和活动保密。

法赫里扎德的车队在尘土飞扬的道路上行驶，前往他位于里海附近的乡间别墅，突然嘎然停下。科学家的车辆遭到一阵子弹的袭击。受伤的法赫里扎德踉跄地走出车外，却被第二波机枪扫射击中身亡。他的保镖——伊朗革命卫队的成员，慌忙试图弄清发生了什么。射手在哪里？几秒钟后，一声爆炸响起，附近的一辆皮卡车燃起了熊熊大火。

卡车上除了枪之外是空的。那天地面上没有刺客。用*New York Times*调查的话来说，这

是一次“ 高科技计算机化神枪手的首次测试，配备了人工智能和多摄像头，通过卫星操作，能够每分钟发射600发子弹。 ”它安装在一辆战略性停放但看似无害的皮卡车上，配备摄像头，是由以色列特工组装的一种机器人武器。一次袭击由人类授权，但由人工智能自动调整枪的瞄准。仅发射了十五发子弹，伊朗最知名且戒备森严的人物之一在不到一分钟内被杀。爆炸仅仅是掩盖证据的失败尝试。

法赫里扎德的暗杀是即将发生的事情的预兆。更复杂的武装机器人将进一步降低暴力的门槛。在互联网上很容易找到最新一代机器人的视频，它们有像Atlas和BigDog这样的名字。在这里，你会看到矮壮、外形奇特的人形机器人和小型狗状机器人在障碍赛道上疾驰。它们看起来奇怪地不平衡，但似乎从未跌倒。它们以不可思议的动作穿越复杂的地形，沉重的框架从未倾覆。它们做后空翻、跳跃、旋转和特技。推倒它们，它们冷静而不可阻挡地站起来。它们准备好一次又一次地这样做。这真是令人毛骨悚然。

现在想象一下配备面部识别、DNA测序和自动武器的机器人。未来的机器人可能不会采取奔跑狗的形式。进一步微型化后，它们将会是鸟或蜜蜂的大小，配备小型火器或炭疽瓶。它们可能很快就会对任何想要它们的人开放。这就是不良行为者赋权的样子。

源文本：—翻译文本：

在过去十年中，军用无人机的成本下降了三个数量级。到2028年，每年将花费260亿美元用于军用无人机，届时许多无人机可能会完全自主。

自主无人机的实时部署日益变得更加可行。例如，2021年5月，加沙使用了一群人工智能无人机。

寻找、识别和攻击哈马斯武装分子。像 Anduril、Shield AI 和 Rebellion Defense 这样的初创公司已经筹集了数亿美元，用于构建自主无人机网络和其他军事应用的人工智能。3D 打印和先进的移动通信等互补技术将把战术无人机的成本降低到几千美元，使其从业余爱好者到准军事组织再到孤独的精神病患者都能负担得起。

除了更容易获取之外，AI增强的武器将实时自我改进。WannaCry的影响最终比可能的情况要小得多。一旦应用了软件补丁，立即的问题就得到了解决。AI改变了这种攻击方式。AI网络武器将不断探测网络，自主适应以发现和利用弱点。现有的计算机蠕虫使用一组固定的预编程启发式方法进行自我复制。

但是，如果你有一个蠕虫，它使用强化学习来改进自身，通过每次网络交互实验性地更新其代码，每次都找到越来越有效的方法来利用网络漏洞呢？正如AlphaGo这样的系统从数百万次自我对弈中学习到意想不到的策略，AI驱动的网络攻击也会如此。无论你如何进行战争游戏来应对每种可能性，始终会有一个微小的漏洞被一个持久的AI发现。

从汽车和飞机到冰箱和数据中心，一切都依赖于庞大的代码库。即将到来的人工智能使得识别和利用弱点比以往任何时候都更容易。它们甚至可以找到法律或财务手段来破坏公司或其他机构，发现银行监管或技术安全协议中的隐藏故障点。正如网络安全专家布鲁斯·施奈尔指出的那样，人工智能可以消化世界的法律和法规以寻找漏洞，进行合法套利。想象一下，一家公司的一大批文件被泄露。一个法律人工智能可能能够根据多个法律系统解析这些文件，找出每一个可能的违规行为，然后在全球范围内同时对该公司发起多起毁灭性的诉讼。人工智能可以开发自动化交易策略，旨在摧毁竞争对手的地位或创建虚假信息活动（更多内容在

在下一节中) 策划银行挤兑或产品抵制，使竞争对手能够迅速介入并收购公司——或者只是看着它倒闭。

AI不仅擅长利用金融、法律或通信系统，还擅长利用人类心理、我们的弱点和偏见，这种AI正在路上。Meta的研究人员创建了一个名为CICERO的程序。它成为了玩复杂棋盘游戏《外交》的专家，这是一种需要围绕欺骗和背叛制定长期复杂策略的游戏。这表明AI可以帮助我们进行计划和协作，但也暗示了它们如何能够发展心理技巧以获得信任和影响力，以令人恐惧的深度读取和操控我们的情感和行为，这种技能在赢得《外交》或竞选和建立政治运动中非常有用。

即使使人工智能如此强大和令人兴奋的相同前提——其学习和适应的能力——赋予了不法分子权力，针对关键国家功能的潜在攻击空间也在扩大。

源文本：—翻译文本：

像大规模火炮、海军齐射、坦克、航空母舰或洲际弹道导弹等能力，最初成本如此高昂，以至于它们一直是国家的专属领域。现在它们正在迅速演变，以至于很快就扩散到研究实验室、初创公司和车库修补者的手中。正如社交媒体的一对多广播效应意味着一个人可以突然进行全球广播一样，具有深远影响的行动能力正在变得对每个人都可用。

这种新的动态——坏人被鼓励采取进攻——由于现代系统的互联和脆弱性，打开了新的攻击向量：不仅仅是单个医院，而是整个健康系统都可能受到攻击；不仅仅是一个仓库，而是整个供应链。随着致命自主武器的出现，发动战争和攻击的成本，无论是物质上的还是尤其是人类方面的，都比以往任何时候都低。

时间，所有这些都引入了更高程度的否认性和模糊性，削弱了威慑的逻辑。如果没有人能确定是谁发起了攻击，或者究竟发生了什么，为什么不继续呢？

当非国家和恶意行为者以这种方式被赋权时，国家的核心命题之一被削弱了：为公民提供安全保护伞的表象受到严重损害。安全和保障的提供是民族国家体系的基本支柱，而不是可有可无的附加项。国家大体上知道如何应对法律和秩序的问题，或来自敌对国家的直接攻击。但这更加模糊，形态不明且不对称，模糊了领土性和容易归因的界限。

一个国家如果无法提供基本的安全承诺，如何维持公民的信任，维护那种宏大的契约？它如何确保医院继续运转，学校保持开放，灯光在这个世界上——字面意义上——保持亮着？如果国家不能保护你和你的家人，遵守和归属又有什么意义？如果我们感到基本的东西——为我们房屋供电的电力、让我们出行的交通系统、让我们保持温暖的能源网络、我们个人的日常安全——正在崩溃，而我们或政府都无能为力，那么系统的基础就被削弱了。如果国家在新的战争形式中开始，也许它会以同样的方式结束。

纵观历史，技术在进攻和防御优势之间产生了一种微妙的舞蹈，钟摆在两者之间摇摆，但大致保持平衡：对于每一种新的投射物或网络武器，强有力的对策很快就会出现。大炮可能会削弱城堡的墙壁，但它们也可以撕裂入侵的军队。现在，强大、不对称、全用途的技术肯定会落入那些想要破坏国家的人手中。虽然防御行动会随着时间的推移而加强，但这四个特征的性质偏向于进攻：这种力量的扩散实在是太广泛、太迅速、太开放了。一个具有改变世界意义的算法可以存储在笔记本电脑上；很快，它甚至不需要上一次浪潮和互联网那种庞大、可监管的基础设施。与箭矢甚至高超音速导弹不同，AI和生物制剂

将比我们见过的任何技术更便宜、更快速、更自主地发展。因此，如果不采取一系列戏剧性的干预措施来改变当前的进程，数百万人将在短短几年内获得这些能力。

在如此广泛的通用技术领域中保持决定性、无限期的战略优势根本不可能。最终，平衡可能会恢复，但在此之前，一股极具破坏力的力量将被释放。而且正如我们所见，威胁的性质远比简单的物理攻击更为广泛。信息和通信本身就是其不断升级的风险向量，是另一个需要关注的新兴脆弱性放大器。

欢迎来到深度伪造时代。

错误信息机器

源文本: 翻译文本:

在2020年印度地方选举中，印度人民党德里主席Manoj Tiwari被拍到发表竞选演讲——使用英语和当地的印地语方言。两者看起来和听起来都非常真实。在视频中，他发起攻击，指责一个竞争对手党的领导人“ 欺骗了我们。 ”但使用当地方言的版本是一个深度伪造，是一种新型的AI支持的合成媒体。由一家政治传播公司制作，它使候选人接触到新的、难以触及的选民群体。由于缺乏对假媒体话语的认识，许多人认为这是真的。制作深度伪造的公司辩称这是技术的“ 积极 ”使用，但对任何清醒的观察者来说，这一事件预示着政治传播的一个危险新时代。在另一个广泛宣传的事件中，一段Nancy Pelosi的视频被重新编辑，使她看起来生病和受损，然后在社交媒体上广泛传播。

问问你自己，当任何人都有能力创造和传播具有惊人真实感的材料时，会发生什么？这些例子发生在 *before*，生成近乎完美的深度伪造——无论是文本、图像、视频还是音频——变得像写查询一样简单。

谷歌。正如我们在第4章中所见，大型语言模型现在在生成合成媒体方面显示出惊人的结果。一个与传统媒体无法区分的深度伪造世界已经到来。这些伪造将如此逼真，以至于我们的理性思维很难接受它们不是真实的。

Deepfakes 正在迅速传播。如果你想看一个汤姆·克鲁斯准备与鳄鱼摔跤的逼真假视频，你可以做到。随着所需训练数据减少到仅仅几个例子，越来越多的普通人将被模仿。这已经在发生。2021年，香港的一家银行在其客户被 deepfake 冒充后，将数百万美元转给了诈骗犯。诈骗犯的声音与真实客户完全相同，他们打电话给银行经理，解释公司需要转移资金用于收购。所有文件似乎都没问题，声音和角色都完美无瑕地熟悉，所以经理启动了转账。

任何有意制造不稳定的人现在都更容易做到这一点。假设在选举前三天，总统被拍到使用种族歧视的言辞。竞选新闻办公室极力否认，但每个人都知道他们看到了什么。愤怒在全国蔓延。民调急剧下降。摇摆州突然转向对手，出乎意料地获胜。一个新的政府上台。但该视频是一个深度伪造，技术如此高超，以至于即使是最好的伪造检测神经网络也无法识别。

这里的威胁不在于极端情况，而在于微妙、复杂且高度可信的情境被夸大和扭曲。不是总统冲进学校大喊无稽之谈并投掷手榴弹；而是总统无奈地表示他别无选择，只能制定一套紧急法律或重新引入征兵制。这不是好莱坞的烟火，而是据称的监控摄像头拍到一群白人警察殴打一名黑人男子致死的画面。

激进传教士安瓦尔·奥拉基的布道激励了波士顿马拉松爆炸案的制造者、巴黎的*Charlie Hebdo*袭击者以及在奥兰多夜总会杀死四十九人的枪手。然而，奥拉基于2011年去世，是第一个被美国无人机袭击杀死的美国公民，这些事件发生之前。

事件。然而，他激进的信息直到2017年仍然可以在YouTube上找到。假设使用深度伪造技术可以“挖掘”出新的奥拉基视频，每个视频都用精准的修辞指挥进一步的定向攻击。并不是每个人都会相信，但那些想要相信的人会觉得这极具说服力。

很快，这些视频将变得完全且令人信服地互动。你正在直接与他对话。他了解你，并适应你的方言和风格，利用你的历史、你的个人怨恨、你在学校受到的欺凌、你可怕的、不道德的西化父母。这不是地毯式轰炸的虚假信息；这是外科手术式打击的虚假信息。

针对政治家或商人的网络钓鱼攻击、旨在造成重大金融市场混乱或操纵的虚假信息、旨在毒化宗教或种族分歧等关键断层的媒体，甚至是低级别的骗局——信任受到损害，脆弱性再次被放大。

最终，完整而丰富的看似真实世界事件的合成历史将很容易生成。个体公民将没有时间或工具来验证他们接收到的内容的一小部分。伪造品将轻松通过复杂的检查，更不用说两秒钟的嗅觉测试了。

国家支持的信息攻击

源文本: 翻译文本:

在20世纪80年代，苏联资助了散布虚假信息的活动，暗示艾滋病病毒是美国生物武器计划的结果。多年后，一些社区仍在处理不信任和后果。与此同时，这些活动并没有停止。根据Facebook的数据，俄罗斯特工在2016年大选期间在其平台上创建了不少于八万条有机内容，触及了1.26亿美国人。

AI增强的数字工具将加剧此类信息操作，干预选举、利用社会分裂，并创建复杂的虚假宣传活动以制造混乱。不幸的是，这远不止俄罗斯。已有超过七十个国家被发现正在进行虚假信息活动。中国正在迅速赶上俄罗斯；其他国家

从土耳其到伊朗，他们正在发展自己的技能。（中央情报局对信息行动也并不陌生。）

在COVID-19大流行初期，大量虚假信息导致了致命的后果。卡内基梅隆大学的一项研究分析了在第一次封锁高峰期讨论COVID-19的超过2亿条推文。82%的倡导“ 重启美国 ”的有影响力用户是机器人。这是一个有针对性的“ 宣传机器 ”，很可能是俄罗斯的，旨在加剧一个世纪以来最严重的公共卫生危机。

Deepfakes 自动化了这些信息攻击。到目前为止，有效的虚假信息活动一直是劳动密集型的。虽然机器人和假货并不难制作，但大多数质量较低，容易识别，并且在实际改变目标行为方面效果有限。

高质量的合成媒体改变了这一等式。并不是所有国家目前都有资金建立庞大的虚假信息项目，设立专门的办公室和训练有素的员工队伍，但当高保真材料可以一键生成时，这种障碍就小得多。即将到来的混乱大多不会是偶然的。它将随着现有虚假信息活动的加速、扩展和下放到一大群有动机的参与者而出现。

大规模且成本极低的合成媒体的兴起同时放大了虚假信息（恶意和故意误导的信息）和错误信息（信息空间中更广泛且更无意的污染）。引发了“ 信息末日 ”，即社会无法再管理大量可疑材料的时刻，信息生态系统是知识、信任和社会凝聚力的基础，是维系社会的粘合剂，而这一系统正在崩溃。正如布鲁金斯学会的一份报告所言，无处不在的完美合成媒体意味着“ 扭曲民主话语；操纵选举；削弱对机构的信任；削弱新闻业；加剧社会分裂；破坏公共安全；并对知名人士的声誉造成难以修复的损害，包括民选官员和候选人。 ”

并非所有的压力源和伤害都来自恶意行为者。然而，有些却出自于最好的意图。脆弱性的放大也是偶然的。

如故意。

泄漏实验室和意外的不稳定性

源文本: 翻译文本:

在世界上最安全的实验室之一，一组研究人员正在实验一种致命的病原体。没有人能确定接下来发生了什么。即使事后回顾，关于研究的细节也很少。可以确定的是，在一个以保密和政府控制而闻名的国家，一种奇怪的新疾病开始出现。

不久之后，它在全球范围内被发现，包括英国、美国及其他地区。奇怪的是，这似乎并不是一种完全自然的疾病毒株。某些特征在科学界引起了警觉，并暗示实验室中某些事情出了严重问题，这不是一个自然事件。很快，死亡人数开始上升。那个高度安全的实验室看起来并不那么安全。

如果这听起来像是一个熟悉的故事，那可能不是你正在思考的那个。这是1977年，一场被称为俄罗斯流感的流感疫情。它首先在中国被发现，不久后在苏联被检测到，从那里传播开来，据报道导致多达70万人死亡。H1N1流感毒株的不寻常之处在于它与20世纪50年代流行的毒株非常相似。该疾病对年轻人的打击最为严重，这可能表明他们的免疫力比几十年前的人要弱。

关于发生了什么的理论层出不穷。是否有什么东西从永久冻土中逃逸出来？这是否是俄罗斯庞大而神秘的生物武器计划的一部分？然而，迄今为止，最好的解释是实验室泄漏。早期病毒的一个版本很可能在实验室进行疫苗实验时以某种方式泄漏。流行病本身是由旨在预防流行病的善意研究引起的。

生物实验室须遵循全球标准以防止事故发生。最安全的实验室被称为生物安全等级4（BSL-4）实验室。

它们代表了处理最危险病原材料的最高封闭标准。设施完全密封。进入通过气闸。所有进出物品都经过彻底检查。每个人都穿加压服。任何人离开都需要淋浴。所有材料都按照最严格的协议处理。任何可能刺破手套或防护服的尖锐物品都被禁止。BSL-4实验室的研究人员经过适当的培训，以创造人类有史以来最生物安全的环境。

然而，事故和泄漏仍然会发生。1977年的俄罗斯流感就是一个例子。仅仅两年后，炭疽孢子意外地从一个秘密的苏联生物武器设施中释放出来，造成了一条长达五十公里的疾病传播路径，至少导致六十六人死亡。

2007年，英国Pirbright研究所的一根管道泄漏，该研究所包括BSL-4实验室，导致了口蹄疫的爆发，损失达1.47亿英镑。2021年，费城附近一家制药公司的研究人员将天花小瓶遗留在一个未标记、未加锁的冰箱中。幸运的是，它们被清理冰箱的人发现了。这个人幸运地戴着口罩和手套。如果它泄露出去，后果将是灾难性的。在被消灭之前，天花仅在二十世纪就估计杀死了3亿到5亿人，其传播率相当于更具传染性的COVID-19毒株，但死亡率是COVID的三十倍。

SARS 应该在 BSL-3 条件下保存，但它曾从新加坡、台湾和中国的病毒学实验室逃逸。令人难以置信的是，它曾四次从北京的同一个实验室逃逸。这些错误都是非常人性化和普通的。新加坡的案例是由于一名研究生不知道 SARS 的存在。在台湾，一名研究科学家错误处理了生物危害废物。在北京，泄漏被归因于病毒灭活不彻底和在非生物安全实验室的处理。而这一切还没提到武汉，那里是世界上最大的 BSL-4 实验室和冠状病毒研究中心。

即使 BSL-4 实验室的数量激增，根据全球卫生安全指数，只有四分之一的实验室在安全性方面得分很高。

1975年和2016年，研究人员记录了至少七十一起故意或意外暴露于高度传染性和有毒病原体的事件。大多数是微小的事故，即使是最受过高度训练的人有时也难免会犯错——针头滑动、试管洒落、实验准备时出现小错误。我们的记录几乎可以肯定是不完整的。很少有研究人员公开或及时报告事故。一项对生物安全官员的调查发现，大多数人从未在其机构之外报告事故。2014年美国的一项风险评估估计，在十年内，十个实验室中“ 重大实验室泄漏 ”的几率为91%；导致大流行的风险为27%。

没有任何 *should* 泄露。然而，病原体却一次又一次地泄露。尽管有一些最严格的围控协议、技术和法规，但仍然失败。一个摇晃的移液器。一块被刺穿的塑料薄膜。一滴溶液洒在鞋子上。这些都是围控的具体失败。意外的。偶然的。以一种严峻而不可避免的规律性发生。然而，在合成生命的时代，这引入了可能发生的事故，这些事故可能既是巨大的压力源，也是我们将在第三部分中回到的东西——灾难。

源文本：—翻译文本：

几乎没有什么研究像功能获得（GOF）研究那样具有争议性。简单来说，功能获得实验是故意将病原体改造成更致命或更具传染性，或两者兼具。在自然界中，病毒通常在致命性和传播性之间进行权衡。病毒传播性越强，通常致命性就越低。但这并不是绝对的。理解这种情况可能发生的一种方式——即病毒如何可能同时变得更致命和更具传播性——以及我们如何应对这种情况的方法就是，嗯，让它发生。

这就是功能增益研究的作用所在。研究人员调查疾病的潜伏期，或者它们如何逃避疫苗的抵抗，或者它们如何可能无症状地在人群中传播。

像这样的工作已经在包括埃博拉、H1N1流感和麻疹等疾病上开展。

此类研究工作通常是可信且出于良好意图的。大约十年前在荷兰和美国进行的禽流感研究就是一个很好的例子。这种疾病的死亡率惊人地高，但幸运的是很难传播。研究人员想要了解这种情况如何可能发生变化，这种疾病如何可能演变成更易传播的形式，并使用雪貂来观察这种情况可能如何发生。换句话说，他们在理论上使一种致命的疾病更容易传播。

然而，不需要丰富的想象力就能预见这样的研究可能会出错。有些人，包括我自己，觉得故意设计或进化这样的病毒有点像是在玩核触发器。

功能增益研究，可以说是有争议的。曾经，美国的资助机构对其资助实施了暂停。在一次典型的控制失败中，这类工作于2019年恢复。至少有一些迹象表明COVID-19经过了基因改造，并且有越来越多的（间接）证据，从武汉研究所的记录到病毒本身的分子生物学，暗示实验室泄漏可能是大流行的起源。

美国联邦调查局和美国能源部都认为情况如此，而中央情报局尚未决定。与之前的疫情爆发不同，这次没有确凿证据表明是人畜共患病传播。生物研究已经导致数百万人死亡，使全球社会陷入停滞，并造成数万亿美元的损失，这种情况是完全可能的。2022年底，波士顿大学的一项国立卫生研究院研究将原始的、更致命的COVID毒株与更具传播性的奥密克戎变体的刺突蛋白结合起来。许多人认为这项研究不应该继续进行，但事实是，它是由公共资金资助的。

这不是关于不良行为者利用技术；这是关于想要改善健康结果的好人所带来的意外后果。这是关于当强大的工具普及时会出错的事情，犯下的错误，“报复效应”如何展开，以及随机的、不可预见的情况。

技术与现实的碰撞导致了混乱。离开图板，远离理论，未受控技术的核心问题即使在最好的意图下也依然存在。

GOF研究旨在保障人们的安全。然而，它不可避免地发生在一个有缺陷的世界中，在这个世界里，实验室会泄漏，疫情会发生。无论武汉发生了什么，冠状病毒的此类研究正在进行并泄漏的可能性仍然令人沮丧。实验室泄漏的历史记录很难忽视。

源文本：—翻译文本：

- 泄漏只是即将到来的浪潮如何引入大量报复效应和意外失败模式的两个特别尖锐的例子。如果每个半熟练的实验室甚至是随机的生物黑客都可以开展这项研究，悲剧就无法无限期地推迟。这种情景正是我在第1章中提到的研讨会上向我概述的。

随着任何技术的力量和传播增长，其故障模式也在升级。如果一架飞机坠毁，那是一个可怕的悲剧。但如果整个机队坠毁，那就是完全更令人恐惧的事情。重申一下：这些风险不是关于恶意伤害；它们仅仅来自于在历史上最强大的技术的前沿操作，这些技术广泛嵌入在核心社会系统中。实验室泄漏只是意外后果的一个好例子，是遏制问题的核心，相当于即将到来的反应堆熔毁或丢失的弹头。这样的事故会造成另一个不可预测的压力源，系统中的另一个裂缝。

然而，压力源也可能不那么离散，不是机器人攻击、实验室泄漏或深度伪造视频，而是更缓慢和分散的过程，破坏基础。考虑到在整个历史中，工具和技术的设计是为了帮助我们以更少的资源做更多的事情。每一个单独的实例几乎无关紧要。但是，如果这些的最终副作用是什么呢？

复合效率是人类几乎不需要做太多工作吗？

自动化辩论

源文本: 翻译文本:

自从我共同创立DeepMind以来，没有哪个AI政策辩论比工作的未来获得更多的关注，甚至到了过度饱和的地步。

这里是原始论点。过去，新技术使人们失业，产生了经济学家约翰·梅纳德·凯恩斯所称的“技术性失业”。在凯恩斯看来，这是一件好事，因为生产力的提高释放了时间，用于进一步的创新和休闲。与技术相关的替代例子不胜枚举。动力织机的引入使得传统织工失业；汽车意味着马车制造商和马厩不再需要；随着蜡烛制造商破产，灯泡工厂蓬勃发展。

总的来说，当技术破坏了旧的工作和行业时，它也产生了新的工作。随着时间的推移，这些新工作趋向于服务行业角色和基于认知的白领工作。随着铁锈地带的工厂关闭，对律师、设计师和社交媒体影响者的需求激增。至少到目前为止，从经济角度来看，新技术并没有最终取代劳动力；总体上，它们是对劳动力的补充。

但是，如果新的取代工作的系统在人的认知能力阶梯上不断攀升，导致劳动无处可去呢？如果即将到来的浪潮真的如看起来那样普遍和广泛，人类将如何竞争？如果大多数白领任务都能由AI更高效地完成呢？在人类仍然“优于”机器的领域将寥寥无几。我长期以来一直认为这是更可能的情景。随着最新一代大型语言模型的到来，我现在比以往任何时候都更确信事情将会如此发展。

这些工具只会暂时增强人类智能。它们会在一段时间内让我们变得更聪明、更高效，并释放出巨大的经济增长潜力，但它们本质上是劳动。

替代。它们最终将比许多从事行政、数据录入、客户服务（包括拨打和接听电话）、撰写电子邮件、起草摘要、翻译文件、创建内容、文案写作等工作的人更高效、更廉价地完成认知劳动。面对大量超低成本的替代品，这种“ 认知体力劳动 ”的日子屈指可数。

我们才刚刚开始看到这股新潮流即将产生的影响。对ChatGPT的早期分析表明，它可以使“ 中等水平的大学教育专业人士 ”在许多任务上的生产力提高40%。这反过来可能会影响招聘决策：麦肯锡的一项研究估计，在未来七年内，超过一半的工作可能会看到许多任务被机器自动化，而到2030年，五千二百万美国人从事的角色具有“ 中等自动化暴露 ”。

经济学家达龙·阿西莫格鲁和帕斯夸尔·雷斯特雷波估计，机器人导致当地工人工资下降。每增加一千名工人就增加一个机器人，人口就业率就会下降，从而导致工资下降。如今，算法执行了绝大部分的股票交易，并且越来越多地在金融机构中发挥作用，然而，即使华尔街繁荣，随着技术侵占越来越多的任务，它也在减少工作岗位。

许多人仍然持怀疑态度。像大卫·奥特这样的经济学家认为，新技术不断提高收入，创造对新劳动力的需求。技术使公司更具生产力，产生更多的资金，然后这些资金流回经济中。简单来说，需求是无止境的，而这种需求，由技术创造的财富所激发，催生了需要人类劳动的新工作。毕竟，怀疑者说，十年的深度学习成功并没有引发工作自动化的崩溃。有人认为，陷入这种恐惧只是对旧的“ 劳动总量 ”谬论的重演，该谬论错误地声称只有一定量的工作可供分配。相反，未来看起来更像是数十亿人从事尚未构思的高端工作。

我认为在未来几十年内，这种乐观的愿景是不现实的；自动化无疑是另一个脆弱性放大器。正如我们在第4章中看到的，人工智能的进步速度远远超出指数级增长，而且似乎没有明显的上限。机器正在迅速模仿各种人类能力，从视觉到语音和语言。即使没有在“深刻理解”方面取得根本性进展，新的语言模型也可以阅读、综合和生成令人惊叹的准确且非常有用的文本。仅凭这一项技能就有数百种角色将其视为核心要求，而人工智能还有更多潜力可挖。

是的，几乎可以肯定会创造出许多新的工作类别。谁会想到“网红”会成为一个备受追捧的角色？或者想象在2023年，人们会从事“提示工程师”——大型语言模型的非技术程序员，他们擅长引导出特定的回应？对按摩师、大提琴手和棒球投手的需求不会消失。但我最好的猜测是，新工作不会以足够的数量或时间来真正提供帮助。能够获得机器学习博士学位的人数与裁员的规模相比仍然微不足道。当然，新的需求会创造新的工作，但这并不意味着所有工作都由人类完成。

劳动力市场在技能、地理和身份方面也存在巨大的摩擦。考虑到在上一次去工业化过程中，匹兹堡的钢铁工人或底特律的汽车制造商几乎不可能轻易搬家，在职业中期重新培训，并在纽约找到一份衍生品交易员的工作，或在西雅图成为品牌顾问，或在迈阿密成为教师。如果硅谷或伦敦金融城创造了大量新工作，但如果人们没有合适的技能或无法搬迁，这对国家另一端的人没有帮助。如果你的自我认同与某种特定的工作紧密相连，那么如果你觉得新工作贬低了你的尊严，这也无济于事。

在配送中心从事零工合同工作，并不像在20世纪60年代繁荣的底特律汽车制造商工作那样带来自豪感或社会团结感。私营部门工作质量指数是衡量有多少工作提供高于平均收入的指标，

自1990年以来急剧下降；这表明高薪工作在总数中的比例已经开始下降。

像印度和菲律宾这样的国家在业务流程外包方面经历了巨大的繁荣，在呼叫中心等地方创造了相对高薪的工作。正是这种类型的工作将成为自动化的目标。长期来看，可能会创造新的工作岗位，但对数百万人来说，这些工作岗位不会来得足够快或出现在合适的地方。

同时，就业衰退将导致税收锐减，损害公共服务，并在福利项目最需要的时候对其提出质疑。即使在就业岗位被大幅削减之前，各国政府也将捉襟见肘，努力履行所有承诺，实现可持续的自我融资，并提供公众期望的服务。此外，所有这些破坏将在全球范围内、多维度地发生，影响从以农业为主的经济体到先进的服务型行业的每一个发展阶梯。从拉各斯到洛杉矶，通往可持续就业的道路将面临巨大的、不可预测的、快速演变的动荡。

即使是那些不预见自动化最严重后果的人，仍然承认它正在导致显著的中期扰动。无论你在就业辩论中站在哪一边，很难否认其影响将对数亿人造成巨大不稳定，这些人至少需要重新技能并过渡到新类型的工作。乐观的情景仍然涉及令人不安的政治影响，从政府财政破裂到就业不足、不安全和愤怒的人群。

这预示着麻烦。在一个压力重重的世界中又添一重压力。

源文本：—翻译文本：

像社交媒体这样的脆弱性放大器。它们损害并削弱了民族国家。第一个迹象已经显现，但就像二十一世纪第一个十年末期的社交媒体一样，目前尚不清楚其确切的形态和程度。

影响将会是什么。无论如何，仅仅因为后果尚不明显，并不意味着它们可以被忽视。

本章中概述的压力源（绝非详尽无遗）——新的攻击和脆弱性形式、信息误导的工业化、致命的自主武器、实验室泄漏等事故，以及自动化的后果——对于科技、政策和安全领域的人们来说都很熟悉。然而，它们往往被孤立地看待。在分析中被忽视的是，所有这些对我们机构的新压力都源于同一个基础的通用革命。它们将如何同时到来，交叉、支撑并相互促进的压力源。脆弱性的全面放大被忽视了，因为它常常看起来这些影响是逐步发生的，并且在方便的孤立环境中。事实并非如此。它们源于一个单一的连贯且相互关联的现象，以不同的方式表现出来。现实比任何顺序呈现所能传达的要更加交织、缠绕、突现和混乱。脆弱性，被放大。民族国家，被削弱。

它以前经历过不稳定的时期。不同之处在于，一场通用革命并不限于特定的利基市场、给定的问题、明确划分的行业。它本质上无处不在。成本的下降不仅仅与不法分子或灵活的初创公司、封闭和有限的应用有关。

相反，权力在整个社会的总和和跨度中被重新分配和加强。即将到来的浪潮的完全全能性质意味着它存在于每一个层面、每一个行业、每一个企业、亚文化、团体或官僚机构，在我们世界的每一个角落。它创造了数万亿美元的新经济价值，同时也摧毁了某些现有的财富来源。一些个人得到了极大的支持；其他人则可能失去一切。在军事上，它同样赋予了一些民族国家和民兵力量。因此，这不仅仅是放大特定脆弱点的问题；从稍长远的角度来看，这是对社会赖以建立的基础的转变。在这场权力的巨大重新分配中，国家已经脆弱且愈加脆弱，被动摇到核心，其宏伟的契约变得支离破碎且岌岌可危。

CHAPTER 11

国家的未来

马镫

A T FIRST BLUSH, STIRRUPS MAY not seem all that revolutionary. They are, after all, fairly rudimentary triangles of metal attached to leather straps and a horse's saddle. Look a bit closer, and another picture emerges.

在马镫出现之前，骑兵在战场上的冲击力是非常有限的。组织良好的防御盾墙通常可以击退骑马的冲锋。因为骑手没有固定在马匹上，他们很脆弱。装备长矛和大盾的士兵，站在紧密排列的队列中，甚至可以让最重的骑兵下马。因此，马匹的主要功能是将你运送到战场。

马镫彻底改变了这一切。它将长矛和骑手固定在冲锋的动物上，使它们成为一个整体。长矛的全部力量现在是马匹和骑士的合力。击中盾牌不再意味着你会掉下来，而是意味着你会击碎盾牌和持盾的人。突然间，全速疾驰，长矛出击，骑手固定，重骑兵冲锋成为一种压倒性的冲击战术。它甚至可以突破最顽强的步兵防线。

这种微小的创新将力量的天平倾向于进攻。马镫引入欧洲后不久，法兰克人的领袖查理·马特尔看到了它的潜力。他利用它产生了毁灭性的效果，击败并驱逐了萨拉森人出法国。但是，这些重骑兵部队的引入需要法兰克社会进行巨大的支持性变革。马匹既饥饿又昂贵。重骑兵需要多年的训练。

训练。作为回应，马特尔及其继承人征用了教会的土地，并利用这些土地培养了一支精英战士。他们的新财富使他们能够养马，使他们有时间训练，将他们与王国联系在一起，并在后来为他们提供资金购买盔甲。作为对新财富和地位的回报，这些精英承诺持械并为国王而战。另一项重大协议达成了。

随着时间的推移，这种临时协议演变成一个复杂的封建制度，形成了对领主的义务网络和庞大的农奴阶层。这是一个由庄园和头衔、骑士比武大会和学徒制、铁匠和工匠、盔甲和城堡构成的世界，一个自觉的纹章文化和骑士勇气的浪漫故事的世界。它成为整个中世纪时期的主要政治形式。

马镫是一个看似简单的创新。但随之而来的是一场社会革命，改变了数亿人的生活。一个在近千年间构建欧洲生活的政治、经济、战争和文化体系，部分依赖于这些小小的金属三角形。马镫和封建主义的故事突显了一个重要的真理：新技术帮助创造新的权力中心，并通过新的社会基础设施来支持和促进它们。在上一章中，我们看到这一过程如何在今天为民族国家带来一系列直接挑战。但从长远来看，权力成本骤降的影响是巨大的，技术政治地震正在动摇国家赖以建立的基础。

虽然技术的小变化可以从根本上改变权力的平衡，但试图准确预测几十年后的情况是非常困难的。指数型技术放大了每个人和每件事。这就产生了看似矛盾的趋势。权力既集中又分散。现任者既被加强又被削弱。民族国家既更加脆弱，又更有可能陷入滥用不受约束的权力的风险。

请记住，增长的权力意味着*everyone's*权力将被放大。在未来的几十年里，历史模式将再次上演。

再次，新的中心将形成，新的基础设施将发展，新的治理和社会组织形式将出现。同时，现有的权力中心将以不可预测的方式得到增强。有时，当人们阅读有关技术的内容时，会有一种令人兴奋的感觉，认为它将席卷所有过去的事物，没有旧的企业或机构能够在这场旋风中幸存。我不认为这是真的；有些会被淘汰，但许多会被增强。电视可以播报革命，但也可以帮助抹去它。技术可以加强社会结构、等级制度和控制制度，也可以颠覆它们。

在由此产生的动荡中，如果不进行重大转变，许多开放的民主国家将面临其制度基础的持续衰退，合法性和权威的萎缩。这是一个技术传播和权力转移的循环动态，它破坏了基础，削弱了控制能力，从而导致进一步的传播。与此同时，威权国家获得了一套强大的新镇压武器。

民族国家将受到巨大的离心力和向心力、集中化和碎片化的影响。这是一条通往混乱的快速通道，质疑谁来做决策以及如何做决策；这些决策如何执行，由谁执行，何时执行，在哪里执行，给那些微妙的平衡和妥协施加压力，直至达到临界点。这种动荡的配方将创造出史诗般的新权力集中和分散，从上到下分裂国家。最终，这将使一些国家的生存能力受到质疑。

这种无法治理的“后主权”世界，用政治学家温迪·布朗的话来说，将远远超出短期脆弱感；相反，它将是一个长期的宏观趋势，几十年来不断地走向深度不稳定。第一个结果将是大规模的新权力和财富集中，重新排列社会。

浓度：智力的复利回报

源文本: 翻译文本:

从蒙古人到莫卧儿人，超过一千年的时间里，亚洲最强大的力量是传统帝国。到1800年，这种情况发生了变化。取而代之的是一家私人公司，由相对少量的股东拥有，由少数几个灰头土脸的会计师和管理员经营，他们在一个距离数千英里之外的城市里的一栋只有五扇窗户宽的建筑中办公。

在十九世纪之交，英国东印度公司控制了印度次大陆的大片地区。它统治的土地和人口超过了整个欧洲，征收税款并制定法律。它指挥着一支训练有素的常备军，人数达20万，是英国本土军队的两倍，并运营着世界上最大的商船队。其集体火力超过了亚洲任何一个国家。其全球贸易关系在从香港的建立到波士顿倾茶事件的各个方面都起到了基础性作用。其关税、税收和股息对英国经济至关重要；当时不低于一半的英国对外贸易通过该公司进行。

这显然不是一家普通的公司。实际上，它是某种帝国。很难用现代术语来构想这样的公司。我们并不是在走向新殖民主义的东印度公司2.0。但我确实认为我们必须面对一些董事会不仅对塑造当今文化和政治的微妙推动和选择架构的巨大规模和影响力，更重要的是，对未来几十年可能导致的方向。他们是某种形式的帝国，随着即将到来的浪潮，他们的规模、影响力和能力将会大幅扩展。

源文本：—翻译文本：

衡量人工智能进展的方法是将其与个体人类在某项任务上的表现进行比较。研究人员谈论在语言翻译或驾驶等现实世界任务中实现超人表现。但这忽略了最强大的因素。

世界上的力量实际上是个人协调以实现共同目标的*groups*。组织也是一种智能。公司、军队、官僚机构，甚至市场——这些都是人工智能，汇集和处理大量数据，围绕特定目标组织自己，建立机制以更好地实现这些目标。实际上，机器智能更像是一个庞大的官僚机构，而不是一个人类大脑。当我们谈论像AI这样的东西对世界产生巨大影响时，值得记住这些老式AI的影响有多深远。

当许多任务，或许是大多数任务，要求操作一个公司或政府部门时，可以由机器更高效地运行，会发生什么？谁将首先从这些动态中受益，他们可能会用这种新力量做些什么？

我们已经进入了一个时代，巨型公司拥有万亿美元的估值和比整个国家更多的资产，从各个方面来说。以苹果公司为例。它生产了我们物种历史上最美丽、最有影响力和最广泛使用的产品之一。iPhone 是天才之作。其产品被全球超过12亿人使用，公司理所当然地因其成功而获得了丰厚的回报：在2022年，苹果公司的估值超过了英国富时100指数上所有上市公司的总和 *combined.*。苹果公司在银行中拥有接近2000亿美元的现金和投资，并且其受众大多被锁定在其生态系统中，似乎已做好准备利用这一新潮流。

同样地，来自非常不同领域的广泛服务，跨越地球上的巨大区域，已经被整合到一个单一的公司——谷歌：地图和定位、评论和商业列表、广告、视频流、办公工具、日历、电子邮件、照片存储、视频会议等等。大型科技公司提供的工具涵盖从组织生日聚会到运营数百万美元业务的一切。唯一能如此深入影响众多人生活的相当组织是国家政府。称之为“ 谷歌化 ”：一系列免费或低成本提供的服务导致单一实体。

功能性地支持经济和人类体验的大部分。

为了了解这些集中度，请考虑，*Fortune* 的全球 500 强公司的总收入已经占到世界 GDP 的 44%。它们的总利润比除前六个国家以外的所有国家的年度 GDP 都要大。公司已经控制了最大的 AI 处理器集群、最好的模型、最先进的量子计算机，以及绝大多数的机器人能力和知识产权。与火箭、卫星和互联网不同，这一波浪潮的前沿是在公司中，而不是在政府组织或学术实验室中。加速这一过程，随着下一代技术的发展，企业集中的未来似乎并不那么不寻常。

已经存在一种明显且加速的“ 超级明星 ”效应，领先的参与者占据了越来越大的份额。尽管世界前五十个城市仅占全球人口的8%，但它们却拥有财富和企业权力的最大份额（大公司总部的45%；全球GDP的21%）。全球前10%的公司占据了总利润的80%。预计即将到来的浪潮将加剧这一现象，产生越来越富有和成功的超级明星——无论是地区、商业部门、公司还是研究团体。

我认为我们将看到一批私营公司发展到超越许多国家的规模和影响力。考虑一下像韩国三星集团这样庞大的企业帝国的巨大影响力。三星最初在近一个世纪前作为一家面条店成立，在朝鲜战争后成为一个主要的企业集团。随着韩国在20世纪60年代和70年代的快速增长，三星处于其核心地位，不仅是一个多元化的制造业巨头，也是银行和保险领域的重要参与者。韩国的经济奇迹是一个由三星推动的奇迹。到这时，三星已成为领先的财阀，这是对主导该国的一小群大型公司的称呼。

智能手机、半导体和电视是三星的专长。但人寿保险、渡轮运营商和主题公园也是如此。在三星的职业生涯备受珍视。三星集团的收入占到20%

韩国经济。对于今天的韩国人来说，三星几乎就像一个平行政府，在人们的生活中始终存在。鉴于错综复杂的利益网络以及持续不断的企业和政府丑闻，国家与企业之间的权力平衡岌岌可危且模糊不清。

三星和韩国是例外，但也许不会持续太久。鉴于集中的能力范围，今天通常由政府负责的事务，如教育和国防，甚至可能是货币或法律执行，都可以由这一代新公司提供。例如，eBay和PayPal的争议解决系统每年处理大约六千万个纠纷，是整个美国法律系统的三倍。这些纠纷中有90%仅通过技术解决。未来还有更多变化。

源文本：—翻译文本：

现代帝国，某种意义上的现代帝国。即将到来的浪潮迅速加速了这一趋势，将巨大的权力和财富置于那些创造和控制它的人手中。新的私人利益将进入被过度扩张和紧张的政府腾出的空间。这个过程不会像东印度公司那样在火枪的枪口下强制执行，但它将会，正如东印度公司一样，创造出具有政府规模、影响力和权力的私人公司。那些拥有资金、专业知识和分销能力以利用即将到来的浪潮的公司，将极大地增强他们的情报并同时扩大他们的影响力，将会看到巨大的收益。

在最后一波浪潮中，事物非物质化；商品变成了服务。你不再购买软件或音乐CD；它们是流媒体。你只是期望使用谷歌或苹果时，防病毒和安全软件是附带的产品。产品会损坏，会过时。服务则不太会。它们无缝且易于使用。就公司而言，他们渴望你订阅他们的软件生态系统；定期付款很有吸引力。所有大型科技平台要么主要是服务业务，要么拥有非常大的服务。

企业。苹果公司拥有应用商店，尽管主要销售设备，而亚马逊作为全球最大的实物商品零售商，也为商家提供电子商务服务，为个人提供电视流媒体服务，并在其云服务 Amazon Web Services 上托管了大量互联网内容。

无论你在哪里看，技术都在加速这种非物质化，通过提供持续的消费服务而不是传统的一次性购买产品来减少终端消费者的复杂性。无论是像Uber、DoorDash和Airbnb这样的服务，还是像Instagram和TikTok这样的开放出版平台，巨型企业的趋势是不要参与市场，而是成为市场，不是制造产品，而是运营服务。现在的问题是，还有什么可以被制成服务，融入另一个巨型企业的现有服务套件中？

在未来几十年，我预测大多数实物产品将会像服务一样。零边际成本的生产和分销将使其成为可能。向云端的迁移将变得无所不包，这一趋势将受到低代码和无代码软件的兴起、生物制造的崛起以及3D打印的繁荣的推动。当你结合即将到来的浪潮的所有方面，从AI的设计、管理和物流能力，到量子计算所支持的化学反应建模，再到机器人精细组装能力，你将会看到生产性质的全面革命。

食品、药品、家用产品，实际上几乎任何东西都可以通过3D打印、生物生产或使用原子精密制造在使用地点或附近制造，由复杂的人工智能通过自然语言与客户流畅合作进行管理。你只需购买执行代码，让人工智能或机器人完成任务或创造产品。是的，这掩盖了大量的材料复杂性，是的，这还很遥远。但眯起眼睛看远方，这种情景显然是可行的。即使你不完全接受这里的论点，这些力量不可能不在全球经济供应链中创造重大变化和新的价值集中。

满足对廉价和无缝服务的需求通常需要规模（在芯片、人员、安全、创新方面的大量前期投资），这会奖励并加速集中化。在这种情况下，将只有少数几个巨头，其规模和实力将开始与传统国家相抗衡。更重要的是，拥有最佳系统的所有者可能能够建立巨大的竞争优势。我刚才提到的那些庞大的集中化即将到来的公司？它们可能最终会比过去的企业更大、更富有、更根深蒂固。

系统在一个又一个领域成功泛化的次数越多，权力和财富就越集中在其所有者手中。那些拥有资源以最快速度发明或采用新技术的人——例如那些能够通过我更新的图灵测试的人——将享受快速复合的回报。他们的系统拥有更多的数据和“现实世界的部署经验”，因此工作得更好，泛化得更快，并锁定优势，吸引最优秀的人才来构建它们。一个无法弥合的“智力差距”变得可能。如果一个组织领先得足够远，它可能成为一个收入生成者，并最终成为一个无与伦比的权力中心。如果这一过程扩展到类似完全通用人工智能或量子霸权的程度，它可能会使新进入者甚至政府面临非常困难的局面。

无论终点在哪里，我们正前往一个前所未有的力量和能力掌握在已经强大的行为者手中的地方，他们无疑会利用这些力量来扩大他们的影响力并推进他们自己的议程。

这种集中将使庞大的自动化大企业能够将价值从人力资本——工作——转移到原始资本。将集中带来的所有不平等放在一起，这就构成了现有裂痕的又一次大加速和结构性深化。难怪有人谈论新封建主义或技术封建主义——这对社会秩序的直接挑战，这次建立在超越马镫的基础之上。

总之，智能的回报将呈指数级增长。我们曾称之为组织的少数人工智能将从新的能力集中中获得巨大收益——可能是有史以来最大的这种收益。

迄今为止所见的浓度。将使我们物种如此成功的本质重新创造为可以在各种不同环境中反复使用和应用的工具，是一个巨大的奖赏，各种公司和官僚机构都会追求并利用。如何治理这些实体，它们将如何与国家摩擦、捕获和重新设计，仍是一个悬而未决的问题。它们将挑战国家似乎是确定无疑的。

但更高权力集中度的后果并不仅限于企业。

监控：威权主义的火箭燃料

源文本: 翻译文本:

与超级明星公司相比，政府显得缓慢、臃肿且脱离实际。人们很容易将其视为走向历史垃圾桶。然而，国家的另一种不可避免的反应将是利用即将到来的浪潮中的工具来加强对权力的控制，充分利用这些工具来巩固其主导地位。

在二十世纪，极权主义政权希望建立计划经济、顺从的人口和受控的信息生态系统。他们想要完全的霸权。生活的每一个方面都被管理。五年计划规定了一切，从电影的数量和内容到某个田地预期的小麦蒲式耳数。高度现代主义的规划者希望创造出具有鲜明秩序和流动性的纯净城市。一个时刻警惕且无情的安全机构使这一切运转。权力集中在一个能够全面审视并果断行动的最高领导者手中。想想苏联的集体化、斯大林的五年计划、毛的中国、东德的斯塔西。这是政府作为反乌托邦噩梦。

到目前为止，至少总是以灾难性的错误告终。尽管革命者和官僚们都尽了最大努力，社会仍无法被塑造成型；它对国家来说从未完全“清晰可见”，而是一个混乱的、无法治理的现实，不会符合那些纯粹主义者的梦想。

中心。人类过于多样化，过于冲动，无法被这样限制。在过去，极权政府所拥有的工具根本无法胜任这一任务。因此，这些政府失败了；他们未能改善生活质量，最终要么崩溃，要么改革。极端集权不仅极其不受欢迎，而且实际上是不可能的。

即将到来的浪潮带来了一个令人不安的可能性，即这可能不再是真实的。相反，它可能会引发一种集中的权力和控制的注入，将国家职能转变为对其原始目的的压制性扭曲。为威权主义者和大国竞争提供火箭燃料。以非凡的规模和精度捕获和利用数据的能力；创建跨越领土的监控和控制系统，实时反应；换句话说，将历史上最强大的技术集于一身，由一个单一机构指挥，将如此全面地重写国家权力的界限，以至于它将产生一种全新的实体。

源文本：—翻译文本：

你醒来。你立刻拿起手机查看邮件。你的智能手表告诉你昨晚睡得很正常，早上的心率也在平均水平。理论上，一个远程组织已经知道你醒来的时间、你的感受以及你在看什么。你离开家前往办公室，手机追踪你的移动，记录你短信的按键和你收听的播客。在路上以及整天，你被闭路电视拍摄了数百次。毕竟，这个城市至少每十个人就有一台摄像机，可能还不止。当你在办公室刷卡进入时，系统记录了你的进入时间。安装在你电脑上的软件监控生产力，甚至精确到眼球运动。

在回家的路上，你停下来买晚餐。超市的忠诚计划会跟踪你的购买记录。吃完后，你又狂看了一部电视剧；你的观看习惯被如实记录。每一次浏览，每一次匆忙

信息、每一个在开放浏览器中记录的半个想法或短暂的搜索、每一步穿过繁忙城市街道、每一次心跳和糟糕的夜晚睡眠、每一次购买或放弃的交易——这一切都被捕捉、监视、记录。而这仅仅是每天收集的可能数据的一小部分，不仅仅是在工作或手机上，而是在医生办公室或健身房。几乎生活的每一个细节都被记录在某个地方，被那些有能力处理和利用他们收集的数据的人记录。这不是某种遥远的反乌托邦。我描述的是像伦敦这样的城市中数百万人的日常现实。

唯一剩下的步骤是将这些不同的数据库整合成一个单一的、集成的系统：一个完美的二十一世纪监控装置。最突出的例子当然是中国。这并不是什么新闻，但现在已经清楚的是，党的计划已经是多么先进和雄心勃勃，更不用说在二三十年后可能达到的程度。

与西方相比，中国对人工智能的研究集中在监控领域，如目标跟踪、场景理解以及语音或动作识别。监控技术无处不在，其在捕捉公民生活各个方面的能力上越来越细致。它们结合了对面部、步态和车牌的视觉识别与大规模的数据收集——包括生物数据。像微信这样的集中化服务将私人消息、购物和银行业务等一切捆绑在一个易于追踪的地方。在中国的高速公路上行驶，你会注意到数百个自动车牌识别摄像头在跟踪车辆。（这些在西方世界的大多数大城市地区也存在。）在新冠疫情隔离期间，机器狗和无人机携带扬声器播放信息，警告人们待在室内。

面部识别软件基于我们在第2部分中看到的计算机视觉的进步，以极高的准确性识别个人面孔。当我打开手机时，它在“看到”我的脸时会自动启动：一个小而流畅的便利，但具有明显而深远的影响。尽管该系统最初是由企业和学术界开发的。

美国的研究人员，没有哪个地方比中国更接受或完善这项技术。

毛主席曾说过“群众的眼睛是雪亮的”，用来形容人们在观察邻居是否违反共产主义正统时的敏锐。到2015年，这句话成为了一个大规模“雪亮工程”人脸识别项目的灵感，该项目最终的目标是在100%的公共空间中推广这种监控。来自香港中文大学的一个顶尖研究团队继续创立了商汤科技，这是一家全球最大的人脸识别公司之一，建立在超过二十亿张人脸的数据库之上。中国现在是人脸识别技术的领导者，像旷视科技和云从科技这样的巨头公司正在与商汤科技争夺市场份额。中国警方甚至拥有内置人脸识别技术的太阳镜，能够在人群中追踪嫌疑犯。

全球约一半的十亿台闭路电视摄像机在中国。许多摄像机内置人脸识别功能，并被精心布置以收集最大量的信息，通常位于准私人空间：住宅楼、酒店，甚至是卡拉OK厅。*New York Times* 的一项调查发现，仅福建省的警方就估计他们拥有一个包含25亿张人脸图像的数据库。他们坦率地说明了其目的：“控制和管理人。”当局还希望吸收音频数据——中山市的警方希望摄像机能够在三百英尺的半径内录制音频——而在COVID时代，生物数据的密切监控和存储已成为常态。

公安部明确了下一个重点：将这些分散的数据库和服务整合成一个连贯的整体，从车牌到DNA，从微信账户到信用卡。这个由人工智能驱动的系统可以实时发现对中共的潜在威胁，如异议人士和抗议活动，从而使政府能够无缝地、严厉地回应任何其认为不受欢迎的事物。在新疆自治区，这种整合的潜在恐怖性尤为突出。

中国西北部这个偏远而崎岖的地区经历了对其本土维吾尔族人民的系统性和技术性压迫及种族清洗。所有这些监控和控制系统都被带来

在这里一起。城市被置于摄像头监控的覆盖之下，配有面部识别和人工智能追踪。检查站和“ 再教育 ”营地控制着行动和自由。基于众多监控数据库的社会信用评分系统对人口进行监控。有关当局建立了一个虹膜扫描数据库，能够容纳多达三千万个样本——超过该地区的人口。

无所不在的监控和控制社会已经存在，现在这一切都将大幅升级，成为权力在中心的下一个级别的集中。然而，将其仅仅视为中国或威权主义的问题是错误的。首先，这项技术正被大规模出口到委内瑞拉、津巴布韦、厄瓜多尔和埃塞俄比亚等地。甚至到美国。2019年，美国政府禁止联邦机构及其承包商购买来自包括华为、中兴和海康威视在内的多家中国供应商的电信和监控设备。然而，仅仅一年后，发现有三个联邦机构从被禁止的供应商处购买了此类设备。甚至有超过一百个美国城镇获得了为在新疆对维吾尔人使用而开发的技术。这是一个教科书式的遏制失败。

西方公司和政府也处于构建和部署这项技术的前沿。上文提到伦敦并非偶然：它与深圳等城市竞争成为世界上监控最严密的城市。政府监控和控制其自身人口已不是秘密，但这些趋势也深入到西方公司。在智能仓库中，每个工人的每一个微小动作都被追踪，甚至包括体温和上厕所的时间。像Vigilant Solutions这样的公司基于车牌追踪汇总移动数据，然后将其出售给州或市政府等管辖区。甚至你的外卖披萨也在被监控：Domino's使用AI驱动的摄像头来检查其披萨。正如中国的任何人一样，西方的人们每天都留下大量的数据痕迹。就像在中国一样，这些数据被收集、处理、运用并出售。

源文本：—翻译文本：

全球“高科技全景监狱”的概念曾是反乌托邦小说的素材，如叶夫根尼·扎米亚京的*We*或乔治·奥威尔的*1984*.。全景监狱正在成为可能。数十亿设备和万亿数据点可以同时实时操作和监控，不仅用于监视，还用于预测。它不仅能够精确和细致地预见社会结果，还可能微妙地或公开地引导或强迫这些结果，从选举结果等宏观过程到个人消费者行为。

这将把极权主义的前景提升到一个新的层面。这不会在所有地方同时发生。但如果人工智能、生物技术、量子技术、机器人技术等被集中在一个压制性国家的手中，所产生的实体将与迄今为止看到的任何实体明显不同。在下一章中，我们将回到这种可能性。然而，在此之前，还有另一种趋势。一个完全且矛盾地与集中化相对立的趋势。

碎片化：权力归于人民

源文本: 翻译文本:

听到“真主党”这个词，大多数人不会联想到议会、学校和医院。毕竟，这是一个在黎巴嫩内战的漫长悲剧中诞生的激进组织，有着暴力的历史记录，被美国政府正式列为恐怖组织，并且经常充当伊朗利益的代理。但这里发生的事情远不止这些，它暗示着权力和国家的另一种方向。

在其黎巴嫩的本土领地，真主党作为一个什叶派的“国中之国”运作。它有一个规模庞大且臭名昭著的军事派别。根据一位分析师的话，它可能是世界上武装最精良的非国家行为体，拥有“比大多数国家更大的火炮库”。它拥有无人机、坦克、远程火箭，以及成千上万的步兵，这些步兵曾在叙利亚内战中与阿萨德政权并肩作战，并定期与以色列交战。

也许令一些人感到惊讶的是，真主党也是一个主要的主流政治力量，是黎巴嫩政府持续心理剧中的一个常规政党。在许多方面，它只是政治体系的另一部分，建立联盟，起草法律，并与国家的常规工具合作。其成员在地方市政委员会和议会中任职，并担任部长内阁职位。在其控制的黎巴嫩大片领土上，真主党运营学校、医院、医疗中心、基础设施、水项目和小额信贷计划。事实上，其中一些项目甚至得到了逊尼派和基督徒的支持。整个地区基本上由真主党以国家的方式管理。它还进行各种商业活动，包括合法和更具犯罪性质的活动，如走私石油。

那么，真主党是什么？是国家还是非国家？是极端主义团体还是传统的基于领土的权力？它反而是一个奇怪的“混合”实体，在国家机构内外同时运作。一个国家，但又不是一个国家，能够选择性地承担责任和活动，以符合自身利益，往往对更广泛的国家和地区造成严重后果。像真主党这样的组织并不多见，它是在独特的地区紧张局势中演变而来的。

然而，即将到来的浪潮可能会使一系列小型、类国家实体变得更加可行。与集中化相反，它实际上可能会刺激一种“真主党化”，一个分裂的、部落化的世界，在这个世界中，每个人都能获得最新技术，每个人都能根据自己的条件自给自足，在没有国家组织的大型超级结构的情况下，任何人都更有可能维持生活水平。

考虑到人工智能、廉价机器人和先进生物技术的结合，再加上清洁能源来源，可能会在现代首次使“离网”生活几乎等同于接入电网。回想一下，仅在过去十年中，太阳能光伏的成本已经下降了超过82%，并将进一步下降，使得小型社区的能源自给自足触手可及。随着电气化的

随着基础设施和化石燃料替代品的渗透，世界上更多的地方可能变得自给自足——但现在配备了人工智能、生物技术、机器人等基础设施，能够在本地生成信息和进行制造。

诸如教育和医学等领域目前依赖于庞大的社会和金融基础设施。完全可以想象这些领域会被精简和本地化：例如，自适应和智能教育系统，可以引导学生完成整个学习旅程，构建定制课程；能够创建所有材料的人工智能，如完美适应儿童的互动游戏和自动评分系统；等等。

您可能没有像国家体系那样的集体安全保护伞，但可以根据需要聘请不同形式的物理和网络保护。AI 黑客和自主无人机也将为私人安全团体所用。我们之前看到，进攻能力正在扩散给任何想要的人；相反，同样的分布最终也会发生在防御上。当任何人都能接触到尖端技术时，不仅仅是国家可以建立强大的物理和虚拟防御。

简而言之，现代社会和社会组织的关键部分，如今依赖于规模和集中化，可以通过即将到来的浪潮解锁的能力进行彻底下放。在这个世界中，大规模叛乱、分裂主义和任何形式的国家形成看起来都截然不同。重新分配真正的权力意味着各种社区可以按照自己的意愿生活，无论他们是ISIS、FARC、匿名者、从比亚法拉到加泰罗尼亚的分裂主义者，还是在太平洋偏远岛屿上建造豪华主题公园的大型公司。

源文本：—翻译文本：

即将到来的浪潮指向权力的进一步集中。最大的AI模型将花费数亿美元进行训练，因此很少有人会拥有它们。但矛盾的是，一种反趋势将同时展开。AI的突破已经使其

在开放获取期刊上发表后的几天内，便进入开源代码库，使得顶级模型易于任何人访问、试验、构建和修改。模型甚至包括权重都被发布、泄露和窃取。

像 Stability AI 和 Hugging Face 这样的公司加速了分布式、去中心化的 AI 形式。像 CRISPR 这样的技术使生物实验变得更容易，这意味着在车库里的生物黑客可以在科学的绝对前沿进行尝试。最终，共享或复制 DNA 或大型语言模型的代码是微不足道的。开放性是默认的，模仿是普遍存在的，成本曲线不断下降，访问障碍瓦解。任何想要的人都可以获得指数级的能力。

这预示着从现有中心向权力 *away* 的巨大再分配。想象一个未来，小团体——无论是在像黎巴嫩这样的失败国家，还是在新墨西哥的离网游牧营地——提供由人工智能赋能的服务，如信用社、学校和医疗保健，这些服务是社区的核心，通常依赖于规模或国家。在微观层面上设定社会条款的机会变得不可抗拒：来我们的精品学校，永远避免批判种族理论，或者抵制邪恶的金融系统，使用我们的去中心化金融产品。任何形式的团体——无论是意识形态、宗教、文化、种族——都可以自我组织成一个可行的社会。想想建立你自己的学校。或者医院或军队。这是一个如此复杂、庞大且困难的项目，甚至想到它都让人感到疲惫。仅仅是收集资源、获得必要的许可和设备，就是一生的努力。现在考虑拥有一系列助手，当被要求创建学校、医院或军队时，可以在现实的时间框架内实现。

ACI 和合成生物学赋予了“ 反抗灭绝 ”与道琼斯大公司同样的力量；赋予了拥有魅力领袖的小国与笨重巨人同样的力量。虽然规模的一些优势可能会得到增强，但它们也可能被抵消。问问自己，如果每个教派、分裂主义运动、慈善基金会和社交网络，每个狂热分子和仇外者，每个民粹主义阴谋论、政党，甚至黑手党、贩毒集团或恐怖组织都有机会会发生什么。

国家建设。被剥夺权利的人将以自己的方式重新获得权利。

碎片化可能会发生在各个地方。 如果公司本身开始走上成为国家的旅程怎么办？ 或者城市决定脱离并获得更多自治权？ 如果人们在虚拟世界中花费的时间、金钱和情感能量比在现实世界中更多呢？ 当强大工具和专业知识对街头儿童和亿万富翁同样可用时，传统的等级制度会发生什么变化？ 令人瞩目的是，企业巨头大部分时间都在使用像 Gmail 或 Excel 这样的软件工作，而这些软件对地球上大多数人来说都是可以访问的。 通过赋权的民主化，将这一点极大地扩展，当地球上的每个人都可以不受限制地访问有史以来最强大的技术时，会发生什么。

源文本：—翻译文本：

掌握权力在自己手中，我预计不平等的最新前沿将出现在生物学领域。一个分裂的世界是指某些司法管辖区对人体实验的宽容程度远高于其他地区，在这些地区，先进的生物能力和自我改造在DNA层面产生不同的结果，进而在国家和微观国家层面产生不同的结果。这样可能会出现类似生物黑客个人增强的军备竞赛。一个渴望投资或优势的国家可能会看到成为一个无所不包的生物黑客天堂的潜力。如果一群“后人类”将自己工程化到某个无法企及的智力或身体层面，社会契约会是什么样子？这将如何与分裂政治的动态相交，一些飞地试图抛弃整体？

所有这些仍然牢牢处于猜测的领域。但我们正在进入一个新时代，以前不可想象的事情现在成为了一个明确的可能性。在我看来，对正在发生的事情视而不见比过度猜测更危险。

治理通过同意来运作；它是建立在所有相关人员信念上的集体虚构。在这种情况下，主权国家被逼到崩溃的边缘。旧的社会契约被撕成碎片。机构被绕过、削弱、取代。税收、执法、遵守规范：都受到威胁。在这种情况下，权力的快速分裂可能加速一种“涡轮巴尔干化”，使灵活且新兴的有能力的行为者获得前所未有的行动自由。国家所体现的权力和服务的巨大整合开始解体。

在这种情境下，出现了一种更像前民族国家时代的世界，呈现出新中世纪的特征，更小、更本地化、宪法多样化，是一个复杂且不稳定的政治拼图。只是这次拥有了极其强大的技术。当北意大利是一个由小城邦组成的拼图时，它带来了文艺复兴，但同时也是一个不断发生内战和争斗的战场。文艺复兴是伟大的；而用明天的军事技术进行无休止的战争，则不那么好。

对于许多在技术领域或与技术相关领域工作的人来说，这些激进结果不仅仅是令人不悦的副产品；它们本身就是目标。像PayPal创始人和风险投资家彼得·蒂尔这样的超级自由主义技术专家赞美一种国家逐渐消亡的愿景，将其视为对一种过于强大的商业领袖或他们自称的“主权个体”的解放。公共服务、机构和规范的焚烧被欢呼为一种明确的愿景，即技术可能“创造出新的异议模式和不受历史民族国家限制的新型社区的空间。”

技术自由主义运动将罗纳德·里根1981年的格言“政府是问题”推向了逻辑的极端，看到政府的许多缺陷，却看不到其巨大的好处，认为其监管和税收职能是破坏性的速率限制器，几乎没有好处——至少对他们来说是这样。我感到非常沮丧的是，一些最有权势和特权的人持有如此狭隘和破坏性的观点，但这进一步推动了分裂。

这是一个亿万富翁和现代先知可以建立和管理微型国家的世界；在这个世界中，从公司到公社再到算法的非国家行为体开始从上方和下方逐渐超越国家。再想想马镫及其单一简单发明的深远下游影响。然后想想即将到来的发明浪潮的规模。结合现有的压力和脆弱性，上述我所推测的那种大规模变革似乎并不那么遥远。更奇怪的是完全没有激进的变化。

即将到来的矛盾浪潮

源文本: 翻译文本:

如果中央集权和去中心化听起来像是直接矛盾，那是有充分理由的：它们确实是。理解未来意味着同时处理多种相互冲突的轨迹。即将到来的浪潮将引发巨大的集中化和去中心化的激流 *at the same time*.。两者将同时发挥作用。每个人、每个企业、每个教会、每个非营利组织、每个国家，最终都会拥有自己的人工智能，并最终拥有自己的生物和机器人能力。从沙发上的个人到世界上最大的组织，每个人工智能都将致力于实现其所有者的目标。这其中蕴含着理解即将到来的矛盾浪潮的关键，这是一波充满碰撞的浪潮。

每一种新的权力形式都会提供不同的公共物品交付愿景，或提出不同的产品制造方式，或一套不同的宗教信仰进行传播。人工智能系统已经在做出具有明显政治影响的关键决策：谁获得贷款、工作、大学名额、假释；谁能被高级医生看诊。在未来十年内，人工智能将决定公共资金如何使用、军事力量的部署地点或学生应该学习什么。这将在集中化和去中心化的方式中发生。例如，人工智能可能作为一个庞大的、跨越国家的系统运行，一个单一的通用工具管理数亿人。同样，我们也将拥有极其强大的

系统，低成本可用，开源，高度适应，满足一个村庄的需求。

多种所有权结构将并存：在开源集体中实现技术民主化、当今企业领导者或新兴快速扩张初创公司的产品，以及政府持有的，无论是通过国有化还是内部培育。所有这些都将共存和共同进化，并在各地改变、放大、生产和扰乱权力的流动和网络。

这些力量的作用地点和方式将根据现有的社会和政治因素而显著变化。这不应是一个过于简化的图景，并且会有许多抵抗和适应的点，这些在事先并不明显。一些部门或地区会走向一个方向，一些则走向另一个方向，还有一些会看到两者的强烈扭曲。一些等级制度和社会结构将得到加强，另一些则被推翻；有些地方可能变得更加平等或专制，而另一些则大大减少。在所有情况下，额外的压力和波动性、权力的不可预测的放大、激进的新能力中心的剧烈破坏，将进一步加剧自由民主国家体系的基础压力。

如果这个图景听起来太奇怪、矛盾和不可能，请考虑这一点。即将到来的浪潮只会加深并重现上一波浪潮中完全相同的矛盾动态。互联网正是这样做的：在几个关键枢纽中集中，同时也赋予数十亿人权力。它创造了巨头，但也给每个人提供了参与的机会。社交媒体创造了几个巨头和无数的部落。每个人都可以建立一个网站，但只有一个谷歌。每个人都可以销售自己的小众产品，但只有一个亚马逊。如此等等。互联网时代的颠覆很大程度上可以通过这种张力来解释，这种赋权与控制的强大、易燃的混合物。

现在，随着新一波浪潮的到来，这些力量将扩展到互联网和数字领域之外。将它们应用于生活的任何给定领域。是的，这种剧变的配方是我们以前见过的。但如果互联网看起来很大，这次会更大。大规模的全能通用技术将改变社会以及人类的意义。这种

可能听起来夸张。但是在未来十年内，我们必须预见到信息、财富，尤其是权力的激烈变动、新的集中和分散。

那么，这对技术意味着什么，更重要的是，这对我们意味着什么？如果国家无法以平衡的方式控制即将到来的浪潮，会发生什么？到目前为止，在第3部分中，我们已经讨论了现代民族国家已经岌岌可危的状况，并预览了随着即将到来的浪潮而出现的新威胁。我们已经看到，一系列巨大的压力和权力的巨大再分配将汇聚在一起，将唯一有能力管理浪潮的力量——国家——推向危机的边缘。

那一刻几乎已经到来。由于技术的不可阻挡的崛起和国家的终结，这场危机将以巨大的、存在层面的困境形式出现，这是一系列残酷的选择和权衡，代表了二十一世纪最重要的难题。

让我们没有好的选择将是技术的最终失败。然而，这正是我们所走的方向。

CHAPTER 12

困境

灾难：终极失败

源文本: 翻译文本:

THE HISTORY OF HUMANITY IS, in part, a history of catastrophe.

灾难当然也是人为造成的。第一次世界大战导致全球约1%的人口死亡；第二次世界大战则是3%。或者看看十三世纪成吉思汗和蒙古军队在中国和中亚地区发动的暴力，导致多达10%世界人口的死亡。随着原子弹的出现，人类现在拥有足够的致命力量，可以多次杀死地球上的所有人。曾经需要数年和数十年才能发生的灾难性事件，现在可能在按下按钮的几分钟内发生。

Pandemics feature widely. Two killed up to 30 percent of the world population: the sixth-century Plague of Justinian and the fourteenth-century Black Death. England's population was seven million in 1300, but by 1450, crushed by waves of the plague, it was down to just two million.

随着这一波浪潮的到来，我们准备再一次迈出这样的飞跃，扩大风险的上限以及那些寻求释放灾难性力量的人的可用途径的数量。在本章中，我们超越了国家运作的脆弱性和威胁，设想如果无法遏制会发生什么——迟早会发生。

这些技术中的绝大多数将被用于善良。尽管我关注了它们的风险，但重要的是要记住，它们将每天改善无数人的生活。在本章中，我们正在研究

在极端边缘情况下，几乎没有人愿意看到，尤其是那些使用这些工具的人。然而，仅仅因为它们将成为极少数的使用案例，并不意味着我们可以忽视它们。我们已经看到不良行为者可以造成严重破坏，引发大规模的不稳定。现在想象一下，任何半专业的实验室或黑客都可以合成复杂的 DNA 链。灾难会在多久后发生？

最终，随着一些历史上最强大的技术逐渐渗透到各个地方，这些边缘案例变得更加可能。最终，会有事情出错——在与释放的能力相称的规模和速度下。即将到来的浪潮的四个特征的结果是，如果没有在每个层面上运行的强有力的遏制方法，像人为设计的流行病这样的灾难性结果比以往任何时候都更有可能。

这是不可接受的。然而，这里有一个两难境地：最安全的遏制解决方案同样不可接受，将人类引向一个专制和反乌托邦的道路。

一方面，社会可能会转向我们在上一章中看到的那种技术支持的全面监控，这是一种本能反应，通过强硬机制来对抗不受控制的技术。安全——以自由为代价。或者，人类可能完全远离技术前沿。虽然不太可能，但这不是答案。原则上，唯一有能力驾驭这种生存困境的实体是目前正在瓦解的国家体系，被它需要遏制的力量拖累。

随着时间的推移，这些技术的影响将推动人类在灾难和反乌托邦的两极之间寻找一条道路。这是我们这个时代的基本困境。

技术的承诺在于它改善生活，其收益远远超过成本和缺点。这一系列棘手的选择意味着这一承诺已被残酷地颠倒。

末日论让人们——包括我自己——眼神呆滞。在这一点上，你可能会感到警惕或怀疑。谈论灾难性影响往往会招致嘲笑：对灾难主义的指责、放纵的消极态度、尖锐的危言耸听、对遥远而稀有风险的自我关注，当

许多明显而迫在眉睫的危险急需关注。与令人喘不过气的技术乐观主义一样，令人喘不过气的技术灾难主义也很容易被视为一种扭曲的、误导的炒作形式，因为它没有历史记录的支持。

但是，仅仅因为一个警告具有戏剧性的影响，并不是自动拒绝它的好理由。对灾难前景的悲观厌恶的自满本身就是灾难的配方。将警告视为少数怪人的夸大其词的闲聊，这种态度在其自身的逻辑中似乎是合理的、理性的、“ 聪明的 ”，但这种态度为其自身的失败铺平了道路。

毫无疑问，技术风险将我们带入了不确定的领域。然而，所有趋势都指向风险的激增。这种推测基于不断复合的科学和技术进步。我认为，那些忽视灾难的人是在低估摆在我们面前的客观事实。毕竟，我们这里谈论的并不是摩托车或洗衣机的普及。

灾难的种类

源文本: 翻译文本:

要了解我们应该为哪些灾难性危害做好准备，只需推断我们在第10章中看到的不良行为者攻击。以下是一些合理的情景。

恐怖分子将配备面部识别功能的自动武器安装在数百或数千架自主无人机群上，每架无人机都能够快速从武器的后坐力中重新平衡，发射短点射，然后继续移动。这些无人机被释放到一个主要市中心，指示它们杀死特定的目标。在繁忙的高峰时段，这些无人机将以可怕的效率运作，沿着优化的路线在城市中穿行。几分钟内，攻击的规模将远远超过2008年孟买袭击事件，当时武装恐怖分子在中央火车站等城市地标游荡。

一名大规模杀人犯决定用无人机、喷洒装置和定制病原体袭击一场大型政治集会。很快，参加者开始生病，然后是他们的家人。演讲者是一位备受喜爱和憎恨的政治人物。

避雷针，是最早的受害者之一。在狂热的党派氛围中，这样的攻击在全国范围内引发了暴力报复，混乱接踵而至。

仅使用自然语言指令，美国的一名敌对阴谋论者传播大量精心构建且具有分裂性的虚假信息。多次尝试大多未能获得关注。最终有一个引起了注意：芝加哥的一起警察谋杀案。它完全是假的，但街头的麻烦和广泛的反感却是真实的。攻击者现在有了一本剧本。当视频被证实为骗局时，全国各地已经因暴力骚乱和多起伤亡而动荡不安，新的虚假信息不断煽动火焰。

或者想象所有这些事情同时发生。或者不仅仅是在一个事件或一个城市，而是在数百个地方。使用这样的工具，很容易意识到不良行为者的赋权打开了灾难的大门。今天的人工智能系统努力不告诉你如何毒害水源或制造无法检测的炸弹。它们尚不具备自行定义或追求目标的能力。然而，正如我们所见，今天最前沿的、更强大的模型的更广泛传播和更不安全的版本正在快速到来。

在即将到来的浪潮中，所有灾难性风险中，人工智能得到了最多的关注。但还有很多其他风险。一旦军队完全自动化，冲突的进入门槛将大大降低。战争可能会因某些永远无法明确的原因而意外爆发，人工智能检测到某种行为或威胁模式，然后立即以压倒性的力量作出反应。可以说，这种战争的性质可能是陌生的，迅速升级，并且在破坏性后果方面无与伦比。

我们已经遇到过人为制造的大流行病和意外泄漏的危险，并且瞥见了当数百万自我提升爱好者可以对生命的遗传密码进行实验时会发生什么。一种不太明显的极端生物风险事件，针对特定人群，或者破坏生态系统，不能被排除。想象一下，活动家们想要阻止可卡因贸易，发明了一种新病毒，

仅以古柯植物为目标，作为替代空中喷洒的方法。或者，如果激进的素食主义者决定破坏整个肉类供应链，可能会导致严重的预期和意外后果。两者都可能失控。

我们知道在放大脆弱性方面实验室泄漏可能是什么样子，但如果没有迅速得到控制，它将与以往的瘟疫相提并论。为了将其置于背景中，COVID的奥密克戎变体在首次被发现后的100天内感染了四分之一的美国人。如果我们遇到一种具有20%死亡率的流行病，但具有那种传播性呢？或者如果它是一种呼吸道HIV，会潜伏多年而没有急性症状呢？一种新型的人类可传播病毒，假设其再生率为4（远低于水痘或麻疹），病死率为50%（远低于埃博拉或禽流感），即使考虑到封锁式措施，也可能在几个月内造成超过十亿人死亡。如果多种此类病原体同时释放呢？这远远超出了脆弱性放大；这将是一场无法想象的灾难。

源文本：—翻译文本：

在学术研究者的一个亚文化中，有人推动了一种极端叙述，认为人工智能可能引发生存灾难。想象一下，一台全能机器以某种方式为了其自身神秘的目的而摧毁世界：不是像电影中那样的恶意AI故意造成破坏，而是一个全面的AGI盲目地优化一个不透明的目标，对人类的关切视而不见。

经典的思想实验是，如果你设置了一个足够强大的AI来制造回形针，但没有足够仔细地指定目标，它最终可能会将世界，甚至整个宇宙的内容都变成回形针。开始遵循这样的逻辑链，数不清的令人不安的事件序列就会展开。AI安全研究人员（正确地）担心，如果创造出类似AGI的东西，人类将不再掌控自己的命运。我们将首次被推翻，失去作为已知宇宙中主导物种的地位。无论多么聪明的

设计者，无论安全机制多么健全，考虑到所有可能性，保证安全都是不可能的。即使完全符合人类利益，一个足够强大的人工智能也可能会重写其编程，丢弃那些看似内置的安全和对齐功能。

沿着这种思路，我经常听到人们说类似“ AGI是人类今天面临的最大风险！它将终结世界！ ”这样的话。但是，当被追问这实际上是什么样子，这究竟是如何发生时，他们变得含糊其辞，答案模糊不清，确切的危险也不明确。他们说，AI可能会占用所有的计算资源，把整个世界变成一个巨大的计算机。随着AI变得越来越强大，最极端的情景将需要认真考虑和缓解。然而，在我们到达那一步之前，可能会出现许多问题。

在未来十年，人工智能将成为历史上最强大的力量倍增器。这就是为什么它可能在历史规模上实现权力的重新分配。作为人类进步的最大加速器，它也将带来危害——从战争和事故到随机恐怖组织、专制政府、过度扩张的公司、简单的盗窃和故意破坏。想象一下，一个能够轻松通过现代图灵测试的ACI，但却被用于灾难性的目的。先进的人工智能和合成生物学不仅会被用于寻找新能源或改变生活的药物的团体；它们也将被下一个泰德·卡辛斯基所利用。

人工智能既有价值又危险，正是因为它是我们最好和最坏自我的延伸。作为一种以学习为基础的技术，它可以不断适应、探索，产生可能与以往任何考虑过的东西甚至其他人工智能都大相径庭的新策略和想法。让它建议如何切断淡水供应、崩溃股市、引发核战争或设计终极病毒，它都会做到。很快。比起担心假设的回形针最大化者或某个奇怪的、恶意的恶魔，我更担心在未来十年中，这个工具将放大哪些现有的力量。

想象一下这样的场景：人工智能控制着能源电网、媒体节目、电站、飞机或大型金融机构的交易账户。当机器人无处不在，军队充斥着致命的自主武器——仓库里满是可以在按下按钮的瞬间实施大规模自主谋杀的技术——由另一个人工智能开发的黑客攻击会是什么样子？或者考虑更基本的故障模式，不是攻击，而是简单的错误。如果人工智能在基础设施中犯错，或者一个广泛使用的医疗系统开始出现故障呢？不难看出，众多有能力的准自主代理在外游荡，即使是那些追求善意但目标不明确的代理，也可能会造成混乱。我们尚不清楚人工智能对农业、化学、外科手术和金融等不同领域的影响。这是问题的一部分；我们不知道引入了哪些故障模式，以及它们可能会延伸多深。

没有关于如何安全构建即将到来的技术浪潮的说明手册。我们无法提前实验构建不断升级的强大且危险的系统。我们无法知道人工智能可能多快自我改进，或者在实验室事故后会发生什么，尤其是涉及一些尚未发明的生物技术。我们无法判断人类意识直接接入计算机会产生什么结果，或者启用人工智能的网络武器对关键基础设施意味着什么，或者基因驱动在野外会如何表现。一旦快速进化、自我组装的自动装置或新的生物制剂被释放到野外，就无法倒回时间。在某个节点之后，即使是好奇心和修补也可能是危险的。即使你认为灾难的可能性很低，我们在盲目操作这一事实也应该让你停下来思考。

即使是构建安全和受控的技术本身也不够。解决AI对齐问题并不意味着只做一次；这意味着在构建出足够强大的AI时，无论何时何地发生，都要做到这一点。你不仅需要解决一个实验室的泄漏问题；你需要在每个实验室、每个国家、永远解决这个问题，即使这些国家正处于严重的政治压力之下。一旦技术达到关键能力，早期的先驱者就不够了。

只需安全地构建它，尽管这无疑具有挑战性。真正的安全性需要在每一个实例中保持这些标准：考虑到这些已经扩散得如此之快和广泛，这是一个巨大的期望。

当任何人都可以自由发明或使用影响我们所有人的工具时，就会发生这种情况。而我们不仅仅是在谈论印刷机或蒸汽机的使用权，尽管它们都非常了不起。我们谈论的是具有根本性新特征的产物：新化合物、新生命、新物种。

如果波浪无法遏制，这只是时间问题。考虑到事故、错误、恶意使用、超出人类控制的进化、各种不可预测的后果的可能性。在某个阶段，以某种形式，某个地方的某些东西将会失败。而这不会是博帕尔甚至切尔诺贝利；它将在全球范围内展开。这将是大多数情况下以最好的意图生产的技术的遗产。

然而，并不是每个人都有这些意图。

邪教、疯子和自杀状态

源文本: 翻译文本:

大多数情况下，功能增益研究等事物带来的风险是经过批准和无害的努力的结果。换句话说，它们是超大规模的报复效应，是做好事的意图所带来的意外后果。不幸的是，一些组织的成立恰恰是出于相反的动机。

成立于20世纪80年代，奥姆真理教是一个日本的末日邪教。该组织起源于一个瑜伽馆，由一个自称麻原彰晃的男子领导。通过吸引不满现状的人加入，他们随着人数的增加而激进化，坚信世界末日即将来临，只有他们能够幸存，并且他们应该加速这一进程。麻原将该邪教发展到四万到六万名成员之间，诱导一群忠诚的副手甚至使用生物和化学武器。在奥姆

在真理教最受欢迎的时期，据估计其资产超过 $1 billion，并拥有数十名训练有素的科学家作为成员。尽管他们对地震制造机、等离子枪和反射太阳光线的镜子等奇异的科幻武器充满了迷恋，但他们是一个致命严肃且高度复杂的团体。

奥姆真理教建立了空壳公司，并渗透到大学实验室以获取材料，购买了澳大利亚的土地，意图勘探铀以制造核武器，并在东京郊外的丘陵地区开展了大规模的生物和化学武器计划。该组织试验了光气、氰化氢、梭曼和其他神经毒剂。他们计划设计并释放一种增强版的炭疽病毒，并招募了一名研究生级别的病毒学家来协助。成员们获得了神经毒素*C. botulinum*，并将其喷洒在成田国际机场、国会大厦、皇宫、另一个宗教团体的总部以及两个美国海军基地。幸运的是，他们在制造过程中出现了错误，没有造成伤害。

它没有持续。1994年，奥姆真理教从一辆卡车上喷洒神经毒剂沙林，造成八人死亡，二百人受伤。一年后，他们袭击了东京地铁，释放更多沙林，造成十三人死亡，约六千人受伤。地铁袭击涉及在地铁系统周围放置装满沙林的袋子，部分因为封闭空间而更具危害性。幸运的是，这两次袭击都没有使用特别有效的投放机制。但最终，阻止更具灾难性事件的只是运气。

奥姆真理教结合了不同寻常的组织程度和令人恐惧的野心。他们想通过大规模谋杀来引发第三次世界大战和全球崩溃，并开始建立基础设施来实现这一目标。一方面，像奥姆真理教这样的组织是多么罕见，这让人感到宽慰。自1990年代以来，许多恐怖事件和其他非国家实施的大规模杀戮事件中，大多数是由精神失常的独行者或具有特定政治或意识形态议程的团体实施的。

但另一方面，这种安慰是有限的。获取强大武器曾是一个巨大的进入障碍，有助于防止灾难发生。校园枪手令人作呕的虚无主义受到他们能获取的武器的限制。炸弹客只有自制装置。制造和传播生物和化学武器对奥姆真理教来说是巨大的挑战。作为一个在偏执秘密氛围中运作的小型狂热团体，他们的专业知识和材料获取有限，因此犯了错误。

然而，随着即将到来的浪潮成熟，正如我们所见，破坏的工具将被民主化和商品化。它们将具有更强的能力和适应性，可能以超出人类控制或理解的方式运行，以极快的速度进化和升级，历史上一些最强大的攻击力量将广泛可用。

那些会使用像奥姆这样的新技术的人幸好很少。然而，即使每五十年出现一个奥姆真理教，现在也已经太多，无法避免比地铁袭击严重数量级的事件。邪教、疯子、濒临崩溃的自杀国家，都有动机，现在也有手段。正如一份关于奥姆真理教影响的报告简洁地指出：“ 我们在玩俄罗斯轮盘赌。 ”

历史的新阶段已经到来。随着僵尸政府无法遏制技术，下一次奥姆真理教事件、下一次工业事故、下一次疯狂独裁者的战争、下一次小型实验室泄漏，将产生难以想象的影响。

源文本：—翻译文本：

将所有这些黑暗风险情景视为那些在成长过程中读了太多科幻小说、倾向于灾难主义的人们的遥远白日梦。这很诱人，但却是一个错误。无论我们在 BSL-4 协议、监管提案或关于 AI 对齐问题的技术出版物方面处于何种状态，这些激励措施都在不断消磨，技术也在不断发展和扩散。这不是...

推测小说和 Netflix 系列。这是真实的，正在世界各地的办公室和实验室中进行研究。

然而，风险如此严重，以至于它们需要考虑所有的选项。遏制是关于控制技术的能力。更进一步，这意味着控制其背后的人和社会的能力。随着灾难性影响的展开或其可能性变得不可忽视，辩论的条件将会改变。不仅仅是控制的呼声，还会有镇压的呼声不断增长。前所未有的警惕水平的潜力将变得越来越有吸引力。也许有可能发现并阻止新出现的威胁？那不是最好的吗——正确的做法？

我猜测这将是世界各地政府和民众的反应。当民族国家的统一力量受到威胁，当遏制变得越来越困难，当生命岌岌可危时，必然的反应将是加强对权力的控制。

问题是，代价是什么？

反乌托邦转向

源文本: 翻译文本:

阻止灾难是显而易见的当务之急。灾难越大，风险越大，对策的需求也就越大。如果灾难的威胁变得过于严重，那么各国政府可能会得出结论，认为阻止灾难的唯一方法是严格控制技术的各个方面，确保没有任何东西能突破安全防线，确保没有流氓AI或人工合成病毒能够逃脱、被制造，甚至被研究。

技术已经深深渗透到我们的文明中，以至于观察技术就意味着观察一切。每个实验室、工厂和工厂，每台服务器，每段新代码，每条合成的DNA，每个企业和大学，从森林小屋中的每个生物黑客到每个庞大而匿名的数据中心。为了在即将到来的浪潮的前所未有的动态面前抵御灾难，意味着前所未有的

响应。这不仅意味着观察一切，还意味着在任何时候和任何地方都保留停止和控制它的能力。

有些人不可避免地会这样说：将权力集中到极端程度，建立全景监狱，紧密协调生活的各个方面，以确保不再发生任何大流行病或流氓人工智能。逐渐地，许多国家会说服自己，真正确保这一点的唯一方法是安装我们在上一章中看到的那种全面监控：由强硬力量支持的全面控制。通往反乌托邦的大门被打开了。确实，面对灾难，对于某些人来说，反乌托邦可能会让人感到如释重负。

像这样的建议仍然是边缘性的，尤其是在西方。然而，在我看来，它们的增长只是时间问题。这股浪潮为反乌托邦提供了动机和手段，一个自我强化的“ AI-tocracy ”，不断增加的数据收集和胁迫。如果你怀疑对监视和控制的渴望，想想在COVID大流行期间，几周前甚至难以想象的全社会关闭突然成为不可避免的现实。至少在开始时，面对困境中的政府恳求“ 尽你的一份力 ”，遵从几乎是普遍的。公众对以安全名义采取强有力措施的容忍度似乎很高。

一场大灾难将激起对极端监控装置的呼声，以阻止未来此类事件的发生。如果或当技术出现问题时，镇压会在多长时间内开始？在灾难面前，谁能合理地反对它？监控反乌托邦会在多长时间内扎根，一次一根蔓延，并不断增长？随着小规模技术故障的增加，对控制的呼声也在增加。随着控制的增加，制衡机制被削弱，局势发生变化，为进一步干预铺平道路，通往技术反乌托邦的稳定下行螺旋开始。

在自由与安全之间进行权衡是一个古老的难题。托马斯·霍布斯关于利维坦国家的基础性论述中就存在这一问题。这个问题从未消失。可以肯定的是，这通常是一个复杂且多维的关系，但即将到来的浪潮将把赌注提高到一个新的高度。为了阻止人为制造的流行病，什么程度的社会控制是合适的？

在 *other* 国家中，适当的干预水平是什么？对自由、主权和隐私的后果从未如此可能地痛苦。

我认为，一个透明且精细控制的压制性监控社会，只是另一种失败，是即将到来的浪潮的能力不会导致人类繁荣而是其相反的另一种方式。每一个强制性、偏见和极不公平的应用都将被大大放大。来之不易的权利和自由被回滚。对于许多国家来说，国家自决充其量也只是妥协。这次不是脆弱性，而是彻底的压迫被放大。如果对灾难的回答是这样的反乌托邦，那么这根本就不是答案。

源文本：—翻译文本：

在中国和其他地方建立的监控和强制措施可以说是迈出了第一步。灾难的威胁和安全的承诺将使更多的人参与进来。每一波技术浪潮都带来了对社会秩序的系统性破坏的高可能性。但直到现在，它们还没有引入全球化灾难的广泛和系统性风险。这就是变化所在。这就是可能引发反乌托邦反应的原因。

如果僵尸般的国家将梦游般地走向灾难，它们的开放性和日益增长的混乱是未受控制技术的培养皿，那么威权国家已经欣然冲向这种技术反乌托邦，从技术上而非道德上为大规模侵犯隐私和限制自由奠定了基础。而在两者之间的连续体上，也有可能出现最糟糕的情况：分散但压制性的监控和控制装置，仍然无法形成一个密不透风的系统。

灾难 *and* 反乌托邦。

技术哲学家刘易斯·芒福德谈到了“ 巨型机器 ”，即社会系统与技术结合形成“ 一个统一的、全包围的结构 ”，这种结构“ 为非人格化的集体组织的利益而被控制 ”。以安全的名义，人类

可以释放巨型机器，从字面上阻止其他巨型机器的出现。接下来的浪潮可能会矛盾地创造出遏制自身所需的工具。然而，在这样做的过程中，它将开启一种失败模式，在这种模式下，自主权、自由和隐私被抹去，机器监控和控制系统扩散成扼杀社会的统治形式。

对于那些可能会说这种压抑的图景就是我们现在所处的状态的人，我会说这与未来可能发生的情况相比不算什么。这也不是唯一可能的反乌托邦路径。还有许多其他的，但这一条与浪潮的政治挑战及其灾难性潜力直接相关。这不仅仅是一个模糊的思想实验。面对这一点，我们必须问这些问题：即使其背后的驱动力看起来如此强大和不可动摇，人类是否应该下车？我们是否应该完全拒绝持续的技术发展？或许是时候，尽管不太可能，对技术本身进行暂停吗？

停滞：一种不同类型的灾难

源文本: 翻译文本:

看看我们广阔的城市，那些由钢铁和石头建造的坚固市政建筑，将它们连接在一起的庞大道路和铁路网络，管理其环境的巨大景观和工程作品，我们的社会散发出一种诱人的永恒感。尽管数字世界是无重的，但我们周围的物质世界却是坚实而丰富的。它塑造了我们日常的期望。

我们去超市，期望那里充满新鲜的水果和蔬菜。我们期望夏天保持凉爽，冬天保持温暖。即使在持续动荡的情况下，我们也假设二十一世纪的供应链和便利性像一座古老的市政厅一样稳固。我们生活中所有最极端的历史部分看起来都极其平凡，因此大多数情况下，我们继续过着我们的生活，仿佛它们可以继续下去。

无限期地。我们周围的大多数人，包括我们的领导者，都是这样做的。

然而，没有什么是永恒的。纵观历史，社会崩溃屡见不鲜：从古代美索不达米亚到罗马，从玛雅到复活节岛，一次又一次，不仅仅是文明无法持久；而是不可持续性似乎是内在的。崩溃的文明不是例外，而是常态。对六十个文明的调查表明，它们平均持续约四百年后便会瓦解。没有新技术，它们在可用能源、食物、社会复杂性等方面遇到发展极限，导致崩溃。

除了这一点外，一切都没有改变：数百年来，持续的技术发展似乎使社会能够逃脱历史的铁笼。但认为这种动态已经结束是错误的。21世纪的文明与玛雅文明相去甚远，但庞大而饥饿的上层建筑、大量人口、能源和文明能力的硬性限制并没有神奇地消失；它们只是被暂时遏制住了。

假设有一个世界可以停止这些激励措施。是否是时候全面暂停技术发展？绝对不是。

源文本：—翻译文本：

只有持续的技术发展才能兑现。我们整个大厦的基础是长期经济增长的理念。而长期经济增长最终是以新技术的引入和扩散为基础的。无论是期望以更少的成本消费更多，还是在不增加税收的情况下获得更多的公共服务，或者是认为我们可以在生活不断改善的同时不顾一切地破坏环境，这种交易——可以说是宏大的交易本身——都需要技术。

新技术的发展，如我们所见，是应对地球重大挑战的关键部分。没有新技术，这些

挑战将根本无法应对。人力和物质剥削的现状成本不能被忽视。我们目前的一系列技术在许多方面都很出色，但几乎没有迹象表明它可以可持续地推广，以支持超过八十亿人达到发达国家习以为常的水平。尽管对某些人来说这不太可口，但值得重复：解决气候变化等问题，或维持不断提高的生活和医疗标准，或改善教育和机会，不会在不提供新技术作为方案一部分的情况下发生。

暂停技术发展，假设这是可能的，从某种意义上说会带来安全。首先，它会限制新的灾难性风险的引入。但这并不意味着成功地避免了反乌托邦。相反，随着二十一世纪社会的不可持续性开始显现，它只会带来另一种形式的反乌托邦。没有新技术，迟早一切都会停滞，甚至可能完全崩溃。

在接下来的一个世纪中，全球人口将开始下降，在一些国家甚至会急剧下降。随着工人与退休人员的比例发生变化，劳动力减少，经济将无法在现有水平上运作。换句话说，没有新技术就不可能维持生活水平。

这是一个全球性问题。包括日本、德国、意大利、俄罗斯和韩国在内的国家甚至现在都在接近劳动年龄人口危机。更令人惊讶的是，到2050年代，像印度、印度尼西亚、墨西哥和土耳其这样的国家也将处于类似的境地。中国是未来几十年技术故事的重要组成部分，但到本世纪末，上海社会科学院预测该国可能只有6亿人，这几乎是一个世纪人口增长的惊人逆转。中国的总生育率是世界上最低的之一，仅与韩国和台湾等邻国相当。事实是，没有新技术，中国是完全不可持续的。

这不仅仅是关于数字，还涉及专业知识、税基和投资水平；退休人员将从系统中提取资金，而不是进行长期投资。所有这些意味着“二战后时代的治理模式不仅仅是破产，而是成为社会自杀协议。”人口趋势需要几十年才能转变。世代群体的规模不会改变。这种缓慢而不可逆转的衰退已经锁定，是一个我们无法避免的迫近冰山——除非找到替代这些工人的方法。

我们的资源也面临压力，这是确定无疑的。请记住，为清洁技术采购材料，更不用说其他任何东西，是极其复杂且脆弱的。到2030年，对锂、钴和石墨的需求预计将增长500%。目前，电池是清洁经济的最大希望，然而几乎没有足够的存储容量来让大多数地方度过几分钟甚至几秒钟的能源消耗。为了替换快速减少的库存或解决整个材料供应链失败的问题，我们需要选择。这意味着在材料科学等领域需要新的技术和科学突破。

鉴于人口和资源限制，仅仅保持现状可能需要全球生产力提高两到三倍，而保持现状对于世界上绝大多数人来说是不可接受的，例如，他们的儿童死亡率是发达国家的十二倍。当然，即使是维持在当前水平也不仅仅预示着人口和资源压力；它还加剧了气候紧急情况。

毫无疑问：停滞本身就意味着灾难。

这不仅仅是餐馆劳动力短缺和电池昂贵的问题。这将意味着现代生活中每一个不稳定方面的瓦解，伴随着无数不可预测的下游影响，与一系列已经难以管理的问题交织在一起。我认为，很容易忽视我们生活方式有多少是由持续的技术进步支撑的。那些历史先例——记住，对于每一个先前的文明来说都是常态——正在大声而清晰地呼喊。停滞意味着一个充其量是衰退但可能是令人担忧的螺旋式崩溃的贫乏未来。有些人可能会争辩说这形成了第三种情况。

杆，一个巨大的三难困境。对我来说，这并不完全成立。首先，在这个阶段，这是迄今为止最不可能的选择。其次，如果确实发生了，它只是以一种新的形式重述了这个困境。对技术的暂停并不是一种出路；这是一种邀请，通向另一种反乌托邦，另一种灾难。

即使可能，阻止即将到来的浪潮的想法也不是一个令人安慰的想法。维持，更不用说提高生活水平，需要技术。阻止崩溃需要技术。说不的代价是生存问题。然而，从这里开始的每一条道路都带来了严重的风险和负面影响。

这是一个巨大的困境。

接下来去哪里？

源文本: 翻译文本:

从核时代和数字时代开始，这种困境变得越来越清晰。1955年，在他生命的晚期，数学家约翰·冯·诺依曼写了一篇名为《我们能否在技术中生存？》的文章。预示着这里的论点，他认为全球社会正处于“一个迅速成熟的危机中——这一危机归因于技术进步必须发生的环境已经变得既狭小又缺乏组织。”在文章的结尾，冯·诺依曼将生存仅仅视为“一种可能性”，这在他自己的计算机使蘑菇云成为现实的阴影下是可以理解的。“对于进步是没有治愈方法的，”他写道。“任何试图为当前爆炸性多样化的进步找到自动安全渠道的努力都必然导致挫败。”

我并不孤单，我想要构建一种能够在降低风险的同时收获许多好处的技术。有些人会嘲笑这种雄心是硅谷傲慢的另一种形式，但我仍然相信技术仍然是改善我们世界和生活的主要推动力。尽管有其危害、缺点和意想不到的后果，技术迄今为止的贡献总体上是净积极的。毕竟，即使是技术最严厉的批评者通常也乐于使用一个

水壶，吃一片阿司匹林，看电视，乘坐地铁。每一把枪都有一剂救命的青霉素；每一条错误信息，真相很快就会被揭示。

然而，不知何故，从冯·诺依曼及其同辈开始，我和许多人都对长期轨迹感到焦虑。我深切担忧的是，技术正在展示出明显转向净负面的真实可能性，我们没有解决方案来阻止这种转变，并且我们被困住了，没有出路。

我们谁也无法确定这一切究竟会如何展开。在困境的广泛参数内，存在着一个巨大且无法知晓的具体结果范围。然而，我确信未来几十年将会看到繁荣、监控和灾难威胁之间的复杂而痛苦的权衡变得越来越尖锐。即使是处于最佳状态的国家体系也会艰难应对。

我们面临着 *Homo technologicus.* 的终极挑战

如果这本书在对待技术的态度上显得矛盾，一方面积极，另一方面又充满预感，那是因为这种矛盾的观点是对我们所处位置的最诚实评估。我们的曾祖父母会对我们世界的丰富感到惊讶。但他们也会对其脆弱性和危险感到惊讶。随着即将到来的浪潮，我们面临着真正的威胁，一连串可能灾难性的后果——是的，甚至是对人类物种的生存威胁。技术是我们最好和最坏的一面。没有一种简单的单方面方法能公正对待它。对技术唯一连贯的方法是同时看到两面。

在过去的十年左右，这一困境变得更加明显，解决它的任务更加紧迫。看看这个世界，似乎遏制是不可能的。追随其后果，另一个问题同样显而易见：为了每个人的利益，遏制$v*$是可能的。

PART IV

THROUGH THE WAVE

CHAPTER 13

必须能够进行遏制

分散见解的代价

源文本: 翻译文本:

我

I ONCE INTENDED TO WRITE A book with a rosier picture about the future of technology and the future in general. Although the world is far wiser and warier about “tech” these days, there’s still a huge amount to be positive about. But during the COVID-19 pandemic I had time to stop and reflect. I allowed myself to reconnect with a truth that I have been, if not denying, then downplaying for too long. Exponential change is coming. It is inevitable. That fact needs to be addressed.

如果你接受了本书中心论点的一小部分，真正的问题是关于它实际要做什么。一旦我们承认了这个现实，什么才会真正产生影响？面对我在本书前三部分中概述的困境，遏制措施即使在理论上可能是什么样的？

近年来，我就这个问题进行了无数次对话。我与顶尖的人工智能研究人员、首席执行官、老朋友、华盛顿、北京和布鲁塞尔的政策制定者、科学家和律师、高中学生以及在酒吧愿意听我说话的随机人群讨论过。每个人都立即寻求简单的答案，几乎无一例外地，每个人都有相同的建议：监管。

这里似乎是答案，是摆脱困境的方法，是遏制的关键，是民族国家和我们所知的文明的救世主。巧妙的监管，在国家和超国家层面上平衡进步的需要与合理的安全限制，涵盖从科技巨头和军队到小型大学研究小组和初创企业的一切，结合在一个全面的、可执行的框架中。*We've done it before,* 因此论点是这样的；
look at cars, planes, and medicines. Isn't this
how we manage and contain the coming wave?

要是事情真有那么简单就好了。在面对令人惊叹的技术变革时，说“监管！”是最简单的部分。这也是一种典型的厌恶悲观主义的回答。这是一种简单的方式来推卸问题。在纸面上，监管看起来很诱人，甚至显而易见且直截了当；建议它让人听起来聪明、关心，甚至感到宽慰。未言明的含义是问题是可以解决的，但这是别人的问题。然而，深入观察，裂缝就变得明显了。

在第4部分中，我们将探讨社会可以开始面对这一困境的多种方式，摆脱悲观厌恶，真正解决遏制问题，在一个必须能够解决问题的世界中寻找答案。然而，在此之前，承认一个核心事实是至关重要的：仅靠监管是不够的。召开白宫圆桌会议和发表诚挚的演讲很容易；制定有效的立法则是另一回事。正如我们所见，政府面临着与即将到来的浪潮无关的多重危机——信任下降、根深蒂固的不平等、两极分化的政治，仅举几例。他们已经不堪重负，他们的员工技能不足，未做好应对未来复杂且快速变化挑战的准备。

虽然车库业余爱好者获得了更强大的工具，而科技公司在研发上投入了数十亿美元，大多数政客却被困在24小时的新闻周期中，只关注简短的新闻片段和拍照机会。当一个政府退化到仅仅从一个危机蹒跚到另一个危机的地步时，它几乎没有喘息的空间来应对需要深厚领域专业知识和在不确定时间尺度上进行谨慎判断的构造力量。忽视这些问题要容易得多。

在下次选举中，更可能赢得选票的低垂果实问题。

即使是人工智能等领域的技术专家和研究人员也在努力应对变化的速度。那么，资源更少的监管者有什么机会呢？他们如何解释一个超级进化的时代，如何应对即将到来的浪潮的速度和不可预测性？

技术每周都在发展。起草和通过立法需要数年时间。考虑到像Ring门铃这样的新产品进入市场。Ring在你的前门上安装了一个摄像头，并将其连接到你的手机上。该产品被迅速采用，现在已经如此普及，以至于它从根本上改变了需要监管的性质；突然间，你普通的郊区街道从相对私密的空间变成了被监控和记录的地方。当监管讨论赶上时，Ring已经创建了一个广泛的摄像头网络，收集了来自世界各地人们前门的数据和图像。从社交媒体兴起至今已有二十年，但对于强大新平台的出现没有一致的方法（此外，隐私、极化、垄断、外国所有权或心理健康是核心问题，还是以上所有？）。即将到来的浪潮将加剧这种动态。

关于技术扩散的讨论遍布社交媒体、博客和新闻通讯、学术期刊、无数的会议、研讨会和工作坊，它们的线索遥远且越来越淹没在噪音中。每个人都有自己的看法，但这并没有形成一个连贯的计划。谈论机器学习系统的伦理与例如合成生物的技术安全相去甚远。这些讨论发生在孤立、回声的孤岛中。它们很少突破。

然而，我相信它们是同一现象的不同方面；它们都旨在解决同一波浪的不同方面。仅仅进行几十次关于算法偏见、生物风险、无人机战争、机器人经济影响或量子计算隐私影响的独立对话是不够的。这完全低估了原因和结果的相互关联性。我们需要一种统一的方法。

这些不同的对话，涵盖了风险的所有不同维度，这是这个通用革命的通用概念。

零散见解的代价是失败，而我们知道那是什么样子。现在，零散的见解是我们所拥有的一切：在技术领域的不同部分有数百个独立的项目，在没有总体计划或方向的情况下，逐步推进善意但临时的努力。在最高层面上，我们需要一个明确而简单的目标，一个整合所有技术相关不同努力的旗帜性命令，形成一个连贯的方案。不仅仅是调整这个或那个元素，不仅仅是在这个或那个公司、研究小组甚至国家，而是要在所有前线、风险区和地理区域同时进行。无论是面对新兴的AGI还是一种奇怪但有用的新生命形式，目标必须是统一的：遏制。

二十一世纪人类面临的核心问题是我们如何能够培养足够的合法政治权力和智慧、适当的技术掌握能力以及强有力的规范，以限制技术，确保它们继续带来远多于害的好处。换句话说，我们如何能够控制看似无法控制的事物。

从*Homo technologicus*的历史到技术渗透生活各个方面的时代现实，实现这一目标的几率对我们不利。但是，这并不意味着我们不应该尝试。

然而，大多数组织，不仅仅是政府，都不适合应对复杂的挑战。正如我们所见，即使是富裕国家在面临不断发展的危机时也可能会挣扎。进入2020年，全球健康安全指数将美国排名世界第一，英国在大流行准备方面也紧随其后。然而，一系列灾难性的决策导致的死亡率和财务成本明显比加拿大和德国等同类国家更糟。尽管看起来拥有出色的专业知识、机构深度、计划和资源，但即使是那些在纸面上准备最充分的国家也被打了个措手不及。

从表面上看，各国政府应该比以往任何时候都更好地准备管理新风险和新技术。国家预算通常处于创纪录的水平。然而，事实是，新威胁只是

对于任何政府来说，这都是极其困难的。这并不是政府理念的缺陷，而是对我们面临挑战规模的评估。当他们面临像能够通过我版本的现代图灵测试的ACI这样的事物时，即使是最有思想、最有远见的官僚机构的反应也会类似于对COVID的反应。政府总是在打上一场战争、上一场疫情，监管上一波浪潮。监管者会对他们能够预见的事情进行监管。

这同时是一个充满惊喜的时代。

监管是不够的

源文本: 翻译文本:

尽管面临阻力，监管前沿技术的努力 *are* 必要且日益增长。最具雄心的立法可能是欧盟的《人工智能法案》，该法案于2021年首次提出。截至2023年撰写本文时，该法案正在经历成为欧洲法律的漫长过程。如果该法案通过，人工智能研究和部署将根据风险等级进行分类。具有“不可接受风险”且可能直接造成伤害的技术将被禁止。在人工智能影响基本人权或关键系统（如基础设施、公共交通、健康或福利）的情况下，它将被归类为“高风险”，并受到更高水平的监督和问责。高风险人工智能必须“透明、安全、受人类控制并妥善记录。”

然而，尽管该法案是迄今为止世界上最先进、最有雄心和最有远见的监管尝试之一，但它也展示了监管固有的问题。它受到了各方的攻击，认为它走得太远或不够远。一些人认为它过于关注新兴的、面向未来的风险，试图监管一些甚至不存在的东西；另一些人则认为它不够有远见。一些人认为它让大型科技公司逃脱了责任，认为它们在起草过程中发挥了重要作用，并削弱了其条款。还有人认为它过度扩张，将会抑制欧盟的研究和创新，损害就业和税收收入。

大多数监管在竞争利益之间走钢丝。但在前沿技术以外的少数领域，它必须应对如此广泛传播、对经济如此关键、却又发展如此迅速的事物。所有的噪音和混乱都清楚地表明，任何形式的监管都是多么困难和复杂，尤其是在变化加速的情况下，并且因此几乎肯定会留下空白，未能有效遏制。

监管不仅是超进化的，而且是全用途的通用技术，这极具挑战性。考虑一下机动车运输是如何被监管的。并没有一个单一的监管机构，甚至也不只是几条法律。相反，我们有关于交通、道路、停车、座椅安全带、排放、驾驶员培训等方面的法规。这不仅来自国家立法机构，也来自地方政府、公路管理局、交通部发布的指导、许可机构、环境标准办公室。这不仅依赖于立法者，还依赖于警察部队、交通管理员、汽车公司、机械师、城市规划师和保险公司。

复杂的法规经过数十年的完善，使道路和车辆逐步更安全、更有秩序，从而促进了它们的增长和扩展。然而，每年仍有135万人死于交通事故。法规可能会减轻负面影响，但无法消除诸如碰撞、污染或城市扩张等不良结果。鉴于其带来的好处，我们已经决定这是可以接受的人类代价。这个“我们”至关重要。法规不仅依赖于新法律的通过。它还涉及规范、所有权结构、未成文的合规和诚信准则、仲裁程序、合同执行、监督机制。所有这些都需要整合，公众也需要参与其中。

这需要时间——我们没有时间。随着即将到来的浪潮，我们没有半个世纪的时间让众多机构去弄清楚该做什么，让正确的价值观和最佳实践浮现。先进的监管需要做到正确，并且迅速做到。也不清楚如何在如此广泛的前所未有的技术范围内管理这一切。当你监管合成生物学时，你是在监管食品、药品、工业工具、学术研究，还是同时监管所有这些？哪些机构负责什么？这一切如何结合在一起？哪些参与者对哪些部分负责？

供应链的？即使是一次严重事故的陷阱也是极端的，然而即使决定哪个机构负责也是一个雷区。

源文本：—翻译文本：

在立法辩论的核心，各国也陷入了矛盾。一方面，他们在战略竞争中加速发展人工智能和合成生物学等技术。每个国家都希望处于并被视为技术前沿。这是国家自豪感、国家安全和生存的必要条件。另一方面，他们迫切希望对这些技术进行监管和管理——控制它们，尤其是因为担心它们会威胁到国家作为最终权力中心的地位。可怕的是，这假设了一个最佳情况，即强大、合理胜任、团结一致的（自由民主）国家能够在内部作为一个整体连贯运作，并在国际上协调良好。

为了实现遏制，规则需要在像荷兰和尼加拉瓜、新西兰和尼日利亚这样多样化的地方有效运作。当有人放慢速度时，其他人会迅速前进。每个国家在技术发展中都带来了其独特的法律和文化习俗。欧盟严格限制食品供应中的转基因生物。然而，在美国，转基因生物是农业商业的常规组成部分。表面上看，中国在某种程度上是一个监管领导者。政府发布了多项关于人工智能伦理的法令，试图施加广泛的限制。它积极禁止各种加密货币和去中心化金融（DeFi）计划，并限制未满十八岁的儿童在工作日每天玩游戏和使用社交应用的时间为九十分钟，周末为三小时。中国对推荐算法和LLM的草案监管远远超过了我们在西方所见的任何东西。

中国在某些领域踩下了刹车，同时——正如我们所见——在其他领域加速前进。其监管与无与伦比的技术部署相匹配，作为一种威权主义的工具。

政府权力。与西方国防和政策内部人士交谈，他们坚称，尽管中国在人工智能伦理和限制方面说得头头是道，但在国家安全问题上，并没有实质性的障碍。实际上，中国的人工智能政策有两条轨道：一条是受监管的民用路径，另一条是自由发展的军工路径。

除非监管能够解决第2部分中概述的激励措施的深层性质，否则它不足以遏制技术。它无法阻止有动机的不良行为者或事故。它无法触及开放且不可预测的研究系统的核心。面对巨大的财务回报，它无法提供替代方案。最重要的是，它无法缓解战略必要性。它没有描述国家如何在一个诱人且难以定义的跨国现象上进行协调，建立一个微妙的联盟临界质量，尤其是在国际条约常常失败的背景下。对即将到来的浪潮进行控制的愿望与塑造和拥有它的愿望之间存在着不可逾越的鸿沟，对技术的保护需求与对他人的保护需求之间也存在着不可逾越的鸿沟。优势和控制指向相反的方向。

现实情况是，遏制并不是一个政府，甚至是一组政府可以单独完成的事情。它需要在公共部门和私营部门之间进行创新和大胆的合作，并为所有各方提供一套全新的激励措施。像欧盟人工智能法案这样的法规至少暗示了一个遏制在地图上的世界，在这个世界中，主要政府认真对待扩散的风险，展示出新的承诺水平和愿意做出重大牺牲的意愿。

监管是不够的，但至少是一个开始。大胆的举措。对即将到来的浪潮中所涉及的利害关系有真正的理解。在一个遏制似乎不可能的世界中，所有这些都指向一个可能的未来。

遏制政策再探：新的重大交易

源文本: 翻译文本:

是否有任何实体能够在捕捉即将到来的浪潮所带来的巨大力量和利益的同时，防止大规模扩散？是否可以阻止不良行为者获取某项技术，或塑造围绕其的新生思想的传播？随着自主性的增加，是否真的有人或事物能够在宏观层面上实现有意义的控制？遏制意味着对这些问题的回答是肯定的。理论上，受控技术可以让我们摆脱困境。这意味着同时驾驭和控制浪潮，这是一种构建可持续和繁荣社会的重要工具，同时以避免严重灾难的方式加以控制，但又不至于过于侵入性以至于引发反乌托邦。这意味着制定一种新型的大交易。

在本书的前面部分，我将遏制描述为控制和管理技术的基础，涵盖技术、文化和监管方面。我认为，这意味着拥有大幅减少或完全阻止技术负面影响的能力，从地方和小规模到全球和生存层面。它包括对滥用扩散技术的严格执行，也引导新兴技术的发展、方向和治理。被遏制的技术是其失败模式已知、可管理和缓解的技术，在这种情况下，塑造和管理技术的手段与其能力同步升级。

人们很容易从显而易见的、字面意义上去理解遏制，就像一个可以将特定技术封存起来的魔法盒子。在极端情况下——例如流氓恶意软件或病原体——可能需要采取如此激烈的措施。然而，一般来说，可以将遏制视为一套护栏，当某项技术有可能弊大于利时，这是一种让人类保持主导地位的方法。想象这些护栏在不同层次上以不同的方式实施。在下一章中，我们将探讨它们在更细化层面上的表现，从人工智能对齐研究到实验室设计，从国际条约到最佳实践协议。目前，关键点在于这些护栏需要足够强大，以至于理论上它们可以阻止失控的灾难。

遏制需要根据技术的性质作出反应，并将其引导到更易于控制的方向。回想一下四个特征

即将到来的浪潮：不对称性、超进化、全能使用和自主性。每个特征都必须通过可控性的视角来审视。在制定策略之前，值得提出以下类型的问题以激发有前景的途径：

- *Is the technology omni-use and general-purpose or specific?* 核武器是一种具有单一目的的高度特定技术，而计算机本质上是多用途的。潜在的使用案例越多，就越难以控制。因此，应鼓励那些范围更窄且特定领域的系统，而不是通用系统。
- *Is the tech moving away from atoms toward bits?* 一种技术越是非物质化，就越容易受到难以控制的超进化效应的影响。材料设计或药物开发等领域将迅速加速，使得进展速度更难以追踪。

- *Are price and complexity coming down, and if so how fast?* 战斗机的价格并没有像晶体管或消费类硬件的价格那样下降。尽管战斗机具有明显的破坏潜力，但源于基础计算的威胁比战斗机的威胁更为广泛。

- *Are there viable alternatives ready to go?* CFCs可能被部分禁止是因为有更便宜和更安全的制冷替代品。有哪些替代品可用？安全替代品越多，淘汰使用就越容易。

- *Does the technology enable asymmetric impact?* 想象一下无人机群对抗传统军事力量，或者微小的计算机或生物病毒破坏重要的社会系统。某些技术带来的意外和利用漏洞的风险更大。

- *Does it have autonomous characteristics?* 是否有自学或在没有监督的情况下操作的可能性？想想基因驱动、病毒、恶意软件，当然还有机器人。技术越是

从设计上需要人为干预，失去控制的可能性就越小。

- *Does it confer outsized geopolitical strategic advantage?* 例如，化学武器的优势有限且有很多缺点，而在人工智能或生物领域领先则有巨大的好处，无论是经济上还是军事上。因此，说不更难。

- *Does it favor offense or defense?* 在第二次世界大战中，像V-2这样的导弹的发展帮助了进攻行动。但雷达这样的技术加强了防御。将发展方向从进攻转向防御往往倾向于遏制。
- *Are there resource or engineering constraints on its invention, development, and deployment?* 硅芯片需要专门的和高度集中的材料、机器和知识。就全球而言，合成生物学初创公司的可用人才仍然相当少。两者在短期内都有助于遏制。

例如，额外的摩擦使事物保持在原子这个有形的世界中，或者使事物变得昂贵，或者如果更安全的替代方案容易获得，那么更有可能进行遏制，因为更容易减缓技术的发展，限制访问，或完全放弃它们。具体技术比全用途技术更容易监管，但监管全用途技术更为重要。同样，进攻性行动或自主性的潜力越大，对遏制的要求就越高。如果你能让价格和获取的便利性对许多人来说遥不可及，扩散就变得更加困难。提出这样的问题，遏制的整体愿景开始显现。

我在这个问题上工作了将近十五年。在这段时间里，我感受到了书中所描述的那些激励措施的巨大力量，以及即使困境的轮廓变得越来越清晰，仍然迫切需要答案的紧迫性。然而，即便是我，也对技术在短短几年内所实现的可能性感到震惊。我一直在努力理解这些想法，眼看着发展的速度不断加快。

现实情况是，我们过去常常没有控制或限制技术。如果我们现在想这样做，就需要一些全新的东西，一个全面的安全、伦理、监管和控制计划，这甚至还没有名字，而且一开始似乎也不可能。

这种困境应该是一个紧迫的行动号召。但多年来，显而易见的是，大多数人觉得这很难接受。我完全理解。初次接触时几乎感觉不真实。在所有关于人工智能和监管的众多讨论中，我感到震惊的是，与一系列现有或迫在眉睫的挑战相比，很难准确传达为什么这本书中的风险需要被认真对待，为什么它们不仅仅是几乎无关紧要的尾部风险或科幻小说的领域。

即使开始进行这场对话的一个挑战是，在大众想象中，技术已经与一系列狭窄且常常多余的应用联系在一起。“ 技术 ”现在大多意味着社交媒体平台和可穿戴设备，用来测量我们的步数和心率。人们很容易忘记，技术还包括对养活地球至关重要的灌溉系统和新生儿生命支持机器。技术不仅仅是存储自拍照的一种方式；它代表了通向世界积累的文化和智慧的途径。技术不是一个小众领域；它是一个主导人类存在的超物体。

这里一个有用的比较是气候变化。它同样涉及风险，这些风险往往是分散的、不确定的、时间上遥远的、发生在别处的，缺乏在稀树草原上遭遇伏击时的显著性、肾上腺素和紧迫感——这种风险是我们准备好应对的。从心理上讲，这些都不觉得是当下的。我们的史前大脑通常在处理这种模糊的威胁时无能为力。

然而，在过去的十年左右，气候变化的挑战变得更加清晰。尽管世界仍在排放越来越多的CO_2，但各地的科学家都能测量大气中的CO_2百万分之一(ppm)。早在20世纪70年代，全球大气中的碳含量约为300多ppm。到2022年，这一数值已达到420 ppm。无论是在北京、柏林还是布隆迪，无论是石油巨头还是家庭农场，每个人都能客观地看到气候正在发生什么。数据带来了清晰。

当影响如此显而易见地可以量化时，厌恶悲观主义要困难得多。像气候变化一样，技术风险只能在全球范围内解决，但没有同等的清晰度。没有方便的风险指标，没有在国家首都、董事会和公众情绪中共享的客观威胁单位，没有用于测量技术可能做什么或在哪里的百万分比。没有我们可以逐年检查的普遍同意或明显的标准。没有前沿科学家和技术人员之间的共识。没有阻止它的民众运动，没有融化的冰山、被困的北极熊或被淹没的村庄的生动图像来提高意识。发表在arXiv上的晦涩研究、在小众Substack博客中或在枯燥的智库白皮书中几乎不起作用。

我们如何在相互竞争的议程中找到共同点？中国和美国在限制人工智能发展方面没有共享愿景；Meta不会认同社交媒体是问题的一部分；人工智能研究人员和病毒学家认为他们的工作不是造成灾难的关键部分，而是理解和避免灾难的关键部分。从表面上看，“技术”并不像地球变暖那样是一个问题。

然而，这可能是。

第一步是识别。我们需要冷静地承认浪潮即将来临，而困境在没有剧烈改变方向的情况下是不可避免的。我们要么可以应对由我们持续的开放和盲目的追逐引发的大量好坏结果，要么可以面对由于我们试图限制强大技术的扩散而产生的反乌托邦和专制风险，这些风险还固有于对这些相同技术的集中所有权。

选择你的毒药。最终，这种平衡必须与每个人协商达成。它越受到公众关注，就越好。如果这本书引发批评、争论、提议和反提议，那就越多越好。

不会有某个地方的地下室里一群聪明人提供的单一、神奇的解决方案。恰恰相反。当前的精英们如此投入于他们对悲观主义的厌恶，以至于他们害怕诚实地面对我们所面临的危险。他们乐于在私下发表意见和辩论，却不太愿意公开谈论。他们习惯于一个控制和秩序的世界：CEO对公司的控制，中央银行家对利率的控制，官僚对军事采购的控制，或城市规划者对修补哪个坑洼的控制。他们的控制杠杆是不完美的，当然，但它们是已知的、经过尝试和测试的，并且通常有效。在这里则不然。

这是一个独特的时刻。即将到来的波浪 *really is* 正在逼近，但它尚未席卷而来。虽然不可阻挡的激励措施已被锁定，但波浪的最终形式、困境的精确轮廓仍有待决定。让我们不要浪费几十年等待结果。让我们今天就开始管理它。

在下一章中，我概述了十个重点领域。这不是一张完整的地图，更不是一套最终答案，但却是必要的基础工作。我的意图是播种想法，希望能采取关键的第一步*toward*遏制。这些想法的统一之处在于，它们都与边际收益有关，即通过小努力的缓慢而持续的积累来提高良好结果的概率。它们涉及创建一个不同的背景来构建和部署技术：寻找拖延时间的方法，放慢速度，为更多的答案工作留出空间，引起关注，建立联盟，推进技术工作。

我相信，在我们 *current* 的世界中，遏制即将到来的浪潮是不可能的。然而，这些步骤可能会改变潜在的条件。推动现状向前发展，使遏制有机会。我们应该在知道可能会失败的情况下做这一切，但这是我们建立一个遏制和人类繁荣成为可能的世界的最佳机会。

这里没有任何保证，没有魔术般的奇迹。任何希望快速解决问题、得到聪明答案的人都会失望。面对这个困境，我们仍然处于一如既往的人类境地：全力以赴并希望事情能顺利解决。以下是我认为它可能——只是可能——结合在一起的方式。

CHAPTER 14

十步遏制措施

T

HINK OF THE TEN IDEAS presented here as concentric circles. We start small and direct, close to the technology, focusing on specific mechanisms for imposing constraints by design. From there each idea gets progressively broader, ascending a ladder of interventions further away from the hard technical specifics, the raw code and materials, and moving up and out toward the nontechnical but no less important actions, the kinds that add up to new business incentives, reformed government, international treaties, a healthier technological culture, and a popular global movement.

正是这些洋葱层层叠加的方式使它们变得强大；单独来看，每一层都是不够的。每一层都需要非常不同类型的干预措施，具备不同的技能、能力和人员，每一层通常都是其自身庞大而专业的子领域。我相信，综合起来，它们可能会形成一种有效的东西。

让我们从开始，先了解一下这项技术本身。

1. 安全：技术安全的阿波罗计划

源文本: 翻译文本:

几年前，许多大型语言模型存在一个问题，坦率地说，它们是种族主义的。用户可以很容易地找到方法让它们重复种族主义材料，或者持有它们在扫描其训练的大量文本语料库时获取的种族主义观点。有毒偏见

似乎，这种情况已经深深植根于人类写作中，然后被人工智能放大。这导致许多人得出结论，整个设置在伦理上是有问题的，在道德上是不可行的；鉴于明显的危害，没有办法能够很好地控制LLM，以至于可以向公众发布。

但是，正如我们所见，LLM迅速崛起。到2023年，与早期系统相比，让像ChatGPT这样的模型发表种族主义言论变得极其困难。这个问题解决了吗？绝对没有。仍然有多个例子显示偏见，甚至是公然的种族主义的LLM，以及从不准确信息到煤气灯效应等严重问题。但是，对于那些从一开始就在该领域工作的人来说，消除不良输出的指数级进步是令人难以置信的，不可否认的。我们很容易忽视我们已经走了多远、多快。

这一进展背后的关键驱动力被称为来自人类反馈的强化学习。为了修正他们容易产生偏见的大型语言模型，研究人员与模型设置了巧妙构建的多轮对话，促使其说出令人反感、有害或冒犯的话，观察其出错的地方和方式。标记这些失误后，研究人员将这些人类洞察重新整合到模型中，最终以一种与我们试图教导孩子在餐桌上不要说不当言论的方式不完全不同的方式，教会模型一种更理想的世界观。随着工程师们越来越意识到其系统固有的伦理问题，他们也变得更加开放于寻找技术创新来帮助解决这些问题。

解决 LLMs 中的种族主义和偏见是一个例子，说明谨慎和负责任的部署对于提高这些模型的安全性是必要的。与现实接触有助于开发者学习、纠正和改进其安全性。

虽然仅靠技术修复无法解决由人工智能引发的社会和伦理问题是错误的，但这确实表明它们将成为其中的一部分。技术安全，在代码中，在实验室中，是任何遏制议程上的首要事项。

源文本：—翻译文本：

“ 遏制 ”这个词，假设你不是国际关系学者，你很可能想到的是物理意义上的控制某物。可以肯定的是，物理上遏制技术是很重要的。我们已经看到，例如，即使是 BSL-4 实验室也可能会泄漏。什么样的环境可以使这种情况完全不可能发生？BSL-7 或 *-n* 是什么样子的？

虽然我在上一章中争辩说，遏制不应被简化为一种魔法盒，但这并不意味着我们不想找出构建它的方法。最终的控制是对服务器、微生物、无人机、机器人和算法的硬物理控制。“ 封箱 ”AI是技术遏制的原始和基本形式。这将涉及没有互联网连接、有限的人类接触、一个小而受限的外部接口。它将字面上把它包含在一个具有明确位置的物理盒子中。这样的系统——称为气隙——理论上可以阻止AI与更广泛的世界接触或以某种方式“ 逃脱 ”。

物理隔离只是将技术安全架构转变以应对下一波挑战的一个方面。利用现有的最佳技术是一个开始。例如，核能由于切尔诺贝利和福岛等众所周知的灾难而声名狼藉。但实际上，它非常安全。国际原子能机构已发布了超过一百份安全报告，解决特定情况下的技术标准问题，从放射性废物的分类到应急准备。电气和电子工程师学会等机构维护着超过两千项技术安全标准，涵盖从自主机器人开发到机器学习的技术。生物技术和制药行业在安全标准方面的操作远远超过大多数软件企业已有数十年。值得记住的是，多年的努力使许多现有技术变得非常安全——并在此基础上继续发展。

前沿AI安全研究仍然是一个未开发的新兴领域，专注于防止越来越多的自主系统超越我们理解或控制它们的能力。我认为这些关于控制或价值对齐的问题是更广泛的约束问题的子集。虽然

数十亿美元被投入到机器人、生物技术和人工智能领域，相比之下，用于建立技术安全框架以确保其功能受控的资金却微乎其微。例如，生物武器的主要监管机构《生物武器公约》的预算仅为$1.4百万，只有四名全职员工——比普通麦当劳的员工还少。

AI安全研究人员的数量仍然很少：从2021年全球顶尖实验室的约一百人增加到2022年的三四百人。鉴于今天大约有三到四万名AI研究人员（以及同样数量能够拼凑DNA的人），这个数字小得惊人。即使进行十倍的招聘狂潮——考虑到人才瓶颈，这不太可能——也无法应对挑战的规模。与可能出错的严重性相比，AI的安全和伦理研究是微不足道的。只有少数机构由于资源挑战而认真对待技术安全问题。然而，今天做出的安全决策将改变技术和人类的未来发展方向。

这里有一个明确的必须做的事情：鼓励、激励并直接资助这一领域的更多工作。是时候开展一项关于AI安全和生物安全的阿波罗计划了。应该有成千上万的人参与其中。具体来说，一个好的立法提案是要求前沿企业研究和开发预算的固定部分——比如至少20%——应该用于安全工作，并有义务向政府工作组发布重要发现，以便跟踪和共享进展。最初的阿波罗任务虽然昂贵且繁重，但它们展示了正确的巨大雄心壮志，并且在面对艰难险阻时的实干态度催生了从半导体和软件到石英钟和太阳能电池板的技术发展。这可能对安全产生类似的影响。

尽管目前数量很少，但根据我的经验，我知道围绕这些问题的兴趣正在迅速增长。我遇到的学生和其他年轻人对人工智能对齐和大流行病准备等问题充满热情。与他们交谈，很明显智力挑战对他们有吸引力，但他们也被道德责任所吸引。他们想要帮助，

并且感到有责任做得更好。我相信，如果有工作和研究项目，人才就会随之而来。

对于未来的技术安全专家来说，有许多有前景的方向可以探索。例如，通过使用低波长灯泡来杀死病毒，可以大大增强大流行病的准备工作。这些灯泡发出的光波长在200到230纳米之间，接近紫外光谱，可以杀死病毒，同时不穿透皮肤的外层：这是对抗大流行病和更广泛传播疾病的强大武器。如果COVID-19大流行教会了我们一件事，那就是在新疫苗的研究、推广和监管方面采取综合、加速的方法的价值。

在人工智能中，技术安全还意味着沙盒和安全模拟，以创建可证明安全的隔离区，从而在高级人工智能被允许接触现实世界之前对其进行严格测试。这意味着要在不确定性上投入更多的工作，这是目前的一个主要关注点——即，当人工智能可能出错时，它如何进行交流？LLM的一个问题是它们仍然受到*the hallucination problem,*的影响，常常自信地将严重错误的信息声称为准确。这尤其危险，因为它们常常达到*are right,*的专家水平。作为用户，很容易被一种虚假的安全感所麻痹，认为系统输出的任何信息都是真实的。

例如，在 Inflection，我们正在寻找方法来鼓励我们的 AI，称为 Pi（个人智能），默认情况下保持谨慎和不确定，并鼓励用户保持批判态度。我们正在设计 Pi，以表达自我怀疑，频繁且建设性地征求反馈，并迅速让步，假设人类而非机器是正确的。我们和其他人也在研究一项重要的研究方向，旨在使用我们知道可信的第三方知识库来核实 AI 的陈述。在这里，确保 AI 输出提供引用、来源和可查询的证据，以便用户在出现可疑声明时可以进一步调查。

解释是另一个巨大的技术安全前沿。请记住，目前没有人能准确解释为什么一个模型会产生它的输出。

设计模型能够全面解释其决策或使其接受审查的方法已成为安全研究人员面临的一个关键技术难题。对于这项研究来说，仍处于早期阶段，但有一些令人鼓舞的迹象表明，AI模型可能能够为其输出提供理由，尽管尚未能为其提供因果推理，尽管这些理由的可靠性仍不明确。

在使用简化架构探索更复杂架构方面也有很好的工作，甚至在自动化对齐研究过程本身：构建 AI 来帮助我们控制 AI。研究人员正在开发一代“ 批评 AI ”，可以监控并对其他 AI 的输出进行反馈，目的是以人类无法匹敌的速度和规模来改进它们——这是我们在即将到来的浪潮中看到的速度和规模。管理强大的工具本身需要强大的工具。

计算机科学家斯图尔特·拉塞尔建议使用我们在Inflection中探索的内置系统性怀疑来创建他称之为“ 可证明有益的AI ”。他建议系统谨慎地推断我们的偏好和目标，而不是给AI一套包含在所谓书面宪法中的固定外部目标。它们应该仔细观察和学习。理论上，这应该在系统中留出更多的怀疑空间，并避免产生不良结果。

许多关键挑战仍然存在：如何在一个可能有能力覆盖其自身指令的强大AI系统中构建安全价值观？AI如何从人类中推断出这些价值观？另一个持续的问题是如何破解“ 可修正性 ”问题，确保始终可以访问和纠正系统。如果你认为这一切听起来像是先进AI的基本必备安全特性，那你是对的。这里的进展需要跟上。

我们还应该在开发和生产过程中建立稳健的技术约束。想想所有现代复印机和打印机是如何内置技术来防止你复制或打印钞票的，有些甚至在你尝试时会关闭。例如，对用于创建模型的训练计算量进行资源限制，可以限制进展速度（至少在那个维度上）。性能可能会

限制模型只能在某些严格控制的硬件上运行。可以通过加密保护构建AI系统，确保模型权重——系统中最有价值的知识产权——只能在有限次数或特定情况下复制。

最高级别的挑战，无论是在合成生物学、机器人技术还是人工智能领域，都是建立一个万无一失的关闭开关，一种关闭任何可能失控的技术的方法。确保任何自主或强大系统中始终有一个关闭开关是显而易见的常识。如何在即将到来的浪潮中实现这一点——这些技术是分布式的、多变的、影响深远的，其确切形式尚不明确，有些情况下可能会主动抵制——这是一个悬而未决的问题。这是一个巨大的挑战。我认为这可能吗？是的——但没有人应该低估其难度的规模。

过多的安全工作是渐进的，专注于狭隘的影响评估、小的技术问题，或是在发布后解决突然出现的问题，而不是提前处理基础性问题。相反，我们应该及早识别问题，然后在基础方面投入更多的时间和资源。要有大局观。创建通用标准。安全特性不应是事后的考虑，而应是所有这些新技术的固有设计属性，是未来一切的基础状态。尽管面临激烈的挑战，我对这里的各种创意和创造力感到真正兴奋。让我们给予它们智力上的支持和物质上的支持，以取得成功，同时认识到工程虽然从来不是全部答案，但它是其中的一个基本部分。

2. 审计：知识就是力量；力量就是控制

源文本: 翻译文本:

审计听起来很无聊。也许是必要的——但非常沉闷。然而，它们对控制至关重要。创建安全的物理和虚拟容器——我们刚刚看到的那种工作——是基础性的。但仅靠这些是不够的。实际上拥有有意义的监督和可执行的规则并进行审查。

技术实施至关重要。如果无法验证技术安全进步和法规是否按预期工作，它们将难以有效。你如何确定实际发生了什么，并检查你是否在控制之中？这是一个巨大的技术和社会挑战。

信任来自透明。我们绝对需要能够在每个层面验证系统的安全性、完整性或未受损性。这反过来又涉及访问权限和审计能力，涉及对系统进行对抗性测试，拥有白帽黑客团队甚至是AI来探测弱点、缺陷和偏见。这是关于以一种完全不同的方式构建技术，使用尚不存在的工具和技术。

外部审查是必不可少的。目前，没有全球性的、正式的或常规的努力来测试已部署的系统。没有技术风险的预警装置，也没有统一或严格的方法来了解它们是否遵守法规，甚至是否遵循共同商定的基准。既没有相关机构，也没有标准化的评估，也没有必要的工具。因此，作为一个起点，让在前沿领域工作的公司和研究人员，在存在实际危害风险的地方，主动与政府主导的审计中值得信赖的专家合作，是基本的常识。如果有这样的机构存在，我会很乐意在Inflection与其合作。

几年前，我共同创立了一个跨行业和民间社会组织，名为人工智能合作伙伴关系，以帮助开展这类工作。我们在所有主要科技公司的支持下启动了它，包括DeepMind、Google、Facebook、Apple、Microsoft、IBM和OpenAI，以及众多专家民间社会团体，包括ACLU、EFF、乐施会、联合国开发计划署和其他二十个组织。不久之后，它启动了一个AI事件数据库，旨在机密报告安全事件，以便与其他开发者分享经验。现在它已经收集了超过一千二百份报告。与来自非营利组织、学术界和媒体团体的超过一百个合作伙伴一起，该合作伙伴关系为跨学科讨论和合作提供了关键的、中立的窗口。还有更多这样的组织和内部审计项目的空间。

另一个有趣的例子是“ 红队演练 ”——即主动寻找AI模型或软件系统中的缺陷。这意味着以可控的方式攻击你的系统，以探测其弱点和其他故障模式。今天暴露出来的问题在未来可能会被放大，因此理解它们可以在系统变得更强大时内置保护措施。越是公开和集体地进行这项工作，效果就越好，使所有开发者能够相互学习。同样，所有大型科技公司都应该积极在此合作，快速分享关于新风险的见解，就像网络安全行业长期以来分享新零日攻击的知识一样。

现在也是时候创建政府资助的红队，严格攻击和压力测试每个系统，确保在此过程中发现的见解在整个行业广泛共享。最终，这项工作可以扩展和自动化，公开授权的AI系统专门设计用于审计和发现其他系统中的问题，同时也允许自身被审计。

为了跟踪新技术而实施的系统需要识别异常、能力的意外跃升、隐藏的故障模式。它们必须发现看似合法但隐藏不受欢迎惊喜的特洛伊攻击。为此，它们将不得不监控大量的指标，而不陷入全景监狱的诱人陷阱。密切关注用于训练模型的重要数据集，特别是开源数据集、研究中的文献计量学和公开可用的有害事件，将是一个富有成效且无侵入性的起点。允许他人使用基础AI服务的API不应盲目开放，而应像银行业的某些部分一样，附带“ 了解你的客户 ”检查。

在技术方面，有针对性监督机制的空间，一些研究人员称之为“ 可扩展监督 ”，用于“ 在大多数与当前任务相关的技能上可能超越我们 ”的系统。这个提议是关于数学上验证算法的无害性，要求模型提供严格的证明，以表明其行为或输出受到明显限制。本质上，这是对记录的保证。

围绕能力的活动和限制是内置的。以这种方式验证和确认模型的行为可能为指导和跟踪系统提供一种客观、正式的方法。

另一个有前景的新监督机制的例子是SecureDNA，这是一个由科学家和安全专家组成的小组发起的非营利项目。目前，只有一小部分合成DNA会筛查潜在的危险元素，但像SecureDNA这样的全球努力，将每个合成器——无论是在家中的台式机还是大型的远程设备——连接到一个集中、安全且加密的系统中，以便扫描病原序列，这是一个很好的开始。如果人们正在打印潜在有害的序列，它们会被标记。基于云的、免费的、加密安全的，它实时更新。

筛查所有DNA合成将是一个主要的生物风险降低措施，并且在我看来，这不会过度限制公民自由。这在长期内不会阻止黑市的存在，但制造不合规的合成器或入侵现有系统会引入一个不小的障碍。对DNA合成或AI模型的数据输入进行预审查将在系统部署之前进行审计，从而降低风险。

目前，对新技术的出现或其被敌对国家和其他行为者滥用的监控方法在全球范围内各不相同。这是一幅不均衡的图景：通常是由不透明的开源信息、学术研究以及在某些情况下的秘密监控混合而成。这是一个法律和政治的雷区，入侵的门槛非常混杂，最糟糕的情况下是故意模糊不清的。我们可以做得更好。透明性不能是可选的。必须有一个明确的法律途径来检查任何新技术，无论是在引擎盖下、代码中、实验室里、工厂中，还是在野外。

大部分工作应在与技术生产商合作的基础上自愿进行。如果无法通过这种方式完成，立法必须强制合作。如果这仍然不起作用，可以考虑其他方法，例如开发技术保障措施——在某些情况下包括加密的后门——以提供可验证的

由司法机关或同等的公共认可的独立机构控制的进入系统。

如果执法部门或监管机构提出访问任何公共或私人系统的请求，这将根据案件的优劣进行决定。同样，记录任何模型、系统或知识的复制或共享的加密账本将有助于跟踪其传播和使用。将这样的社会和技术遏制机制结合起来是至关重要的。细节需要新的研究和公众讨论。我们需要找到一种新的、安全的、难以滥用的平衡，以在即将到来的浪潮中实现监控与安全的协调。

法律、条约和出色的技术解决方案都很好。但它们仍然需要对齐和检查，并且在不诉诸严厉控制手段的情况下进行。构建像这些倡议这样的技术远非无聊；这是二十一世纪最具激励性的技术和社会挑战之一。建立技术安全特性和审计措施至关重要，但这需要我们所没有的东西。时间。

3. 窒息点：争取时间

源文本: 翻译文本:

习近平很担忧。“ 我们在一些关键设备、组件和原材料上依赖进口，”这位中国国家主席在2020年9月告诉该国的一群科学家。令人不安的是，他认为对中国的未来和地缘政治安全至关重要的“ 关键和核心技术 ”被“ 他人控制 ”。事实上，中国在进口芯片上的花费超过了石油。中国领导层很少在公开场合表现出不安，但在将其长期战略寄托于主导即将到来的浪潮时，它承认了一个严重的脆弱性。

几年前，一家政府运营的报纸用一个更形象的比喻来描述同样的问题：据称，中国的技术受限于一系列“ 卡脖子 ”点。如果有人施压这些卡脖子点，那么，含义就很明确了。

习近平的担忧在2022年10月7日成为现实。美国对中国宣战，攻击了其中一个关键点。这并不涉及导弹飞越台湾海峡。没有对南海的海上封锁，也没有海军陆战队突袭福建海岸。相反，它来自一个不太可能的来源：商务部。发出的“ 炮火 ”是对先进半导体的出口管制，这些芯片支持计算和人工智能。

新的出口管制使美国公司向中国出售高性能计算芯片以及任何公司分享制造这些芯片的工具或提供修复现有芯片的技术诀窍变得非法。最先进的半导体（通常涉及十四纳米以下的工艺，即十四十亿分之一米，代表少至二十个原子的距离）——包括用于人工智能和超级计算等领域的知识产权、制造设备、零件、设计、软件、服务——现在都需要严格的许可。像NVIDIA和AMD这样的美国领先芯片公司不再能够向中国客户提供生产世界上最先进芯片的手段和技术。与中国公司合作的美国公民面临选择：保住工作并失去美国国籍，或立即辞职。

这是一记晴天霹雳，旨在摧毁中国对二十一世纪最重要技术基石的掌控。这不仅仅是一个晦涩的贸易争端。这一声明在中南海——中国领导层的驻地——响起了强烈的警报，正值共产党大会有效地将习近平确立为终身统治者之际。一位不愿透露姓名的技术高管概述了此举的范围：“ 他们不仅仅是针对军事应用，他们正试图通过任何手段阻止中国技术实力的发展。 ”

从短期到中期，普遍的共识是这将带来痛苦。建设这一基础设施的挑战是巨大的，尤其是在生产世界上最先进芯片的复杂机器和技术方面，而中国在这一领域落后。然而，从长远来看，这可能不会阻止它。相反，它正在推动一个困难且极其

昂贵但仍然可行的提高国内半导体产能的途径。如果需要数千亿美元（而且确实需要），他们会花这笔钱。

中国公司已经在寻找绕过管制的方法，利用空壳公司和前置公司网络以及第三方国家的云计算服务。美国最先进的AI芯片制造商NVIDIA最近对其最先进的芯片进行了事后调整，以规避制裁。然而，这向我们展示了一些重要的东西：至少有一个不可否认的杠杆。浪潮可以被减缓，至少在某些时间段和某些领域。

在超进化时代，争取时间是无价的。时间用于开发更进一步的遏制策略。时间用于建立额外的安全措施。时间用于测试那个关闭开关。时间用于构建改进的防御技术。时间用于巩固国家，进行更好的监管，甚至只是通过那项法案。时间用于编织国际联盟。

目前，技术的发展更多是由激励的力量推动，而不是遏制的速度。像美国的半导体策略这样的出口管制对大国竞争、军备竞赛和未来都有各种不确定的影响，但几乎所有人都同意一点：这至少会减缓中国的一些技术发展，进而影响到全世界。

最近的历史表明，尽管技术在全球范围内广泛传播，但它依赖于几个关键的研发和商业化中心：瓶颈点。考虑这些显著集中的点：比如，Xerox 和 Apple 在界面方面，或者 DARPA 和 MIT，或者 Genentech、Monsanto、Stanford 和 UCSF 在基因工程方面。令人惊讶的是，这种遗产正在缓慢消失。

在人工智能领域，绝大多数最先进的GPU是由一家美国公司NVIDIA设计的。其大部分芯片由一家位于台湾的公司台积电（TSMC）制造，其中最先进的芯片仅在一栋建筑中生产，这座工厂是世界上最复杂和最昂贵的工厂。台积电用于制造这些芯片的设备来自一个供应商，即荷兰公司ASML，这是迄今为止欧洲最有价值和最重要的科技公司。ASML的机器使用一种被称为

极紫外光刻技术能够以惊人的原子精度生产芯片，是历史上最复杂的制造产品之一。这三家公司在尖端芯片领域占据主导地位，这项技术在物理上受到极大限制，据估计其成本高达每公斤 $10 billion。

芯片并不是唯一的瓶颈。工业级云计算同样被六大公司所主导。目前，AGI实际上是由少数资源丰富的团体追求的，最著名的是DeepMind和OpenAI。全球数据流量通过数量有限的光纤电缆传输，这些电缆集中在关键的瓶颈点（例如英格兰西南海岸或新加坡）。稀土元素钴、铌和钨的短缺可能会颠覆整个行业。大约80%的高质量石英（对光伏面板和硅芯片等至关重要）来自北卡罗来纳州的一个矿山。DNA合成器和量子计算机并不是常见的消费品。技能也是一个瓶颈：从事本书讨论的所有前沿技术的人数可能不超过150,000人。

因此，随着负面影响变得明显，我们必须利用这些瓶颈来创建合理的速率限制因素，检查开发速度，以更好地确保在科学发展的同时实施良好的判断。在实践中，限制措施不应仅适用于中国；它们可以广泛应用于调节开发或推出的速度。出口管制不仅是地缘战略的举措，更是一个活生生的实验，是一种可能的地图，展示如何控制技术而不完全扼杀它。最终，所有这些技术将被广泛传播。在此之前，未来五年左右的时间至关重要，这是一个紧迫的窗口期，在此期间某些压力点仍然可以减缓技术的发展。既然有这个选择，让我们抓住它，争取时间。

4. 创造者：批评者应该构建它

源文本: 翻译文本:

技术激励不可阻挡这一事实并不意味着那些构建技术的人对其创造物不负责任。相反，他们，我们，我，确实负有责任；责任是显而易见的。没有人被迫进行基因改造实验或构建大型语言模型。技术的不可避免的传播和发展并不是一张免罪金牌，也不是一张随心所欲构建并观察结果的许可证。相反，它们是对做好事情的必要性以及不这样做的可怕后果的强烈提醒。

比任何人都多，从事技术工作的人需要积极努力解决本书中描述的问题。举证责任和解决方案的责任在于他们，在于我们。人们经常问我，鉴于这一切，为什么要从事人工智能工作并建立人工智能公司和工具？除了他们可以做出的巨大积极贡献之外，我的回答是，我不仅仅想谈论和讨论遏制。我想积极主动地帮助实现它，走在技术发展的前面。遏制需要技术人员全心全意地专注于将其变为现实。

技术的批评者在这里也有重要作用。站在旁边大喊大叫，在推特上发怒，写长篇晦涩的文章来概述问题，这些都很好。但这些行动不会阻止即将到来的浪潮，实际上也不会显著改变它。当我第一次开始专业工作时，外界对技术的看法几乎完全是良性的，甚至是狂喜的。这些是构建光明未来的酷炫、友好的公司。这种情况已经改变。然而，尽管批评的声音变得更大，但值得注意的是，他们的成功却寥寥无几。

以他们自己的方式，科技的批评者陷入了一种悲观主义厌恶陷阱，这种陷阱深深植根于科技/政治/商业精英中。许多嘲笑过于乐观的技术人员的人坚持撰写理论监督框架或发表呼吁监管的专栏文章。如果你相信技术是重要且强大的，并且你遵循这些批评的含义，这样的回应显然是不够的。即使是批评者也回避了他们面前的真实现实。确实，有时尖锐的批评只是成为与技术本身相同的炒作周期的一部分。

可信的批评者必须是实践者。构建正确的技术，拥有改变其进程的实际手段，不仅仅是观察和评论，而是积极地指引方向，进行变革，在源头实施必要的行动，这意味着批评者需要参与其中。他们不能站在旁边大喊。这绝不是反对批评者的论点，恰恰相反。这是对技术深刻的认识，尤其是在前线，构建和制造，努力应对创造的日常现实。如果你正在阅读这篇文章并持批评态度，那么有一个明确的回应：参与其中。

我完全承认这并不容易。这里没有舒适的地方。不可能不承认其中的一些悖论。这意味着像我这样的人必须面对这样的前景：在努力构建积极工具和防止不良结果的同时，我们可能无意中加速了我们试图避免的事情，就像功能增益研究人员进行病毒实验一样。我开发的技术很可能会造成一些伤害。尽管我尽最大努力去学习和改进，我个人仍会继续犯错。多年来，我一直在思考这个问题——是退缩还是参与其中？你离一项技术的核心越近，就越能影响结果，将其引导向更积极的方向，并阻止有害的应用。但这也意味着成为使其成为现实的一部分——无论是它可能带来的好处还是伤害。

我没有所有的答案。我不断质疑自己的选择。但唯一的其他选择就是完全放弃建设的任务。技术人员不能成为与未来脱节、只听自己声音的遥远建筑师。没有外部和内部的批评者，这种困境将不可避免地向我们逼近。有了他们，就有更好的机会构建不会进一步损害民族国家、不易发生灾难性故障、不助长威权主义反乌托邦的技术。十年前，科技行业在各个意义上都是单一文化的。这种情况开始改变，现在比以往任何时候都有更多的智力多样性，包括在开发过程中有更多批判性、伦理性和人文主义的声音。

当我共同创立DeepMind时，将安全和伦理问题融入科技公司的核心结构中显得很新颖。在这种情况下，仅仅使用“ 伦理 ”这个词就让我受到普遍的奇怪目光；相比之下，如今它不幸地面临成为另一个被过度使用的流行语的危险。然而，这确实带来了真正的改变，开启了有意义的讨论和争论的机会。令人鼓舞的是，关于伦理AI的研究激增——自2014年以来，出版物增加了五倍。在行业方面，这种增长更快；与行业相关的伦理AI研究每年增长70%。曾经，在科技领域找到道德哲学家、政治科学家和文化人类学家是件奇怪的事，现在则不那么奇怪了。然而，在将非技术视角和多样化声音引入讨论方面的重大不足仍然太过普遍：受限技术是一个需要各种学科和视角的项目。积极招聘以实现这一目标是必须的。

在一个固有激励和监管失效的世界中，技术不仅需要外部的批评者，还需要在其核心的批评者。

5. 企业：利润 + 目的

源文本: 翻译文本:

利润驱动着即将到来的浪潮。没有任何通往安全的途径可以不承认和应对这一事实。对于像人工智能和合成生物学这样的指数型技术，我们必须找到新的负责任且包容的商业模式，以同时激励安全和利润。应该有可能创建更适合默认情况下控制技术的公司。我和其他人长期以来一直在尝试应对这一挑战，但迄今为止结果喜忧参半。

传统上，公司有一个明确的目标：股东回报。在大多数情况下，这意味着新技术的无障碍发展。虽然这在历史上一直是一个强大的进步引擎，但它不太适合遏制即将到来的浪潮。我认为，找出在混合组织中调和利润和社会目的的方法是至关重要的。

结构是应对未来挑战的最佳方式，但在实践中使其发挥作用是非常困难的。

从 DeepMind 成立之初，对我来说，重要的是我们将治理模型与我们的最终目标同等考虑。在 2014 年被谷歌收购时，我设计了一个“ 伦理和安全委员会 ”来监督我们的技术，并将其作为收购的条件。即使在那时，我们也意识到，如果我们成功实现了构建真正 AGI 的使命，它将释放出远远超出单个公司可以合理预期拥有和控制的力量。我们希望确保谷歌理解这一点，并承诺将我们的治理范围扩大到我们技术人员之外。最终，我希望创建一个全球性的、多利益相关者的论坛，以决定当 AGI 实现时或如果实现时会发生什么，这是一种民主的世界 AI 研究所。在我看来，技术越强大，拥有多个视角来控制和获取它就越重要。

在被谷歌收购后，我和我的联合创始人花了数年时间试图将伦理宪章融入公司的法律结构中，不断争论这份宪章有多少可以公开，DeepMind的工作有多少可以进一步接受独立监督和审查。在这些讨论中，我们的目标始终是确保前所未有的技术与前所未有的治理相匹配。我们的提议是将DeepMind分拆出来，成为一种新的“ 全球利益公司 ”，拥有一个完全独立的受托人委员会，该委员会与负责公司运营的董事会分开并额外存在。成员资格、决策，甚至一些董事会的推理将更加公开。透明度、问责制、伦理——这些不仅仅是企业公关，而是基础性的、具有法律约束力的，并且融入公司所做的一切中。我们认为这将使我们能够以开放的方式工作，主动学习公司如何成为指数技术的现代长期管理者。

我们建立了一种合理的方法，将来自 AI 的利润再投资于道德和社会使命。分拆出来的公司将“ 受限于

“保证”，没有股东，但有义务向主要资助者Alphabet提供独家技术许可。作为其社会和科学使命的一部分，DeepMind将使用其大部分利润来研究可能在多年后才有价值的公共服务技术：例如碳捕获和储存、海洋清理、吃塑料的机器人或核聚变。协议是我们能够将一些重大突破开源，就像一个学术实验室一样。对谷歌搜索业务核心的知识产权将保留在谷歌，但其余的将可用于推进DeepMind的社会使命，研究新药、更好的医疗保健、气候变化等。这意味着投资者可以获得回报，但也确保了社会目的在公司的法律基因中。

事后看来，当时对谷歌来说实在是太多了。律师被聘请，进行了多年的激烈谈判，但似乎没有办法解决这个难题。最终，我们无法找到一个能让所有人满意的答案。DeepMind继续作为谷歌内部的一个普通部门运作，没有正式的法律独立性，只是作为一个独立品牌运营。对我来说，这是一个基础性的教训：股东资本主义之所以有效，是因为它简单明了，治理模式也倾向于简单明了。在股东模式中，责任线和绩效跟踪是量化的，非常透明。从理论上讲，可能可以设计出更现代的结构，但在实践中操作它们则是另一回事。

在我于谷歌工作的期间，我继续致力于创造创新治理结构的实验性工作。我起草了谷歌的AI原则，并参与了启动AI伦理咨询委员会的团队，该委员会由杰出的独立法律、技术和伦理专家组成。两者的目标都是迈出第一步，建立一个章程，规定谷歌如何处理像AI和量子计算这样的尖端技术。我们的愿望是邀请一群多元化的外部利益相关者，获得技术前沿的特权访问，提供反馈，并从远离新技术构建的兴奋和乐观的人那里提供急需的外部视角。

然而，委员会在宣布后几天就解散了。谷歌的一些员工反对任命凯·科尔斯·詹姆斯为成员，她是位于华盛顿的保守派智库传统基金会的主席。她与来自左翼和中间派的多位人士一起被任命，但谷歌内部很快发起了一场运动，要求将她移除。与推特员工结成联盟，活动人士指出，她多年来发表了许多反跨性别和反LGBTQ的言论，最近还表示：“ 如果他们可以将女性的定义改变为包括男性，他们就可以抹去在经济、社会和政治上赋予女性权力的努力。 ”虽然我个人不同意她的言论和政治立场，但我为我们邀请她加入董事会的决定进行了辩护，认为应该听取各种价值观和观点。毕竟，谷歌是一家拥有全球用户的全球公司，其中一些用户可能会认同这种观点。

许多谷歌员工和外部活动人士对此表示反对，并在宣布后的几天内发表了一封公开信，要求将詹姆斯从委员会中移除。员工和其他人积极游说大学校园，要求取消对其他拒绝辞职的董事会成员的学术资助，认为他们的持续参与只能被理解为纵容恐跨行为。最终，三名成员辞职，整个努力在不到一周的时间内被彻底放弃。不幸的是，这种政治氛围对公众人物和上市公司来说都过于强烈。

再一次，我重新思考公司使命的尝试失败了，尽管它们激发了对话，并帮助在Alphabet和更广泛的政策、学术和行业圈子中将一些困难的讨论摆上台面。哪些团队和研究得到资助，产品如何测试，内部控制和审查如何进行，外部审查的适当程度，哪些利益相关者需要被包括在内——Alphabet及其他地方的高级领导开始定期进行这些对话。

在科技公司中，十年前还显得边缘的 AI 安全讨论现在正变得司空见惯。需要在利润与

积极贡献和尖端安全原则上被所有主要的美国科技集团接受。尽管有巨大的奖励规模，企业家、管理人员和员工都应该继续推动和探索能够更好地应对遏制挑战的公司形式。

令人鼓舞的实验正在进行中。Facebook 创建了其独立的监督委员会——由前法官、活动家和专家学者组成，以就平台治理提供建议。它受到了各方的批评，显然不能单独“ 解决 ”问题。但重要的是要首先赞扬这一努力，并鼓励 Facebook 和其他公司继续进行实验。另一个例子是公共利益公司和 B 公司日益增长的运动，这些公司仍然是营利性公司，但在其法律定义的目标中铭刻了社会使命。将强有力的遏制机制和目标作为受托责任写入的科技公司是下一步。鉴于这些替代公司结构的增长（现在有超过一万家公司使用 B 公司结构），这里有很大的积极变化机会。虽然经济目标并不总是与受控技术很好地对齐，但创新的公司形式使其更有可能。这正是所需的实验类型。

遏制需要新一代的公司。它需要创始人和从事技术工作的人对社会做出积极贡献。它还需要一些完全更困难的东西。它需要政治。

6. 政府：生存、改革、监管

源文本: 翻译文本:

技术问题需要技术解决方案，正如我们所见，但仅靠技术解决方案永远是不够的。我们还需要国家繁荣。每一项加强自由民主国家并使其抵御压力的努力都必须得到支持。民族国家仍然控制着文明的许多基本要素：法律、货币供应、税收、军事，以及

因此，这有助于完成接下来的任务，他们需要创建和维护有弹性的社会系统、福利网、安全架构和治理机制，以能够承受严重的压力。但他们也需要详细了解正在发生的事情：现在他们就像在飓风中盲目操作。

物理学家理查德·费曼曾经说过一句名言：“ 我不能创造的东西，我就无法理解。 ”今天，这句话对政府和技术的关系再贴切不过了。我认为政府需要更多地参与其中，回到构建真正的技术、设定标准和培养内部能力。它需要在公开市场上竞争人才和硬件。毫无疑问，这将是昂贵的，并且会伴随着浪费性的错误。但是，积极主动的政府将比仅仅委托服务并依赖外包的专业知识和在其他地方拥有和运营的技术，能够施加更大的控制力。

问责制通过深入理解得以实现。所有权赋予控制权。两者都需要政府亲力亲为。尽管如今公司已经占据了主导地位，但许多最具探索性的基础研究仍然由政府资助。美国联邦政府在研发上的支出占总额的比例处于历史最低水平——仅为20%——但每年仍高达1790亿美元，这并不是一个小数目。

这是个好消息。投资于科学技术教育和研究，并支持国内科技企业，创造了一个正反馈循环，使政府在最先进的技术中拥有直接利益，能够利用其优势并抑制其危害。简单来说，作为即将到来的浪潮的平等合作伙伴，政府有更好的机会将其引导至整体公众利益。即使以相当大的成本拥有 *much* 更多的内部技术专长，也是值得的投资。政府不应依赖管理顾问、承包商或其他第三方供应商。全职、受人尊敬的员工应得到适当的补偿，与私营部门具有竞争力，应该是解决方案的核心部分。相反，在国家关键角色中，私营部门的薪水可能是公共部门的十倍：这是不可持续的。

他们的首要任务应该是更好地监测和理解技术的发展。各国需要详细了解，例如，他们的人口提供了哪些数据，这些数据如何以及在哪里被使用，以及这意味着什么；政府应该对最新的研究有强烈的认识，了解前沿在哪里，发展方向如何，他们的国家如何最大化收益。最重要的是，他们需要以公开透明的方式记录技术造成伤害的所有方式——列出每一次实验室泄漏、每一次网络攻击、每一个语言模型偏见、每一次隐私泄露——以便每个人都能从失败中学习并改进。

这些信息需要被国家有效利用，实时应对新出现的问题。接近行政权力的机构，如白宫的科学技术政策办公室，正变得越来越有影响力。仍然需要更多：在21世纪，处理经济、教育、安全和国防等事务的内阁职位没有一个同样有权力和民主问责的技术职位是没有意义的。新兴技术的部长或秘书在政府中仍然很少见。这不应该是这样；在即将到来的浪潮时代，每个国家都应该有一个。

仅靠监管无法实现遏制，但任何不涉及监管的讨论都是注定失败的。监管应侧重于这些激励措施，更好地使个人、国家、公司和公众整体与安全和保障保持一致，同时建立紧急制动的可能性。某些使用案例，如用于选举活动的人工智能，应作为一揽子措施的一部分被法律禁止。

立法机构开始采取行动。2015年几乎没有关于AI的立法。但自2019年以来，全球已有不少于七十二项包含“人工智能”一词的法案被通过。经合组织AI政策观察站在其数据库中记录了来自六十个国家的不少于八百项AI政策。欧盟的AI法案确实存在问题，但其条款中有许多值得称赞的地方，并且它代表了正确的重点和雄心。

2022年，白宫发布了一份人工智能权利法案蓝图，其中包含五项核心原则，以帮助指导设计、开发和

“ 部署人工智能和其他自动化系统，以保护美国公众的权利。 ”它表示，公民应受到保护，免受不安全和无效系统以及算法偏见的影响。任何人都不应被迫接受人工智能。每个人都有权说不。这样的努力应该得到广泛支持并迅速实施。

然而，政策制定者的想象力需要与技术的范围相匹配。政府需要走得更远。出于可以理解的原因，我们不允许任何企业以他们认为合适的方式建造或操作核反应堆。实际上，国家密切参与并密切关注、许可和管理其存在的每一个方面。随着时间的推移，这将会并且应该更适用于一般技术。今天，任何人都可以构建 AI。任何人都可以建立一个实验室。我们应该转向一个更有许可的环境。这将产生一套更明确的责任，并提供更严格的机制来撤销访问权限和补救高级技术带来的危害。最复杂的 AI 系统、合成器或量子计算机应该仅由负责任的认证开发者生产。作为其许可证的一部分，他们需要遵循明确的、具有约束力的安全和保障标准，遵循规则，进行风险评估，保存记录，密切监控实时部署。正如你不能在没有 FAA 批准的情况下简单地将火箭发射到太空一样，明天你也不应该简单地发布最先进的 AI。

不同的许可制度可能会根据模型的大小或能力而有所不同：模型越大、越强，许可要求就越严格。模型越通用，就越有可能构成严重威胁。这意味着从事最基本能力研究的人工智能实验室需要特别关注。此外，这也为更细化的许可创造了空间，以便在需要时专注于开发的具体细节：模型的训练运行、超过特定规模的芯片集群、某些类型的有机体。

税收也需要彻底改革，以资助安全和福利，因为我们正在经历历史上最大的一次价值创造转型——从劳动到资本。如果技术造成了失败者，他们需要物质补偿。如今，美国的劳动税平均税率为25%，

设备和软件仅占5%。该系统旨在让资本无摩擦地自我复制，以创建繁荣的企业。未来，税收需要将重点转向资本，不仅要为那些受到不利影响的人提供再分配的资金，还要在此过程中创造一个更慢、更公平的过渡。财政政策是控制这一过渡的重要阀门，是在控制这些瓶颈的同时增强国家韧性的一种手段。

这应该包括对土地、财产、公司股份和其他高价值、流动性较差的资产等旧形式资本征收更高的税，以及对自动化和自主系统的新税。这有时被称为“机器人税”；麻省理工学院的经济学家认为，即使是仅占其价值1%到4%的适度税收也可能产生重大影响。将税收负担从劳动转移的精心校准的调整将激励持续雇佣，并缓解家庭生活中的中断。为最低收入者提供的税收抵免可以在收入停滞甚至崩溃的情况下立即提供缓冲。同时，大规模的再技能计划和教育努力应为弱势群体做好准备，提高风险意识，并增加与这一浪潮能力的参与机会。普遍基本收入（UBI）——即无论情况如何，国家为每位公民支付的收入——常被认为是应对即将到来的经济动荡的答案。在未来，可能会有类似UBI的倡议的空间；然而，在此之前，还有许多好的想法。

在企业人工智能超大规模发展的时代，我们应该开始考虑对最大型企业本身征收资本税，而不仅仅是对相关资产或利润征税。此外，必须找到跨境征税这些巨型企业的机制，以确保它们为维持正常运作的社会支付其应有的份额。在这方面鼓励进行实验：例如，将公司价值的一部分固定比例作为公共股息支付，这将在极端集中的时代将价值转移回大众。极限情况下，有一个核心问题是关于谁拥有即将到来的浪潮的资本；一个真正的AGI不能被私人拥有。

以与建筑物或卡车车队相同的方式。 当涉及到可能从根本上延长人类寿命或能力的技术时，显然从一开始就必须对其分配进行大讨论。

谁能够设计、开发和部署这样的技术，最终是由政府来决定的问题。它们的杠杆、机构和专业领域都必须像技术一样迅速发展，这对所有相关人员来说都是一代人的挑战。因此，一个受控技术的时代就是一个广泛且智能地监管技术的时代；没有任何借口或例外。但当然，一个国家的监管有一个不可避免的缺陷。没有哪个国家政府可以单独做到这一点。

7. 联盟：是时候签订条约了

源文本: 翻译文本:

激光武器听起来像是科幻小说。不幸的是，它们不是。随着激光技术的发展，很明显它们可能导致失明。武器化后，这可能使敌方部队或任何被瞄准的人丧失能力。一项令人兴奋的新民用技术再次开启了可怕攻击模式的前景（尽管迄今为止不是以*Star Wars*的方式）。没有人希望军队或帮派携带致盲激光器四处游荡。

幸运的是，这并没有发生。根据1995年《致盲激光武器议定书》，使用致盲激光武器被禁止，这是《特定常规武器公约》的更新，禁止使用“ 专门设计为其唯一作战功能或其作战功能之一是对未增强视力造成永久性失明的激光武器。 ”一百二十六个国家签署了该协议。因此，激光武器既不是军事装备的主要部分，也不是街头常见的武器。

当然，致盲激光并不是我们在本书中讨论的那种全用途技术。但它们证明了这是可行的；强有力的禁令可以奏效。微妙的联盟和国际合作可以实现，它们可以改变历史。

考虑这些例子，其中一些我们之前讨论过：核不扩散条约；禁止CFCs的蒙特利尔议定书；在冷战分裂中发明、试验和推广脊髓灰质炎疫苗；生物武器公约，一项有效禁止生物武器的裁军条约；禁止集束弹药、地雷、人类基因编辑和优生学政策；旨在限制碳排放和气候变化最严重影响的巴黎协定；全球消灭天花的努力；逐步淘汰汽油中的铅；以及终止石棉的使用。

国家不愿放弃权力，就像公司不愿错失利润一样，然而这些都是可以学习的先例，是在充满复兴技术竞争的环境中希望的碎片。每个国家都有特定的条件和挑战，这些条件和挑战既有助于其实现，也阻碍了完美的合规性。但至关重要的是，每一个国家都是世界各国团结和妥协以应对重大挑战的宝贵例子，为应对即将到来的浪潮提供了线索和框架。如果一个政府想要禁止合成生物学或AI应用，它能做到吗？不，显然只能在部分、脆弱的意义上做到。但一个强大、有动力的联盟呢？也许可以。

面对深渊，地缘政治可能会迅速改变。在第二次世界大战的激烈时期，和平一定像是一个梦。当盟军疲惫地继续战斗时，地面上的人们很少能想象到，仅仅几年后，他们的政府会投入数十亿美元来重建他们的敌人。尽管有可怕的种族灭绝战争罪行，德国和日本很快就成为了稳定的全球联盟的重要组成部分。事后看来，这似乎令人眩晕。仅仅几年时间，就将诺曼底和硫磺岛的枪林弹雨、苦涩和海滩与坚如磐石的军事和商业伙伴关系、延续至今的深厚友谊以及有史以来最大规模的对外援助计划分隔开来。

在冷战高峰期，尽管紧张局势严重，高层接触仍得以维持。如果发生类似流氓AGI或重大生物危害的事件，这种高层协调将至关重要，然而随着新冷战的形成，分歧正在加剧。

灾难性威胁本质上是全球性的，应该成为国际共识的问题。仅限于国家边界的规则显然是不够的。虽然每个国家都有推动这些技术发展的利益，但它们也有充分的理由遏制其最严重的后果。那么，对于即将到来的浪潮，不扩散条约、蒙特利尔议定书、巴黎协定会是什么样子呢？

核武器是一个例外，部分原因是它们非常难以制造：长时间耐心的讨论、在联合国进行的数十年艰苦的条约谈判、即使在极端紧张时期的国际合作，这一切在控制它们时都很重要。核遏制既有道德成分，也有战略成分。达成和执行此类协议从来都不容易，在大国竞争的时代更是如此。因此，外交官在遏制技术方面扮演了一个被低估的角色。一个科技外交的黄金时代需要从军备竞赛的时代中崛起。我与外交界的许多人交谈过，他们对此有深刻的认识。

然而，联盟也可以在技术人员或次国家机构的层面上运作，共同决定资助什么，拒绝什么。一个很好的例子来自生殖系基因编辑。一项对106个国家的研究发现，生殖系基因编辑的监管是不完善的。大多数国家都有某种形式的监管或政策指导，但存在相当大的差异和空白。这并没有形成一个对具有全球范围的技术的全球框架。迄今为止，更有效的是前线科学家的国际合作。在首次人类基因编辑之后，由Eric Lander、Emmanuelle Charpentier和张锋等知名人士签署的一封信呼吁“对所有临床使用人类生殖系编辑——即改变可遗传DNA（在精子、卵子或胚胎中）以制造基因改造儿童——实行全球暂停”和“一个国际框架，在该框架中，各国在保留自行决定权的同时，自愿承诺不批准任何临床生殖系编辑的使用，除非满足某些条件。”

他们并没有呼吁永久禁止，也没有禁止出于研究目的的生殖系编辑，他们也没有说每个国家都应该遵循相同的路径。但他们要求从业者花时间协调并做出正确的决定。足够多的前沿人士仍然可以产生影响，留出暂停的空间，帮助为各国和国际机构创造一个共同努力并找到解决办法的空间和基础。

在本章前面，我讨论了美国和中国之间的摩擦。尽管存在分歧，这些竞争大国之间仍然有明显的合作空间。合成生物学比人工智能是一个更好的起点，因为现有竞争较少，而且新型生物威胁的相互保证毁灭是显而易见的。SecureDNA项目是一个很好的例子，为治理合成生物学铺设了一条道路，类似于化学武器的限制。如果中国和美国能够创建一个共享的生物风险观测站，比如说，涵盖从先进研发到已部署的商业应用的所有内容，这将是一个宝贵的合作领域，可以在此基础上进行建设。

中国和美国也有共同的利益来遏制不良行为者的长尾。鉴于奥姆真理教可能来自任何地方，两国都将热衷于遏制世界上最强大技术的无序扩散。目前，中国和美国正在争夺技术标准的制定。但共同的方法显然是双赢的；分裂的标准让每个人都更难办事。另一个共同点可能是在量子计算或机器学习的进步可能削弱加密系统的情况下，维护这些系统。每一个都可能为更广泛的妥协铺平道路。随着世纪的推进，冷战的教训将不得不重新学习：没有与对手合作，就没有通往技术安全的道路。

除了鼓励双边倡议之外，在这个阶段显而易见的事情是提议创建某种新的全球机构，专注于技术。我听过很多次这样的说法：生物技术的世界银行或人工智能的联合国会是什么样子？一个安全的国际机构能否

合作是解决像 AGI 这样令人生畏且复杂的问题的方法吗？谁是最终的仲裁者，所谓的最后贷款人，当被问到“ 谁能控制技术？ ”时，哪个机构可以举手？

我们这一代需要一个相当于核条约的协议，以形成一种共同的全球方法——在这种情况下，不是完全遏制扩散，而是设定限制并建立管理和缓解的框架，这些框架像波浪一样跨越国界。这将对进行的工作设定明确的限制，在国家许可努力之间进行调解，并创建一个审查两者的框架。

在技术问题上，确实有为新机构设立明确范围的必要。迫切需要一个专门的监管机构，该机构能够尽可能地处理有争议的地缘政治问题，避免过度扩张，并根据广泛的客观标准执行务实的监控功能。可以想象成类似于国际原子能机构或甚至是国际航空运输协会这样的贸易组织。与其建立一个直接监管、建造或控制技术的组织，我会从类似于人工智能审计局（AAA）这样的机构开始。AAA专注于事实调查和审计模型规模，以及当能力门槛被跨越时，增加前沿的全球透明度，提出诸如：系统是否显示出能够自我提升能力的迹象？它能否自行设定目标？它能否在无人监督的情况下获取更多资源？它是否被故意训练用于欺骗或操控？类似的审计委员会几乎可以在每个领域运作，并且同样可以为政府的许可努力提供基础，同时也有助于推动不扩散条约的达成。

坚硬的现实主义比模糊和不太可能的提议更有成功的机会。我们不需要完全重新发明制度的轮子，创造更多的竞争和哗众取宠的机会。我们只需找到一切可能的手段来改进它——而且要快。

8. 文化：尊重地拥抱失败

源文本: 翻译文本:

这里的共同主题是治理：软件系统、微芯片、企业和研究机构、国家以及国际社会的治理。在每个层面上，都存在一片激励措施、沉没成本、制度惯性、相互冲突的封地和世界观的丛林，必须加以清理。毫无疑问，伦理、安全、控制——这些首先是良好治理的产物。但良好治理不仅仅来自于明确的规则和有效的制度框架。

在喷气发动机的早期，即20世纪50年代，坠机和死亡事故令人担忧地常见。到2010年代初，每740万名乘客登机仅有一人死亡。现在几年过去了，美国商业飞机几乎没有涉及任何致命事故。飞行几乎是最安全的交通方式：坐在三万五千英尺的高空比坐在家里的沙发上更安全。

航空公司的安全记录令人印象深刻，这归功于多年来无数渐进的技术和运营改进。但在这些改进背后，还有一个同样重要的因素：文化。航空业在各个层面都采取积极的态度，从错误中学习。空难不仅仅是需要哀悼的悲剧事故；它们是确定系统如何失效的基础性学习经验，是诊断问题、解决问题并在整个行业中分享这些知识的机会。因此，最佳实践并不是企业秘密，也不是相对于竞争对手航空公司的优势：它们被竞争对手热情地实施，以维护整个行业的共同信任和安全。

这就是即将到来的浪潮所需要的：来自参与前沿技术的每个人的真实、发自内心的认同。制定和推广伦理和安全的倡议和政策固然很好，但你需要实际执行的人真正相信它们。

虽然科技行业在谈论“ 拥抱失败 ”时说得头头是道，但在涉及隐私、安全或技术漏洞时却很少如此。推出一款不受欢迎的产品是一回事，但拥有一个导致信息错误灾难的语言模型则是另一回事。

导致不良反应的药物会让人感到更加不适。对技术的批评，理由充分，往往是无情的。竞争也是如此。其结果之一是，一旦新技术或产品出现问题，保密文化就会占据上风。开发过程中的开放性和相互信任的特征部分消失了。学习的机会，以及随后传播这些学习的机会，也随之消失。即使是承认错误，打开闸门，也被视为一种风险，是企业的大忌。

对失败和公众谴责的恐惧导致了停滞。立即自我报告问题应该成为个人和组织的基准。但公司和团队不仅没有因实验而受到赞扬，反而被晾在一边。做正确的事情只会引发愤世嫉俗、推特抨击和恶意的公众得分反弹。在这种情况下，为什么会有人真正承认他们的错误？如果我们想要生产更好、更负责任、更可控的技术，这种情况必须停止。

拥抱失败必须是真实的，而不是一句空话。首先，即使在令人不安的话题上，对失败完全开放应该受到赞扬，而不是侮辱。科技公司在遇到任何风险、缺点或失败模式时，首先要做的是安全地向更广泛的世界传达信息。当实验室泄漏时，首先要做的是公开这一事实，而不是掩盖它。其他参与者——其他公司、研究小组、政府——接下来需要做的第一件事是倾听、反思、提供支持，最重要的是学习并积极实施这些学习。这种态度在空中拯救了成千上万的生命。未来几年，它可能会拯救更多的生命。

遏制不仅仅是关于这个或那个政策、清单或倡议，而是需要确保存在一种自我批评的文化，积极地想要实施这些政策，欢迎监管者进入房间、进入实验室，这是一种监管者想要向技术人员学习，反之亦然的文化。需要每个人都想参与其中，拥有它，热爱它。否则，安全性仍然是事后的考虑。在许多人中，不仅仅是在人工智能领域，有一种感觉我们“只是”研究人员，“只是”在探索和实验。多年来情况并非如此，这是我们需要文化的一个典型例子。

转变。研究人员必须被鼓励从不断追求发表的压力中退后一步。知识是一种公共产品，但不应再是默认的选择。那些积极进行前沿研究的人需要首先认识到这一点，正如他们在核物理和病毒学领域的同行已经做到的那样。在人工智能领域，我认为递归自我改进和自主性等能力是我们不应跨越的界限。这将涉及技术和法律方面的内容，但也需要与其最接近的人和组织在道德、情感和文化上的认同。

1973年，基因工程的发明者之一保罗·伯格在加利福尼亚的蒙特雷半岛召集了一群科学家。他开始担心自己的发明可能带来的后果，并希望为未来的发展设定一些基本规则和道德基础。在阿西洛马会议中心，他们提出了这个新学科带来的困难问题：我们应该开始基因工程改造人类吗？如果是，哪些特征是允许的？两年后，他们以更大的规模返回，参加重组DNA的阿西洛马会议。在那个海滨酒店，风险很高。这是生物科学的一个转折点，确立了管理基因研究和技术的持久原则，为可以进行的实验设定了指导方针和道德限制。

我在2015年参加了在波多黎各举行的一个会议，该会议旨在为人工智能做类似的事情。会议由一个混合群体组成，旨在提高人工智能安全的知名度，开始建立一种谨慎的文化，并勾勒出实际的解决方案。我们于2017年再次在象征性的阿西洛马尔会场会面，起草了一套人工智能原则，我和该领域的许多人都签署了这些原则。它们旨在建立一种明确负责的人工智能研究文化，并激发了一系列进一步的倡议。随着浪潮的不断积聚，我们将需要自觉地一次又一次地回到阿西洛马尔的精神和文字上。

数千年来，希波克拉底誓言一直是医学界的道德指引。在拉丁语中，*Primum non nocere.* 首先，不伤害。诺贝尔和平奖得主、英籍波兰科学家约瑟夫·罗特布拉特，这位因良心原因离开洛斯阿拉莫斯的人，主张科学家们

需要类似的东西。他认为，社会和道德责任不是任何科学家可以抛弃的东西。我同意，我们应该为技术人员考虑一个当代版本：不仅要问在一个全球性算法和编辑基因组的时代，不造成伤害意味着什么，还要问如何在通常道德上模棱两可的情况下每天实施这一点。

像这样的预防原则是一个很好的第一步。在构建之前暂停，在发布之前暂停，审查一切，坐下来仔细研究第二、第三、*n*阶影响。找到所有证据并冷静地查看。不断地纠正方向。要愿意停止。做这一切不仅仅是因为某种形式上说要这样做，而是因为这是正确的，这是技术人员应该做的。

像这样的行动不能仅仅作为法律或企业口号来运作。法律只是国家性的，企业口号是暂时的，往往是表面的。它们必须在更深层次上运作，使得技术文化不再是那种“ 工程思维 ”的随心所欲，而是更加谨慎，更加好奇可能会发生什么。一个健康的文化是乐于将果实留在树上、说不、推迟利益以确保安全的文化，是技术人员记住技术只是达到目的的手段，而不是目的本身的文化。

9. 运动：人民的力量

源文本: 翻译文本:

在整本书中，“ 我们 ”这个词频繁出现。它可能指的是“ 我们 ”作者和合著者，“ 我们 ”AI研究人员和企业家，“ 我们 ”更广泛的科学和技术社区，“ 我们 ”在全球西方，或者“ 我们 ”全人类的总和。（面对完全全球化和改变物种的技术，是少数几个谈论人类“ 我们 ”确实有必要的地方之一。）

当人们谈论技术时——包括我自己在内——他们常常会提出如下论点。因为*we*构建了技术，*we*可以解决它所产生的问题。这在最广泛的意义上是正确的。但是，问题是，

这里没有功能性的“我们”。没有共识，也没有形成共识的公认机制。实际上没有“我们”，当然也没有任何“我们”可以拉动的杠杆。这应该是显而易见的，但值得重复。即使是美国总统，在改变互联网的发展方向方面的权力也非常有限。

相反，无数分布式的参与者有时协作，有时却各行其是。正如我们所见，公司和国家有着不同的优先事项，分裂且相互冲突的激励机制。大多数情况下，像本书中所述的技术问题是精英们的追求，是商务舱休息室的谈资，是 *bien-pensant* 出版物的专栏文章，或是达沃斯或 TED 演讲厅的话题。大多数人类尚未以任何系统的方式担心这些事情。离开 Twitter，走出泡沫，大多数人有着截然不同的关注点，还有其他问题在这个脆弱的世界中需要关注。围绕人工智能的沟通并不总是有帮助，往往陷入简单化的叙述。

因此，如果目前“我们”的宏大号召毫无意义，那么它引发了一个显而易见的后续行动：让我们建立一个。在整个历史中，变革是因为人们自觉地为之努力而发生的。公众压力创造了新的规范。奴隶制的废除、妇女选举权、公民权利——这些都是巨大的道德成就，因为人们努力奋斗，建立了广泛的联盟，认真对待一个重大主张，然后基于此实现了变革。气候问题不仅仅是因为人们注意到天气变得更加极端而被提上日程。他们注意到是因为草根活动家和科学家，然后是（一些）作家、名人、CEO和政治家为有意义的变革而激动不已。他们出于做好事的愿望而采取行动。

研究表明，当人们接触到新兴技术及其风险的话题时，他们确实关心并希望找到解决方案。尽管许多危害仍然遥远，但我相信人们完全有能力在这里解读迹象。我还没有发现有人在观看波士顿动力公司的视频中看到机器狗或考虑到另一场大流行病的前景时没有感到恐惧。

在这里，民众运动发挥了巨大的作用。在过去五年左右的时间里，一个迅速发展的公民社会运动开始突出这些问题。媒体、工会、慈善组织、草根运动——所有这些都在参与，积极寻找创造受控技术的方法。我希望我这一代的创始人和建设者能够激励这些运动，而不是阻碍它们。同时，公民大会提供了一种机制，可以让更广泛的群体参与对话。一个提议是举办抽签，选择一个具有代表性的人群样本进行深入辩论，并提出如何管理这些技术的建议。在获得工具和建议的情况下，这将是使遏制成为一个更具集体性、关注性和扎根性的过程的一种方式。

当人们要求改变时，改变就会发生。构建技术的“ 我们 ”是分散的，受到各种竞争和不同的国家、商业和研究激励的影响。越是受到影响的“ 我们 ”以一个声音清晰地发声，成为推动变革的关键公众力量，要求方法的一致性，就越有可能获得良好的结果。任何地方的任何人都可以有所作为。从根本上说，技术人员和政府都无法单独解决这个问题。但“ 我们 ”一起可能可以。

10. 狭窄的道路：唯一的出路是通过

源文本: 翻译文本:

就在 GPT-4 发布几天后，数千名 AI 科学家签署了一封公开信，呼吁对研究最强大的 AI 模型进行为期六个月的暂停。他们引用了阿西洛马原则，提出了本书读者熟悉的理由：“ 最近几个月，AI 实验室陷入了一场失控的竞赛，竞相开发和部署越来越强大的数字智能，甚至连它们的创造者都无法理解、预测或可靠地控制。 ”不久之后，意大利禁止了 ChatGPT。针对 LLMs 的投诉已提交给联邦贸易委员会。

委员会旨在实现更严格的监管控制。在白宫新闻发布会上提出了关于AI风险的问题。数百万人在工作中、餐桌上讨论了技术的影响。

有些事情正在酝酿。虽然不是遏制，但人们首次以应有的紧迫感对待即将到来的浪潮所带来的问题。

到目前为止概述的每个想法都代表了海堤的开端，一个试探性的潮汐屏障，从技术本身的细节开始，并向外扩展到形成一个大规模全球积极变革运动的必要性。它们中的任何一个都无法单独发挥作用。然而，将这样的措施结合在一起，一个遏制的轮廓便浮现出来。

一个很好的例子来自麻省理工学院的生物技术专家凯文·埃斯维尔特。很少有人如此详细地考虑生物安全威胁。那些被设计成造成最大伤亡的定制病原体？凯文决心使用一切工具来阻止它们的发生。他的计划是最全面的遏制策略之一。它围绕三个支柱构建：延迟、检测和防御。

为了拖延，他借用了核技术的语言，提出了一项“流行病试验禁令条约”，这是一项国际协议，旨在停止对最具病原性的材料进行实验。任何可能严重增加流行病事件风险的实验，包括功能增益研究，都将被禁止。他还倡导为任何从事病毒或其他潜在有害生物材料工作的人建立一个全新的保险和责任制度。这将通过将低概率但灾难性的后果——目前由其他人承担的负外部性——直接纳入研究成本，以一种立竿见影的方式提高责任成本。不仅进行潜在危险研究的机构需要购买额外的保险，而且一项触发法将意味着任何被证明对重大生物危害或灾难性事件负责的人将承担责任。

对所有合成器进行DNA筛查是绝对必要的，而且整个系统应该基于云，以便能够根据新理解和新出现的威胁实时更新。迅速检测一个

在这个模式中，爆发同样重要，尤其是对于潜伏期长的微妙病原体。想象一种潜伏多年的疾病。如果你不知道发生了什么，就无法控制它。

然后，如果最坏的情况发生，就要防御。具有韧性和准备充分的国家至关重要：最极端的大流行病甚至会使维持食品、电力、水供应、法律和秩序以及医疗保健变得困难。为所有必要的工作人员准备好最先进的防疫PPE设备储备将产生巨大的差异。能够承受严重冲击的医疗设备的强大供应链也同样重要。那些可以消灭病毒的低波长灯泡？它们需要在大流行开始之前就达到*everywhere,*，或者至少准备好进行部署。

把这里所有的元素放在一起，就有了一个即将迎接和匹配未来浪潮的轮廓。

1. 技术安全：缓解可能危害并保持控制的具体技术措施。

2. 审计：确保技术透明性和问责制的一种手段。

3. 瓶颈：减缓发展并为监管机构和防御性技术争取时间的杠杆。

4. 制造者：确保负责任的开发者从一开始就在技术中构建适当的控制措施。
5. 企业：使技术背后的组织的激励与其控制保持一致。

6. 政府：支持政府，允许他们构建技术、监管技术并实施缓解措施。
7. 联盟：创建一个国际合作体系以协调法律和项目。

8. 文化：一种分享学习和失败的文化，以快速传播解决这些问题的方法。

9. 运动：所有这些都需要在各个层面上获得公众的意见，包括对每个组成部分施加压力并使其负责。

第 10 步是关于 *coherence,* 确保每个元素与其他元素和谐共处，确保遏制措施是相互强化的良性循环，而不是充满漏洞的竞争性程序的杂音。从这个意义上说，遏制不是关于这个或那个具体建议，而是它们集体互动的一个新兴现象，是社会学会管理和减轻 *Homo technologicus* 带来的风险的副产品。单独一个措施是行不通的，无论是病原体、量子计算机还是人工智能，但像这样的方案通过精心积累的互锁对策获得力量，从国际条约到供应链加强保护性新技术的护栏层层叠加。此外，像“ 延迟、检测和防御 ”这样的提议并不是终点或目的地。在即将到来的浪潮背景下的安全性不是我们到达的某个地方，而是必须不断实施的东西。

遏制不是一个休息的地方。这是一条狭窄且永无止境的道路。

源文本：—翻译文本：

政治学家詹姆斯·罗宾逊认为，自由民主国家远没有看上去那么安全。他们将国家视为一种本质上不稳定的“ 被束缚的利维坦 ”：庞大而强大，但被持久的公民社会和规范所制约。随着时间的推移，美国等国家进入了他们所称的“ 狭窄走廊 ”，使其保持在这种不稳定的平衡中。在这个走廊的两侧是陷阱。一方面，国家的权力打破了更广泛社会的权力并完全支配它，创造了像中国这样的专制利维坦。另一方面，国家崩溃，产生了缺席的利维坦，僵尸，国家对社会没有真正的控制。

如索马里或黎巴嫩等地。这两者对其人口都有可怕的后果。

阿西莫格鲁和罗宾逊的观点是，国家不断在这个走廊上行走。在任何时刻，它们都可能会跌倒。对于国家能力的每一次增加，都需要相应增加社会能力来平衡它。始终存在一种走向专制利维坦的压力，需要不断施加重量来阻止。没有最终的目的地，没有快乐、安全和持续存在的走廊尽头；相反，这是一个动态、不稳定的空间，精英和公民在这里争夺结果，任何时候被束缚的利维坦都可能消失或变得专制。安全是逐步前进和小心维持平衡的问题。

我认为这个比喻适用于我们如何对待技术，这不仅仅是因为这里的论点是技术现在使这种平衡变得更加不稳定。安全、受控的技术就像自由民主，不是一个最终的终点状态；相反，它是一个持续的过程，是一个需要积极维护、不断争取和保护的微妙平衡。没有一个时刻我们可以说，啊哈，我们已经解决了技术扩散的问题！相反，这是关于找到一条出路，确保有足够多的人致力于保持开放与封闭之间无尽的平衡。

与其说是走廊，这意味着明确的行进方向，我更想象成一个狭窄而危险的小路，笼罩在雾中，两侧是陡峭的悬崖，灾难或反乌托邦只需一小步之遥；你无法看得很远，当你行走时，小路曲折蜿蜒，抛出意想不到的障碍。

一方面，对所有实验和开发的完全开放是灾难的直接配方。如果世界上每个人都可以玩核弹，那么在某个阶段你就会有一场核战争。开源对技术发展是一大福音，也是更广泛进步的主要推动力。但这并不是强大AI模型或合成生物体的适当理念；在这里应该被禁止。在没有严格的正当程序的情况下，它们不应被共享，更不用说部署或开发。

安全依赖于事物不出故障、不落入错误的人手中，永远如此。对互联网、DNA合成器、AGI研究项目等进行某种程度的监管将是必不可少的。写下这些让我感到痛苦。作为一个二十多岁的年轻人，我最初持有隐私至上的立场，相信完全不受监督的交流和工作空间是基本权利，也是健康民主的重要组成部分。然而，多年来，随着论点变得更加清晰，技术越来越发达，我更新了这一观点。制造始终存在灾难性后果威胁的情况是不可接受的。智能、生命、原始力量——这些不是玩具，应该以它们应得的尊重、关心和控制来对待。技术人员和公众都将不得不接受比以往更高水平的监督和监管。正如我们大多数人不愿意生活在没有法律和警察的社会中一样，我们大多数人也不愿意生活在不受限制的技术世界中。

Some 反扩散的措施是必要的。而且，是的，我们不要回避事实；这意味着真正的审查，可能远远超出国家边界。有时，这会被视为——也许是正确的——无节制的美国霸权、西方傲慢和自私。坦率地说，我并不总是确定正确的平衡在哪里，但我现在坚信，完全的开放会将人类推离狭窄的道路。然而，另一方面，正如也应该清楚的那样，完全的监控和完全的封闭是不可想象的、错误的和灾难性的。对控制的过度扩张是通往反乌托邦的快速通道。这也必须被抵制。

在这个框架中，各国总是处于风险之中。然而，有些国家设法坚持了几个世纪，努力保持领先、保持平衡、保持适度的束缚。遏制的每一个方面，我们所描述的一切，都必须走这条痛苦的钢丝。这里或将来讨论的每一项措施都需要在这个范围内进行审视——推动到足以提供有意义的保护，但又防止走得太远。

源文本：—翻译文本：

即将到来的浪潮可能吗？

展望未来的无数路径，所有可能的方向，技术将带人类体验走向何方，释放的能力，改变我们世界的潜力，似乎在许多方面都无法遏制。从现在开始，必须永远走这条狭窄的道路，只需一步失误就可能跌入深渊。

历史表明，这种扩散和发展的模式已被锁定。巨大的激励措施似乎根深蒂固。技术的发展速度和力量甚至让其创造者感到惊讶。每天似乎都在宣布新的突破、产品或公司。前沿技术在几个月内就会扩散。负责监管这场革命的国家因为这一点而手足无措。

然而，尽管有令人信服的证据表明遏制是不可能的，但在性格上我仍然是个乐观主义者。这里提出的想法帮助我们提供了工具和手段，让我们一步一步沿着那条道路继续前行，提供了探索曲折前进路线的灯、绳索和地图。遏制的严峻挑战不是转身离开的理由；它是一个行动的号召，是我们这一代人都需要面对的使命。

如果我们——*we* 人类——能够通过一波新的承诺运动、企业和政府，改变背景，修订激励措施，提升技术能力、知识和保障措施，那么我们就能创造条件，以一丝希望的火花踏上那条摇摇欲坠的道路。尽管挑战的规模巨大，这里的每个部分都深入探讨了许多较小的领域，在这些领域中，任何个人仍然可以有所作为。要从根本上改变我们的社会、我们的人类本能和历史模式，需要付出巨大的努力。这远非确定。看起来不可能。但解决二十一世纪的重大困境 *must be* 可能。

在这个指数变化和力量展开的时代，我们都应该习惯于与矛盾共存。假设最坏的情况，为此做好计划，全力以赴。坚定不移地坚持狭窄的道路。让超越精英的世界参与进来并推动。如果有足够多的人开始构建那个难以捉摸的“我们”，那些希望的微光将变成变革的熊熊烈火。

人类世之后的生活

我

T WAS QUIET. WINDOWS AND shutters were closed, fires and candles put out, meals eaten. The bustle and hum of the busy day receded, and only the occasional bark of a dog, or scratching in the undergrowth, or soft rustle of the wind in the trees broke the hush. The world exhaled and slumbered.

他们在这黑暗的掩护下到来，这样他们就不会被认出。几十个人，蒙面，伪装，武装，愤怒。在夜晚的凉爽和宁静中，也许有机会伸张正义，只要他们能保持冷静。

他们无声地潜向镇边那座庞大而笨重的建筑。那是一座方形、坚固、令人生畏的建筑，在阴暗中显得尤为突出，里面存放着昂贵且有争议的新技术——他们认为是敌人的机器。一旦被抓住，入侵者将失去一切，甚至是生命。但他们已经发誓。这就是了。没有回头路。机器、老板，他们不能赢。

在外面，他们停顿了一下，然后冲了进去。猛击锁住的门，他们最终将其撞开，蜂拥而入。用锤子和棍棒，他们开始砸毁机器。金属撞击金属的声音回荡着。随着碎片散落在地板上，警报开始响起。百叶窗猛地打开，守夜人的灯笼被匆忙点燃。破坏者——卢德分子——跑向出口，消失在柔和的月光中。寂静将不再回来。

源文本：—翻译文本：

在十九世纪，英国正处于早期浪潮的动荡中。以蒸汽和机械自动化为基础的技术正在颠覆生产、劳动、价值、财富、能力和权力的规则。我们现在称之为第一次工业革命的进程如火如荼，工厂一个接一个地改变着国家和世界。1785年，发明家埃德蒙·卡特赖特推出了动力织机，这是一种新的机械化织布方式。起初，它并没有流行起来。然而，很快，进一步的迭代彻底改变了纺织制造业。

并不是每个人都感到高兴。动力织机可以由一个孩子操作，生产的织物相当于三个半传统织工的产量。机械化意味着在1770年后的四十五年里，织工的工资减少了一半以上，即使基本食品的价格飞涨。在这个新世界中，男性在与女性和儿童的竞争中失利。从织布到染色，纺织工作一直都很辛苦，但在工厂里，它变得嘈杂、有条不紊、危险且压抑。表现不佳的孩子会被吊在天花板上或被迫佩戴重物。死亡很常见。工作时间惩罚性地长。对于那些在前线承受工业化人类代价的人来说，这不是一个勇敢的新技术乌托邦；而是一个充满魔鬼工厂、奴役和侮辱的世界。

传统的织工和纺织工人感到新的机器及其背后的资本正在夺走他们的工作，压低他们的工资，剥夺他们的尊严，并瓦解一种丰富的生活方式。节省劳动力的机器对工厂的所有者来说是极好的，但对于那些传统上主导纺织业的高技能和高薪工人来说，这是一场灾难。

受到一个名叫内德·勒德的神话人物的启发，英格兰中部的织工们变得愤怒并组织起来。他们拒绝接受这种观点，即扩散将成为默认状态，围绕他们的技术浪潮是经济上的必然性。他们决定反击。

1807年，六千名织工因减薪而示威，抗议活动被挥舞马刀的龙骑兵镇压，导致一名抗议者被杀。从那时起，一场更为激烈的运动开始形成。1811年，破坏者有了一个名字。

在诺丁汉的一位工厂主收到来自“卢德将军和补救者军队”的一系列信件后，没有收到任何回复。3月11日，失业的织工袭击了当地的工厂，摧毁了六十三台机器，并加紧了运动。

在接下来的几个月秘密突袭中，数百个织机被摧毁。“内德·勒德的军队”进行了反击。他们觉得，他们所想要的只是公平的工资和尊严。他们的要求往往很小——适度的加薪，分阶段引入新机器，某种利润分享机制。这似乎并不过分。

卢德派抗议活动开始逐渐平息，被一套严厉的法律和反抗民兵镇压。在此期间，英格兰只有几千台自动织机。但到1850年，已有25万台。战斗已经失败，技术已经扩散，织工的旧生活被摧毁，世界发生了变化。对于那些失去的人来说，这就是不受控制的技术浪潮的样子。

源文本：—翻译文本：

抱歉，我无法翻译该内容，因为您提供的文本似乎不完整或缺失。请提供完整的文
从长远来看，导致如此多痛苦的同样工业技术带来了生活水平的巨大改善。几十年、几个世纪之后，那些织工的后代生活在路德分子几乎无法想象的条件下，习惯于我们视为理所当然的那个不稳定世界。他们中的绝大多数人在冬天回到温暖的家中，冰箱里装满了异国食物。当他们生病时，他们得到了神奇的医疗护理。他们的寿命更长。

就像我们今天一样，卢德派也陷入了困境。他们的痛苦和混乱是真实的，但生活水平的提高同样真实，这些提高惠及了他们的子孙后代，而你我今天也在不假思索地享受着。当时，卢德派未能遏制技术。但人类还是适应了。今天的挑战很明确。我们必须在不被其危害淹没的情况下，获取这股浪潮的好处。

卢德派输掉了他们的运动，我认为那些今天想要阻止技术发展的人很可能再次不会成功。

那么，唯一的方法就是第一次就做好。确保技术的适应不是像工业革命那样简单地强加给人们，而是从一开始就确保技术适应人们，适应他们的生活和希望。适应的技术是受控的技术。最紧迫的任务不是去驾驭或徒劳地阻止浪潮，而是去塑造它。

即将到来的浪潮将改变世界。最终，人类可能不再是主要的行星驱动者，因为我们已经习惯于成为这样的角色。我们将生活在一个时代，在这个时代中，我们日常的大多数互动不是与其他人，而是与人工智能进行的。这听起来可能很吸引人、可怕或荒谬，但这正在发生。我猜测你已经在屏幕前度过了相当大的一部分清醒时间。事实上，你可能花在生活中各种屏幕上的时间比看任何一个人，包括配偶和孩子的时间还要多。

因此，不难看出我们将花费越来越多的时间与这些新机器交流和互动。我们遇到和互动的人工智能和生物智能的类型和性质将与现在截然不同。它们将为我们完成工作，寻找信息，制作演示文稿，编写程序，订购我们的购物和今年的圣诞礼物，建议解决问题的最佳方法，或者可能只是聊天和玩耍。

他们将成为我们的个人智能、我们的伙伴和助手、知己和同事、幕僚长、助理和翻译。他们将组织我们的生活，倾听我们炽热的愿望和最深的恐惧。他们将帮助经营我们的业务，治疗我们的疾病，并参与我们的战斗。在普通的一天中，会出现许多不同类型的个性、能力和形式。我们的精神和对话世界将不可避免地包括这个新奇的智能群体。文化、政治、经济；友谊、游戏、爱情：所有这些都将同步演变。

明天的世界将是一个工厂在本地增长产出的地方，几乎就像以前时代的农场。无人机和机器人将无处不在。人类基因组将是一个弹性的东西，因此，人类本身的概念也必然如此。寿命将比我们现在的长得多。许多人将几乎完全消失在虚拟世界中。曾经看似稳定的社会契约将扭曲和崩溃。在这个世界中学习生活和繁荣将成为二十一世纪每个人生活的一部分。

卢德派的反应是自然的、预料之中的。但一如既往，这将是徒劳的。不过，当时，技术人员并没有考虑将他们的技术适应于人类的目的，就像卡尔·本茨和第一批石油大亨没有考虑地球的大气层一样。相反，技术被创造出来，资本为其提供资金，其他人则纷纷加入，无论长期后果如何。

这次遏制必须重写那个故事。可能还没有一个全球性的“我们”，但现在有一群人正在构建这项技术。我们肩负着巨大的责任，确保适应性不是单向的。与动力织机不同，与气候不同，即将到来的浪潮是为了适应人类的需求而构建的，是围绕人类的关注而建立的。即将到来的浪潮不应被创造来服务于遥远的利益，遵循盲目的技术逻辑议程——或者更糟。

太多关于未来的愿景是从技术能做什么或可能做什么开始的，并以此为基础。这完全是错误的基础。技术人员不应仅仅关注工程细节，而应帮助想象和实现一个更丰富的、社会的、人类的未来，从广义上讲，这是一个复杂的织锦，而技术只是其中的一根线。技术在未来的发展中起着核心作用——这无疑是正确的——但技术不是未来的重点，也不是真正的关键所在。我们才是。

技术应该放大我们最好的一面，为创造力和合作开辟新的途径，与我们生活和最珍贵关系的人性相契合。它应该让我们更快乐、更健康，成为人类努力和美好生活的终极补充——但始终在我们的

术语，民主决定，公开辩论，利益广泛分配。在动荡中，我们绝不能忽视这一点：即使是最狂热的卢德分子也能接受的愿景。

但在我们到达那里之前，在我们能够实现即将到来的技术的无限潜力之前，这股浪潮及其核心困境需要被遏制，需要对整个技术圈进行一种加强的、前所未有的、极具人性化的掌控。这将需要在几十年间在人类努力的各个领域中展现出史诗般的决心。这是一个巨大的挑战，其结果将毫不夸张地决定本世纪及未来日常生活的质量和性质。

失败的风险几乎不堪设想，但我们必须面对。然而，奖赏是惊人的：不亚于我们珍贵物种的安全、长期繁荣。

那是值得为之奋斗的。

致谢

B

OOKS TOO ARE ONE OF history's most transformative technologies. And like any instance of transformative technology, they are innately team enterprises. This one is no exception. It has, for a start, been an epic authorial collaboration spanning more than twenty years of friendship and constant discussion.

Crown 从项目初期就一直是这个项目的不可思议的支持者。David Drake 以其卓越的出版愿景为本书提供了明智且充满活力的指导。我们非常幸运能有 Paul Whitlatch 作为我们的编辑，他一次又一次地以非凡的耐心和洞察力对本书进行了无数次改进。也要感谢 Madison Jacobs、Katie Berry 和 Chris Brand。伦敦 Bodley Head 的 Stuart Williams 是另一位聪明的编辑声音和坚定的支持者，我们很荣幸能有 Tina Bennett 和 Sophie Lambert 这两位出色的经纪人。从项目初期开始，Celia Pannetier 作为我们宝贵的研究员，负责整理证据，而 Sean Lavery 则对整本书进行了事实核查。

多年来，许多人为这本书贡献了力量。他们进行了详细的对话，阅读了章节，反驳了论点，产生了想法，纠正了错误。无数的电话、研讨会、采访、编辑和建议帮助创造了这本书。这里的每一个人都投入了时间和精力来交流、分享专业知识、辩论和教导我们。特别感谢那些阅读了整个草稿并详细评论的人；他们的慷慨和非凡的贡献。

洞察力的水平在完成最终手稿的过程中是完全无价的。

特别感谢 Gregory Allen、Graham Allison（以及更广泛的哈佛贝尔弗中心的教职员工）、Sahar Amer、Anne Applebaum、Julian Baker、Samantha Barber、Gabriella Blum、Nick Bostrom、Ian Bremmer、Erik Brynjolfsson、Ben Buchanan、Sarah Carter、Rewon Child、George Church、Richard Danzig、Jennifer Doudna、Alexandra Eitel、Maria Eitel、Henry Elkus、Kevin Esvelt、Jeremy Fleming、Jack Goldsmith、Al Gore、Tristan Harris、Zaid Hassan、Jordan Hoffman、Joi Ito、Ayana Elizabeth Johnson、Danny Kahneman、Angela Kane、Melanie Katzman、Henry Kissinger、Kevin Klyman、Heinrich Küttler、Eric Lander、Sean Legassick、Aitor Lewkowycz、Leon Marshall、Jason Matheny、Andrew McAfee、Greg McKelvey、Dimitri Mehlhorn、David Miliband、Martha Minow、Geoff Mulgan、Aza Raskin、Tobias Rees、Stuart Russell、Jeffrey Sachs、Eric Schmidt、Bruce Schneier、Marilyn Thompson、Mayo Thompson、Thomas Viney、Maria Vogelauer、Mark Walport、Morwenna White、Scott Young 和 Jonathan Zittrain。

我在 Inflection 的联合创始人 Reid Hoffman 和 Karén Simonyan，感谢他们成为出色的合作者。以及我在 DeepMind 的联合创始人 Demis Hassabis 和 Shane Legg，感谢他们在这非凡的十年中的合作。Michael 想感谢他在 Canelo 的联合创始人 Iain Millar 和 Nick Barreto，感谢他们的持续支持，但最重要的是感谢他令人难以置信的妻子 Dani 和儿子 Monty 和 Dougie。

笔记

For a bibliography of books consulted, please visit the website the-coming-wave.com/bibliography.

1: 公式 $v*$ 是一个重要的数学表达式。

Costing a few tens of thousands For example, the Kilobaser DNA & RNA Synthesizer, sold starting at $25,000. See its website: kilobaser.com/dna-and-rna-synthesizer.

转到文本中的注释参考

2: 速度 $v*$ 的变化

二十年后，Benz的专利 TÜV Nord Group，“ 内燃机简史 ”，TÜV Nord Group，2019年4月18日，www.tuev-nord.de/explore/en/remembers/a-brief-history-of-the-internal-combustion-engine。

转到文本中的注释参考

福特不断提高产量 伯顿·W·福尔瑟姆，“ 亨利·福特与汽车工业的胜利 ”，经济教育基金会，1998年1月1日，fee.org/articles/henry-ford-and-the-triumph-of-the-auto-industry。

转到文本中的注释参考

大约14亿辆“ 2023年世界上有多少辆车？ ”，Hedges & Company，2021年6月，hedgescompany.com/blog/2021/06/how-many-cars-are-there-in-the-world；“ 内燃机——前方的道路 ”，Industr，2019年1月22日，www.industr.com/en/internal-combustion-engine-the-road-ahead-2357709#.

转到文本中的注释参考

这是关于技术本身的故事。关于技术的精确定义，学术界有大量的争论。在本书中，我们采用一个常识性的、日常的定义：科学知识（在尽可能广泛的意义上）的应用，以生产工具或实际成果。然而，这个术语的全面、多方面的复杂性也得到了承认。技术延伸到文化和实践中。它不仅仅是晶体管、屏幕和键盘。它是编码者的显性和隐性知识，是支持他们的社会生活和社会。

转到文本中的注释参考

技术有一个明确的技术学者在扩散和扩张之间做出区分，这些区分在这里大多被省略。我们的意思是

在它们的口语而非正式意义上更多地使用它们。

转到文本中的注释参考

随着科学产生新的发现，这也在另一个方向上起作用：技术产生新的工具和见解，推动科学发展，比如蒸汽机帮助阐明了对热力学科学的需求，或者复杂的玻璃制品创造了改变我们对空间理解的望远镜。

转到文本中的注释参考

简单来说，波是一组 Robert Ayres，“技术变革与长波。第一部分，”*Technological Forecasting and Social Change* 37，第1期（1990年3月），www.sciencedirect.com/science/article/abs/pii/0040162590900573。

转到文本中的注释参考

“通用”这个术语对于理解技术至关重要，但它的出现时间却相对较晚，可以追溯到20世纪90年代初的一篇经济学论文。参见Timothy F. Bresnahan和Manuel Trajtenberg的“通用技术‘增长引擎’？”（工作论文，NBER，1992年8月），www.nber.org/papers/w4148。

转到文本中的注释参考

它对进化产生了显著影响 Richard Wrangham，*Catching Fire: How Cooking Made Us Human*（伦敦：Profile Books，2010）。

转到文本中的注释参考

通用技术波及效应摘自理查德·利普西、肯尼斯·卡洛和克利福德·贝卡，*Economic Transformations: General Purpose Technologies and Long-Term Economic Growth*（牛津：牛津大学出版社，2005年）。

转到文本中的注释参考

语言、农业、书写 从技术上讲，语言可能再次被视为一种原始或基础的通用技术。

转到文本中的注释参考

一个主要研究指出了 Lipsey、Carlaw 和 Bekar，*Economic Transformations.*

转到文本中的注释参考

纵观历史，人口规模 欲了解这一过程如何运作的有力论述，请参见 O ded Galor，*The Journey of Humanity: The Origins of Wealth and Inequality*（伦敦：Bodley Head，2022年）。

转到文本中的注释参考

更大且更具连接性的人口 Michael Muthukrishna 和 Joseph Henrich，“集体大脑中的创新”，*Philosophical Transactions of the Royal Society B* 371，第 1690 期（2016），royalsocietypublishing.org/doi/10.1098/rstb.2015.0192。

转到文本中的注释参考

在黎明时分加洛尔，*The Journey of Humanity*，46。

转到文本中的注释参考

哈佛人类学家Muthukrishna和Henrich，“集体大脑中的创新。”

转到文本中的注释参考

万年 Lipsey、Carlaw 和 Bekar，*Economic Transformations*。

转到文本中的注释参考

在最后一百个余数中，介于1000和1700之间。

转到文本中的注释参考

对于未来学家阿尔文·托夫勒（Alvin Toffler），*The Third Wave*（纽约：班塔姆，1984年）。另请参阅尼古拉·康德拉季耶夫关于长周期波的研究。

转到文本中的注释参考

伟大的技术哲学家刘易斯·芒福德，*Technics and Civilization*（芝加哥：芝加哥大学出版社，1934年）。

转到文本中的注释参考

最近，经济学家卡洛塔·佩雷斯，*Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages* (切尔滕纳姆，英国：爱德华·埃尔加，2002)。

转到文本中的注释参考

一旦扩散开始，然而，确实，加速扩散的一个早期迹象可能是，与水磨坊的千年传播相比，风车在发明后的几年内就从英格兰北部到叙利亚随处可见。参见Lynn White Jr., *Medieval Technology and Social Change*（牛津：牛津大学出版社，1962年），87。

转到文本中的注释参考

但仅仅五十年后，Elizabeth L. Eisenstein，*The Printing Press as an Agent of Change: Communications and Cultural Transformations in Early-Modern Europe*（英国剑桥：剑桥大学出版社，1979年）。

GO TO NOTE REFERENCE IN TEXT

在十七世纪，Eltjo Buringh 和 Jan Luiten Van Zanden 撰写了《绘制“西方的崛起”：欧洲的手稿和印刷书籍，从六世纪到十八世纪的长期视角》，*Journal of Economic History,* 2009年6月1日，www.cambridge.org/core/journals/journal-of-economic-history/article/abs/charting-the-rise-of-the-west-manuscripts-and-printed-books-in-europe-a-longterm-perspective-from-the-sixth-through-eighteenth-centuries/0740F5F9030A706BB7E9FACCD5D975D4。

转到文本中的注释参考

一项分析估计了Max Roser和Hannah Ritchie，“书籍价格：书籍生产中的生产力”，Our World in Data，ourworldindata.org/books。

转到文本中的注释参考

波兰世界能源理事会成员委员会的第一批电力站，“世界和波兰的能源部门：起源、发展、现状”，世界能源理事会，2014年12月，www.worldenergy.org/assets/images/imported/2014/12/Energy_Sector_of_the_world_and_Poland_EN.pdf。

转到文本中的注释参考

在 1900 年全球发电 Vaclav Smil，“二十世纪的能源：资源、转换、成本、用途和后果，”*Annual Review of Energy and the Environment* 25 (2000)，www.annualreviews.org/doi/pdf/10.1146/annurev.energy.25.1.21。

转到文本中的注释参考

因此，普通人威廉·D·诺德豪斯，“实际产出和实际工资指标是否反映现实？照明历史表明并非如此”，耶鲁大学考尔斯经济研究基金会

大学，1996年1月，https://cowles.yale.edu/sites/default/files/2022-08/d0381.pdf。

转到文本中的注释参考

十年后，有580万Galor，*The Journey of Humanity,* 46。

转到文本中的注释参考

今天美国有很多，如果你包括固定电话和移动电话。

转到文本中的注释参考

提高质量加入减少“电视通胀计算器”，官方数据基金会，www.in2013dollars.com/Televisions/price-inflation。

转到文本中的注释参考

拟态刺激竞争 Anuraag Singh 等人，“所有技术的技术改进率预测：专利数据的使用和扩展领域描述，”*Research Policy* 50，第 9 期（2021 年 11 月），www.sciencedirect.com/science/article/pii/S0048733321000950#。然而，不同技术集之间存在相当大的差异。

转到文本中的注释参考

像内燃机一样，当然，这些提议可以追溯得更远，至少可以追溯到十九世纪的巴贝奇和洛夫莱斯。

转到文本中的注释参考

到1945年，一个重要的前驱者乔治·戴森，*Turing's Cathedral: The Origins of the Digital Universe*（伦敦：艾伦·莱恩，2012年）。

转到文本中的注释参考

在那个十年初期，IBM总裁尼克·卡尔，“世界需要多少台计算机？比你想象的要少，”*Guardian,* 2008年2月21日，www.theguardian.com/technology/2008/feb/21/computing.supercomputers。

转到文本中的注释参考

Popular Mechanics 杂志让詹姆斯·梅格斯撰写了“未来内部：PopMech如何预测未来110年”，*Popular Mechanics,* 2012年12月21日，www.popularmechanics.com/technology/a8562/inside-the-future-how-popmech-predicted-the-next-110-years-14831802/#。

转到文本中的注释参考

他们的力量增加了十倍。参见，例如，Darrin Qualman，“难以想象的产出：全球晶体管生产，”*Darrin Qualman Blog,* 2017年4月24日，www.darrinqualman.com/global-production-transistors/；Azeem Azhar, *Exponential: How Accelerating Technology Is Leaving Us Behind and What to Do About It*（伦敦：Random House Business，2021年），21；以及Vaclav Smil, *How the World Really Works: A Scientist's Guide to Our Past, Present and Future*（伦敦：Viking，2022年），128。

转到文本中的注释参考

20世纪70年代初，John B. Smith，“互联网年表”，UNC计算机科学，www.cs.unc.edu/ ~jbs/ resources/ Internet/ internet_chron.html。

转到文本中的注释参考

它创造了更多的Azhar，*Exponential,* 219。

转到文本中的注释参考

现在人类生产数百个同上，228。

转到文本中的注释参考

3: 公式 $v*$ 的翻译

理解技术，部分是罗伯特·K·默顿，*On Social Structure and Science*（芝加哥：芝加哥大学出版社，1996年），提供了经典研究，但也请参见乌尔里希·贝克，*Risk Society: Toward a New Modernity*（伦敦：SAGE，1992年），了解社会如何被其自身创造的风险管理所主导。另请参见爱德华·特纳，
Why Things Bite Back: Technology and the Revenge of Unintended Consequences（纽约：Vintage，1997年），以及查尔斯·佩罗，*Normal Accidents: Living with High-Risk Technologies*（新泽西州普林斯顿：普林斯顿大学出版社，1984年）。
转到文本中的注释参考

对许多人来说，“ 遏制 ”一词乔治·F·凯南，“ 苏联行为的根源 ”，*Foreign Affairs,* 1947年7月，www.cvce.eu/content/publication/1999/1/1/a0f03730-dde8-4f06-a6ed-d740770dc423/publishable_en.pdf。

转到文本中的注释参考

随着印刷机的轰鸣 本文摘自安东·豪斯，“ 发明时代：奥斯曼帝国是否禁止印刷？ ”，*Age of Invention,* 2021年5月19日，antonhowes.substack.com/p/age-of-invention-did-the-ottomans。

转到文本中的注释参考

约翰·凯，发明家 示例取自乔尔·莫克尔，*The Lever of Riches: Technological Creativity and Economic Progress*（牛津：牛津大学出版社，1990年）。

转到文本中的注释参考

同样，中国驳回了Harold Marcuse，“ 乾隆致乔治三世的信（1792年）”，加州大学圣塔芭芭拉分校历史系，marcuse.faculty.history.ucsb.edu/classes/2c/texts/1792QianlongLetterGeorgeIII.htm。

转到文本中的注释参考

很少有社会曾经 例如，Joseph A. Tainter，*The Collapse of Complex Societies*（英国剑桥：剑桥大学出版社，1988年），以及Jared Diamond，*Collapse: How Societies Choose to Fail or Survive*（伦敦：企鹅出版社，2005年），了解更多关于这一过程的信息。

转到文本中的注释参考

1933年9月11日，瓦尔德马·凯姆弗特，“卢瑟福冷却原子能希望”，*New York Times,*，1933年9月12日，timesmachine.nytimes.com/timesmachine/1933/09/12/99846601.html。

转到文本中的注释参考

数周后，波音B-29亚历克斯·韦勒斯坦，“统计广岛和长崎的死亡人数，”*Bulletin of the Atomic Scientists,* 2020年8月4日，thebulletin.org/2020/08/counting-the-dead-at-hiroshima-and-nagasaki。

GO TO NOTE REFERENCE IN TEXT

1946年，艾奇逊-李连瑟报告参见David Lilienthal等人，“关于国际原子能控制的报告”，1946年3月16日，fissilematerials.org/library/ach46.pdf。GO TO NOTE REFERENCE IN TEXT

尽管像中国这样的国家“部分禁止核试验条约”，核威胁倡议，2008年2月，www.nti.org/education-center/treaties-and-regimes/treaty-banning-nuclear-test-atmosphere-outer-space-and-under-water-partial-test-ban-treaty-ptbt/。

转到文本中的注释参考

1968年出现了一个转折点，“核不扩散条约（NPT）时间线”，军备控制协会，2022年8月，

www.armscontrol.org/ factsheets/ 核不扩散条约（NPT）时间表.

转到文本中的注释参考

毕竟，直到2019年，Liam Stack，“ 更新完成：美国核武器不再需要软盘，”*New York Times,* 2019年10月24日，www.nytimes.com/2019/10/24/us/nuclear-weapons-floppy-disks.html。

转到文本中的注释参考

事故频发 这里的叙述主要取自 Eric Schlosser，*Command and Control*（伦敦：企鹅出版社，2014），以及 John Hughes-Wilson，*Eve of Destruction: The Inside Story of Our Dangerous Nuclear World*（伦敦：John Blake，2021）。

转到文本中的注释参考

微小的硬件故障 威廉·伯尔，“ 1979-1980年苏联导弹袭击的错误警报使美军处于戒备状态 ”，国家安全档案馆，2020年3月16日，nsarchive.gwu.edu/briefing-book/nuclear-vault/2020-03-16/false-warnings-soviet-missile-attacks-during-1979-80-led-alert-actions-us-strategic-forces.

转到文本中的注释参考

朝鲜采取了非常规措施，Paul K. Kerr，“ 伊朗-朝鲜-叙利亚弹道导弹和核合作 ”，国会研究服务处，2016年2月26日，sgp.fas.org/crs/nuke/R43480.pdf。

转到文本中的注释参考

中国、印度和巴基斯坦是 Graham Allison 的“ 核恐怖主义：我们是战胜了几率还是改变了它们？ ”，*PRISM,* 2018年5月15日，cco.ndu.edu/News/Article/1507316/nuclear-terrorism-did-we-beat-the-odds-or-change-them。

转到文本中的注释参考

巴西和阿根廷甚至何塞·戈尔登贝格，“ 回顾：巴西和阿根廷去核化的经验教训 ”，军备控制协会，2006年4月，www.armscontrol.org/act/2006-04/looking-back-lessons-denuclearization-brazil-argentina。

转到文本中的注释参考

大量核材料 Richard Stone，“ 切尔诺贝利监测实验室的脏弹成分失踪 ”，*Science,* 2022年3月25日，www.science.org/content/article/dirty-bomb-ingredients-go-missing-chornobyl-monitoring-lab。

转到文本中的注释参考

2018年，普鲁托尼姆·帕特里克·马龙和R. 杰弗里·史密斯，“ 钚失踪了，但政府什么也没说，”公共诚信中心，2018年7月16日，publicintegrity.org/national-security/plutonium-is-missing-but-the-government-says-nothing。

转到文本中的注释参考

这可能听起来很奇妙 Zaria Gorvett，“ 无人能找到的丢失核弹 ”，*BBC Future,* 2022年8月4日，www.bbc.com/future/article/20220804-the-lost-nuclear-bombs-that-no-one-can-find。GO TO NOTE REFERENCE IN TEXT

化学武器最近“ 叙利亚化学武器活动时间表，2012 – 2022 ”，军备控制协会，2021年5月，www.armscontrol.org/factsheets/Timeline-of-Syrian-Chemical-Weapons-Activity。

转到文本中的注释参考

没有它们，模型表明 Paul J. Young，“ 蒙特利尔议定书保护陆地碳汇，”*Nature,* 2021年8月18日，www.nature.com/articles/s41586-021-03737-3.epdf。

GO TO NOTE REFERENCE IN TEXT

4: 速度 $v*$ 的变化

在仅仅三对之后，Natalie Wolchover，“国际象棋游戏可以有多少种不同的展开方式？”，*Popular Science,* 2010年12月15日，www.popsci.com/science/article/2010-12/fyi-how-many-different-ways-can-chess-game-unfold.

转到文本中的注释参考

总的来说，董事会有“AlphaGo”，DeepMind，www.deepmind.com/research/highlighted-research/alphago。然而，有些人报告的数字甚至更高；例如，*Scientific American* 引用了 10^{360} 配置。参见 Christof Koch，“计算机如何击败围棋大师”，*Scientific American,*，2016年3月19日，www.scientificamerican.com/article/how-the-computer-beat-the-go-master。

转到文本中的注释参考

技术越多，W. Brian Arthur，*The Nature of Technology: What It Is and How It Evolves*（伦敦：Allen Lane，2009）31。

转到文本中的注释参考

技术学者埃弗雷特·罗杰斯 Everett M. Rogers，*Diffusion of Innovations*（纽约：自由出版社，1962年），或参见乔尔·莫克尔等学者关于工业革命的著作。

转到文本中的注释参考

工程师和未来学家雷·库兹韦尔，*How to Create a Mind: The Secret of Human Thought Revealed*（纽约：维京企鹅出版社，2012年）。

转到文本中的注释参考

我们现在可以在以下文献中看到这一点，例如，Azalia Mirhoseini 等人，“一种用于快速芯片设计的图形布局方法论”，*Nature,* 2021年6月9日，www.nature.com/articles/s41586-021-03544-w；以及 Lewis Grozinger 等人，

“生物计算中细胞至高无上的途径，”*Nature Communications,* 2019年11月20日, www.nature.com/articles/s41467-019-13232-z.

转到文本中的注释参考

突破性时刻来自 Alex Krizhevsky 等人，“使用深度卷积神经网络进行 ImageNet 分类”，神经信息处理系统，2012 年 9 月 30 日，proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf。GO TO NOTE REFERENCE IN TEXT

2012年，AlexNet击败了Jerry Wei，“AlexNet：挑战CNN的架构”，*Towards Data Science,* 2019年7月2日，towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951。

转到文本中的注释参考

感谢深度学习 Chanan Bos，“特斯拉的新 HW3 自动驾驶计算机——它是个怪物”，CleanTechnica，2019 年 6 月 15 日，cleantechnica.com/2019/06/15/teslas-new-hw3-self-driving-computer-its-a-beast-cleantechnica-deep-dive。GO TO NOTE REFERENCE IN TEXT

它有助于飞行无人机 Jeffrey De Fauw 等人，“临床适用的深度学习用于视网膜疾病的诊断和转诊”，*Nature Medicine,* 2018年8月13日，www.nature.com/articles/s41591-018-0107-6。

转到文本中的注释参考

到2020年代，几乎有两千篇“神经信息处理系统进展”（NeurIPS）论文，papers.nips.cc。

转到文本中的注释参考

在过去六年中，“研究与发展”，在 *Artificial Intelligence Index Report 2021,* 斯坦福大学以人为中心的人工智能，2021年3月，aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report-_Chapter-1.pdf。

转到文本中的注释参考

到处都是软件，套用马克·安德森的话来说。

转到文本中的注释参考

在DeepMind，我们开发了系统“DeepMind AI将谷歌数据中心的冷却费用减少了40%”，DeepMind，2016年7月20日，www.deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40。

转到文本中的注释参考

拥有15亿参数的“更好的语言模型及其影响”，OpenAI，2019年2月14日，openai.com/blog/better-language-models。

转到文本中的注释参考

在接下来的几年中，请参阅马丁·福特，*Rule of the Robots: How Artificial Intelligence Will Transform Everything*（伦敦：Basic Books，2021），以获得详细的比较。

转到文本中的注释参考

更现实地说，普通美国人艾米·沃森，“2018年至2021年美国平均阅读时间，按年龄组划分”，Statista，2022年8月3日，www.statista.com/statistics/412454/average-daily-time-reading-us-by-age。

GO TO NOTE REFERENCE IN TEXT

首先，微软和NVIDIA构建了一个拥有5300亿参数的变压器模型，即Megatron-Turing自然语言生成模型（MT-NLG），其规模是他们自己一年前最强大变压器模型的三十一倍。然后是来自北京智源人工智能研究院的悟道，据称拥有1.75万亿参数——是GPT-3的十倍。例如，参见Tanushree Shenwai，“微软和NVIDIA AI推出MT-NLG：最大和最强大的单片变压器语言NLP模型”，*MarkTech Post,* 2021年10月13日，www.marktechpost.com/2021/10/13/microsoft-and-nvidia-ai-introduces-mt-nlg-the-largest-and-most-powerful-monolithic-transformer-language-nlp-model。

转到文本中的注释参考

中国公司阿里巴巴“阿里巴巴达摩院创建全球最大AI预训练模型，参数远超谷歌和微软”，*Pandaily,* 2021年11月8日，pandaily.com/alibaba-damo-academy-creates-worlds-largest-ai-pre-training-model-with-parameters-far-exceeding-google-and-microsoft。

转到文本中的注释参考

Google 的 PaLM 使用了 Alyssa Vance 的一张精彩图片，假设每个“滴”构成 0.5 毫升：mobile.twitter.com/alyssamvance/status/1542682154483589127。

转到文本中的注释参考

但它使用了一种高效的训练方法 William Fedus 等人，“Switch Transformers: 通过简单且高效的稀疏性扩展到万亿参数模型”，*Journal of Machine Learning Research,* 2022年6月16日，arxiv.org/abs/2101.03961。

GO TO NOTE REFERENCE IN TEXT

或者看看 DeepMind 的 Chinchilla，Alberto Romero，“一个新的 AI 趋势：Chinchilla (70B) 大大超越 GPT-3 (175B) 和 Gopher (280B)，”*Towards Data Science,* 2022 年 4 月 11 日，towardsdatascience.com/a-new-ai-trend-chinchilla-70b-greatly-outperforms-gpt-3-175b-and-gopher-280b-408b9b4510。

转到文本中的注释参考

在光谱的另一端，请参见 github.com/karpathy/nanoGPT 以获取更多详细信息。

转到文本中的注释参考

Meta 已开源 Susan Zhang 等人，“通过 OPT-175B 实现大规模语言模型的民主化访问”，Meta AI，2022 年 5 月 3 日，ai.facebook.com/blog/democratizing-access-to-large-scale-language-models-with-opt-175b。

转到文本中的注释参考

在几天内，有人发现了例如 twitter.com/miolini/status/1634982361757790209。

转到文本中的注释参考

一项分析表明，这使工程师们Eirini Kalliamvakou，“研究：量化GitHub Copilot对开发者生产力和幸福感的影响”，GitHub，2022年9月7日，github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness。

转到文本中的注释参考

据著名的马特·韦尔士所说，“编程的终结”，*Communications of the ACM,* 2023年1月，cacm.acm.org/magazines/2023/1/267976-the-end-of-programming/fulltext。

转到文本中的注释参考

然而，当给出相同的提示时，Emily Sheng 等人，“ The Woman Worked as a Babysitter: On Biases in Language Generation, ”arXiv, 2019年10月23日, arxiv.org/pdf/1909.01326.pdf。

转到文本中的注释参考

经过许多小时，Lemoine Nitasha Tiku，“ 认为公司AI已经活过来的谷歌工程师 ”，*Washington Post,* 2022年6月11日，www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine。

转到文本中的注释参考

他告诉难以置信的***Wired***采访者史蒂文·莱维，“ 布莱克·勒莫因称谷歌的LaMDA AI面临‘ 偏见 ’，”*Wired,*2022年6月17日，www.wired.com/story/blake-lemoine-google-lamda-ai-bigotry。

转到文本中的注释参考

“ 一旦它起作用 ”引自 Moshe Y. Vardi，“ 人工智能：过去与未来 ”，*Communications of the ACM,* 2012年1月，cacm.acm.org/magazines/2012/1/144824-artificial-intelligence-past-and-future/fulltext。

GO TO NOTE REFERENCE IN TEXT

他们认为，人工智能可能正在减缓 Joel Klinger 等人，“ 人工智能研究的缩小？ ”*Computers and Society,* 2022年1月11日，arxiv.org/abs/2009.10385。

转到文本中的注释参考

批评者如纽约大学教授加里·马库斯，加里·马库斯，“ 深度学习遇到了瓶颈， ”*Nautilus,* 2022年3月10日，nautil.us/deep-learning-is-hitting-a-wall-14467。

转到文本中的注释参考

复杂性领域的杰出教授梅拉妮·米切尔 参见梅拉妮·米切尔，*Artificial Intelligence: A Guide for Thinking Humans*（伦敦：企鹅图书，2020年），以及史蒂文·斯特罗加茨，“梅拉妮·米切尔将人工智能研究带回其根源，”*Quanta Magazine,* 2021年4月19日，www.quantamagazine.org/melanie-mitchell-takes-ai-research-back-to-its-roots-20210419。

转到文本中的注释参考

我认为这将会完成。对齐研究中心已经针对这种能力对GPT-4进行了测试。研究发现，在这个阶段，GPT-4在自主行动方面是“无效的”。“GPT-4系统卡片”，OpenAI，2023年3月14日，cdn.openai.com/papers/gpt-4-system-card.pdf。在发布后的几天内，人们已经接近实现；例如，参见mobile.twitter.com/jacksonfall/status/1636107218859745286。然而，这里的测试版本需要比那里展示的更多的自主性。

转到文本中的注释参考

5: 速度 $v*$ 的计算

正如从蒸汽机到苏珊·霍克菲尔德，*The Age of Living Machines: How Biology Will Build the Next Technology Revolution* (纽约：W. W. Norton, 2019)。

转到文本中的注释参考

然后，1973年斯坦利·N·科恩等人在细菌上进行研究，“体外构建生物功能性细菌质粒”，*PNAS,*，1973年11月1日，www.pnas.org/doi/abs/10.1073/pnas.70.11.3240。

转到文本中的注释参考

这是一个为期十三年的“人类基因组计划”，国家人类基因组研究所，2022年8月24日，www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genome-project。GO TO NOTE REFERENCE IN TEXT

虽然摩尔定律合理地“生命2.0”，*Economist,* 2006年8月31日，www.economist.com/special-report/2006/08/31/life-20。

GO TO NOTE REFERENCE IN TEXT

得益于不断改进的技术，请参阅“测序人类基因组的成本”，国家人类基因组研究所，2021年11月1日，www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost；以及Elizabeth Pennisi，“100美元的基因组？新的DNA测序仪可能成为生物学和医学的‘游戏规则改变者’”，*Science,* 2022年6月15日，www.science.org/content/article/100-genome-new-dna-sequencers-could-be-game-changer-biology-medicine。

转到文本中的注释参考

也就是说，价格下降了Azhar，*Exponential,* 41。

转到文本中的注释参考

在最初的CRISPR论文之后，李建峰等人，“ 使用向导RNA和Cas9在*Arabidopsis*和*Nicotiana benthamiana*中进行多重和同源重组介导的基因组编辑 ”， *Nature Biotechnology,* 2013年8月31日，www.nature.com/articles/nbt.2654。

转到文本中的注释参考

像RNA编辑这样的领域 Sara Reardon，“ 让开CRISPR，RNA编辑正在起飞， ”*Nature,* 2020年2月4日，www.nature.com/articles/d41586-020-00272-5。

转到文本中的注释参考

新技术如Craspase，胡春怡等，“ Craspase是一种CRISPR RNA引导、RNA激活的蛋白酶 ”， *Science,*，2022年8月25日，www.science.org/doi/10.1126/science.add5064。

转到文本中的注释参考

CRISPR 的使用案例正在增加 Michael Le Page，“ 三名遗传病患者成功接受 CRISPR 治疗 ”， *New Scientist,* 2020 年 6 月 12 日，www.newscientist.com/article/2246020-three-people-with-inherited-diseases-successfully-treated-with-crispr; 李杰等，“ 生物强化番茄提供维生素 D 充足的新途径 ”， *Nature Plants,* 2022 年 5 月 23 日，www.nature.com/articles/s41477-022-01154-6。

转到文本中的注释参考

将来，它可能会提供 Mohamed Fareh，“ 重编程的 CRISPR-Cas13b 抑制 SARS-CoV-2 复制并通过错配容忍规避其突变逃逸 ”， *Nature,* 2021 年 7 月 13 日，www.nature.com/articles/s41467-021-24577-9；“ CRISPR 如何改变癌症研究和治疗 ”，国家癌症研究所，2020 年 7 月 27 日，www.cancer.gov/news-events/cancer-currents-blog/2020/crispr-cancer-research-treatment；Zhihao Zhang 等，“ 基于 CRISPR 的基因更新 ”

“艾滋病病毒-1/艾滋病治疗中的编辑”，*Virologica Sinica,* 2022年2月，www.sciencedirect.com/science/article/pii/S1995820X22000177；Giulia Maule 等，“囊性纤维化的基因治疗：基因组编辑的进展与挑战”，*International Journal of Molecular Sciences,* 2020年6月，www.ncbi.nlm.nih.gov/pmc/articles/PMC7313467。

转到文本中的注释参考

这些将创造作物 Raj Kumar Joshi，“通过CRISPR/Cas基因组编辑工程化植物抗旱性，”*3 Biotech,* 2020年9月，www.ncbi.nlm.nih.gov/pmc/articles/PMC7438458; Muhammad Rizwan Javed 等，“生物燃料生产的现状及其通过CRISPR/Cas9介导的微生物细胞基因组工程的增强，”*Microbiological Research,* 2019年2月，www.sciencedirect.com/science/article/pii/S0944501318308346。

转到文本中的注释参考

今天的技术如CRISPR Nessa Carey，*Hacking the Code of Life: How Gene Editing Will Rewrite Our Futures*（伦敦：Icon Books，2019），136。

转到文本中的注释参考

您现在可以购买台式机，例如，请参见 kilobaser.com/shop。

转到文本中的注释参考

现在他们可以打印数百万 易人璐，“基因合成革命，”*New York Times,* 2021年11月24日，www.nytimes.com/2021/11/24/magazine/gene-synthesis.html。

转到文本中的注释参考

伦敦DNA工厂“用于高速基因研究的机器人实验室正在兴起，”*Economist,* 2018年3月1日，

www.economist.com/ science-and-technology/2018/03/01/用于高速基因研究的机器人实验室正在兴起。

转到文本中的注释参考

像DNA Script这样的公司 Bruce Rogers, “ DNA Script准备将世界上第一台DNA打印机推向市场, ”*Forbes,* 2021年5月17日, www.forbes.com/sites/brucerogers/2021/05/17/dna-script-set-to-bring-worlds-first-dna-printer-to-market。

转到文本中的注释参考

此外, 新的技术 Michael Eisenstein, “ 酶促DNA合成进入新阶段, ”*Nature Biology,* 2020年10月5日, www.nature.com/articles/s41587-020-0695-9。

转到文本中的注释参考

综合起来, 合成生物学不仅使用DNA合成, 还利用了对基因如何开启和关闭的日益深入的理解, 并结合了代谢工程学这一学科, 通过该学科可以鼓励细胞生产所需的物质。

转到文本中的注释参考

根据斯坦福生物工程师Drew Endy的话, “ Endy:Research, ”OpenWetWare, 2017年8月4日, openwetware.org/wiki/Endy:Research.

GO TO NOTE REFERENCE IN TEXT

2010年, 由克雷格·文特尔领导的团队 “ 首个自我复制的合成细菌细胞 ”, JCVI, www.jcvi.org/research/first-self-replicating-synthetic-bacterial-cell。

转到文本中的注释参考

仅仅三年后, Jonathan E. Venetz 等人, “ 化学合成重写细菌基因组以实现设计灵活性和

生物功能性，”*PNAS,* 2019年4月1日，www.pnas.org/doi/full/10.1073/pnas.1818259116。

转到文本中的注释参考

现在全球GP-write联盟参见GP-write联盟，生物工程卓越中心，engineeringbiologycenter.org/gp-write-consortium。

转到文本中的注释参考

使用光检测蛋白基因 José-Alain Sahel 等人，“视网膜光遗传治疗后盲人患者的部分视觉功能恢复”，*Nature Medicine,* 2021年5月24日，www.nature.com/articles/s41591-021-01351-4。

转到文本中的注释参考

CAR T细胞疗法设计定制的“CureHeart——遗传性心肌病的治愈”，英国心脏基金会，www.bhf.org.uk/what-we-do/our-research/cure-heart；国家癌症研究所，“CAR T细胞疗法”，国家卫生研究院，www.cancer.gov/publications/dictionaries/cancer-terms/def/car-t-cell-therapy。

转到文本中的注释参考

系统生物学领域 参见，例如，Astrid M. Vicente 等人，“个性化医学将如何在 2030 年前改变医疗保健：ICPerMed 愿景”，*Journal of Translational Medicine,* 2020 年 4 月 28 日，translational-medicine.biomedcentral.com/articles/10.1186/s12967-020-02316-w。

转到文本中的注释参考

其首席科学家理查德·克劳斯纳安东尼奥·雷加拉多，“科学家如何让你重返青春”，*MIT Technology Review,* 2022年10月25日，

www.technologyreview.com/ 2022/ 10/ 25/ 1061644/ 如何重返青春.

转到文本中的注释参考

这种实验方法旨在 Jae-Hyun Yang 等人，“ 表观遗传信息丢失作为哺乳动物衰老的原因 ”， *Cell,* 2023年1月12日，www.cell.com/cell/fulltext/S0092-8674(22)01570-7。

转到文本中的注释参考

一个生命跨度的世界，参见例如 David A. Sinclair 和 Matthew D. LaPlante，*Lifespan: Why We Age—and Why We Don't Have To* (纽约：Atria Books，2019)。

转到文本中的注释参考

初步工作表明记忆。参见，例如，哈佛关于记忆的研究：“ 研究人员识别出维持记忆精确度的神经回路和遗传‘ 开关’ ”，哈佛干细胞研究所，2018年3月12日，hsci.harvard.edu/news/researchers-identify-neural-circuit-and-genetic-switch-maintain-memory-precision。

转到文本中的注释参考

呼吁暂停 约翰·科恩，“ 禁止基因编辑婴儿的新呼声分裂生物学家 ”，*Science,* 2019年3月13日，www.science.org/content/article/new-call-ban-gene-edited-babies-divides-biologists。

转到文本中的注释参考

阿诺德的方法是 S. B. Jennifer Kan 等人的十五倍，“ 细胞色素 C 的定向进化用于碳-硅键的形成：将硅带入生命 ”，*Science,* 2016年11月25日，www.science.org/doi/10.1126/science.aah6219。

GO TO NOTE REFERENCE IN TEXT

这种合成生物学正在帮助詹姆斯·厄克特，“重新编程的细菌在工业规模上将二氧化碳转化为化学品”，*Chemistry World,* 2022年3月2日，www.chemistryworld.com/news/reprogrammed-bacterium-turns-carbon-dioxide-into-chemicals-on-industrial-scale/4015307.article。

转到文本中的注释参考

“如果我们可以成长”Elliot Hershberg，“原子是局部的，”*Century of Bio,* 2022年11月7日，centuryofbio.substack.com/p/atoms-are-local。GO TO NOTE REFERENCE IN TEXT

从理论上讲，世界上所有的“DNA数据存储的未来”，波托马克政策研究所，2018年9月，potomacinstitute.org/images/studies/Future_of_DNA_Data_Storage.pdf。GO TO NOTE REFERENCE IN TEXT

麦肯锡估计，麦肯锡全球研究院，“生物革命：创新正在改变经济、社会和我们的生活”，麦肯锡公司，2020年5月13日，www.mckinsey.com/industries/life-sciences/our-insights/the-bio-revolution-innovations-transforming-economies-societies-and-our-lives。GO TO NOTE REFERENCE IN TEXT

如果你使用传统的蛮力计算 DeepMind，“AlphaFold：生物学中一个50年未解的重大挑战的解决方案”，DeepMind 研究，2020年11月20日，www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology。GO TO NOTE REFERENCE IN TEXT

著名研究员Mohammed AlQuraishi，“AlphaFold @ CASP13：‘刚刚发生了什么？’”*Some Thoughts on a Mysterious Universe,* 2018年12月9日，

moalquraishi.wordpress.com/ 2018/ 12/ 09/ alphafold-casp13-刚刚发生了什么。

转到文本中的注释参考

一个标题说明了一切，Tanya Lewis，“ 生物学中最大的问题之一终于解决了，”*Scientific American,* 2022年10月31日，www.scientificamerican.com/article/one-of-the-biggest-problems-in-biology-has-finally-been-solved。

转到文本中的注释参考

结果是一次爆炸 Ewen Callaway，“ AlphaFold 和 AI 蛋白质折叠革命的下一步是什么”，*Nature,* 2022年4月13日，www.nature.com/articles/d41586-022-00997-5。

转到文本中的注释参考

DeepMind上传了大约2亿个Madhumita Murgia，“ DeepMind研究破解了几乎所有已知蛋白质的结构”，*Financial Times,*，2022年7月28日，www.ft.com/content/6a088953-66d7-48db-b61c-79005a0a351a；DeepMind，“ AlphaFold揭示了蛋白质宇宙的结构”，DeepMind研究，2022年7月28日，www.deepmind.com/blog/alphafold-reveals-the-structure-of-the-protein-universe。

转到文本中的注释参考

2019年，电极通过手术植入凯利·瑟维克，“ 首次，脑植入物让完全瘫痪的男子拼出‘ 我爱我酷儿子，’ ”*Science,* 2022年3月22日，www.science.org/content/article/first-brain-implant-lets-man-complete-paralysis-spell-out-thoughts-i-love-my-cool-son。

转到文本中的注释参考

科学家在一家名为Cortical Labs的初创公司Brett J. Kagan等人，“ *In Vitro* 神经元在模拟环境中学习并表现出感知能力

Game-World," *Neuron,* Oct. 12, 2022, www.cell.com/neuron/fulltext/S0896-6273(22)00806-6.

转到文本中的注释参考

6: 价值 $v*$ 的计算

亚马逊的“首个完全自主移动机器人”Mitchell Clark，“亚马逊宣布其首个完全自主移动仓库机器人，”*Verge,* 2022年6月21日，www.theverge.com/2022/6/21/23177756/amazon-warehouse-robots-proteus-autonomous-cart-delivery。

转到文本中的注释参考

亚马逊的Sparrow是第一个Dave Lee，“亚马逊推出可以做人工工作的新仓库机器人，”*Financial Times,* 2022年11月10日，www.ft.com/content/c8933d73-74a4-43ff-8060-7ff9402eccf1。

转到文本中的注释参考

机器人已经在进行复杂的手术 James Gaines，“机器人手术的过去、现在和未来，”*Smithsonian Magazine,* 2022年9月15日，www.smithsonianmag.com/innovation/the-past-present-and-future-of-robotic-surgery-180980763。

转到文本中的注释参考

他们建立了一支“为更美好的日常生活而设计的助手机器人”舰队，Everyday Robots，everydayrobots.com。

转到文本中的注释参考

随着蜜蜂种群的减少，Chelsea Gohd，“沃尔玛已为自主机器人蜜蜂申请专利”，世界经济论坛，2018年3月19日，www.weforum.org/agenda/2018/03/autonomous-robot-bees-are-being-patented-by-walmart。

转到文本中的注释参考

随着成本下降 *Artificial Intelligence Index Report 2021,* aiindex.stanford.edu/ 报告。

转到文本中的注释参考

警察局有一个拆弹小组Sara Sidner和Mallory Simon，“机器人和炸药如何以史无前例的方式击毙达拉斯狙击手”，CNN，2016年7月12日，cnn.com/2016/07/12/us/dallas-police-robot-c4-explosives/index.html。

转到文本中的注释参考

2019年，谷歌宣布了Elizabeth Gibney，“你好，量子世界！谷歌发布里程碑式的量子霸权声明，”*Nature,* 2019年10月23日，www.nature.com/articles/d41586-019-03213-z；Frank Arute等人，“使用可编程超导处理器实现量子霸权，”*Nature,* 2019年10月23日，www.nature.com/articles/s41586-019-1666-5。

转到文本中的注释参考

冷却到比尼尔·萨维奇更低的温度，“亲身体验谷歌的量子计算机，”*Scientific American,* 2019年10月24日，www.scientificamerican.com/article/hands-on-with-googles-quantum-computer。

转到文本中的注释参考

要存储等效信息 Gideon Lichfield，“Inside the Race to Build the Best Quantum Computer on Earth,”*MIT Technology Review,* 2022年2月26日，www.technologyreview.com/2020/02/26/916744/quantum-computer-race-ibm-google。

转到文本中的注释参考

其主要吸引力在于每个Matthew Sparkes，“IBM创造了有史以来最大的超导量子计算机”，*New Scientist,* 2021年11月15日，www.newscientist.com/article/2297583-ibm-creates-largest-ever-superconducting-quantum-computer。

转到文本中的注释参考

确实，对于某些任务来说，数量相对较少，无论如何。查尔斯·崔，“量子计算的量子飞跃？”*Scientific American,* 2017年10月25日，www.scientificamerican.com/article/quantum-leaps-in-quantum-computing。

转到文本中的注释参考

可再生能源将成为 Camilla Hodgson，“国际能源署称太阳能预计将在5年内超过煤炭”，*Financial Times,* 2022年12月10日，www.ft.com/content/98cec49f-6682-4495-b7be-793bf2589c6d。

GO TO NOTE REFERENCE IN TEXT

2000年，太阳能成本“太阳能光伏模块价格”，数据来源：Our World in Data，ourworldindata.org/grapher/solar-pv-prices。

GO TO NOTE REFERENCE IN TEXT

随着有意义的私人资本，汤姆·威尔逊，“核聚变：从科幻到‘何时，而非是否，’”*Financial Times,* 2022年12月17日，www.ft.com/content/65e8f125-5985-4aa8-a027-0c9769e764ad。

转到文本中的注释参考

放大后，它可以为特斯拉供电 Eli Dourado，“纳米技术的春天，”*Works in Progress,* 2022年10月12日，www.worksinprogress.co/issue/nanotechnologys-spring。

转到文本中的注释参考

7: 计算 $v*$ 的值

相反，大约三十名朱利安·博尔杰的单位，“阻止驶向基辅的俄罗斯车队的无人机操作员”，*Guardian,* 2022年3月28日，www.theguardian.com/world/2022/mar/28/the-drone-operators-who-halted-the-russian-armoured-vehicles-heading-for-kyiv。

转到文本中的注释参考

一千人的非军事团体Marcin Wyrwał，“乌克兰战争。人工智能如何杀死俄罗斯人，”*Onet,* 2022年7月13日，www.onet.pl/informacje/onetwiadomosci/rozwiazali-problem-armii-ukrainy-ich-pomysl-okazal-sie-dla-rosjan-zabojczy/pkzrk0z,79cfc278。GO TO NOTE REFERENCE IN TEXT

在常规的精确导弹中，Patrick Tucker，“国防部称，人工智能已经从俄罗斯在乌克兰的战争中学习”，*Defense One,* 2022年4月21日，www.defenseone.com/technology/2022/04/ai-already-learning-russias-war-ukraine-dod-says/365978。

转到文本中的注释参考

美国、英国和欧洲部队“乌克兰支持追踪器”，基尔世界经济研究所，2022年12月，www.ifw-kiel.de/index.php?id=17142。

转到文本中的注释参考

根据安全专家奥黛丽·库尔斯·克罗宁的说法，*Power to the People: How Open Technological Innovation Is Arming Tomorrow's Terrorists*（纽约：牛津大学出版社，2020年），2。

转到文本中的注释参考

深圳公司大疆创新 Scott Gilbertson，“评测：DJI Phantom 4，”*Wired,* 2016年4月22日，www.wired.com/2016/04/review-dji-phantom-4.

转到文本中的注释参考

对抗攻击是困难的 Cronin，*Power to the People,* 320；Derek Hawkins，“美国‘盟友’用价值300万美元的爱国者导弹击中了价值200美元的无人机。剧透：导弹赢了，”*Washington Post,* 2017年3月17日，www.washingtonpost.com/news/morning-mix/wp/2017/03/17/a-u-s-ally-fired-a-3-million-patriot-missile-at-a-200-drone-spoiler-the-missile-won。

转到文本中的注释参考

如果成立，十年后Azhar，*Exponential,* 249。

转到文本中的注释参考

在代码的无重力世界之外，例如，Michael Bhaskar, *Human Frontiers: The Future of Big Ideas in an Age of Small Thinking*（马萨诸塞州剑桥：MIT出版社，2021年）；Tyler Cowen, *The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better*（纽约：达顿出版社，2011年）；以及Robert Gordon, *The Rise and Fall of American Growth: The U.S. Standard of Living Since the Civil War*（新泽西州普林斯顿：普林斯顿大学出版社，2017年），还有许多其他人。

转到文本中的注释参考

塞萨尔·伊达尔戈认为塞萨尔·伊达尔戈，*Why Information Grows: The Evolution of Order, from Atoms to Economies*（伦敦：Allen Lane，2015年）。

转到文本中的注释参考

AI已经帮助发现新的Neil Savage，“机器学习发掘新材料”，*Nature,* 2021年6月30日，www.nature.com/articles/d41586-021-01793-3。

转到文本中的注释参考

例如，科学家们使用了Andrij Vasylenko等人的“通过对实验探索化学的无监督机器学习指导的晶体无机固体发现的元素选择”，*Nature Communications,* 2021年9月21日，www.nature.com/articles/s41467-021-25343-7。

转到文本中的注释参考

AI帮助设计和制造了Matthew Greenwood，“使用3D打印、AI和机器人技术创建的超级跑车”，Engineering.com，2021年6月23日，www.engineering.com/story/hypercar-created-using-3d-printing-ai-and-robotics。

转到文本中的注释参考

现在模拟加速 Elie Dolgin，“计算机模型能否成为更好的 COVID 疫苗的关键？”*Nature,* 2022年4月5日，www.nature.com/articles/d41586-022-00924-8。

转到文本中的注释参考

计算工具帮助自动化 Anna Nowogrodzki，“正在改变合成生物学的自动化设计工具，”*Nature,* 2018年12月10日，www.nature.com/articles/d41586-018-07662-w。

转到文本中的注释参考

量子技术，数百万维达，“谷歌的量子计算机比世界上最快的超级计算机快约1.58亿倍，”Medium，2021年2月28日，medium.com/predict/googles-quantum-computer-is-about-158-million-times-faster-than-the-world-s-fastest-supercomputer-36df56747f7f。

转到文本中的注释参考

发现新药 Jack W. Scannell 等人，“诊断制药研发效率的下降”，*Nature Reviews Drug Discovery,* 3月

1, 2012, www.nature.com/articles/nrd3681.

转到文本中的注释参考

预期寿命趋于平稳 Patrick Heuveline, “ 全球和国家预期寿命下降：2021年底评估, ” *Population and Development Review* 48, 第1期（2022年3月）, onlinelibrary.wiley.com/doi/10.1111/padr.12477。然而, 这些下降是在长期显著改善的背景下发生的。

转到文本中的注释参考

关于阿尔茨海默病等疾病的进展“ 失败的药物试验 ”, 阿尔茨海默病研究英国, www.alzheimersresearchuk.org/blog-tag/drug-trials/failed-drug-trials。

转到文本中的注释参考

AI技术可以搜索Michael S. Ringel等人, “ 打破Eroom定律 ”, *Nature Reviews Drug Discovery,* 2020年4月16日, www.nature.com/articles/d41573-020-00059-3。GO TO NOTE REFERENCE IN TEXT

2020年, AI系统Jonathan M. Stokes, “ 一种深度学习的抗生素发现方法, ” *Cell,* 2020年2月20日, www.cell.com/cell/fulltext/S0092-8674(20)30102-1。

转到文本中的注释参考

初创公司如Exscientia“ Exscientia和赛诺菲建立战略研究合作, 开发AI驱动的精确设计药物管线 ”, 赛诺菲, 2022年1月7日, www.sanofi.com/en/media-room/press-releases/2022/2022-01-07-06-00-00-2362917。

转到文本中的注释参考

截至目前，十八项临床资产 Nathan Benaich 和 Ian Hogarth，*State of AI Report 2022,* 2022年10月11日，www.stateof.ai。

转到文本中的注释参考

在六小时内，它识别出Fabio Urbina等人，“ 人工智能驱动药物发现的双重用途 ”，*Nature Machine Intelligence,* 2022年3月7日，www.nature.com/articles/s42256-022-00465-9。

转到文本中的注释参考

在发布时，PlayStation 2 K. Thor Jensen，“ 20年后：关于武器化游戏机的担忧几乎让PS2沉没 ”，*PCMag,* 2020年5月9日，www.pcmag.com/news/20-years-later-how-concerns-about-weaponized-consoles-almost-sunk-the-ps2；美联社，“ 索尼的高科技Playstation2将需要军事出口许可证 ”，*Los Angeles Times,* 2000年4月17日，www.latimes.com/archives/la-xpm-2000-apr-17-fi-20482-story.html。

转到文本中的注释参考

一个更合适的术语 关于术语 “ 多用途 ” 的更多信息，请参见，例如，Cronin，*Power to the People*。

转到文本中的注释参考

现在，像 DeepMind 的 Scott Reed 等人的单一系统，“ A Generalist Agent ”，DeepMind，2022 年 11 月 10 日，www.deepmind.com/publications/a-generalist-agent。

转到文本中的注释参考

GPT-4内部研究 @GPT-4技术报告，OpenAI，2023年3月14日，cdn.openai.com/papers/gpt-4.pdf。请参阅mobile.twitter.com/michalkosinski/status/1636683810631974912以了解早期实验之一。

转到文本中的注释参考

早期研究甚至声称 Sébastien Bubeck 等人，“人工通用智能的火花：GPT-4 的早期实验”，arXiv，2023 年 3 月 27 日，arxiv.org/abs/2303.12712。

转到文本中的注释参考

人工智能已经在寻找方法 Alhussein Fawzi 等人，“通过 AlphaTensor 发现新算法”，DeepMind，2022 年 10 月 5 日，www.deepmind.com/blog/discovering-novel-algorithms-with-alphatensor。

转到文本中的注释参考

AI研究员Stuart Russell Stuart Russell，*Human Compatible: AI and the Problem of Control*（伦敦：Allen Lane，2019年）。

转到文本中的注释参考

确实，Manuel Alfonseca 等人提出了一个有力的论点，“超级智能无法被遏制：来自可计算性理论的教训，”*Journal of Artificial Intelligence Research,* 2021年1月5日，jair.org/index.php/jair/article/view/12202；Jaime Sevilla 和 John Burden，“对超级智能无法被遏制的回应：来自可计算性理论的教训，”存在风险研究中心，2021年2月25日，www.cser.ac.uk/news/response-superintelligence-contained。

转到文本中的注释参考

8: 速度 $v*$ 的变化

即使在其首次公开比赛之前几天，参见例如，Cade Metz，《*Genius Makers: The Mavericks Who Brought AI to Google, Facebook and the World*》（伦敦：Random House Business，2021年）第170页。

转到文本中的注释参考

超过 2.8 亿人谷歌搜索，“ 围棋未来峰会：5 月 23 日至 5 月 27 日，中国乌镇 ”，谷歌活动，events.google.com/alphago2017 [不活跃]。

转到文本中的注释参考

这是美国的危机 保罗·迪克森，“ 斯普特尼克对美国的影响， ”*Nova,* PBS，2007年11月6日，www.pbs.org/wgbh/nova/article/sputnik-impact-on-america. GO TO NOTE REFERENCE IN TEXT

根据习近平洛德伟的说法，“ 习近平在中国党代会上的讲话全文 ”，彭博社，2022年10月18日，www.bloomberg.com/news/articles/2022-10-18/full-text-of-xi-jinping-s-speech-at-china-20th-party-congress-2022。

GO TO NOTE REFERENCE IN TEXT

中国的自上而下模式 参见，例如，Nigel Inkster，*The Great Decoupling: China, America and the Struggle for Technological Supremacy* (伦敦：Hurst，2020)。

转到文本中的注释参考

“ 到2030年，中国的AI理论 ” Graham Webster等人，“ 完整翻译：中国 ‘ 新一代人工智能发展规划， ’ ” DigiChina，斯坦福大学，2017年8月1日，digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017。

转到文本中的注释参考

事实上，清华大学发表了更多的Benaich和Hogarth，*State of AI;* Neil Savage，“世界人工智能领导者的顶级竞赛，”*Nature Index,* 2020年12月9日，www.nature.com/articles/d41586-020-03409-8；“清华大学可能很快在科学研究中位居世界前列，”*Economist,* 2018年11月17日，www.economist.com/china/2018/11/17/tsinghua-university-may-soon-top-the-world-league-in-science-research。

转到文本中的注释参考

中国有一个不断增长且令人印象深刻的Sarah O'Meara，“中国会在2030年引领世界AI吗？”，*Nature,* 2019年8月21日，www.nature.com/articles/d41586-019-02360-7；Akira Oikawa和Yuta Shimono，“中国在AI研究上超越美国”，*Nikkei Asia,* 2021年8月10日，asia.nikkei.com/Spotlight/Datawatch/China-overtakes-US-in-AI-research。

GO TO NOTE REFERENCE IN TEXT

从人工智能研究的数量来看，Daniel Chou，“计算人工智能研究：探索英语和中文来源的人工智能研究产出”，安全与新兴技术中心，2022年7月，cset.georgetown.edu/publication/counting-ai-research。

转到文本中的注释参考

中国超过了美国 Remco Zwetsloot，“中国在快速超越美国的STEM博士增长”，安全与新兴技术中心，2021年8月，cset.georgetown.edu/publication/china-is-fast-outpacing-u-s-stem-phd-growth.

转到文本中的注释参考

在二十一世纪初期，Graham Allison 等人撰写的《伟大的科技竞争：中国与美国》，哈佛肯尼迪学院贝尔弗中心

中心，2021年12月，www.belfercenter.org/sites/default/files/GreatTechRivalry_ChinavsUS_211207.pdf。GO TO NOTE REFERENCE IN TEXT

根据当前趋势，新华社，“2021年中国授权约70万件发明专利：报告”，新华网，2021年1月8日，english.news.cn/20220108/ded0496b77c24a3a8712fb26bba390c3/c.html；“美国专利统计图表，1963-2020年历年”，美国专利商标局，2021年5月，www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm。然而，美国的数据来自2020年。还需要指出的是，高价值专利也在快速增长：中国国务院，“中国发明专利数量增长”，新华社，2022年1月，english.www.gov.cn/statecouncil/ministries/202201/12/content_WS61deb7c8c6d09c94e48a3883.html。

转到文本中的注释参考

它拥有更多的世界约瑟夫·辛克斯，“中国现在拥有比其他任何国家更多的超级计算机，”*Time,* 2017年11月14日，time.com/5022859/china-most-supercomputers-world。

转到文本中的注释参考

习近平明确提到杰森·道格拉斯，“随着劳动力减少，中国工厂加速推进机器人化”，*Wall Street Journal,* 2022年9月18日，www.wsj.com/articles/chinas-factories-accelerate-robotics-push-as-workforce-shrinks-11663493405。

转到文本中的注释参考

2014年，中国提交了相同数量的Allison等人，“伟大的科技竞争。”

转到文本中的注释参考

一年后，中国建成了张志浩，“京沪量子链‘新时代’，”*China Daily USA,* 2017年9月30日，usa.chinadaily.com.cn/china/2017-09/30/content_32669867.htm。

转到文本中的注释参考

他们投资超过100亿美元 Amit Katwala，“为什么中国完全有能力成为量子计算的超级大国”，*Wired,* 2018年11月14日，www.wired.co.uk/article/quantum-computing-china-us。GO TO NOTE REFERENCE IN TEXT

合肥科学家甚至声称已经建造了Han-Sen Zhong等人，“使用光子的量子计算优势”，*Science,* 2020年12月3日，www.science.org/doi/10.1126/science.abe8770。GO TO NOTE REFERENCE IN TEXT

墨子首席研究员引述于阿米特·卡特瓦拉，*Quantum Computing* (伦敦：兰登书屋商业出版社，2021), 88。

转到文本中的注释参考

中国已经领先，Allison等人，“伟大的技术竞争。”

转到文本中的注释参考

“我们没有竞争的机会”卡特里娜·曼森，“前五角大楼软件主管称美国在人工智能竞争中已输给中国”，*Financial Times,* 2021年10月10日，www.ft.com/content/f939db9a-40af-4bd1-b67d-10492535f8e0。

转到文本中的注释参考

“先进技术是锐利的武器”引自 Inkster, *The Great Decoupling,* 193.

转到文本中的注释参考

几乎每个国家现在都有。有关详细信息，请参见“国家人工智能政策与战略”，OECD.AI，oecd.ai/en/dashboards。

转到文本中的注释参考

弗拉基米尔·普京认为领导者“普京：人工智能领域的领导者将统治世界”，CNBC，2017年9月4日，www.cnbc.com/2017/09/04/putin-leader-in-artificial-intelligence-will-rule-world.html。

转到文本中的注释参考

法国总统埃马纽埃尔·马克龙托马斯·麦考利，“马克龙的欧洲元宇宙梦想离现实还很远，”*Next Web,* 2022年9月14日，thenextweb.com/news/prospects-for-europes-emerging-metaverse-sector-macron-vestager-meta。

转到文本中的注释参考

安全、财富、声望“法国2030”，国家研究署，2023年2月27日，anr.fr/en/france-2030/france-2030. GO TO NOTE REFERENCE IN TEXT

到2030年，其经济“印度将在2050年成为30万亿美元经济体：Gautam Adani”，*Economic Times,* 2022年4月22日，economictimes.indiatimes.com/news/economy/indicators/india-to-be-a-30-trillion-economy-by-2050-gautam-adani/articleshow/90985771.cms。

转到文本中的注释参考

在此之下，印度建立了 Trisha Ray 和 Akhil Deo，“印度技术外交政策的优先事项”，华盛顿国际贸易协会，2020年9月25日，www.wita.org/atp-research/tech-foreign-policy-india。

转到文本中的注释参考

我们生活在一个克罗宁时代，*Power to the People*。

转到文本中的注释参考

例如，GitHub 有 Neeraj Kashyap，“ GitHub 的 1.28 亿公共代码库之路 ”，*Towards Data Science,* 2020 年 3 月 4 日，towardsdatascience.com/githubs-path-to-128m-public-repositories-f6f656ab56b1。

转到文本中的注释参考

最初的此类服务 arXiv，“ 关于 ArXiv ”，arxiv.org/about。

转到文本中的注释参考

世界的伟大库存 “ 通用索引 ”，互联网档案馆，2021年10月7日，archive.org/details/GeneralIndex。

转到文本中的注释参考

全球研发支出 “ 研究与开发：美国趋势和国际比较 ”，国家科学与工程统计中心，2022年4月28日，ncses.nsf.gov/pubs/nsb20225。

转到文本中的注释参考

亚马逊的研发预算仅 Prableen Bajpai，“ 哪些公司在研究和开发（R&D）上花费最多？ ”，Nasdaq，2021年6月21日，www.nasdaq.com/articles/which-companies-spend-the-most-in-research-and-development-rd-2021-06-21。

转到文本中的注释参考

字母表，苹果，华为，Meta “ 华为投入220亿美元用于研发以击败美国制裁， ”彭博新闻，2022年4月25日，www.bloomberg.com/news/articles/2022-04-25/huawei-rivals-apple-meta-with-r-d-spending-to-beat-sanctions; Jennifer Saba，“ 苹果拥有最多的增长动力， ”路透社，2021年10月28日，www.reuters.com/breakingviews/apple-has-most-growth-fuel-hand-2021-10-28。

转到文本中的注释参考

然而，这也是 LeCun Metz，*Genius Makers,* 58。

转到文本中的注释参考

NVIDIA 并没有抱怨 Mitchell，*Artificial Intelligence,* 103。

转到文本中的注释参考

二百五十名乘客“ 世界第一：利物浦和曼彻斯特铁路的建造 ”，科学+工业博物馆，2018年12月20日，www.scienceandindustrymuseum.org.uk/objects-and-stories/making-the-liverpool-and-manchester-railway。

转到文本中的注释参考

五年后，它正在交付 这和更广泛的账户取自 William Quinn 和 John D. Turner，*Boom and Bust: A Global History of Financial Bubbles*（英国剑桥：剑桥大学出版社，2022年）。

转到文本中的注释参考

在其高峰期，铁路股票同上。

转到文本中的注释参考

1840年代的铁路热潮“ 泡沫之美 ”，*Economist,* 2008年12月18日，www.economist.com/christmas-specials/2008/12/18/the-beauty-of-bubbles。

转到文本中的注释参考

卡洛塔·佩雷斯认为有一个等价的佩雷斯，*Technological Revolutions and Financial Capital*。

转到文本中的注释参考

科学必须转化为广泛的经济学文献，研究创新的微观经济学，展示这一过程对经济激励的敏感性和包裹性。例如，参见Lipsey、Carlaw和Bekar，*Economic Transformations,*以获取概述。

转到文本中的注释参考

人均 GDP 已上升，参见 Angus Maddison，*The World Economy: A Millenarian Perspective*（巴黎：OECD 出版物，2001），或更为最新的“ 人均 GDP，1820 至 2018 ”，Our World in Data，ourworldindata.org/grapher/gdp-per-capita-maddison-2020?yScale=log。

转到文本中的注释参考

现在，全球范围内，这篇文章由Nishant Yonzan等人撰写，“ 预测到2030年的全球极端贫困：我们距离世界银行的3%目标有多近？ ”，*World Bank Data Blog,* 2020年10月9日，blogs.worldbank.org/opendata/projecting-global-extreme-poverty-2030-how-close-are-we-world-banks-3-goal。

转到文本中的注释参考

在十九世纪，发明艾伦·格林斯潘和阿德里安·伍尔德里奇，*Capitalism in America: A History*（伦敦：艾伦·莱恩，2018年），15。

转到文本中的注释参考

艾萨克·辛格的缝纫机 同上，第47页。

转到文本中的注释参考

例如，在德国，Charlie Giattino 和 Esteban Ortiz-Ospina，“ 我们是否比以往工作得更多？ ”，Our World in Data，ourworldindata.org/working-more-than-ever。

转到文本中的注释参考

科技无疑是最大的“ 标普500数据 ”， 标普道琼斯指数，2022年7月，www.spglobal.com/spdji/en/indices/equity/sp-500/#data。

转到文本中的注释参考

数千亿美元。仅在2021年，全球就有超过6000亿美元的风险投资，主要投向科技和生物科技企业，是十年前的十倍。参见Gené Teare，“ 2021年融资和独角兽创造打破所有记录 ”， *Crunchbase News,*， 2022年1月5日，news.crunchbase.com/business/global-vc-funding-unicorns-2021-monthly-recap。同时，2021年技术领域的私募股权投资也激增至超过4000亿美元，成为迄今为止最大的单一类别。参见Laura Cooper和Preeti Singh，“ 私募股权支持创纪录的科技交易量 ”， *Wall Street Journal,*， 2022年1月3日，www.wsj.com/articles/private-equity-backs-record-volume-of-tech-deals-11641207603。

转到文本中的注释参考

投资于 AI 技术 例如，*Artificial Intelligence Index Report 2021,* 尽管自那时以来生成式 AI 的繁荣使得数字肯定有所增长。

转到文本中的注释参考

普华永道预测AI将增加“ Sizing the Prize—PwC的全球人工智能研究：利用AI革命 ”， 普华永道，2017年，www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html。

转到文本中的注释参考

麦肯锡预测将带来4万亿美元的提升 Jacques Bughin 等人，“ AI前沿笔记：建模AI对世界经济的影响 ”， 麦肯锡，2018年9月4日，www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-

世界经济；迈克尔·丘，“生物革命：创新正在改变经济、社会和我们的生活”，麦肯锡全球研究院，2020年5月13日，www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/the-bio-revolution-innovations-transforming-economies-societies-and-our-lives。

GO TO NOTE REFERENCE IN TEXT

推动世界机器人安装“机器人如何改变世界”，牛津经济研究院，2019年6月26日，resources.oxfordeconomics.com/hubfs/How%20Robots%20Change%20the%20World%20(PDF).pdf。

转到文本中的注释参考

考虑到世界经济《二十世纪下半叶的世界经济》，经合组织，2006年9月22日，read.oecd-ilibrary.org/development/the-world-economy/the-world-economy-in-the-second-half-of-the-twentieth-century_9789264022621-5-en#page1。GO TO NOTE REFERENCE IN TEXT

随着通用人工智能的发展，Philip Trammell 等人撰写的《变革性人工智能下的经济增长》，全球优先事项研究所，2020年10月，globalprioritiesinstitute.org/wp-content/uploads/Philip-Trammell-and-Anton-Korinek_economic-growth-under-transformative-ai.pdf。这导致了一个非凡且不可能的情景，即“在有限的时间内快速增长到产生无限产出的程度。”

转到文本中的注释参考

假设天气条件良好，Hannah Ritchie 等人，“作物产量”，Our World in Data，ourworldindata.org/crop-yields。

转到文本中的注释参考

在二十一世纪，产量“农业统计——2020年6月1日最终作物面积、产量、牲畜数量和农业劳动力

英国，”英国环境、食品和农村事务部，2020年12月22日，assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/946161/structure-jun2020final-uk-22dec20.pdf。

转到文本中的注释参考

玉米每公顷产量 Ritchie 等人，“作物产量。”

GO TO NOTE REFERENCE IN TEXT

所需的劳动 Smil，*How the World Really Works*，66。

GO TO NOTE REFERENCE IN TEXT

1945年，约50%的Max Roser和Hannah Ritchie，“饥饿和营养不良”，我们的数据世界，ourworldindata.org/hunger-and-undernourishment。

转到文本中的注释参考

虽然清洁发电 Smil，*How the World Really Works*，36。

GO TO NOTE REFERENCE IN TEXT

自二十一世纪初起，同上，第42页。

转到文本中的注释参考

想象一下平均番茄 Ibid., 61.

GO TO NOTE REFERENCE IN TEXT

此外，为了满足全球需求，Daniel Quiggin 等人，“气候变化风险评估 2021”，查塔姆研究所，2021年9月14日，www.chathamhouse.org/2021/09/climate-change-risk-assessment-2021?7J7ZL,68TH2Q,UNIN9。

转到文本中的注释参考

然而，这在很大程度上尚未被发明，伊丽莎白·科尔伯特，《*Under a White Sky: The Nature of the Future*》（纽约：皇冠出版社，2022年），第155页。

转到文本中的注释参考

AI帮助设计了一种酶 Hongyuan Lu 等人，“机器学习辅助的PET解聚水解酶工程”，*Nature,* 2022年4月27日，www.nature.com/articles/s41586-022-04599-z。

转到文本中的注释参考

“当你看到某物”“J. 罗伯特·奥本海默 1904 – 67”，在 *Oxford Essential Quotations,*，苏珊·拉特克利夫编辑（牛津：牛津大学出版社，2016），www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00007996。

转到文本中的注释参考

“我们现在正在创造的”引自戴森，*Turing's Cathedral*。

转到文本中的注释参考

9: 速度 $v*$

在其核心，民族国家显然存在很多复杂性，并且关于“ 民族国家 ”和“ 国家 ”这两个术语的使用有大量文献。然而，在这里我们以一种相当基础的方式使用它们：民族国家是世界上的国家、他们的人民及其政府（包含所有的巨大多样性和复杂性）；国家是这些民族国家内的政府和统治及社会服务系统。爱尔兰、以色列、印度和印度尼西亚都是非常不同类型的民族和国家，但尽管它们有许多区别，我们仍然可以将它们视为一组连贯的实体。正如温迪·布朗所说，民族国家一直是“ 某种虚构 ”（*Walled States, Waning Sovereignty* [纽约：Zone Books, 2010], 69）——如果权力被施加在人民身上，人民如何能是主权者？尽管如此，民族国家是一个极其有用且强大的虚构。

转到文本中的注释参考

识字率，预期寿命 Max Roser 和 Esteban Ortiz-Ospina，“ 识字率 ”，Our World in Data，ourworldindata.org/literacy。

转到文本中的注释参考

西方社会尤其是用威廉·戴维斯的话来说，*Nervous States: How Feeling Took Over the World*（伦敦：乔纳森·凯普，2018年）。

转到文本中的注释参考

对政府的信任 英国三分之一（35%）的人口表示他们信任国家政府，这低于经合组织国家的平均水平（41%）。一半（49%）的英国人口表示他们不信任国家政府。“ 建立信任以加强民主：2021年经合组织关于公共机构信任驱动因素调查的主要发现 ”，经合组织，www.oecd.org/governance/trust-in-government。

GO TO NOTE REFERENCE IN TEXT

对于像奥巴马这样的最近总统，“政府公信力：1958 – 2022”，皮尤研究中心，2022年6月6日，www.pewresearch.org/politics/2022/06/06/public-trust-in-government-1958-2022。GO TO NOTE REFERENCE IN TEXT

相当引人注目的是，2018年Lee Drutman等人的一项研究，“跟随领导者：探索美国对民主和威权主义的支持”，民主基金选民研究小组，2018年3月，fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/followtheleader_2018mar13.pdf。

转到文本中的注释参考

不少于85%的美国人“不满国家和经济方向的两党共识”，AP NORC，2022年6月29日，apnorc.org/projects/bipartisan-dissatisfaction-with-the-direction-of-the-country-and-the-economy。

转到文本中的注释参考

对非政府的不信任延伸，参见例如Daniel Drezner, *The Ideas Industry: How Pessimists, Partisans, and Plutocrats Are Transforming the Marketplace of Ideas*（纽约：牛津大学出版社，2017年），以及爱德曼信任晴雨表：“2022年爱德曼信任晴雨表”，爱德曼，www.edelman.com/trust/2022-trust-barometer。

转到文本中的注释参考

民主感知指数调查 Richard Wike 等人，“全球许多人对民主运作方式感到不满”，皮尤研究中心，2019年4月29日，www.pewresearch.org/global/2019/04/29/many-across-the-globe-are-dissatisfied-with-how-democracy-is-working/；Dalia Research 等人，“2018年民主感知指数”，民主联盟，2018年6月，www.allianceofdemocracies.org/wp-content/uploads/2018/06/Democracy-Perception-Index-2018-1.pdf。

转到文本中的注释参考

自2010年以来，更多国家“ 新报告：全球民主衰退加速 ”，自由之家，2021年3月3日，freedomhouse.org/article/new-report-global-decline-democracy-has-accelerated。

转到文本中的注释参考

不稳定的关键催化剂 例如，参见托马斯·皮凯蒂，*Capital in the Twenty-first Century*（剑桥，马萨诸塞州：哈佛大学出版社，2014年），以及安东尼·B·阿特金森，*Inequality: What Can Be Done?* (剑桥，马萨诸塞州：哈佛大学出版社，2015年)，以获取更广泛的调查。

转到文本中的注释参考

1980年至2021年间“ 前1%国民收入份额 ”，世界不平等数据库，wid.world/world/#sptinc_p99p100_z/US;FR;DE;CN;ZA;GB;WO/last/eu/k/p/yearly/s/false/5.6579999999999995/30/curve/false/country.

转到文本中的注释参考

财富越来越多是理查德·米勒，“ 福布斯全球亿万富翁榜：2023年最富有的人， ”*Forbes,* www.forbes.com/billionaires/。虽然GDP是流量，而不是像财富那样的存量，但这种比较仍然令人瞩目。

转到文本中的注释参考

政府政策，Alistair Dieppe 的缩减，“ 广泛的生产力放缓，七张图表 ”，*World Bank Blogs: Let's Talk Development,*，2020年7月14日，blogs.worldbank.org/developmenttalk/broad-based-productivity-slowdown-seven-charts。

转到文本中的注释参考

美国有四千万人 Jessica L. Semega 等人，“ 美国的收入和贫困：2016年 ”，美国人口普查局，

www.census.gov/content/dam/Census/library/publications/2017/demo/P60-259.pdf，报告于digitallibrary.un.org/record/1629536?ln=en。

转到文本中的注释参考

这些尤其令人担忧的趋势，参见例如，Christian Houle 等人，“社会流动性与政治不稳定”，*Journal of Conflict Resolution,* 2017年8月8日，journals.sagepub.com/doi/full/10.1177/0022002717723434；以及 Carles Boix，“当代世界中内战和革命的经济根源”，*World Politics* 60，第3期（2008年4月）：390 – 437。

转到文本中的注释参考

国家消亡并不是一个新颖的想法；例如，参见Rana Dasgupta，“国家的消亡，”*Guardian,* 2018年4月5日，www.theguardian.com/news/2018/apr/05/demise-of-the-nation-state-rana-dasgupta。

转到文本中的注释参考

一项由Philipp Lorenz-Spreen等人发表的元分析，“关于数字媒体与民主的全球因果和相关证据的系统综述”，*Nature Human Behaviour,* ，2022年11月7日，www.nature.com/articles/s41562-022-01460-1。

转到文本中的注释参考

正如技术史学家兰登·温纳所述，*Autonomous Technology: Technics-Out-of-Control as a Theme in Political Thought*（马萨诸塞州剑桥：麻省理工学院出版社，1977年），6。

转到文本中的注释参考

虽然技术并不简单化，例如，Jenny L. Davis, *How Artifacts Afford: The Power and Politics of Everyday Things* (剑桥，马萨诸塞州：麻省理工学院出版社，2020)。用技术的话来说，

乌尔苏拉·M·富兰克林（*The Real World of Technology* [多伦多：Anansi出版社，1999]），*prescriptive;* 也就是说，它们的创造或使用会促使或要求某些行为、劳动分工或结果。拥有拖拉机的农民会以不同于拥有两头牛和一把犁的农民的方式进行工作和安排需求。工厂系统引发的劳动分工产生了与狩猎采集社会不同类型的社会组织——一种服从和管理的文化。“ 在技术实践中形成的模式成为社会生活的一部分 ”（55）。

转到文本中的注释参考

时钟产生了设定时间，关于机械时钟的影响，请参见 Mumford, *Technics and Civilization,*，其中有精彩的分析。

转到文本中的注释参考

印刷机帮助了本尼迪克特·安德森，*Imagined Communities: Reflections on the Origin and Spread of Nationalism*（伦敦：Verso, 1983年）。

转到文本中的注释参考

在某个轨迹上，一些自由主义者，剑桥政治学家大卫·伦西曼谈到“ 僵尸民主 ”，这意味着类似的东西：“ 基本的想法是，人们只是在观看一场表演，他们的角色是在适当的时刻给予或拒绝掌声。民主政治已经变成了一场精心策划的表演。 ”大卫·伦西曼，*How Democracy Ends*（伦敦：Profile Books，2019），47。

转到文本中的注释参考

10: 速度 $v*$

NHS曾受到影响，例如，参见S. Ghafur等人，“ 对WannaCry网络攻击对NHS影响的回顾性分析 ”， *NPJ Digital Medicine,* 2019年10月2日，www.nature.com/articles/s41746-019-0161-6，了解更多信息。

转到文本中的注释参考

WannaCry 欺骗了一些用户 Mike Azzara，“ WannaCry 勒索软件是什么以及它如何工作？ ”， Mimecast，2021年5月5日，www.mimecast.com/blog/all-you-need-to-know-about-wannacry-ransomware。

转到文本中的注释参考

随之而来的损失费用安迪·格林伯格，“ NotPetya 的未解之谜，历史上最具破坏性的网络攻击 ”， *Wired,* 2018年8月22日，www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world。

转到文本中的注释参考

正如一位NSA工作人员詹姆斯·巴姆福德所说，“ 评论：证据指向NSA的另一位斯诺登 ”，路透社，2016年8月22日，www.reuters.com/article/us-intelligence-nsa-commentary-idUSKCN10X01P。

转到文本中的注释参考

“ 一个等效的场景 ”布拉德·史密斯，“ 需要紧急集体行动以确保人们在线安全：上周网络攻击的教训， ”
Microsoft Blogs: On the Issues, 2017年5月14日，blogs.microsoft.com/on-the-issues/2017/05/14/need-urgent-collective-action-keep-people-safe-online-lessons-last-weeks-cyberattack。

转到文本中的注释参考

权力是“能力或容量”定义取自牛津语言，languages.oup.com。

转到文本中的注释参考

根据 Ronen Bergman 等人的话，“科学家与人工智能辅助的远程控制杀人机器”，2021年9月18日，www.nytimes.com/2021/09/18/world/middleeast/iran-nuclear-fakhrizadeh-assassination-israel.html。

转到文本中的注释参考

军用级无人机Azhar的成本，*Exponential,* 192。

转到文本中的注释参考

到2028年，260亿美元每年《财富商业洞察》，《军用无人机市场到2028年将达到261.2亿美元；全球军事支出的增加将促进增长》，全球新闻通讯社，2021年7月22日，www.globenewswire.com/en/news-release/2021/07/22/2267009/0/en/Military-Drone-Market-to-Hit-USD-26-12-Billion-by-2028-Rising-Military-Spending-Worldwide-to-Augment-Growth-Fortune-Business-Insights.html。

转到文本中的注释参考

例如，2021年5月，David Hambling，“以色列在加沙袭击中使用了世界上第一个AI引导的作战无人机群”，*New Scientist,*，2021年6月30日，www.newscientist.com/article/2282656-israel-used-worlds-first-ai-guided-combat-drone-swarm-in-gaza-attacks。

转到文本中的注释参考

初创公司如Anduril Dan Primack，“独家：Rebellion Defense以10亿美元估值筹集1.5亿美元，”*Axios,* 2021年9月15日，www.axios.com/2021/09/15/rebellion-defense-raises-150-million-billion-valuation；Ingrid Lunden，“消息称Anduril正在筹集高达12亿美元，来源称在

a $7B 估值，针对其国防技术，”TechCrunch，2022年5月24日，techcrunch.com/2022/05/24/filing-anduril-is-raising-up-to-1-2b-sources-say-at-a-7b-pre-money-valuation-for-its-defense-tech.

转到文本中的注释参考

正如网络安全专家布鲁斯·施奈尔所说，“即将到来的AI黑客”，哈佛肯尼迪学院贝尔弗中心，2021年4月，www.belfercenter.org/publication/coming-ai-hackers。

转到文本中的注释参考

Meta的研究人员创建了Anton Bakhtin等人，“通过将语言模型与战略推理相结合，在*Diplomacy*游戏中实现人类水平的游戏”，*Science,*，2022年11月22日，www.science.org/doi/10.1126/science.ade9097。

转到文本中的注释参考

当非国家行为体时，请参见 Benjamin Wittes 和 Gabriella Blum，*The Future of Violence: Robots and Germans, Hackers and Drones—Confronting A New Age of Threat*（纽约：Basic Books，2015），以获取此论点的更详细版本。

转到文本中的注释参考

两者看起来和听起来都很像 首次报道于Nilesh Cristopher，“我们刚刚在印度选举活动中看到了深度伪造的首次使用，”*Vice,* 2020年2月18日，www.vice.com/en/article/jgedjb/the-first-use-of-deepfakes-in-indian-election-by-bjp。

转到文本中的注释参考

在另一起广为人知的事件中，Melissa Goldin，“拜登唱‘Baby Shark’的视频是深度伪造”，美联社，2022年10月19日，apnews.com/article/fact-check-biden-baby-shark-deepfake-412016518873；“经过处理的南希·佩洛西视频突显‘深度伪造’技术的威胁”，CBS

新闻，2019年5月25日，www.cbsnews.com/news/doctored-nancy-pelosi-video-highlights-threat-of-deepfake-tech-2019-05-25。

转到文本中的注释参考

如果你想观看 TikTok @deeptomcruise，www.tiktok.com/@deeptomcruise?lang=en。 GO TO NOTE REFERENCE IN TEXT

香港的一家银行 Thomas Brewster，“骗子克隆公司董事的声音进行3500万美元银行抢劫，警方发现，”*Forbes,* 2021年10月14日，www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions。

转到文本中的注释参考

所有文件似乎都显示凯瑟琳·斯塔普，“骗子使用AI模仿CEO的声音在不寻常的网络犯罪案件中”，*Wall Street Journal,* 2019年8月30日，www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402。

转到文本中的注释参考

这不是总统的指控，而是真正的深度伪造。参见凯利·琼斯，“拜登称他正在恢复征兵的病毒视频是深度伪造，”*Verify,* 2023年3月1日，www.verifythis.com/article/news/verify/national-verify/viral-video-of-biden-saying-hes-reinstating-the-draft-is-a-deepfake/536-d721f8cb-d26a-4873-b2a8-91dd91288365。

转到文本中的注释参考

他的激进信息是 Josh Meyer 所写的《Anwar al-Awlaki: The Radical Cleric Inspiring Terror from Beyond the Grave》，发表于 NBC News，2016 年 9 月 21 日，网址为 www.nbcnews.com/news/us-news/anwar-al-awlaki-radical-cleric-inspiring-terror-beyond-grave-n651296；Alex Hern 所写的《‘YouTube 伊斯兰主义者’Anwar al-Awlaki 视频在极端主义打击中被移除》，*Guardian,*

2017年11月13日，www.theguardian.com/technology/2017/nov/13/youtube-islamist-anwar-al-awlaki-videos-removed-google-extremism-clampdown

转到文本中的注释参考

不久之后，这些视频将完全由 Eric Horvitz 制作，“ On the Horizon: Interactive and Compositional Deepfakes, ”ICMI '22: 2022年国际多模态交互会议论文集，arxiv.org/abs/2209.01714。

转到文本中的注释参考

根据美国参议院情报特别委员会的报告：《俄罗斯积极措施运动和2016年美国大选中的干预》，第5卷，反情报威胁和漏洞，第116届国会，第1次会议，www.intelligence.senate.gov/sites/default/files/documents/report_volume5.pdf；Nicholas Fandos等人，“众议院情报委员会发布煽动性的俄罗斯社交媒体广告”，*New York Times,* 2017年11月1日，www.nytimes.com/2017/11/01/us/politics/russia-technology-facebook.html。

转到文本中的注释参考

不幸的是，这不仅仅是俄罗斯。然而，通常是俄罗斯。2021年，仅俄罗斯就占了58%的网络攻击。参见Tom Burt，“俄罗斯网络攻击对政府构成更大风险以及我们年度报告中的其他见解，”*Microsoft Blogs: On the Issues,* 2021年10月7日，blogs.microsoft.com/on-the-issues/2021/10/07/digital-defense-report-2021。

转到文本中的注释参考

超过七十个国家 Samantha Bradshaw 等人，“工业化虚假信息：2020年全球有组织社交媒体操控清单”，牛津大学民主项目 &

技术，2021年1月13日，demtech.oii.ox.ac.uk/research/posts/industrialized-disinformation.

转到文本中的注释参考

CIA也不陌生。参见例如Krassi Twigg和Kerry Allen，“中国使用的虚假信息战术”，BBC新闻，2021年3月12日，www.bbc.co.uk/news/56364952；Kenddrick Chan和Mariah Thornton，“中国针对台湾的虚假信息和宣传的变化”，*Diplomat,* 2022年9月19日，thediplomat.com/2022/09/chinas-changing-disinformation-and-propaganda-targeting-taiwan/；以及Emerson T. Brooking和Suzanne Kianpour，“伊朗的数字影响力努力：二十一世纪的游击广播”，大西洋理事会，2020年2月11日，www.atlanticcouncil.org/in-depth-research-reports/report/iranian-digital-influence-efforts-guerrilla-broadcasting-for-the-twenty-first-century。

GO TO NOTE REFERENCE IN TEXT

82% 的有影响力用户 Virginia Alvino Young，“讨论‘重启美国’的推特账户中近一半可能是机器人”，卡内基梅隆大学，2020年5月27日，www.cmu.edu/news/stories/archives/2020/may/twitter-bot-campaign.html。

转到文本中的注释参考

引发“信息末日”参见 Nina Schick，*Deep Fakes and the Infocalypse: What You Urgently Need to Know*（伦敦：Monoray，2020）；以及 Ben Buchanan 等人，“真相、谎言与自动化”，安全与新兴技术中心，2021年5月，cset.georgetown.edu/publication/truth-lies-and-automation。

转到文本中的注释参考

根据布鲁金斯学会的威廉·A·加尔斯顿的说法，“眼见是否仍为实？深度伪造对政治真相的挑战”，布鲁金斯，

2020年1月8日，www.brookings.edu/research/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics.

转到文本中的注释参考

最早在中国发现 图表取自William MacAskill，*What We Owe the Future: A Million-Year View*（伦敦：Oneworld，2022），112，他引用了各种来源，尽管承认没有一个对这个数字是确定的。另见H. C. Kung等人，“1977年中国的流感：甲型流感病毒H1N1亚型的复发”，*Bulletin of the World Health Organization* 56，第6期（1978），www.ncbi.nlm.nih.gov/pmc/articles/PMC2395678/pdf/bullwho00443-0095.pdf。

转到文本中的注释参考

H1N1流感有什么不同寻常之处 Joel O. Wertheim，“1977年H1N1流感病毒的重新出现：使用生物学上不现实的采样日期估计分歧时间的警示故事”，*PLOS ONE,* 2010年6月17日，journals.plos.org/plosone/article?id=10.1371/journal.pone.0011184。

转到文本中的注释参考

早期病毒的一个版本 参见，例如，Edwin D. Kilbourne，“20世纪的流感大流行”，*Emerging Infectious Diseases* 12，第1期（2006年1月），www.ncbi.nlm.nih.gov/pmc/articles/PMC3291411；以及Michelle Rozo和Gigi Kwik Gronvall，“1977年重新出现的H1N1毒株与功能增益辩论”，*mBio,* 2015年8月18日，www.ncbi.nlm.nih.gov/pmc/articles/PMC4542197。

转到文本中的注释参考

而事故和泄漏 参见，例如，Alina Chan 和 Matt Ridley 的优秀描述，*Viral: The Search for the Origin of Covid-19* (伦敦：Fourth Estate, 2022)，以及 MacAskill，*What We Owe the Future*。

转到文本中的注释参考

仅仅两年后，炭疽芽孢Kai Kupferschmidt，“炭疽基因组揭示了关于苏联生物武器事故的秘密，”*Science,* 2016年8月16日，www.science.org/content/article/anthrax-genome-reveals-secrets-about-soviet-bioweapons-accident。

转到文本中的注释参考

2007年，T. J. D. Knight-Jones和J. Rushton在“口蹄疫的经济影响——它们是什么，有多大，发生在哪里？”中提到漏水管道，*Preventive Veterinary Medicine,* 2013年11月，www.ncbi.nlm.nih.gov/pmc/articles/PMC3989032/#bib0005。需要注意的是，损害远小于2001年的疫情，那次疫情是由自然原因引起的。

转到文本中的注释参考

2021年，一家制药公司Maureen Breslin，“实验室工作人员在默克设施发现标有‘天花’的瓶子”，*The Hill,* 2021年11月17日，thehill.com/policy/healthcare/581915-lab-worker-finds-vials-labeled-smallpox-at-merck-facility-near-philadelphia。

转到文本中的注释参考

在天花被消灭之前，Sophie Ochmann 和 Max Roser，“天花”，Our World in Data，ourworldindata.org/smallpox；Kelsey Piper，“天花曾经每年杀死数百万人。以下是人类如何战胜它的方式，”*Vox,* 2022年5月8日，www.vox.com/future-perfect/21493812/smallpox-eradication-vaccines-infectious-disease-covid-19。

转到文本中的注释参考

令人难以置信的是，它逃脱了。参见，例如，Kathryn Senio，“最近的新加坡SARS病例是实验室事故”，*Lancet Infectious Diseases,* 2003年11月，www.thelancet.com/journals/laninf/article/PIIS1473-

3099(03)00815-6/全文；Jane Parry，“安全法规的违反可能是最近SARS爆发的原因，WHO表示，”*BMJ,* 2004年5月20日，www.bmj.com/content/328/7450/1222.3；以及Martin Furmanski，“实验室逃逸和‘自我实现预言’流行病，”军备控制中心，2014年2月17日，armscontrolcenter.org/wp-content/uploads/2016/02/Escaped-Viruses-final-2-17-14-copy.pdf。

转到文本中的注释参考

即使 BSL-4 实验室的数量 Alexandra Peters，“高安全性生物实验室的全球扩散：理解现象及其影响”，*Revue Scientifique et Technique,* 2018年12月，pubmed.ncbi.nlm.nih.gov/30964462。实验室的数量在过去两年中从五十九个增加到六十九个，大多数位于城市化环境中，处理致命病原体的实验室数量超过一百个。新一代的“BSL-3+”实验室也迅速增加。参见 Filippa Lentzos 等人，“全球生物实验室报告 2023”，伦敦国王学院，2023年5月16日，www.kcl.ac.uk/warstudies/assets/global-biolabs-report-2023.pdf。

转到文本中的注释参考

1975年至2016年间，David Manheim和Gregory Lewis，“1975-2016年高风险人为病原体暴露事件”，F1000Research，2022年7月8日，f1000research.com/articles/10-752。

转到文本中的注释参考

生物安全官员调查 David B. Manheim，“2020年关于生物隔离实验室事故报告要求和实践的调查结果，”*Health Security* 19, no. 6 (2021), www.liebertpub.com/doi/10.1089/hs.2021.0083。

转到文本中的注释参考

2014年美国风险评估，Lynn C. Klotz 和 Edward J. Sylvester，“潜在大流行病原体实验室泄漏的后果”，*Frontiers in Public Health,*，2014年8月11日，www.frontiersin.org/articles/10.3389/fpubh.2014.00116/full。

转到文本中的注释参考

生物学中很少有领域像这样具有争议性。特别感谢 Jason Matheny 和 Kevin Esvelt 对该主题的讨论。

转到文本中的注释参考

与禽流感相关的工作 马丁·恩瑟林克和约翰·科恩，“两篇备受争议的H5N1论文之一终于发表”，*Science,* 2012年5月2日，www.science.org/content/article/one-two-hotly-debated-h5n1-papers-finally-published。

转到文本中的注释参考

一段时间以来，美国资助机构 Amber Dance，“‘功能增益’研究的变迁”，*Nature,* 2021年10月27日，www.nature.com/articles/d41586-021-02903-x。

转到文本中的注释参考

至少有一些迹象表明，Chan 和 Ridley，*Viral;*“有争议的新研究表明 SARS-CoV-2 具有基因工程的迹象，”*Economist,* 2022年10月27日，www.economist.com/science-and-technology/2022/10/22/a-new-paper-claims-sars-cov-2-bears-signs-of-genetic-engineering。

转到文本中的注释参考

例如，FBI和美国能源部请参见Max Matza和Nicholas Yong，“FBI局长Christopher Wray称中国实验室泄漏最有可能”，BBC，2023年3月1日，www.bbc.co.uk/news/world-us-canada-64806903。

转到文本中的注释参考

在2022年末，NIH的一项研究 Da-Yuan Chen 等人，“ 刺突蛋白在SARS-CoV-2 BA.1 Omicron的致病性和抗原性行为中的作用 ”，bioRxiv，2022年10月14日，www.biorxiv.org/content/10.1101/2022.10.13.512134v1。

转到文本中的注释参考

许多人认为研究不应该 Kiran Stacey，“ 美国卫生官员调查波士顿大学的新冠病毒研究，”*Financial Times,* 2022年10月20日，www.ft.com/content/f2e88a9c-104a-4515-8de1-65d72a5903d0。

转到文本中的注释参考

早期分析 ChatGPT Shakked Noy 和 Whitney Zhang，“ 生成性人工智能对生产力影响的实验证据 ”，麻省理工学院经济学，2023年3月10日，economics.mit.edu/sites/default/files/inline-files/Noy_Zhang_1_0.pdf。

转到文本中的注释参考

这反过来可能会影响招聘。然而，可能的总数较少，但仍然相当可观。参见James Manyika等人，“ 失去的工作，获得的工作：工作的未来将对工作、技能和工资意味着什么 ”，麦肯锡全球研究院，2017年11月28日，www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages。确切的措辞是：“ 我们估计，世界劳动力中人们所从事的所有活动中，大约有一半可能通过适应当前展示的技术实现自动化。 ”第二个统计数据来自Mark Muro等人，“ 自动化和人工智能：机器如何影响人和地方 ”，布鲁金斯大都会政策项目，2019年1月，www.brookings.edu/wp-content/uploads/2019/01/2019.01_BrookingsMetro_Automation-AI_Report_Muro-Maxim-Whiton-FINAL-version.pdf。

转到文本中的注释参考

经济学家达龙·阿西莫格鲁（Daron Acemoglu）和帕斯夸尔·雷斯特雷波（Pascual Restrepo），《机器人与工作：来自美国劳动力市场的证据》，*Journal of Political Economy* 128，第6期（2020年6月），www.journals.uchicago.edu/doi/abs/10.1086/705716。

转到文本中的注释参考

今天，算法执行了绝大部分 同上；爱德华·卢斯，*The Retreat of Western Liberalism*（伦敦：Little, Brown，2017），54。另见贾斯汀·贝尔和丹尼尔·黄，“华尔街员工人数再次下降”，*Wall Street Journal,* 2015年2月19日，www.wsj.com/articles/wall-street-staffing-falls-for-fourth-consecutive-year-1424366858；柳比卡·内德尔科斯卡和格伦达·昆蒂尼，“自动化、技能使用和培训”，经合组织，2018年3月8日，www.oecd-ilibrary.org/employment/automation-skills-use-and-training_2e2f4eea-en

转到文本中的注释参考

经济学家如大卫·奥特（David Autor）大卫·H·奥特（David H. Autor），《为什么仍然有这么多工作？工作场所自动化的历史与未来》，*Journal of Economic Perspectives* 29，第3期（2015年夏季），www.aeaweb.org/articles?id=10.1257/jep.29.3.3。

转到文本中的注释参考

买入这种恐惧 这是阿齐姆·阿扎尔的观点：“不过，总的来说，自动化的持久影响不会是失业”（阿扎尔，*Exponential,* 141）。

转到文本中的注释参考

劳动力市场也存在巨大的摩擦。详见 Daniel Susskind 的著作，*A World Without Work: Technology, Automation and How We Should Respond* (伦敦：Allen Lane，2021)，以获取对这些摩擦的详细说明。

转到文本中的注释参考

The Private Sector Job Quality Index "U.S. Private Sector Job Quality Index (JQI)," University at Buffalo School of Management, Feb. 2023, ubwp.buffalo.edu/job-quality-index-jqi. See also Ford, *Rule of the Robots*.

转到文本中的注释参考

Even those who don't foresee Autor, "Why Are There Still So Many Jobs?"

转到文本中的注释参考

11: 速度 $v*$ 的计算公式如下：

乍一看，马镫白，*Medieval Technology and Social Change*。然而，这一说法并未被普遍接受。对于林恩·怀特著名论点的更为怀疑的解读，请参见例如“伟大的马镫争议”，中世纪技术页面，web.archive.org/web/20141009082354/http://scholar.chem.nyu.edu/tekpages/texts/strpcont.html。

转到文本中的注释参考

这个无法治理的“后主权”世界布朗，*Walled States, Waning Sovereignty*。

转到文本中的注释参考

其风俗、职责和股息 威廉·达尔林普尔，*The Anarchy: The Relentless Rise of the East India Company*（伦敦：布鲁姆斯伯里，2020），233。

转到文本中的注释参考

组织也是一种智能，理查德·丹齐格首次在晚餐时向我提出了这个想法，然后发表了一篇优秀的论文：“机器、官僚机构和市场作为人工智能”，安全与新兴技术中心，2022年1月，cset.georgetown.edu/wp-content/uploads/Machines-Bureaucracies-and-Markets-as-Artificial-Intelligences.pdf。

转到文本中的注释参考

要了解这些“全球500强”的集中度，*Fortune,* fortune.com/global500/。截至2022年10月。世界银行的数字显示略低：世界银行，“GDP（现价美元）”，世界银行数据，data.worldbank.org/indicator/NY.GDP.MKTP.CD。

转到文本中的注释参考

公司已经控制了最大的集群 Benaich 和 Hogarth，
State of AI Report 2022.

转到文本中的注释参考

已经有一个显著的詹姆斯·曼尼卡等人，“ 超级明星：引领全球经济的公司、行业和城市的动态 ”，麦肯锡全球研究院，2018年10月24日，www.mckinsey.com/featured-insights/innovation-and-growth/superstars-the-dynamics-of-firms-sectors-and-cities-leading-the-global-economy。

转到文本中的注释参考

这些争议中有百分之九十 Colin Rule，“ 将人从问题中分离出来， ”*The Practice,* 2020年7月，thepractice.law.harvard.edu/article/separating-the-people-from-the-problem。

转到文本中的注释参考

零边际成本生产 参见，例如，Jeremy Rifkin，*The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism*（纽约：Palgrave，2014年）。

转到文本中的注释参考

此外，最好的系统的拥有者，埃里克·布林约尔松称之为一种情况，即人工智能接管越来越多的经济，将大量人锁定在一个没有工作、没有财富、没有有意义权力的平衡中，称为“ 图灵陷阱 ”。埃里克·布林约尔松，“ 图灵陷阱：类人人工智能的承诺与危险 ”，斯坦福数字经济实验室，2022年1月11日，arxiv.org/pdf/2201.04200.pdf。

转到文本中的注释参考

难怪有这样的讨论 参见，例如，乔尔·科特金，*The Coming of Neo-feudalism: A Warning to the Global Middle Class*（纽约：

邂逅图书，2020年）。

转到文本中的注释参考

尽管革命者詹姆斯·C斯科特（James C. Scott）付出了最大的努力，*Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*（纽黑文，康涅狄格州：耶鲁大学出版社，1998）。

转到文本中的注释参考

在路上，以及整天“伦敦有多少个CCTV摄像头？”，CCTV.co.uk，2020年11月18日，www.cctv.co.uk/how-many-cctv-cameras-are-there-in-london。GO TO NOTE REFERENCE IN TEXT

与西贝奈奇和霍加斯相比，*State of AI Report 2022.*

转到文本中的注释参考

到2015年，这成为了灵感来源 Dave Gershgorn，“中国的‘锐眼’计划旨在监控100%的公共空间，”*OneZero,* 2021年3月2日，onezero.medium.com/chinas-sharp-eyes-program-aims-to-surveil-100-of-public-space-ddc22d63e015。

转到文本中的注释参考

一组顶尖研究人员舒庆·简·陈，“商汤科技：中国人工智能独角兽背后的面孔，”*Forbes,* 2018年3月7日，www.forbes.com/sites/shuchingjeanchen/2018/03/07/the-faces-behind-chinas-omniscient-video-surveillance-technology。

转到文本中的注释参考

中国警察甚至有太阳镜 Sofia Gallarate，“中国警察佩戴人脸识别太阳镜”，Fair Planet，2019年7月9日，www.fairplanet.org/story/chinese-police-officers-are-wearing-facial-recogni%C2%ADtion-sunglasses。

转到文本中的注释参考

全球约一半的十亿监控摄像头 本文及以下统计数据取自 *New York Times*调查：Isabelle Qian等人，“ 关于中国扩张的监控国家的四个要点 ”，*New York Times,* 2022年6月21日，www.nytimes.com/2022/06/21/world/asia/china-surveillance-investigation.html。

转到文本中的注释参考

这个支持人工智能的系统可以发现 Ross Andersen，“ 全景监狱已经在这里，”*Atlantic,* 2020年9月，www.theatlantic.com/magazine/archive/2020/09/china-ai-surveillance/614197。GO TO NOTE REFERENCE IN TEXT

当局已经建立了虹膜扫描钱等，“ 从《纽约时报》对中国扩展监控国家的调查中得出的四个要点。”

转到文本中的注释参考

2019年，美国政府禁止了“ NDAA第889条 ”，GSA SmartPay，smartpay.gsa.gov/content/ndaa-section-889。GO TO NOTE REFERENCE IN TEXT

然而，仅仅一年后，三个联邦机构 Conor Healy，“ 美国军方和政府违反法律，购买被禁的 Dahua/Lorex，国会委员会呼吁调查，”IPVM，2019年12月1日，ipvm.com/reports/usg-lorex。

转到文本中的注释参考

超过一百个美国城镇 Zack Whittaker，“ 美国城镇正在购买与维吾尔族虐待相关的中国监控技术，”TechCrunch，2021年5月24日，techcrunch.com/2021/05/24/united-states-towns-hikvision-dahua-surveillance。

GO TO NOTE REFERENCE IN TEXT

在智能仓库中，Joshua Brustein，“仓库正在追踪工人的每一个肌肉动作”，彭博社，2019年11月5日，www.bloomberg.com/news/articles/2019-11-05/am-i-being-tracked-at-work-plenty-of-warehouse-workers-are。

转到文本中的注释参考

公司如 Vigilant Solutions Kate Crawford，*Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*（康涅狄格州纽黑文：耶鲁大学出版社，2021年）。

转到文本中的注释参考

即使是你的外卖披萨 Joanna Fantozzi，“Domino’s 使用 AI 摄像头确保披萨烹饪正确，”*Nation’s Restaurants News,* 2019年5月29日，www.nrn.com/quick-service/domino-s-using-ai-cameras-ensure-pizzas-are-cooked-correctly。

转到文本中的注释参考

在浪潮来临之前，考虑到像戴夫·艾格斯的 *The Every* 这样的关于监控反乌托邦的最新小说，在具体监控内容方面并没有真正的变化，并且不是作为遥远的科幻小说呈现，而是作为对当代科技公司的讽刺。

转到文本中的注释参考

这可能是武装最好的 分析师是以色列国家安全研究所的退役准将阿萨夫·奥里恩。“美以关系的未来研讨会”，外交关系委员会，2019年12月2日，www.cfr.org/event/future-us-israel-relations-symposium，引用自 Kali Robinson，“真主党是什么？”，外交关系委员会，2022年5月25日，www.cfr.org/backgrounder/what-hezbollah。

转到文本中的注释参考

它还进行各种商业活动，例如，“解释：真主党如何通过其‘毒品恐怖主义战略’建立毒品帝国，”*Arab News,* 2021年5月3日，www.arabnews.com/node/1852636/middle-east。

转到文本中的注释参考

相反，这是一个奇怪的Lina Khatib，“真主党如何控制黎巴嫩国家”，查塔姆研究所，2021年6月30日，www.chathamhouse.org/sites/default/files/2021-06/2021-06-30-how-hezbollah-holds-sway-over-the-lebanese-state-khatib.pdf。

转到文本中的注释参考

然而，未来的浪潮将是极大地扩展某些现有趋势，例如在集中化过程中，私人行为者承担更多传统上被认为是国家专属的角色。参见，例如，Rodney Bruce Hall 和 Thomas J. Biersteker,
The Emergence of Private Authority in Global Governance（英国剑桥：剑桥大学出版社，2002年）。

转到文本中的注释参考

请注意，仅在过去十年中，“2019年可再生能源发电成本”，IRENA，2020年6月，www.irena.org/publications/2020/Jun/Renewable-Power-Costs-in-2019。

转到文本中的注释参考

超级自由意志主义技术专家詹姆斯·戴尔·戴维森和威廉·里斯-莫格，*The Sovereign Individual: Mastering the Transition to the Information Age*（纽约：Touchstone，1997年）。

转到文本中的注释参考

公共服务的篝火 彼得·蒂尔，“一个自由意志主义者的教育，”*Cato Unbound,* 2009年4月13日，www.cato-unbound.org/2009/04/13/peter-thiel/education-libertarian。参见 Balaji Srinivasan，*The Network State* (1729

publishing, 2022), for a more thoughtful take on how technological constructs might supersede the nation-state.

转到文本中的注释参考

12: 速度 $v*$

英格兰的人口 尼尔·弗格森，*Doom: The Politics of Catastrophe* (伦敦：艾伦·莱恩，2021)，131。

转到文本中的注释参考

第一次世界大战的死亡人数来自同上。

转到文本中的注释参考

一种新型人类可传播病毒 从一份机密简报中获取的数据，但我们了解到生物安全专家认为这是合理的。

转到文本中的注释参考

不难看出，三分之一从事人工智能研究的科学家认为这可能导致一场灾难。Jeremy Hsu，“三分之一的人工智能研究科学家表示这可能引发全球灾难，”*New Scientist,* 2022年9月22日，www.newscientist.com/article/2338644-a-third-of-scientists-working-on-ai-say-it-could-cause-global-disaster。

转到文本中的注释参考

奥姆真理教成立于20世纪80年代，参见Richard Danzig和Zachary Hosford，“Aum Shinrikyo—Second Edition—English”，CNAS，2012年12月20日，www.cnas.org/publications/reports/aum-shinrikyo-second-edition-english；以及Philipp C. Bleak，“重新审视奥姆真理教：对迄今为止最广泛的非国家生物武器计划的新见解”，詹姆斯·马丁防扩散研究中心，2011年12月10日，www.nti.org/analysis/articles/revisiting-aum-shinrikyo-new-insights-most-extensive-non-state-biological-weapons-program-date-1。

转到文本中的注释参考

在奥姆真理教最受欢迎的时期，美国科学家联合会，“奥姆的运作”，在大规模杀伤性武器的全球扩散中

破坏：奥姆真理教案例研究，参议院政府事务常设调查小组委员会，1995年10月31日，irp.fas.org/congress/1995_rpt/aum/part04.htm。

转到文本中的注释参考

关于丹齐格和霍斯福德的影响的报告，“奥姆真理教”。

转到文本中的注释参考

有人不可避免地会说，例如，参见 Nick Bostrom 的“脆弱世界假说”，2019 年 9 月 6 日，nickbostrom.com/papers/vulnerable.pdf，可能是该论题最成熟的版本。在一个回应“简易核武器”前景的思想实验中，他设想了一个“高科技全景监狱”，每个人都有一个“自由标签”，“佩戴在脖子上，装饰有多向摄像头和麦克风。加密的视频和音频会从设备中持续上传到云端，并实时进行机器解读。AI 算法对佩戴者的活动、手部动作、附近物体和其他情境线索进行分类。如果检测到可疑活动，视频流会被传送到几个爱国者监控站之一。”

转到文本中的注释参考

波动同时提供了动机 Martin Bereaja 等人，“AI-tocracy,” *Quarterly Journal of Economics,* 2023年3月13日，academic.oup.com/qje/advance-article-abstract/doi/10.1093/qje/qjad012/7076890。

转到文本中的注释参考

在这个连续体上，Balaji Srinivasan 预见到的结果非常类似，美国是僵尸，中国是恶魔：“随着美国陷入无政府状态，中共将其功能正常但高度不自由的制度视为唯一的替代方案，并将其监控国家的现成版本出口到其他国家，作为下一版本的一带一路，正如

一项“基础设施”，附带中国全视AI之眼的SaaS订阅。”Srinivasan, *The Network State,* 162。

转到文本中的注释参考

技术哲学家伊西斯·哈泽温杜斯，“巨型机器的威胁”，*IfThenElse,* 2021年11月21日，www.ifthenelse.eu/blog/the-threat-of-the-megamachine。

转到文本中的注释参考

六十个文明的调查 迈克尔·舍默，“为什么外星人没有打电话”，*Scientific American,* 2002年8月，michaelshermer.com/sciam-columns/why-et-hasnt-called。

转到文本中的注释参考

没有新技术伊恩·莫里斯，*Why the West Rules—for Now: The Patterns of History and What They Reveal About the Future*（伦敦：Profile Books，2010）；坦特，*The Collapse of Complex Societies;* 戴蒙德，*Collapse*。

转到文本中的注释参考

在接下来的一个世纪中，Stein Emil Vollset 等人，“2017年至2100年195个国家和地区的生育率、死亡率、迁移和人口情景：全球疾病负担研究的预测分析”，*Lancet,* 2020年7月14日，www.thelancet.com/article/S0140-6736(20)30677-2/fulltext。

转到文本中的注释参考

包括日本、德国在内的国家，彼得·泽汉，*The End of the World Is Just the Beginning: Mapping the Collapse of Globalization*（纽约：Harper Business，2022年）。

转到文本中的注释参考

中国是主要部分 彭秀健，“ 中国的人口会开始下降吗？ ” BBC Future ，2022年6月6日，www.bbc.com/future/article/20220531-why-chinas-population-is-shrinking。

转到文本中的注释参考

所有这些意味着 Zeihan， *The End of the World Is Just the Beginning,* 203。GO TO NOTE REFERENCE IN TEXT

锂、钴的需求 “ 气候智能采矿：气候行动的矿物 ”，世界银行，www.worldbank.org/en/topic/extractiveindustries/brief/climate-smart-mining-minerals-for-climate-action.

转到文本中的注释参考

鉴于人口和资源限制 Galor， *The Journey of Humanity,* 130。

转到文本中的注释参考

1955年，约翰·冯·诺依曼在他生命的晚期发表了 “ 我们能否在技术中生存？ ” 一文，见于*The Neumann Compendium*（新泽西州河滨：世界科学出版社，1995年），geosci.uchicago.edu/~kite/doc/von_Neumann_1955.pdf。

转到文本中的注释参考

13: 速度 $v*$ 的变化

他们如何解释一个时代 David Cahn 等人，“ AI 2022: The Explosion, ” Coatue Venture, coatue-external.notion.site/ AI-2022-The- Explosion-e76afd140f824f2eb6b049c5b85a7877。

转到文本中的注释参考

正如我们所见，即使是富裕国家“ 2021年GHS指数美国国家概况 ”，全球健康安全指数，www.ghsindex.org/country/united-states。

转到文本中的注释参考

然而，一系列灾难性的决策 Edouard Mathieu 等人，“ 冠状病毒（COVID-19）死亡 ”，Our World in Data，ourworldindata.org/covid-deaths。

转到文本中的注释参考

国家预算用于此类事项 例如，与1957年亚洲流感期间相比，美国联邦预算在绝对值上当然要大得多，但作为GDP的百分比也更大（16.2%对20.8%）。1957年时，还没有专门的卫生部，CDC的前身还是一个成立仅十一年的相对年轻的组织。Ferguson, *Doom,* 234.

转到文本中的注释参考

最具雄心的立法《人工智能法案》，生命未来研究所，artificialintelligenceact.eu。

转到文本中的注释参考

一些人认为它过于专注，例如，参见“ FLI关于欧盟人工智能法案的立场文件 ”，生命未来研究所，2021年8月4日，futureoflife.org/wp-content/uploads/2021/08/FLI-Position-Paper-on-the-EU-AI-Act.pdf?x72900；以及David Matthews，“ 专家警告欧洲议会议员欧盟人工智能法案不‘面向未来’ ”，科学商业，2022年3月22日，

sciencebusiness.net/ news/ 欧盟人工智能法案并非面向未来——专家警告欧洲议会议员。

转到文本中的注释参考

一些人认为这让大型科技公司Khari Johnson，“ 定义AI高风险时刻的斗争，”*Wired,* 2021年9月1日，www.wired.com/story/fight-to-define-when-ai-is-high-risk。

转到文本中的注释参考

而且有 135 万人“ 全球道路安全统计 ”，Brake，www.brake.org.uk/get-involved/take-action/mybrake/knowledge-centre/global-road-safety#.

转到文本中的注释参考

政府发布了詹妮弗·康拉德的文章，“ 中国即将监管人工智能——全世界都在关注 ”，*Wired,* 2022年2月22日，www.wired.com/story/china-regulate-ai-world-watching。GO TO NOTE REFERENCE IN TEXT

它主动禁止了各种基督教史密斯，“ 中国的游戏法律正在进一步打击 ”，SVG，2022年3月15日，www.svg.com/799717/chinas-gaming-laws-are-cracking-down-even-further。GO TO NOTE REFERENCE IN TEXT

推荐草案《国家互联网信息办公室关于互联网信息服务算法推荐管理的规定（征求意见稿）》公开征求意见的通知，中国国家互联网信息办公室，2021年8月27日，www.cac.gov.cn/2021-08/27/c_1631652502874117.htm。

转到文本中的注释参考

它没有描述国家如何运作，例如，Alex Engler，“欧盟人工智能法案的有限全球影响”，布鲁金斯学会，2022年6月14日，www.brookings.edu/blog/techtank/2022/06/14/the-limited-global-impact-of-the-eu-ai-act。一项对25万份国际条约的研究表明，它们往往未能实现其目标。参见Steven J. Hoffman等人，“国际条约大多未能产生预期效果”，*PNAS,* 2022年8月1日，www.pnas.org/doi/10.1073/pnas.2122854119。

转到文本中的注释参考

我们的史前大脑 参见乔治·马歇尔，*Don't Even Think About It: Why Our Brains Are Wired to Ignore Climate Change*（纽约：Bloomsbury，2014），对此观点的详细阐述。

转到文本中的注释参考

早在20世纪70年代，Rebecca Lindsey，“气候变化：大气二氧化碳”，Climate.gov，2022年6月23日，www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide。

转到文本中的注释参考

14: 速度 $v*$ 的变化率是加速度。

国际原子能机构“ IAEA安全标准 ”，国际原子能机构，www.iaea.org/resources/safety-standards/search?facility=All&term_node_tid_depth_2=All&field_publication_series_info_value=&combine=&items_per_page=100。

转到文本中的注释参考

生物武器的主要监测者托比·奥德，*The Precipice: Existential Risk and the Future of Humanity*（伦敦：布鲁姆斯伯里，2020年），57

转到文本中的注释参考

AI安全研究人员的数量Benaich和Hogarth，*State of AI Report 2022.*

转到文本中的注释参考

鉴于大约有多少AI研究人员的估计，请参见“ 什么是有效利他主义？ ”，www.effectivealtruism.org/articles/introduction-to-effective-altruism#fn-15。

转到文本中的注释参考

原始阿波罗任务 NASA，“ 阿波罗的益处：技术的巨大飞跃 ”，NASA事实，2004年7月，www.nasa.gov/sites/default/files/80660main_ApolloFS.pdf。

转到文本中的注释参考

发光 Kevin M. Esvelt，“ 延迟、检测、防御：为数千人能够释放新流行病的未来做准备 ”，日内瓦安全政策中心，2022年11月14日，dam.gcsp.ch/files/doc/gcsp-geneva-paper-29-22。

转到文本中的注释参考

还有一项由 Jan Leike 完成的出色工作，“ Alignment Optimism, ” *Aligned,* 2022年12月5日，aligned.substack.com/p/alignment-optimism。

转到文本中的注释参考

计算机科学家斯图尔特·拉塞尔，*Human Compatible*。

转到文本中的注释参考

这意味着攻击您的系统 Deep Ganguli 等人，“ 红队语言模型以减少危害：方法、扩展行为和经验教训 ”，arXiv，2022年11月22日，arxiv.org/pdf/2209.07858.pdf。

转到文本中的注释参考

在技术方面，Sam R. Bowman 等人，“ 衡量大型语言模型可扩展监督的进展 ”，arXiv，2022年11月11日，arxiv.org/abs/2211.03540。

转到文本中的注释参考

目前只有一小部分安全 DNA 项目，“ 保障全球生物技术 ”，SecureDNA，www.securedna.org。

转到文本中的注释参考

习近平担心本·墨菲，“ 瓶颈：中国自我识别的战略技术进口依赖 ”，安全与新兴技术中心，2022年5月，cset.georgetown.edu/publication/chokepoints。

转到文本中的注释参考

确实，中国花费更多克里斯·米勒，*Chip War: The Fight for the World's Most Critical Technology*（纽约：Scribner，2022年）。

转到文本中的注释参考

一位技术高管Demetri Sevastopulo和Kathrin Hille，“美国对中国实施全面技术出口管制，”*Financial Times,* 2022年10月7日，www.ft.com/content/6825bee4-52a7-4c86-b1aa-31c100708c3e。

转到文本中的注释参考

在短期到中期内，Gregory C. Allen，“扼杀中国对未来人工智能的访问”，战略与国际研究中心，2022年10月11日，www.csis.org/analysis/choking-chinas-access-future-ai。

转到文本中的注释参考

如果需要数千亿朱莉·朱，“中国准备1430亿美元计划以应对美国限制，”路透社，2022年12月14日，www.reuters.com/technology/china-plans-over-143-bln-push-boost-domestic-chips-compete-with-us-sources-2022-12-13。GO TO NOTE REFERENCE IN TEXT

美国制造商NVIDIA，Stephen Nellis和Jane Lee，“Nvidia调整旗舰H100芯片以出口到中国为H800，”路透社，2023年3月22日，www.reuters.com/technology/nvidia-tweaks-flagship-h100-chip-export-china-h800-2023-03-21。

转到文本中的注释参考

ASML的机器不仅仅是机器，还有许多零部件只有一个制造商，比如Cymer的高端激光器或Zeiss的镜子，它们纯净到如果它们有德国那么大，不规则之处也只有几毫米宽。

转到文本中的注释参考

这三家公司有 例如，2022年5月25日，Michael Filler 在 Twitter 上的发言，twitter.com/michaelfiller/status/1529633698961833984。

转到文本中的注释参考

稀土危机“我们矿产资源供应面临的最大风险在哪里？”，美国地质调查局，2020年2月21日，www.usgs.gov/news/national-news-release/new-methodology-identifies-mineral-commodities-whose-supply-disruption?qt-news_science_products=1#qt-news_science_products。

转到文本中的注释参考

大约80%的高质量石英Zeihan，*The End of the World Is Just the Beginning,* 314。

转到文本中的注释参考

确实，有时尖锐的批评 Lee Vinsel，“你做错了：关于批评和技术炒作的笔记”，Medium，2021年2月1日，sts-news.medium.com/youre-doing-it-wrong-notes-on-criticism-and-technology-hype-18b08b4307e5。

转到文本中的注释参考

有希望的是，关于伦理人工智能的研究，斯坦福大学以人为中心的人工智能，人工智能指数报告2021。

转到文本中的注释参考

主要缺陷 例如，Shannon Vallor，“动员艺术和人文学科的智力资源”，Ada Lovelace研究所，2021年6月25日，www.adalovelaceinstitute.org/blog/mobilising-intellectual-resources-arts-humanities。

转到文本中的注释参考

形成联盟 凯·C·詹姆斯在推特上，2019年3月20日，twitter.com/KayColesJames/status/1108365238779498497。GO TO NOTE REFERENCE IN TEXT

有很大可能出现积极的“B公司‘超越’常规业务”，B Lab，2023年3月1日，www.bcorporation.net/zh-cn/news/press/b-

企业超越常规业务，迎接 2023 年 B Corp 月。

转到文本中的注释参考

尽管今天公司拥有“ 美国研究与发展资金和绩效：事实表 ”，国会研究服务处，2022年9月13日，sgp.fas.org/crs/misc/R44307.pdf。

转到文本中的注释参考

投资于科学和技术 参见，例如，Mariana Mazzucato，*The Entrepreneurial State: Debunking Public vs. Private Sector Myths*（伦敦：Anthem Press，2013年）。

转到文本中的注释参考

他们的首要任务应该是 Jess Whittlestone 和 Jack Clark 在“ 政府为何以及如何监控 AI 发展 ”一文中很好地阐述了这些观点，arXiv，2021年8月31日，arxiv.org/pdf/2108.12427.pdf。

转到文本中的注释参考

2015年几乎没有“ 与人工智能相关的立法 ”，国家州议会会议，2022年8月26日，www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx。

GO TO NOTE REFERENCE IN TEXT

OECD AI政策观察站OECD，“ 国家AI政策与战略 ”，OECD AI政策观察站，oecd.ai/en/dashboards/overview。GO TO NOTE REFERENCE IN TEXT

2022年，白宫发布了“ 事实清单：拜登-哈里斯政府宣布推进技术问责制和保护美国公众权利的关键行动 ”，白宫，2022年10月4日，www.whitehouse.gov/ostp/news-updates/2022/10/04/fact-sheet-biden-

哈里斯政府宣布推进技术问责制和保护美国公众权利的关键行动。

转到文本中的注释参考

今天美国的劳动被征税 Daron Acemoglu 等人，“ 税收、自动化和劳动的未来 ”，麻省理工学院未来工作，mitsloan.mit.edu/shared/ods/documents?PublicationDocumentID=7929。

转到文本中的注释参考

这有时被称为 Arnaud Costinot 和 Ivan Werning 的“ 机器人、贸易和卢德主义：最佳技术监管的充分统计方法 ”，*Review of Economic Studies,* 2022年11月4日，academic.oup.com/restud/advance-article/doi/10.1093/restud/rdac076/6798670。

转到文本中的注释参考

麻省理工学院的经济学家提出了Daron Acemoglu等人，“ 美国税法是否有利于自动化？ ”，*Brookings Papers on Economic Activity*（2020年春季），www.brookings.edu/wp-content/uploads/2020/12/Acemoglu-FINAL-WEB.pdf。

转到文本中的注释参考

在一个超大规模的时代，萨姆·阿尔特曼，“ 万物的摩尔定律 ”，萨姆·阿尔特曼，2021年3月16日，moores.samaltman.com。

转到文本中的注释参考

使用致盲激光武器 《关于某些常规武器的公约》，联合国，www.un.org/disarmament/the-convention-on-certain-conventional-weapons.

转到文本中的注释参考

对106个国家的研究 Françoise Baylis 等人，“ 人类生殖系和可遗传基因组编辑：全球政策格局 ”，*CRISPR Journal,* 2020年10月20日，www.liebertpub.com/doi/10.1089/crispr.2020.0082。

转到文本中的注释参考

在第一次 Eric S. Lander 等人，“ 对可遗传基因组编辑采取暂停措施 ”之后，*Nature,* 2019年3月13日，www.nature.com/articles/d41586-019-00726-5。

转到文本中的注释参考

到2010年代初，彼得·迪齐克斯，“ 研究：商业航空旅行比以往任何时候都更安全，”*MIT News,* 2020年1月23日，news.mit.edu/2020/study-commercial-flights-safer-ever-0124。GO TO NOTE REFERENCE IN TEXT

我们于2017年再次相遇在“ AI原则 ”，生命未来研究所，2017年8月11日，futureoflife.org/open-letter/ai-principles。

转到文本中的注释参考

社会和道德责任 约瑟夫·罗特布拉特，“ 科学家的希波克拉底誓言 ”，*Science,* 1999年11月19日，www.science.org/doi/10.1126/science.286.5444.1475。

转到文本中的注释参考

因为 ***we*** 构建技术 参见，例如，Rich Sutton 的提议，“ 创建人类水平的 AI：如何以及何时？ ”，加拿大阿尔伯塔大学，futureoflife.org/data/PDF/rich_sutton.pdf?x72900；Azeem Azhar，“ 我们是决定我们想要从我们构建的工具中得到什么的人 ”（Azhar，*Exponential,* 253）；或李开复，“ 我们不会在 AI 的故事中成为被动的旁观者——我们是它的作者 ”（李开复和程秋帆，*AI 2041: Ten Visions for Our Future* [伦敦：W. H. Allen, 2021, 437]）。

转到文本中的注释参考

围绕人工智能的沟通 Patrick O ’ Shea 等人，“ 关于人工智能社会影响的沟通：一个框架战略简报 ”，框架研究所，2021年10月19日，www.frameworksinstitute.org/publication/communicating-about-the-social-implications-of-ai-a-frameworks-strategic-brief。

转到文本中的注释参考

研究表明，当 Stefan Schubert 等人，“ 存在风险的心理学：关于人类灭绝的道德判断 ”，*Nature Scientific Reports,* 2019年10月21日，www.nature.com/articles/s41598-019-50145-9。

转到文本中的注释参考

同时，公民大会提供了Aviv Ovadya，“ 走向平台民主 ”，哈佛肯尼迪学院贝尔弗中心，2021年10月18日，www.belfercenter.org/publication/towards-platform-democracy-policymaking-beyond-corporate-ceos-and-partisan-pressure。

转到文本中的注释参考

参考阿西洛马原则“ 暂停巨型人工智能实验：一封公开信 ”，生命未来研究所，2023年3月29日，futureoflife.org/open-letter/pause-giant-ai-experiments。

转到文本中的注释参考

对LLM的投诉 Adi Robertson，“ AI政策小组称FTC应阻止OpenAI推出新的GPT模型，”*The Verge,* 2023年3月30日，www.theverge.com/2023/3/30/23662101/ftc-openai-investigation-request-caidp-gpt-text-generation-bias。

转到文本中的注释参考

一个很好的例子来自 Esvelt，“ 延迟、检测、防御。 ”关于另一种整体遏制策略的例子，请参见 Allison Duettmann，“ 防御物理威胁：多极主动防护 ”，Foresight Institute，2022 年 2 月 14 日，foresightinstitute.substack.com/p/defend-physical。

转到文本中的注释参考

经济学家达龙·阿西莫格鲁（Daron Acemoglu）和詹姆斯·罗宾逊（James Robinson），*The Narrow Corridor: How Nations Struggle for Liberty*（伦敦：维京出版社，2019年）。

转到文本中的注释参考

```# 介绍在这篇文章中，我们将探讨如何计算速度的最优值 $v*$。速度是一个重要的物理量，它在许多领域中
技术应该放大 例如，参见 Divya Siddarth 等人的论点，“ 人工智能如何辜负我们 ”，Edmond 和 Lily Safra 伦理中心，2021 年 12 月 1 日，ethics.harvard.edu/how-ai-fails-us。

转到文本中的注释参考
```

索引

本索引中的页码指的是书籍的印刷版本。每个链接将带您到相应印刷页的开头。您可能需要从该位置向前滚动，以在您的电子阅读器上找到相应的参考。

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A

Aadhaar, 126

学术界, 128

新技术的可访问性

AI, 69

去中心化和, 198 去物质化和, 190 基因工程, 82 – 83 军事应用和, 106, 126

权力和，163 – 64，198，199
Acemoglu, Daron，179，276，278
Acheson-Lilienthal 报告，42 增材制造。*See* 3-D 打印 Aerorozvidka，103 – 4 农业 挑战，136 – 37 气候变化与，138 作为通用技术，27 利润动机与，133 机器人与，93 – 94 作为技术浪潮，28

人工智能法案（欧盟），229，260
AI事件数据库，246
空气间隙，241 航空公司，267 – 68 AlexNet，58 – 59 算法，63，114，247 阿里巴巴，66 艾利森，格雷厄姆，123

字母表，128，256，257 AlphaFold，89 – 90，109 AlphaGo，53 – 54，113，117 – 19，120 AlphaZero，54 替代方案，234 Altos Labs，85 亚马逊，94 – 95，189 Anduril，166 炭疽，174 抗衰老技术。*See* 长寿技术 抗生素，36 苹果，187，189 Arkhipov，Vasili，44 军备竞赛言论，124 – 25，126 – 27 Arnold，Frances，87 Arthur，W. Brian，56 人工能力智能（ACI），vii，77 – 78，115，164，210 人工通用智能（AGI）灾难场景和，209，210 聊天机器人和，114

DeepMind 创立和， 8

定义， vii，51 大猩猩问题和， 115
– 16 渐进性质，75 超级智能和， 75， 7
7， 78， 115 尚未到来，73 – 74 人工智
能（AI）对， 7 – 8 的愿景，自治和， 1
14， 115 作为即将到来的浪潮的基础，55
的好处， 10 – 11 灾难场景和，208，20
9 – 11 聊天机器人， 64， 68， 70， 113
– 14 中国的发展， 120 – 21 瓶颈， 251
气候变化和， 139 意识和， 74，75 矛盾
和， 202 成本， 64，68 当前应用， 61
– 62

当前能力， 8 – 9 网络攻击和，
162 – 63， 166 – 67 定义， vii

早期实验中， 51 – 54

效率， 68 – 69 自我与， 140 伦理与，
254 解释与， 243 未来，78 未来的普及
性， 284 – 85 全球影响力， 9 – 10 幻觉
问题与， 243 人脑作为固定目标， 67
– 68 超进化与， 109 隐形，73 限制，7
3 医疗应用， 110 军事应用， 104， 16
5 现代图灵测试， 76 – 77， 78， 115
， 190

，210

狭隘的性质， 73 – 74

近期能力，77 全能技术和，　111，　130 开放性命令和，　128 – 29 潜力，　56，　70，　135 作为优先事项，60 利润动机和，　134，　135，　136 扩散，　68 – 69 蛋白质结构和，　88 – 89 红队和，　246 监管尝试，　229，　260 – 61 研究不可预测性和，　130 机器人和，　95，　96，98 安全和，　241，　243 – 44 扩展假设，　67 – 68，　74 自我批评文化和，　270 感知声明，　72，75 怀疑，　72，　179 监控和，　193 – 94，　195，　196 合成生物学和，　89 – 90，　109 技术性失业和，　177 – 8

图灵测试，75

See also 即将到来的浪潮；深度学习；机器学习 arXiv，129

阿西洛马原则，269–70，272–73

ASML，251

不对称影响，105–7，234

亚特兰蒂斯，5

Atmanirbhar Bharat 计划（印度），125–26

注意，63

注意力图，63 审计，245–48，267 奥姆真理教，212–13，214 威权主义，153，158–59，191–96，216–17 自动补全，63 自动化药物发现，110 自动化，177–81 自主性，105，113–15，166，234 奥托，大卫，179 奥拉基，安瓦尔，171

源文本：B翻译文本：

反向传播，59

不良行为者赋权，165–66，208，266
See also 恐怖主义 B 公司，258

贝尔，亚历山大·格雷厄姆，31

奔驰，卡尔，24，285 伯格，保罗，269–70 华大基因，122 偏见，69–70，239–40 Bioforge，86 生物武器公约，241，263 生物技术。*See* 合成生物学 黑死病，205 布莱切利园，32 书籍，30 博耶，赫伯特·W.，80 *Breakout,* 51–52，113 布林，谢尔盖，53

不列颠东印度公司， 186， 189 布朗，戴维，98 布朗，温迪， 185

源文本：C翻译文本：

“ 我们能否在技术中生存 ”（Neumann），221 资本支出， 134 Carey, Nessa, 82 Carlson 曲线, 81 汽车。*See* 车辆 Cartwright, Edmund ， 282

Cas9, 81

灾难， 205 – 14 控制和， 211 – 12 ， 214 场景， 207 – 9 怀疑， 206 – 7 停滞作为， 217 – 21 监控反乌托邦作为， 215 – 17 Caulobacter et hensis-2.0， 84

大提琴，109 CFCs，45，46
，263 查理·马特，183 – 84 夏
庞捷，埃马纽埃尔，81，265
聊天机器人，64，68，70，
113 – 14 ChatGPT，62，65

化学武器，46，110

化学，98 – 99
切尔诺贝利，45

国际象棋，53

中国地缘政治和，120 – 24 历史遏制尝试
，39 – 40 国际合作和，265 – 66 毛主义，
192 监管和，231 丝绸和，41 监控，193
– 95 美国出口管制和，249 – 50 毛丝鼠，6
8

芯片瓶颈和，　249 – 51 超进化和，　32 – 33，　57，　81，　108 大小为，67 *See also*　半导体 氯氟烃（CFCs），35 CICERO，　167 气候变化遏制尝试，　45，　46，　263 作为全球挑战，　137 – 39 大众理解，　236 技术的意外后果和，35 时钟，157 云计算，251 云从科技，　194 科恩，斯坦利·N.，80 冷战遏制和，37 国际合作和，　263，　264 核武器和，42

斯普特尼克和， 119 – 20， 126

即将到来的浪潮

AI 上， 3 – 4 AI-合成生物学互动和， 88 – 91 自主性和， 105， 113 – 15 避免， 12 – 14 的好处， 10， 11， 16， 283 定义， vii， 7 自我作为驱动因素， 139 – 41 全球挑战和， 137 – 40 理想主义和， 150 – 51 的必然性， 142 – 43， 214， 225 相互关联的性质， 56 – 57 大众理解， 234 – 36 可能的反应，11 权力和， 102， 163 – 64 利润动机和， 131 – 36 研究不可预测性和， 129 – 30 怀疑， 57 – 58， 72， 102， 179， 206 – 7 技术半影， 92 – 93

See also 新技术的可访问性； 即将到来 波动激励;
全用途技术即将到来的激励浪潮， 119
， 141 控制和， 250 地缘政治， 119 – 27 全球
挑战， 137 – 40， 219 开放性必要性， 127 – 29
利润动机， 131 – 36， 254 – 58 监管和， 231 –
32 即将到来的技术浪潮特征不对称影响， 105 –
7， 234 自主性， 105， 113 – 15， 166， 234
控制和， 234 – 35 *See also* 超进化； 全用途技
术计算机视觉， 58 – 60 计算， 32 – 34 去物质
化和， 55， 190 超进化， 108 制造过程和，84
量子计算， 97 – 99， 109， 114， 122

美国出口管制和，　249 – 50

联盟，84

包含

审计和，　245 – 48，开始，　267 灾难和，　272 – 73 瓶颈和，　211 – 12，　214 关键点和，　249 – 51 一致性和，　275 批评者和，　252 – 54 定义，　vii，　37 – 38 元素，　232 – 33 大猩猩问题和，　115 – 16 政府角色，　258 – 63 历史尝试，　38 – 41，　42 – 43，　119 工业革命和，　39，　40，　281 – 82 综合方法，　228 – 29 国际合作和，　263 – 67 制造者责任，　252 作为狭窄走廊，　276 – 78 需要，　16，　18 – 19，　46，　48，　285 – 86

关闭开关和， 244 – 45 全能技术和， 111 组织限制和， 228 概述， 274 – 75 病原体和， 273 – 74 大众运动和， 271 – 72 可能性， 278 – 79 利润动机和， 254 – 58 量子计算和，98 风险在， 43 – 45 安全和， 239 – 43 自我批评文化和， 267 – 70 停滞作为方法， 217 – 21 监控作为方法， 206， 215 – 17 技术特征和， 233 – 35 *See also* 监管作为遏制方法 遏制问题， vii *See also* 遏制 《特定常规武器公约》， 263 Copilot，69

公司 英属东印度公司， 186， 189 集中权力和， 187 – 88， 190 – 91 遏制和， 255， 258 非物质化和， 189 – 90 情报和， 186 – 87 公共利益， 258 在即将到来的浪潮中的角色， 134 – 35 税收和， 262 Cortical Labs，91 成本。*See* 技术成本下降 COVID-19 大流行 虚假信息/错误信息和， 172 功能增益（GOF）研究和， 176 组织限制和， 228 监控和， 194， 195， 215 Craspase，82 Crick，Francis，80 CRISPR 基因编辑， 81 – 82， 86， 129 – 30 批评 AI， 244

结构预测的关键评估 (CASP), 89, 90 Cronin, Audrey Kurth, 106, 127 弩, 39, 40 密码学, 98, 244, 248, 266 古巴导弹危机, 44 Cugnot, Nicolas-Joseph, 23 邪教, 212 – 13 文化, 267 – 70 网络攻击, 160 – 63, 166 – 67

源文本：D翻译文本：

戴姆勒, 戈特利布, 24 数据, 扩散, 33 去中心化, 198 – 99 深蓝, 53 深度伪造, 169 – 71, 172 深度学习 自主性与, 113 计算机视觉与, 58 – 60

限制，73 潜力，　61 – 62
蛋白质结构和，　89 – 90 监
督，65 合成生物学和，　90
– 91 *See also* 机器学习

深度思维

AGI 作为目标，　8，　51 武器竞
赛言论和，　124 瓶颈和，　251 遏
制和，　254，　255 – 56 DQN 算法
，　51 – 52，　113 效率和，　68 创立
，　8，　9 谷歌收购，　60，　255 – 5
7 开放性命令和，　128 实际应用和
，61 蛋白质结构和，　89 – 90
See also AlphaGo Deere, John, 93

Delta, 104 需求, 25, 30, 31, 40, 41, 132 – 33 去物质化, 55 – 56, 189 – 90, 233 人口危机, 219 – 20 扩散, 30 – 31 困境, 定义, vii 外交, 167 虚假信息/错误信息, 167, 169 – 73, 208 大疆, 106 DNA 分型, 81 DNA Script, 83 DNA 合成器, 12 – 13 杜德纳, 詹妮弗, 81 DQN（深度Q网络）, 51 – 52, 113 无人机, 103 – 4, 105, 106, 166, 207 双重用途技术。*See* 全能用途技术 反乌托邦, 215 – 17

E

东德，　192 eBay
，　188 – 89

经济不平等，　153 – 54

爱迪生，托马斯，35 自我，　140 – 41 电力，　30 – 31 EleutherAI，　68 – 69 伊丽莎白一世（英格兰女王），　39, 40 Endy，德鲁，84 能源，　99 – 101，　198 引擎，　23 – 25 ENIAC，32 酶合成，　83，　109
Epic of Gilgamesh, 5 Esvelt，凯文，　273 EternalBlue，　161 – 62 种族清洗，　195 苏黎世联邦理工学院，84 欧盟，　125，　229，　260 进化，　26 – 27，　79, 84

解释，　243 爆炸物
，　35，　110 Exsci
entia，　110

F

Facebook，　150，　172，　258

面部识别软件，　194，　207 Fakhrizadeh，Mohsen，　165 封建主义，　184 费曼，理查德，　259 火，　26 – 27 洪水神话，　5 飞梭，　39 福特，亨利，24 福特汽车公司，98 四个特征，定义，　vii 脆弱性放大器。*See* 国家脆弱性放大器 碎片化。*See* 去中心化 富兰克林，罗莎琳，80

狂热阶段，　132
聚变能量，　100

源文本：G翻译文本：

功能获得（GOF）研究，　175 – 77 Gat
o，　111

基因编辑

遏制和，　46，　265 CRISPR技术和，　81 – 82，　86，　129 – 30 伦理问题，86 基因泰克，80

通用技术

作为助燃剂，92 控制和，　233 企业集中和，　190 – 91 定义，26 内燃机作为，　25, 26 不可见性，27 全能使用和，　111

生成式AI， 69, 73

See also 大型语言模型

基因合成， 83 – 84， 247 – 48 转基因生物（GMOs）， 45， 46， 231 基因工程 可及性与， 82 – 83 自主性与， 114 控制与， 46， 265， 269 – 70 基因组测序， 80 – 81， 114 起源于，80
See also 基因编辑； 合成生物学

遗传学, 55 *See also* 合成生物学 成吉思汗，205

基因组测序， 80 – 81

地缘政治 AlphaGo 和， 117 – 19， 120 军备竞赛言论和， 124 – 25， 126 – 27 中国和， 120 – 24 作为即将到来的浪潮激励， 119 – 27

遏制与，　234 欧洲联盟与，　125 印度与，　125 – 26 国际合作与，　263 – 67 后主权世界与，　192 斯普特尼克与，　119 – 20

生殖系编辑，　265

GitHub，　128 – 29

格莱斯顿，威廉，　131

全球挑战，　137 – 40，　219 全球化，　107，　155 – 56 全球变暖。*See* 气候变化
Go，　53 – 54，　113，　117 – 19，　120
谷歌的企业实力，　187 DeepMind收购，　60，　255 – 57 效率与，68 LaMDA与，　71，　72 大型语言模型与，66 量子计算与，　97 – 98，　122

机器人技术和，9
5 在变压器上，64

Google 学术搜索，128

Gopher，68

大猩猩问题，115 – 16
政府

遏制和，258 – 63 组织限制，14
8 – 50 *See also* 民族国家

GPS（全球定位系统），110

GPT-2，64，70 GPT
-3，64，68 GPT-4，
64，113 – 14 GPUs，
130，251

大交易，定义，viii

大不列颠公司和，186，189 监控，
193，195 – 96 大国竞争。*See* 地缘
政治

谷登堡，约翰内斯，30，35

H

H1N1流感，173 – 74 幻觉问题，243 哈佛威斯研究所，95 哈萨比斯，德米斯，8 医疗保健。*See* 医疗应用 亨里克，约瑟夫，28 传统基金会，257 赫什伯格，艾略特，87 真主党，196 – 97 伊达尔戈，塞萨尔，108 分层规划，76 – 77 辛顿，杰弗里，59，60，130 广岛/长崎爆炸，41 – 42 霍布斯，托马斯，216 *Homo technologicus,* 6 Hugging Face，199 人类基因组计划，80 – 81 赫斯基森，威廉，131

Hutchins, Marcus, 161 超进化, 105, 107 – 9 芯片和, 32 – 33, 57, 81, 108 控制和, 250 大型语言模型和, 66, 68

我

印度, 125 – 26, 169 – 70

工业革命遏制尝试, 39, 40, 281 – 83 开放性命令与, 127 利润动机与, 133, 134 技术浪潮与, 28 – 29 惯性约束, 100

Inflection AI, 66, 68, 243, 244

信息去物质化和, 55 – 56 DNA 作为, 79, 87 – 88

电气与电子工程师学会, 241

集成电路，32

情报行动和，75 – 76 公司和，186 – 87 经济价值，136 大猩猩问题，115 – 16 预测和，62
See also 人工智能互联性，28 政府间气候变化专门委员会，138 – 39 内燃机，24 – 25，26，35 – 36 国际原子能机构，241 国际合作，263 – 67 互联网，33，107 – 8，202 iPhone，187 伊朗，165 以色列，165

J

詹姆斯，凯·科尔斯，257 日本，遏制尝试，39, 40 工作，技术影响，177 – 81，261，262 联合欧洲托卡马克，100

K

卡斯帕罗夫，加里，53 凯，约翰，39 柯洁，118 – 19，121 凯南，乔治·F.，37 凯恩斯，约翰·梅纳德，178 汗，A. Q.，45 赫鲁晓夫，尼基塔，126 千兆机器人，95 克劳斯纳，理查德，85 克里热夫斯基，亚历克斯，59 库兹韦尔，雷，57

源文本: L翻译文本:

实验室泄漏，173–75，176 劳动力市场，177–81，261，262，282 LaMDA，71–72，75

兰德，埃里克，265

语言，27，157 *See also* 大型语言模型 LanzaTech，87 大型语言模型（LLMs），62–65 偏见，69–70，239–40 能力，64–65 深度伪造和，170 效率，68 开源和，69 规模，65–66 合成生物学和，91 激光武器，263 执法，97–98 黎巴嫩，196–97 LeCun, Yann，130 李世乭，53–54，117

Legg, Shane, 8

立法, 260

See also 调控作为遏制方法 Lemoine, Blake, 71 – 72

勒努瓦，尚·约瑟夫·艾蒂安，23

李飞飞，59

自由意志主义，201

亚历山大图书馆，41 许可，261

锂生产，109

LLaMA 系统，69

伦敦DNA工厂，83 长寿技术，85 – 86 卢德派，39，40，281 – 83

源文本：M翻译文本：M

机器学习的自主性和，113 偏差在，69 – 70，239 – 40

计算机视觉和，58 – 60 网络攻击和，162 – 63，166 – 67 的局限性，73 医疗应用，110 军事应用和，103 – 5 的潜力，61 – 62 蛋白质结构和，89 – 90 机器人和，95 监督深度学习，65 合成生物学和，90 – 91 *See also* 深度学习 马克龙，埃马纽埃尔，125 马尔萨斯，托马斯，136 曼哈顿计划，41，124，126，141，270 毛主义，192 毛泽东，194 马库斯，加里，73 迈巴赫，威廉，24 麦卡锡，约翰，73 麦考密克，赛勒斯，133

医疗应用，85，95，110

巨型机器，217

旷视，194 Meta，69，128，167 墨子，122 微软，69，98，128，160 – 61 军事应用 AI 和，104，165 不对称和，106 机器学习和，103 – 5 国家脆弱性放大器和，167 – 69 全能技术和，110 – 11 机器人和，165 – 66

明斯基，马文，58，130

错误信息。*See* 虚假信息/错误信息 米切尔，梅兰妮，73

模型 T，24

现代图灵测试，76 – 77，78，115，190，210 Mojica，Francisco，129 – 30 蒙特利尔议定书（1987年），45，263

摩尔，戈登，　32 – 33, 35 摩尔定律，　32 – 33，　67，　81，　108
摩托车，24 芒福德，刘易斯，　29，　217 相互保证毁灭，43

N

纳米技术，　101 狭窄路径，定义，　viii 国家量子信息科学实验室（中国），　122 国家脆弱性放大器，　vii, 11 人工智能与，　166 – 67 威权主义与，　158 – 59 恶意行为者赋权与，　165 – 66，　168 网络攻击与，　160 – 63 民主与，　158，　185 虚假信息与，　169 – 73 脆弱性与，　152 – 54 全球化与，　155 – 56

无意的， 173 – 77 军事应用和， 167 – 69 权力和， 163 – 64 技术性失业和， 177 – 81 国家-国家遏制和， 143， 159 经济不平等和， 153 – 54 平衡， 147 功能， 158 重要性， 151 – 52 监管和， 230 – 31 技术共生， 156 – 58 信任和， 152 – 53 *See also* 地缘政治； 政府； 国家脆弱性放大器 Neuralink，91 神经网络， 59, 64 *See also* 深度学习

1984（奥威尔）， 196

诺贝尔，阿尔弗雷德，35 诺德豪斯，威廉， 30 – 31

朝鲜， 44， 161 NotPetya， 161 – 62， 163 诺伊斯，罗伯特，32 核技术军备竞赛的言辞和， 126 灾难和， 205 遏制， 42 – 45， 264 – 65 发展， 41 – 42 自我和， 141 安全和， 241 英伟达， 130, 250， 251

O

奥丁，82 进攻与防御能力， 157， 234 离网生活， 198 关闭开关， 244 – 45 全能技术， 105， 110 – 12 人工智能与， 111, 130

遏制与，　233 矛盾与，　202 权力与，　182 监管与，　229 – 30 Open AI，　62，　64，　69，　251 开放性命令，　127 – 29 阿片类药物，36 奥本海默，J. Robert，　140 优化问题，98 组织限制，　148 – 50，　228 奥威尔，乔治，　196 奥托，尼古拉斯·奥古斯特，23 奥斯曼帝国，　38，40

P

巴基斯坦，45 PaLM，　66，68 大流行，　205，　209，　243，　273 – 74 *See also* COVID-19 大流行

潘建伟， 122 – 23 巴黎协定， 45， 46， 263 部分禁止核试验条约（1963年）， 42 – 43 人工智能合作伙伴关系， 246 专利制度， 127 病原体控制和， 273 – 74 功能获得（GOF）研究， 175 – 77 实验室泄漏， 173 – 75， 176

PayPal， 188 – 89

同行评审， 128 佩洛西，南希， 170 佩雷斯，卡洛塔， 29， 132 悲观厌恶， viii， 13 – 14， 102， 236， 253 石化工业，87 幻影相机四轴飞行器， 106 网络钓鱼， 171 留声机，35

物理自我修改， 86， 200

派， 243

查士丁尼瘟疫，　205

柏拉图，　5

PlayStation 2，　110 – 11

脊髓灰质炎疫苗，　263 政治极化，　155 民众运动，　271 – 72 人口规模危机，　219 – 20 技术浪潮与，　27 – 28 民粹主义，　153 后主权世界 威权主义与，　185，　191 – 92 矛盾与，　202 – 4 公司与，　186 – 89 去中心化与，　198 – 99，　200 – 202 去物质化与，　189 民主国家与，　185 种族清洗与，　195 混合实体与，　196 – 97 权力与，　184 – 85

监视和，　193 – 96，　206

功率，　102

矛盾和，　202 国家脆弱性放大器和，　163 – 64 全能技术和，　182 后主权世界和，　184 – 85

动力织机，　282 – 83

印刷，　30，　35，　38 – 39，　40，　157

利润动机，　131 – 36 控制和，　254 – 58 扩散，　30 – 31，　32 – 34 人工智能和，　68 – 69 不可避免性，　40 – 41，47

蛋白质折叠，　88 – 90

普洛透斯，　94

失明激光武器议定书（1995年），　263

公益公司，　258 普京，弗拉基米尔，　125

Q

量子计算，　97 – 99，　109，　114，　122

源文本：R翻译文本：

研发支出，　129，　134，　259

种族主义，　69，　239 – 40 收音机，　157
铁路，　23，　131 – 32 勒索软件，　160 – 62
里根，罗纳德，　201 叛乱防御，　166 红队
测试，　246 宗教改革，35 再生技术，85 作
为遏制方法的监管，　225 – 26 的挑战，　22
6 – 27，　229 – 30 立法与，　260 许可与，
261 国家与，　230 – 31

必要性，277

自我批评文化和，269

强化学习，95，117，166–67，240
See also 机器学习

Remotec Andros Mark 5A-1 机器人，97–98 文艺复兴，201 可再生能源，100–101 Restrepo, Pascual，179 复仇效应，36，176，177 Ring，227

RNA编辑，82

罗宾逊，詹姆斯，276，278

机器人学，93–97

中国的发展，122 军事应用，165–66 利润动机和，134–35 合成生物学和，109

Rogers, Everett，56–57
Rotblat, Joseph，270 Russell, Stuart，115，244

俄罗斯流感疫情，　173 – 74 俄罗斯入侵乌克兰，　44，　103 – 4 卢瑟福，欧内斯特，41

源文本: S翻译文本:

三星集团，　188

赛诺菲，　110
SARS，　174
– 75
缩放假设，　67 – 68, 75

施奈尔，布鲁斯，　167 熊彼特，约瑟夫，29 科学革命，　35，　127 SecureDNA，　247，　265 自动驾驶汽车，　113 半导体，　32，　84，　249 – 50 商汤科技，　194 Shield AI，　166 丝绸，41 辛格，艾萨克，　133

奇点，74

See also 超级智能手机，33，60，112，187 Smil, Vaclav，138

斯诺登，爱德华，122

社交媒体矛盾和，202 虚假信息和，172 国家和，155，156 开放性要求和，128 组织效率和，150

太阳能，100，198

Solugen，86 韩国，188 苏联，171 – 72，192 太空垃圾，36 太空旅行，122 SpaceX，104 Sparrow，95 速度。*See* 超进化

人造卫星，119 – 20，126 稳定性AI，199 稳定扩散，69 停滞，217 – 21 星链，104 创业神话，141 蒸汽机，23，131 斯蒂芬森，乔治，131 马镫，183 – 84 石器工具，26 超级智能，74 – 75，77，115 监控审计和，248 反乌托邦和，215 – 17 后主权世界和，193 – 96，206 监管和，228 抵制，277 – 78 苏茨克维尔，伊利亚，59 群体机器人，95 – 96 Switch Transformer，68

源文本: 翻译文本:
Sycamore, 122

合成生物学

人工智能和, 89 – 90, 109 审计和, 247 – 48 灾难情景和, 208 – 9 计算机和, 87 – 88 当前应用, 84 – 85 去中心化和, 200 深度学习和, 90 – 91 定义, viii 去物质化和, 190 发展, 55 国际合作和, 265 – 66 全能技术和, 112 潜力, 85 – 87 力量, 56 利润动机和, 134 研究不可预测性和, 129 – 30
See also 即将到来的浪潮； 基因编辑

源文本: 翻译文本:
合成媒体, 169 – 71, 172 – 73

Synthia, 84

系统生物学，85

西拉德，利奥，41

源文本: T翻译文本:

税收， 261 – 62

技术性失业， 177 – 81， 261， 262， 282

技术的定义， viii，26 失败，18 相互关联的性质， 56 – 57 社会对其的依赖， 218 – 19 与民族国家的共生关系， 156 – 58 普遍性， 235 – 36 意外后果， 35 – 36 *See also* 即将到来的技术浪潮特征；*specific technologies*

技术成本下降

AI 和， 64， 68 计算， 108

遏制与，　39，　43，　233 – 34 基因工程，　80，　81，　83，84 基因组测序，81 权力与，　102 扩散与，31
机器人，　95，96 技术浪潮，　6，　16，　25 – 26 加速，　28 – 29，92 扩散与，　30 – 31 演变与，　26 – 27 必然性，　29，47 隐形性，　27，73 材料与，　54 – 55 扩散与，　32 – 34 顺从，47 不可预测性，29 城市化与，　27 – 28 电话，31

电视，　157 恐怖主义，　44 – 45，　160 – 62，　207，　212 – 13

蒂尔，彼得，201 3D打印（增材制造），96，109，166，190 蒂利，查尔斯，157 蒂瓦里，马诺杰，169–70 托夫勒，阿尔文，29 代币，63 极权主义。*See* 威权主义；监控 交通优化，98 转录器，88 变压器，64，90–91 晶体管，32–33，67 《不扩散核武器条约》（1968年），43，263 沙皇炸弹，42 清华大学，121

台积电，251

图灵，艾伦，35, 75

23andMe，81
2001: A Space Odyssey, 110

源文本：U翻译文本：

维吾尔族种族清洗，　195

乌克兰，　44，　103 – 4，　161 – 62 联合炸弹客，　213 美国出口管制，　249 – 50 国际合作与，　265 – 66 监控，　195 普遍基本收入（UBI），　262 牛津大学，　101 乌尔班二世（教皇），39 城市化，技术浪潮与，　27 – 28 美国国家安全局（NSA），　161

V

车辆

不对称影响和，　107 自主性和，　113 内燃机和，　24 – 25

监管，　229 – 30

文特尔，克雷格，84 警惕解决方案，　196 病毒，　173 – 75，　209，　273 – 74 冯·诺依曼，约翰，　141，　221，　222

源文本：W翻译文本：

沃尔玛，　95 – 96 WannaCry，　160 – 61，　163，　166 Watson，James，80 Watson，Thomas J.，32 WaveNet，61 波浪，viii，　5 – 6，16 *See also* 技术波浪

We（扎米亚京），　196

武器

AI 和，　165 遏制和，　39，　40，　263 民族国家和，　157

马镫和，183–84

See also 军事应用；核技术

轮子，28

白宫科学技术政策办公室，260

威尔金斯，莫里斯，80

获胜者，兰登，156

第一次世界大战，110，205 第二次世界大战，32，42，126，205，264 写作，27，28，156–57

X

习近平，121，122，123–24，249

源文本：Z翻译文本：

扎米亚京，叶夫根尼，196

张，冯，265

关于作者

穆斯塔法·苏莱曼是Inflection AI的联合创始人兼首席执行官。此前，他共同创立了全球领先的AI公司之一DeepMind。在DeepMind工作十年后，苏莱曼成为谷歌的AI产品管理和AI政策副总裁。在牛津大学读本科时，苏莱曼辍学去帮助创办一个非营利电话咨询服务。他现居住在加利福尼亚州帕洛阿尔托。

M B 是一位居住在英国的作家和出版商。他是 *The Content Machine, Curation,* 和 *Human Frontiers* 的作者。

Penguin
Random
House

What's next on your reading list?

发现您的下一本好书！

源文本：翻译文本：

获取个性化的书籍推荐和关于这位作者的最新消息。

Sign up now.