



智能简史

雨果·德·加里

内容简介

不久的将来，人工智能机器的智能将是人类的万亿个万亿倍，它们面对我们，并不像我们面对狗，而是如同我们面对蚊子、跳蚤甚至岩石，当它们消灭我们的时候，如同我们将蚊子拍死，将臭虫冲进下水道，谁会在消灭跳蚤的时候觉得这样太残忍了呢？！这也许是人类逃脱不了的宿命，这也使德·加里斯教授认为有责任写这本书。我知道如果不让地球上每个人都对这个关乎人类存亡的问题知情并且思考，那么书中所描写的比所有科幻电影更刺激、可怕的事情很有可能发生，。

声明

本书内容系书仓用户从网络中收集，利用书仓在线制作、转换服务生成，存放于书仓，

仅供个人收藏、学习使用。

其制作、收藏者承诺不把此内容用于商业用途。

也请借阅者勿公开传播以及应用于商业用途。

书仓网仅止为个人或者机构提供在线内容制作、转换、存储等服务。

书仓网上所有非 epubsys 帐号下的内容均为个人使用行为。



书仓（shucang.com），旨在为读者弥合数字时代电子化阅读中的各种障碍，从而让个人获得随时、随地、随意阅读与拥有数字内容的惬意体验。

目前，书仓网提供的功能有：

1、在线的电子书报制作工具；2、在线电子书转换；3、在线电子书、文章存储分享；4、在线阅读；5、手机在线阅读；6、支持各种不同阅读软件、阅读器的电子书下载；

可以处理的格式包括 epub、mobi、chm、txt、doc、pdf 等。

在线版：<http://www.shucang.com>

手机版：<http://shucang.com/m>

目录版：<http://www.shucang.com/s/index.php>

注：比如 iphone/itouch 的 stanza，已经把书仓放在默认书库中，如果没有看到，可以手工添加目录版地址。

正文

书名：谁会代替人类：智能简史

作者：雨果 德 加里斯

作者简介

“完全的怪人”和“超越时代的天才”：

“人工大脑之父”——德 加里斯教授

世界人工大脑之父雨果 德 加里斯博士，美国犹他州立大学计算机系教授，他离开美国来到中国，在中国的大学从事数学、物理和计算机科学三方

导读

这本书很可能让你以为又是一个科学幻想，可能你会说，同类的东西已经看得太多了。

因此，你现在也许会错过这本书，但是，将来你一定会拼命去找这本书，而那时就太晚了。

我的同事告诉我，她看过一个关于超级智能机器的影片。一开始，它们随意屠杀人类。我问她，影片后来结局是什么？她说，机器人被人类感动了，成了人类的朋友。于是我再问她，人类会被蚊子感动吗？会和岩石成为朋友吗？想象远远不及真实！于是，她陷入了沉思.....

一般的科幻片中描述的人工智能机器，它们虽然在很多方面强于人类，但是，总体和人类的差距不大，大致也就是人类和狗的差距。那么，它们最后被人类感动，会和人类成为朋友是很合理的。就像人类会被狗感动，和狗成为朋友一样。但是，那也许只是 20 或者 40 年后的智能机器的情况。

不远的将来，人工智能机器的智能将是人类的万亿个万亿倍，它们面对我们，并不像我们面对狗，而是如同我们面对蚊子、跳蚤甚至岩石，当它们消灭我们的时候，如同我们将蚊子拍死，将臭虫冲进下水道，谁会在消灭跳蚤的时候觉得这样太残忍了呢？！

这也许是人类逃脱不了的宿命，也是德·加里斯教授认为有责任写这本书，我们认为有责任出版这本科学伦理图书的原因。

因为，我知道如果不让地球上每个人都对这个关乎人类存亡的问题知情并且思考，那么书中所描写的比所有科幻电影更刺激、可怕的事情很有可能发生，甚至更加糟糕。但是，我们相信，本书的出版以及由此引发的人类广泛的关注和思考，一定可以避免书中所描写的情节的发生！

希望每个人都来关注，希望每个人都来思考，希望这一切都还来得及……

推荐序

推荐序

我和雨果·德·加里斯相识已久。近几年，我们常同时被邀请在国际未来思考者的聚会上演讲。我们都认为，21 世纪晚期，人类将不得不面临被一个新物种，即由超级智慧人工大脑控制的超级智能机器人所取代。德·加里斯和我不同的是，(使用他的术语)他基本上是一个“宇宙主义者”(“cosmist”，即那些认为应该制造“神一样的超级智能机器”而不考虑人类未来生存是否有危险的一类人)。我认为，德·加里斯会把我归类于“地球主义者”(“terran”，即那些反对宇宙主义者去制造这样超级智能机器的一类人)。

我记得，在一次国际会议上，他问我是宇宙主义者还是地球主义者。我说我可能更倾向于地球主义者。德·加里斯半开玩笑地说：“战争就此开始了！”他是指此书中提到的“人工智能战争”。他认为，在本世纪(即 21 世纪)晚期，将在宇宙主义者和地球主义者之间爆发一场关于人类是否应该制造那些神一样的“人工智能机器”的“大战”。

事实上，我认为自己更倾向于一个“半机器主义者”(“cyborgian”，是指那些期待用技术来改良自身成为“半机器人”的人，也就是说一半是机器，一半是人类)而不是“地球主义者”。我希望，通过改良自身，人类可以转化成半人工智能机器，获得德·加里斯所说的神一样的力量，从而不必承担“大战”的威胁，且还能分享一块蛋糕。在某种意义上，我在寻求一种折中方法——不是让超级智能人工大脑反对人类，而是我们加入它们。

几年前，德·加里斯和我在瑞士的一个会议上做客座演讲。为了更加生动，德·加里斯让会议组织者给他一把玩具牛仔手枪，当我们做完演讲后，他半开玩笑地用这把枪向我“射击”(照片见本书文前彩插)。它的标题是“人工智能战争的第一次枪战”。我希望现实不会像这个游戏那样。德·加里斯对于 21 世纪晚期那场战争的场景预测，就是 21 世纪武器的使用造成了几十亿人的死亡，这是非常令人沮丧的。为了人类(和半机器主义者)的生存，我很坚定地希望他是错误的，绝对错误的。

读完本书后，读者可能对我过于乐观的愿景能否实现深感怀疑。德·加里斯的推理令人恐惧，但不得不信服，即使我内心对他所说的很排斥。他在“人工大脑”这个新领

域里是先驱者的事实，只会让他的见解更加可信。如果说世界上有哪个人在预测高级人工大脑对于人类影响方面取得卓越成就的话，那么这个人就是德·加里斯。

我相信本书将有深远的影响。几十年后，如果德·加里斯关于“物种主导问题将在本世纪主导全球政治”的言论被证实是正确的话，那么他将是 21 世纪最主要的思想家之一。对于德·加里斯我没有任何冒犯，我希望他被证实是错误的、夸张的和反应过度的。因为如果不是这样的话，他的预言就会在我们的儿孙辈成为事实，他所说的“大规模死亡”对于人类是一件太可怕的事情。

不管你在宇宙主义者—地球主义者的观点范围中处于何种位置，这都是一本你不能也不应该忽视的书。它太重要了，太轰动了，不应该被草率地忽视。我建议你读它，然后再读一遍，试着找出逻辑和辩论上的漏洞，以便我们可以期待一个和平繁荣的 21 世纪。如果德·加里斯的见解是正确的话，那个像电影《终结者》中描述的地狱般的梦魇将会成为事实。

凯文·沃里克博士

英格兰里丁大学控制论教授

《我，半机器人》、《机器的思想》、《机器的前进》的作者

第一部分 导言 智能简史 导言(1)

第一部分 导言 智能简史 导言(1)

我是雨果·德·加里斯教授，一个设计和制造“人工大脑”的研究机构的负责人。研究这个领域是我的专长。但我不仅是一个研究者和科学家，也是一个有着政治和伦理意识的社会评论家。我十分担心我的研究结果会在 21 世纪下半叶给地球上的人类带来我所恐惧的后果。

你可能会问：“既然你对现在进行的工作可能会给人类带来的负面影响感到如此焦虑，为什么不干脆停下来去做点别的呢？”事实上，我认为我正建造着一个在未来几十年内，将会成为神一样无所不能的东西(虽然在我有生之年可能看不到它的完成)。制造神一般的生物的期望，带给我一种宗教般敬畏的感觉，这感觉进入我的灵魂深处，强烈地激发我继续这项工作，尽管那些可怕的负面效应可能会接踵而来。

对此，我感觉似乎要精神分裂。一方面，我的确非常想制造这些人工大脑，并尽可能将它们制造得极其聪明。我把这当作人类追求的一个伟大目标，我将会在后面的章节里，对这个问题进行深入的讨论。另一方面，我又对人工大脑制造成功将产生的影响感到不寒而栗。因为这个成功同时也意味着人工大脑最终会比我们生物大脑智慧很多倍。这一点，我也同样会在本书后面的章节里进行深入的讨论。

具体说，作为一个职业的人工大脑制造研究者和曾经的理论物理学家，我认为，对于 21 世纪科技所产生的“高智机器”的潜在能力，我比大多数人看得更清楚。所谓“高智”，是指比人类大脑智慧不止 2 倍或者 10 倍，而是万亿个万亿倍的人工大脑，也就是真正神一样的东西。如此巨大的数字对你来说，可能更像是科幻小说而不是任何可能的未来科学。在本书的第 3 章我将解释一些 21 世纪科技的一些基本原理，我相信这些科技将允许人类(如果人类选择这样做的话)来制造神一样的机器。我会让你相信，这不是科幻小说，因为它有着充分的理由迫使你相信这些令人毛骨悚然的巨大数字。我将尽可能地用一种简单清晰的方式来陈述，以便人们不需要成为一个“火箭科学家(rocketscientist)”(美国人的说法，意思是“一个非常聪明的人”)就能够理解它们。这些基本概念可以被任何人所理解，只要他们愿意对这个研究花上一点精力。

第 3 章将介绍一些难以置信的科技，它们使制造神一样无所不能的高度智能机器成为可能。你将接触到许多伦理、哲学和政治问题。因为人类制造这些高智能机器的前景会引发一些非常重要的问题。本书将从各个角度讨论这些问题，我不会装作知晓所有的答案，但我会尽我所能。

在现代历史上，“摩尔定律”一直是一个伟大的科技经济趋势，它描述并解释集成电路或者芯片的运算能力(例如，电子元件密度、电子信号处理速度，等等)每年都成两倍增加。这个趋势自从英特尔微处理器制造公司创始人之一戈登·摩尔 1965 年提出以后，经验证一直是正确的。如果你把一个数一直乘以 2 很多次，你会得出一个非常巨大的数字。举个例子，2 乘以 2 乘以 2 乘以 2.....(10 次)等于 1024。如果你这样乘 20 次会得到 1048576，也就是大概 100 万。如果你这样乘 30 次，你会得到 10 亿，乘 40 次会得到万亿，等等。摩尔定律在过去的 40 年一直是有效的，以至于倍数增加得到的数字非常巨大。我称之为“巨大的摩尔倍增”。

摩尔定律是芯片上电路尺寸不断缩小的结果，以至于电子(一种基本粒子，它在电路中的流动形成电流)在电子元件(例如两个晶体管)间传递的距离被缩短了。根据爱因斯坦的理论，物体可以移动的最快速度是光速(大概 300000 千米/秒)，并且这也是电流必须遵守的自然常量。如果两个电子元件间的距离被缩短了，它们之间电子信号的传递距离也将缩短，因此，通过这个距离所需要的时间就会更少(以恒定的光速)。

芯片制作公司在过去几十年里投入了巨大的精力来使电路变得更小，使之分布得更加密集，以便让它们运行得更快。微处理芯片运行越快，就会越经济。如果你是加州“硅谷”中一家制造芯片公司的 ceo，而你的竞争对手比你早 6 个月在市场上推出一款比你(的产品)快 30% 的芯片，那么你的公司可能会破产，而竞争对手的市场份额会大幅增长，因为每个人都想拥有一台更快的计算机。因此，在过去的几十年，电子线路变得更小更快。

第一部分 导言 智能简史 导言(2)

第一部分 导言 智能简史 导言(2)

摩尔定律还能够保持多长时间的有效性呢？如果一直有效到 2020 年，海量内存芯片上电路尺寸可能允许在一个原子上面存储一个比特(一个“比特”是一个“二进制数字”，一个“0”或者一个“1”，是计算机进行运算所代表的数字和符号)的信息。那么，一个这样等量级的物体，比如说苹果，包含多少个原子(即苹果这么大的存储器可以存储多少比特)呢？答案是令人吃惊的，有万亿万亿个原子(比特)，就是说，1 后面跟 24 个零。

你现在是不是慢慢开始了解，为什么我如此相信 21 世纪晚期高度智能机器会比人聪明万亿个万亿倍。

21 世纪科技不仅将会在一个原子上存储一个比特的信息，今后还会采用一种新的名为“量子计算”的计算方式。这种新的计算方式和人类 20 世纪所通用的“传统计算”完全不同。第 3 章将试着大致介绍一下量子计算原理，因为在不远的将来，它极有可能成为计算机的基础。

在这里先跟大家简单介绍量子计算的一些基本特征。如果用一串 n 比特(在计算机科学里称作一个“寄存器”，例如，001011101111010)进行某种计算操作(先不管目前进行何种的操作)，用“传统计算”会需要一定量的时间。然而在同样的时间里，采用“量子计算”技术，可以进行 $2n$ 这样的运算。当 n 变得很大时， $2n$ 变得令人惊讶的巨大。因此，量子计算的潜能是极大优越于传统计算的。既然摩尔定律很可能带我们进入原子等级，在这个范围里，“量子力学”这个物理规则将会被应用，人类将必须用量子理论进行量子计算。因此，在过去一些年里，理论研究和试验上人们都投入了巨大的努力，以理解和制造“量子计算机”。

在量子计算机推向大众之前，量子计算仍然有很多概念和运用上的困难有待解决。但是面对这些困难，每个月都有进步，所以我相信，拥有实用的量子计算机只是时间上的问题。

如果把每原子存储 1 比特的记忆存储能力和量子计算方式结合起来，其组合能力是“爆炸性”的。21 世纪计算机可能拥有的潜在计算能力会比现在计算机的能力高万亿万亿个万亿倍。

说到这里，希望我所讲的你们都能理解。

论述到现在，你们的理解可能比我所讲的更超前，并且会反对我，因为我似乎在暗示，大规模存储能力和惊人的运算能力就足以产生高智能机器，其他的什么都不需要。我也因为这一点而被我的某些同事指责，所以请先让我陈述一下对于这个问题的一些观点。

有些人(例如，因黑洞理论闻名的罗杰·彭罗斯爵士以及他的主要竞争对手英国宇宙学家斯蒂芬·霍金)认为要产生智能机器需要的远不止是大规模的计算能力。彭罗斯声称意识也是必要的，并且认为需要新的物理知识来理解自然中意识如何产生，并借此来创造机器人工意识。

我并不介意这些反对的声音，又或许这样的批评是对的。但即使这样，他们的反对也无法改变我的基本论题，因为这也许只是延迟几十年，直到意识的本质被更好地理解。我认为，要理解自然是怎样创造我们的，只是一个时间问题。也就是说，我预计科学会逐渐理解“胚胎发育”的过程，这是从一个受精卵细胞到包含万亿个细胞的胚胎然后形成一个婴儿的过程。

我们自己，既有智慧又有意识，就是一个存在的证据，即证明自然界可以以一种适合的方式排列大分子来制造我们。当一个孕妇进食后，她所摄取食物里面的一些大分子会被重新组合，然后自己去组合包含万亿个原子的大分子结构，就是她的婴儿。婴儿就是由于大分子的自我排列组合而形成的一个有效的三维的并且有意识和智慧的生物。

自然界通过进化已经找到了一种解决这一问题的方法，所以这个问题是可以解决的。如果科学家想制造有智慧有意识的机器，一个显而易见的办法就是通过尽可能地复制自然界创造意识的途径来实现。或迟或早，科学会产生一种和人类一样运作的人工生命形式。

第一部分 导言 智能简史 导言(3)

第一部分 导言 智能简史 导言(3)

同样，如果科学对于生物大脑的工作原理有更深入的了解，那么制造人工大脑将变得更加容易。尽管过去一个世纪甚至在更长时间内，神经科学家们投入了大量的精力来试图理解人类大脑工作的基本原理。但不幸的是，当代神经科学对于人类大脑工作原理的了解仍然十分有限。迄今为止，在神经元微电路层次上，对于神经电路怎样高度联系以实现大脑工作方面的了解还是很少。科学家们依然缺少用于仔细研究大脑结构的工具。

然而，随着科技进步人们可以制造越来越小的设备——从微米级达到纳米级。比如说，当设备从百万分之一米(细菌的大小)变得只有十亿分之一米(分子的大小)的时候，制造分子级机器人来探索大脑如何工作将成为可能。

关于生物大脑工作原理的科学知识到现在之所以还非常有限，正是由于我们现在使用的工具非常有限。但是，随着分子级工具(称作纳米科技)的出现，神经学家将会有一套强有力的崭新工具用来探索生物大脑，该领域的研究也会取得很快的进展。

像我这样的人工大脑制作者将会利用这些新创建的神经科学原理，迅速地把它们应用到人工大脑结构上。

可能不久人们便会破解很多生物大脑的工作原理，从而将会产生一种能够在神经元电路基础上解释的——如为什么爱因斯坦的大脑会比绝大多数人聪明很多的“智能理论”。这样的智能理论一旦存在，对于像我这样的神经工程师来说，就可以用一种更工程化的方法来制造大脑。我们将不必去做“神经科学的奴隶”。我们可以使用一种替代方式来制造人工智能机器(虽然一开始是基于神经科学原理的)。

当拥有了纳米科技工具提供的崭新的神经科学知识，量子计算以及一个原子存储一个比特所允许的惊人的计算能力，像我这样的人工大脑制造者才可能具备制造真正有意识的人工智能机器所需的所有要素。

现在有很多问题被提出，我将在本书大部分章节里试着解答它们。让我们把目光投向未来，想象一下上述的科技将怎样影响我们普通人的生活。

很快，购买人工大脑控制机器人来做家务将变为现实。如果这些机器人的价格变得可以承受，那么对于它们的需求将变得非常巨大。我相信，不久，世界经济将依靠这些基于人工大脑的计算机。这样的设备将变得非常有用和流行，以至于地球上的每一个人都想拥有。随着科技和经济的进步，这种设备在全球市场将会快速增长以至于地球上的大多数政治家都会支持它。不仅商务公司会热衷于制造更聪明更有用的机器人和基于人工大脑的各种设备，就连各国军事领域也将广泛采用这种技术。

在未来几十年，地球将真正形成一个全球性国家，但有全球性的警察维护全球性的法律这种可能性是很小的。相反，我认为在下半个世纪为成为世界上最强大国家而产生的政治对立将白热化。这样就会使处于敌对状态的国家不允许对方研发出更聪明的机器士兵和其他基于人工大脑的国防系统。因此，国家政府将会加强为军方服务的人工大脑研究的发展，这将不限于商务应用，就像一个世纪以来其他技术的发展模式一样。

所以，人工大脑的机器人和相应领域的崛起，看起来是不可阻挡的。要阻挡商业及军方的推动力是很难想象的，除非某种大规模的政治运动才可以阻止它。

如何才能引发这样的运动？

试想几十年以后，几百万人都买了家务机器人、性机器人、教学机器人、育儿机器人、伴侣机器人、友谊机器人，等等，并且这些人工大脑控制机器人可以很好地理解人类语言并与人类交谈。一些年之后会发生什么呢？毫无疑问，早年那些型号的机器人会因为老式的而变得没有吸引力。新型号机器人因为有更高质量的语言能力而被认为是更“智能的”。它们会更好理解并与人类沟通。它们的行动指令会更丰富。一句话，它们会很快替代早期型号的机器人。

第一部分导言智能简史导言(4)

第一部分 导言 智能简史 导言(4)

那么大家会怎样呢？当然，他们会丢弃老式机器人去买新型的，或者把他们的老式机器人进行人工神经元电路升级。之后，这样的过程将反复出现，就像 20 世纪 80 年代至 90 年代人们购买个人电脑那样。

然而，有些爱思考的购买者会发现一代一代的家用机器和机器人变得越来越聪明，并且机器人和人类智商的差异越来越小。当机器人真的变得非常聪明后，几百万的机器人拥有者会开始问自己一些有趣的问题。诸如：

“这些人工大脑机器人会变得多聪明呢？”

“它们会变得和人类一样聪明吗？”

“如果可能，那是一件好事吗？”

“这些机器人会变得足够聪明而成为人类的威胁吗？”

“这些机器人会比人类聪明吗？”

“如果能够，会聪明多少呢？”

“人类应不应该允许这些机器人比人类聪明呢？”

“当它们变得比人类聪明时，它们会认为人类是地球上的瘟疫或癌症而把人类灭绝吗？”

“人类应该为这些可能发生的事情冒险吗？”

“机器人人工智商需要设置一个限制吗，使机器人足够聪明从而有利于人类但不至于成为人类的威胁？”

“政治上、军事上、经济上有阻碍机器人逐年变得更聪明的可能性吗？”

“有很多人认为，制造超智能机器是人类的使命。这些人不喜欢对人工智商设置任何限制。这会在人类间产生冲突吗？”

随着人工智能机器的出现以及制造更强能力的人工大脑，会出现更多诸如此类的问题。

我又是怎样看待人类应对人工智能机器的出现所面临的挑战呢？我将会陈述我认为最有可能发生的状况。在这之前，先给大家介绍一个本书中经常用到的新术语，它是表达“神一样超级智能机器”的缩写。这个新术语是“artilect”(人工智能机器)，它是“artificialintellect”的缩写，是本书中最重要的概念和术语。

我相信，人工智能机器在 21 世纪将成为主导问题——具体指人类是否应该制造比人类智能高出万亿个万亿倍的人工智能机器，也就是说，拥有神一样智慧的机器。我预言人类将分裂成两大主要政治集团，并且会随着人工智能机器逐渐变为现实而使得他们之间的斗争更加激烈。

支持制造人工智能机器的集团，我称之为“宇宙主义者”(cosmist)，是基于“cosmos”(宇宙)一词。对于宇宙主义者，制造人工智能机器就像宗教中人类的使命；是真正伟大和值得崇拜的；是值得人们用尽一生的精力来实现的。对于宇宙主义者，不去制造人工智能机器，不去创造更高的进化形式，而停留在人类这样微不足道的形式上，是“宇宙大悲剧”。宇宙主义者会强烈反对任何试图阻碍 21 世纪人工智能机器产生的行为。

强烈反对制造人工智能机器的集团，我称之为“地球主义者”(terran)，是基于“terra”(地球)一词，反映了他们内在的非宇宙主义的看法。我坚信，地球主义者会争论说，允许制造人工智能机器(以非常高级的形式)意味着接受一个冒险——那就是某一天，人工智能机器会认为，无论从什么样的角度来说，人类都是一种瘟疫或有害物。而人工智能机器的智能已远远超过人类，那么对人工智能机器来说，如果它们决定把人类灭绝将是很容易的。

但是你可能会辩论说，如果人工智能机器真的非常聪明，那么它们会认为人类赋予了它们生命，人类就是它们的父母。因此，人工智能机器会尊重我们，并且对我们

很友好。这样的情况可能会发生，但问题是，你无法保证人工智能机器们会像我们期望的那样对待我们。

不要忘记，人工智能机器可能比我们聪明万亿个万亿倍，由于它们是如此的聪明以至于人类对于它们来说是低等甚至是不值一提的，以至于人类存在与否对于它们来说没有任何不同。

毫不夸张地说，人工智能机器们试图和人类交流，就像人类想要和石头说话。

第一部分 导言 智能简史 导言(5)

第一部分 导言 智能简史 导言(5)

换一个比方，考虑一下你对一只停在小臂皮肤上的蚊子的感受。当你打死它的时候，你会停下来想，你刚杀死的小生命是纳米科技工程的杰作，是 20 世纪科学家根本没有办法制造的吗？蚊子拥有几十亿个细胞，每一个细胞都可以看作一种大分子城市，细胞里的每一个分子就像城市里的人一样。大分子相对于细胞的大小就像人相对于城市一样。

尽管蚊子是经过几十亿年进化的极其复杂和不可思议的生物，我们人类根本就不会考虑这些而轻易杀死它们，因为我们从来都只是把它们看成是一种有害生物。当我们漫步在森林里踩死一只蚂蚁，或者把蜘蛛冲到下水道里的时候，我们都会有同样的态度。

谁又会说那些人工智能机器们不会对我们有同样的看法而消灭我们呢？拥有超级的“人造”智慧，做这些事情对它们来说是小菜一碟。

对于地球主义者，这场辩论中最重要的词汇是“风险”。地球主义者争论说，人类从来就不该冒任何风险，比如像制造高级形式的人工智能机器可能会导致人类灭亡这样的风险。唯一让风险降为零的保险办法，就是永远不要制造人工智能机器。

关键时候，如果地球主义者发现宇宙主义者真的要制造高级人工智能机器时，地球主义者可能会为了保证人类的生存而去消灭宇宙主义者。他们会认为，为了保护整个人类，而杀死几百万宇宙主义者是可以接受的。

地球主义者有理由认为这样的牺牲是合理的。举一个历史上类似的例子——在二次世界大战的末期，斯大林的军队为了占领柏林和摧毁杀害了 2000 万俄国人的希特勒纳粹政权，为了夺取纳粹占领的每一个欧洲城市，付出 100000 俄国士兵伤亡的代价。对于斯大林来说，这样的牺牲是合理的，因为可以得到非常大的好处，那就是让俄国人从纳粹大屠杀的恐惧中解救出来。

你可能会问：“真的会有人头脑清醒地选择做一个真正的宇宙主义者，冒着整个人类可能被灭绝的风险吗？”

相信在未来，几百万人将会对这个问题给以肯定答案。我觉得，当更多的人意识到人工智能机器的发展前景，他们最终会支持人工智能机器的制造。本书将会用一个章节，以宇宙主义者的视角来论述为什么要支持制造人工智能机器。

这些人，即宇宙主义者，将会给这些生命(一个人工智能机器等价于一万亿万亿个人类)的制造以更高的优先级别，因为它们是神一样的、永恒的、无所不能的，反而不会去理会人类可能被人工智能机器灭绝的风险。

让我重新解释一下，以便没有人再对宇宙主义者的理念有疑问。宇宙主义者，定义为支持制造人工智能机器的人。人工智能机器如果被制造出来，它们迟早会发现人类是如此的低等，像一个有害物，从而决定来灭绝我们，不论以什么样的理由。因此，宇宙主义者已经准备接受人类被灭绝的风险。如果人类被灭绝，那意味着你我的后代也被消灭。那将是人类历史上最大的灾难，因为再也没有历史了，再也没有人类了。人类也将成为地球上曾经存在但已灭绝的那 99% 的物种之一。

因此，对于地球主义者而言，宇宙主义者是魔鬼附身，比希特勒、日本帝国主义或者其他任何杀害亿万人类的政权都坏得多，因为受害的范围将大得多。现在我们谈论的不是几百万人的屠杀，而是整个人类，即几十亿人的灭绝。

但是对于宇宙主义者来说，在这个无关紧要的小行星上谈论人类生存与否是微不足道的。因为这个行星围绕的恒星也只是我们这个星系上约 2000 亿个恒星中微不足道的一颗，已知宇宙中同样的星系也有这么多(也是以 10 亿计算的)，并且在这个多元宇宙中可能还有很多其他宇宙(根据一些最近的宇宙理论)。我称他们为“宇宙主义者”也正是基于他们的看法是宇宙范围的。他们更注重“大局”，意思就是说，在一个无关紧要的小行星上，灭绝超原始的、非人工智能的物种(也就是人类)，对于人工智能机器的“上帝”来说是不重要的。

第一部分导言智能简史导言(6)

第一部分 导言 智能简史 导言(6)

本书将会用两章来陈述地球主义者和宇宙主义者各自的观点。他们双方的论证都很有说服力，意味着他们之间不可避免的冲突将会随着未来几十年人工智能辩论的白热化而更加敌对。

我对未来潜在冲突的激烈程度感到担心，对这两种观点(地球主义者/宇宙主义者)支持的人数将不相上下。举个例子，当我对关于宇宙主义者/地球主义者/人工智能机器的情景进行讲座的时候，我经常请听众对他们自己是地球主义者或者宇宙主义者进行投票。我发现投票的结果不是我以前所预计的(10%是宇宙主义者，90%是地球主义者)，而竟然是 50：50，60：40，40：60。人们对这个问题的看法分成了对立的两派。

让我更沮丧的是，人工智能问题在 21 世纪将会非常热门。基本上可以肯定地说，在这个问题上的观点分歧会在 21 世纪后半段在地球主义者和宇宙主义者之间引发一场大战。如果你把从 19 世纪初期(拿破仑战争)到 21 世纪末期的历次重大战争的死亡数字列举在图表上的话，你会发现，这场战争带来的死亡数字可能是令人沮丧的几十亿，就是我所说的“大规模死亡”。

地球上的人口数量只是几十亿而已，因此，我们得到了一个悲剧性的结论，那就是为了避免人类被人工智能机器种族灭绝，人类不得不进行内部的战争，并且(或基本上)杀完自己。

这场我所谓的“人工智能战争”将会是历史上最波澜壮阔的，因为从来都没有如此高的赌注，那就是整个人类的生存。战争将用 21 世纪的武器进行，因此伤亡数字也将是 21 世纪的数量级。

悲哀的是，对于这个令人沮丧的情形，尽管我尽了很多的个人努力仍然无法找到脱离困境的办法。我夜不能寐，试图找到可以避免“大规模死亡”的现实办法。我没有成功，这让我感到非常的悲哀。事实上，我是如此的悲观，以至于我很高兴我现在还活着，至少我会寿终正寝。但是我很为我的后代担心，他们可能正好处于这样的恐怖时代，而且随时有可能会被毁灭。

对我这样的年纪来说，我可能会在 30~40 年后死去，这段时间可能看不到人工智能战争——这个潘多拉盒子的开启。我认为，可能需要更长的时间来让人们掌握那些制造超智能的人工大脑或者人工智能机器所必要的知识。然而，我这辈子可以看到的也就是围绕本书所提出的人工智能问题展开的激烈争论。

越来越多像我一样的研究者和教授们开始意识到这个问题，他们在媒体或书本上向大众声称 21 世纪将会看到超级人工智能机器的出现。但是至今我是唯一一个预言超级人工智能会带来大战的人，也就是所谓的“人工智能的战争”。

因此，这个观点开始引起世界媒体的注意，像美国、英国、法国和荷兰都开始关注这个问题。事实上我认为，几年之后，这些问题就会脱离学术范围的局限，而逐渐被普通大众接受，引起政治、宗教、国防等领域的关注。

“人工智能的战争”看起来像科幻小说，是对未来的预言，好像多数人都不会太担心，但是随着机器变得越来越聪明，对此的恐惧将越来越深。

那么我对这个问题的态度呢？我为什么要写这本书呢？在内心深处，我是一个宇宙主义者。我认为如果人类不制造人工智能机器，那么将是宇宙的悲剧。为了解释我为什么是宇宙主义者，先让我来讲一个小故事。

试想一下，你是一个拥有神一样能力的 et(外星人)，你 30 亿年前来到了地球。你观察那个时候地球生命的形式，发现它们还是处在非常原始的单细胞阶段。你挥舞一下手中的魔力小棍，改变了地球上所有细菌的 dna，以至于(根据某个论点)让它们永远都不能进化到多细胞生物。因此，永远都不会有植物、动物，没有人类，没有爱因斯坦，没有贝多芬第九交响曲。那不是灾难是什么？当多细胞生物出现在地球，无数的细菌将被它们吃掉。地球上多细胞生物的进化没有对细菌表示任何怜惜。

第一部分导言智能简史导言(7)

第一部分 导言 智能简史 导言(7)

我希望你理解这个类比。如果我们制造人工智能机器导致人类最终被灭绝，拥有神一样智慧的机器人会创造出怎样的、类似贝多芬第九交响曲一样的奇迹呢？作为人类，我们太愚蠢了，以至于想象不出。我们是如此的低等以至于没有能力去鉴赏这些。就像让老鼠来研究爱因斯坦广义相对论一样，人类根本就不可能去理解，因为我们根本就不具备理解这些所必需的神经电路。

但是你会问，如果我内心里是一个宇宙主义者，那我为什么写这本书呢？答案是，我不是 100% 的宇宙主义者。如果我是个纯粹的宇宙主义者，那么我将会安安静静地从事我的人工大脑制造研究，而不会去向公众敲响人工智能问题的警钟。当我临终之时，我会为我被认为是“人工大脑之父”而自豪，但是如果历史谴责我是“大规模死亡之父”的话，那么这个前景真的让我很害怕。我第二任妻子的母亲仍然对奥斯维辛集中营的纳粹生活非常后怕。我清楚大屠杀给人类情感上带来的阴影是会伴随人一辈子的。

我写本书的目的是给出一个警告，我认为在宇宙主义者的工作发展到一定程度前，应该给予人类机会去选择停止他们的工作，如果那是大多数人的选择。我应该停止我的人工大脑研制工作吗？答案是否定的。我认为制造近于人类程度的人工智能机器是非常困难的事情，需要几十年的时间去解决。在未来 30~40 年内，机器人的人工智能将发展得足够高，而变得对人类非常有用。机器人将负责大多数的工作，它们会从事很多枯燥、肮脏和危险的工作。人类将会从这些工作中解脱出来，转而去从事更有回报的工作，去做更有趣的事情。

现在停止人工大脑的研究是很不明智的。然而，当人工大脑真正开始变得聪明起来，并且非常迅速地变得非常聪明以至于(在处于一个“拐点”的时候)成为威胁时，人类就应该准备决定是否继续。在这个关系到整个人类生存与否的问题上做出正确的决定是非常重要的，而对人工智能问题的必要讨论则应该越早越好。在人工智能时代来临之前，应该有足够的时间去理解这个问题的复杂程度。

在公开场合，我是地球主义者，我在试着警告大家。私底下，我又是一个宇宙主义者。就像我在本书中提到的，这个问题困扰着我，让我左右为难。当人工智能的辩

论真正开始后，数十亿人将会和我有同样的感受。从地球主义者的视角看，作为一个宇宙主义者就像一个“种族怪物”一样(物种杀手)，他会接受人类将被消灭的风险。是这个形式的固有本质。是否要制造这些人工智能机器只有两个答案——制造或者不制造。决定制造它们的同时就是接受它们可能会把我们消灭的风险。另一方面，不去制造它们就是决定不去制造“神”，一种弑杀“神”的行为。从宇宙主义者的观点出发，地球主义者是“弑神怪物”。

之前，有些人认为，通过把人类自己变成人工智能机器，可以化解宇宙主义者和地球主义者之间的冲突。我们可以给人类的头脑等加一些成分，使他们成为“半机器人”(半机器生命体，比如，一部分是人，一部分是机器)。我个人觉得这样的看法很幼稚，除非整个人类向人工智能人的转化是一样的快，然而这显然是不可能发生的。

一颗糖粒大小的人工脑潜在的计算能力可能比人类大脑的计算能力要高几十亿倍。对于地球主义者来说，让这样的一个颗粒融入人脑而使之成为半机器人，是“伪装成人类的人工智能机器”。地球主义者会像对待人工智能机器一样仇恨半机器人，并且会把它们两者都消灭掉。有一个人类的外表，并不会让地球主义者觉得半机器人的威胁会减小。

让我试着用一种更形象的方式来解释地球主义者对半机器人的憎恨，可能女性会比男性更有体会。举个例子，一个刚生了孩子的年轻母亲，决定把自己的孩子变成半机器人。她只需把“一颗糖粒”放在孩子的脑子里，就能把孩子变成人形的人工智能机器。“孩子”将只有万亿分之一思想是人的思维方式，其余的脑容量(也就是大脑的99.9999999999%)都是人工智能机器的思想(无论是什么)。结果，这个母亲“杀死”了她的孩子，因为它再也不是人类了。它只是“伪装成人类的人工智能机器”，是和她不同的。

第一部分导言智能简史导言(8)

第一部分 导言 智能简史 导言(8)

因此对于我来说，选择半机器人不能解决宇宙主义者/地球主义者的冲突。任何轻举妄动，都可能加剧冲突，因为这会让地球主义者因难以区别人类和半机器人而更加偏执和多疑。

过去的十年里，我一直比较骑墙和中立，一直用“一方面，另一方面”的方式来表达我的观点和解释两种看法，一边支持地球主义者，另一边又青睐宇宙主义者。经过一段时间，我的朋友开始指责我是“伪善者”。“德·加里斯，你指望人类在地球主义者与宇宙主义者之间进行选择，但是你自己都没有这样做。”“很公平，”我想，“那么我就选择吧。”在我内心深处，我是宇宙主义者，我将在讨论宇宙主义者观点的第4章陈述我支持宇宙主义者的感受和论据。在那一章我将说明为什么我和其他宇宙主义者会如此热情地支持制造人工智能机器，我们准备接受人类可能会灭绝的可怕威胁。

在论述地球主义者观点的第5章，我将解释为什么地球主义者会觉得制造人工智能机器是一场灾难。

在本书的第6章，我将展现给大家一个画面，那就是我怎样看待冲突的酝酿和可能产生的后果。

现阶段，你们中的某些人可能很难想象，机器怎么可能比人类更有能力并且聪明几万亿个万亿倍。这看起来简直就像科幻小说，不会被人类认真对待。

这类事情的发生已经不是第一次了。

1933年，一个叫利奥·西拉特的匈牙利裔犹太物理学家，正在伦敦的一个旅馆里读报纸上关于著名新西兰物理学家、原子核发现者卢瑟福爵士演讲的报告。一个参加演讲的新闻记者询问卢瑟福是否相信那一天会到来——原子核里释放的不可思议的高能量被用于工业上。卢瑟福的著名回答是“妄想”，也就是“没门”、“不可能”、“废话”。西拉特对此表示怀疑，认为肯定有办法。

当西拉特在伦敦街上漫步的时候，突然产生一个灵感。这年是 1933 年，发现中子的第二年。他意识到中子是一种非常好的子弹，它不会被原子核的电荷偏转出去，因为根据定义，中子没有电荷，它的名字也是因此而得。

西拉特知道铀是化学元素表里最不稳定的核子。西拉特的办法是用中子轰击铀原子核使它变得不稳定，因此原子核可能会裂变成两个更小的原子核。而原子核越小其包含的中子越少，也就意味着裂变时有些中子会被发散出去从而去轰击其他的铀原子核。

因此，西拉特是世界上第一个想到核链式反应的人。作为一个物理学家和爱因斯坦的同事朋友，他能够计算出，如果发生这样的链式反应大概可以释放多少能量。他知道两个原子核和发散出去的中子的质量之和将会小于原先的铀原子核和中子的质量之和。那么不见的质量跑到哪儿去了？

他知道这些质量将会被转化成巨大的能量释放出去。一个西瓜大小的铀，经过链式反应释放出的能量有多大呢？经过计算，发现这个能量将足以毁掉整座城市。

西拉特是一个犹太人，说德语，他读过希特勒慕尼黑暴动失败后在监狱里写的书——*meinkampf*（《我的奋斗》），所以他知道希特勒对犹太人有一个“*endlosung*”（“最终解决办法”），那就是，要把犹太人全部消灭。1933 年，希特勒掌握了德国的大权。

因此，西拉特不仅是世界上第一个想出核链式反应概念的人，也是第一个害怕希特勒首先“拥有核弹”的人。这个想法使他惊恐万分。德国在 20 世纪 20 年代是世界上物理学最先进的国家——拥有众多杰出的科学家，如爱因斯坦、普朗克、海森堡，等等。因此，西拉特觉得德国非常有可能成为世界上第一个拥有核弹的国家。

他跑到美国华盛顿，开始向五角大楼坚持不懈地宣扬自己的观点。五角大楼那些人认为他是疯子。仅仅一颗炸弹就可以摧毁一座城市的概念让他们觉得荒谬之极。在 20 世纪 30 年代，威力最大的炸弹仅仅可以炸毁一个大型建筑物。

第一部分 导言 智能简史 导言(9)

第一部分 导言 智能简史 导言(9)

西拉特志在使美国成为世界上第一个发明出核弹的国家，但这个目标没有实现，所以他改变了策略。他找到当时住在美国的老朋友爱因斯坦。爱因斯坦马上就明白了这个理论，因为是爱因斯坦发现了这个理论的大部分内容，尤其是那个著名的、西拉特用来计算能量的公式 $E=mc^2$ 。由于爱因斯坦也是德国籍犹太人，所以他很了解希特勒搞到第一个核弹后可能带来的后果。

西拉特让爱因斯坦在他起草的给美国富兰克林 d.罗斯福总统的信上签了名，总统批准了在几年内制造出核弹的“曼哈顿”工程。从西拉特最早产生的核链式反应的想法到后来的广岛大爆炸只有短短 12 年。

我非常钦佩西拉特。我认为历史低估了他的作用。在我的眼里，他是 20 世纪没有被歌颂的最伟大的英雄之一，他应该获得更高的历史声誉。

在我看来，西拉特开始是孤独的。当他试图说服其他物理学家时，他预言未来核物理学能够制造出一枚足以荡平整座城市的炸弹。他的先见之明是正确的。他知道自己是正确的，他懂得这个理论，并且坚持劝说那时候的掌权者采取行动。

类似的事情再次发生。越来越多的研究者和教授们预见到这样的未来，那就是：很可能在 21 世纪，科技会制造出神一样无所不能的超级智能机器。这些人已经掌握了原理。

他们预见到即将发生的一切，并且为之忧虑。我也是其中的一员。本书试图针对这个问题向人类敲响警钟。如果人类有技术来制造人工智能机器，人类该怎么做？

如果你觉得人工智能这个概念太具科幻色彩的话，那么，请想想 20 世纪 30 年代西拉特所作的预言。它们之间有很强的可比性。想象一下，一个炸弹可以摧毁整座城市，这个观念在 20 世纪 30 年代是多么疯狂，然而，它成为了现实。假如现在制造比人类聪明几万亿个万亿倍的机器的观念对你来说是荒谬的话，就请记住西拉特和他的预言吧！

我希望本书可以引发您去思考,让您意识到这个问题将会主导 21 世纪的世界政治,“物种主导”这个问题会标识和定义这个时代。“人类是否应该制造人工智能机器?”我相信,在 21 世纪,这个问题将会比 20 世纪的“谁将拥有资本”更明显地把人类划分成更加对立的两大阵营。“谁将是地球上的支配物种,人工智能机器还是人类?”将会决定 21 世纪的全球政治。

我用一个小口号来结束此章,以更简洁地表达人工智能辩论的精髓。

“我们是在制造上帝,还是在制造我们潜在的终结者?”

第二部分我正在进行的工作自传(1)

第二部分 我正在进行的工作 自传(1)

谁是德·加里斯？一位提出大胆构想的人——在 21 世纪末，人工智能机器将会比人类聪明万亿个万亿倍，而且关于物种支配这个问题会引发一场重大的战争，并导致几十亿人的死亡。他是疯子吗？还是一个科幻小说家？他的话值得我们去听吗？人类能够承担忽视他的后果吗？

为了使作者及其观点更为可信，本章将分为三部分。第一部分简单地介绍作者生平，第二部分简要介绍他现在的工作，然后在第三部分陈述他未来的工作目标和理想。

我于 1947 年出生于澳大利亚悉尼，在开始写这本书时，我已经是个中年人了。我离过婚，和我的第一任妻子有两个孩子。由于我第二任妻子的离世使我寡居了很长一段时间。我的性格热情而理性。我有一个私人图书馆，有 6000 多本藏书。我是一个科学家、研究型教授、社会活动分子、作家和社会评论家，也有人说我是思想家。

我的少年时期是在澳大利亚度过的。我觉得我热情而理性的价值观和澳大利亚的冷漠、非理性、粗犷的社会文化格格不入。在悉尼奥运会期间，一位 bbc 的记者曾经说过，澳大利亚人对赢得一枚金牌的渴望胜过赢得诺贝尔奖。当我 23 岁完成应用数学和理论物理的本科学位后，我决定离开这个国家。

当我踏上伦敦的第一天，就被一种感觉所笼罩——被一种更加包容更加博大精深的文化所包围。那天晚上我在电视上看到 bbc 的一个辩论节目，我被它的理性所吸引。我如释重负，找到了自己的家——这里有一种可以衡量我个人价值的文化。

在伦敦居住了一年，由于 20 世纪 70 年代，伦敦空气污染严重，我患了黏膜炎，决定搬到充满学术氛围而又美丽的剑桥。我成了一个辅导数学的自由职业者，辅导对象是六七个剑桥大学本科学生。学生们三三两两来到我的公寓里，让我帮助他们解决教师布置的数学难题。

在剑桥呆了一些年后，我看了曾经给第一任妻子买的世界地图册。我的第一任妻子是澳大利亚人，我们邂逅在从澳大利亚到英格兰的“5 周”邮轮上。我想我应该到国际

化的都市，例如布鲁塞尔去居住，可以从其深邃的文化中汲取营养。我所需要做的就是学几种语言然后搬到那里。

尽管国际大城市的生活很诱人，但我还是很喜欢在剑桥的四年生活。喜欢它的绿色、它的美丽、它的学术传统，特别是它的智慧。那是我一生中最快乐的一段日子。但最终我还是不得不离开，因为我在那里找不到长期的工作机会。

我搬到了布鲁塞尔，学会说流利的法语、德语和荷兰语，正如计划的一样，我开始汲取这些文化，并将其融入我的个性。我变成一个更加丰富的人，一个“多元化”(多种语言、多种文化的人)的人而不是“单元化”。作为一个“多元化”的人，我发现和其他“多元化”的人一起，比和“单元化”的人在一起更让人有灵感。

我很喜欢在布鲁塞尔的新生活。很不幸，我的澳大利亚妻子不是这样。她很想回到她的故乡澳大利亚，因为那里有她熟悉的澳大利亚本国文化和语言根源。她极度想念那里。兴趣上的不同最终让我们分手了。她带着孩子回到了澳大利亚。

离婚以后，我和一个说法语的比利时人生活在一起，后来结婚了。无须惊讶，我的法语水平进步飞快。

我在一家很大的荷兰电子/计算机公司得到了一份工作，但很快就感到非常厌烦和沮丧。我很怀念剑桥的理性生活和学术氛围。在计算机界工作了几年后，我在布鲁塞尔大学开始攻读人工智能和人工生命的博士学位，成为一名研究员。

早在 1992 年，我的第二任妻子和我就离开了欧洲来到日本生活。我获得在筑波“科学城”从事人工智能的博士后奖学金。我那时相信，2000 年之前，日本会超过美国，成为世界上经济最发达的国家。这个没有成为事实。我在日本住了 8 年，致力于制造世界上第一个人工大脑。

第二部分我正在进行的工作自传(2)

第二部分 我正在进行的工作 自传(2)

我在布鲁塞尔的一个研究实验室获得了一个从事同样工作的机会。我一个人回去了，因为我的第二任妻子已经死于肺癌。她在认识我之前烟瘾一直很重，虽然在我的坚持下戒了烟，但是危害早已形成。

布鲁塞尔的私立实验室购买了我的一个人工大脑制造机器(世界上仅有的 4 个机器中的一个，售价 50 万美元)。这个实验室创建于信息泡沫时代(dotcomboom)，当时我决定把我在日本攒的 10 万美元投资于此，希望成为一个百万富翁。

我的人工大脑制造工作和我的机器吸引了世界媒体的关注——一个星期内有两家国际媒体联系我。法国的主流报纸 **lemonde** 写了大约 10 余篇关于人工智能的文章，由于媒体的大量关注以至于引起了一次国会听证会。法国是世界上第一个在政治上讨论宇宙主义的国家(巴黎，2001 年 7 月)。

看起来我的生活似乎非常顺利，直到灾难的到来。信息泡沫变成了信息灾难。投资者不再向高科技有风险的研究室投资。实验室破产后，我损失了 10 万美元，也丢掉了工作。

我的下一份工作是在美国当计算机科学教授。难以置信的巧合是，我到美国工作的第一天正好是 2001 年 9 月 11 号。我的系主任那天早上在大学旅馆里碰见我，说：“嗨！德·加里斯。你看到这个了吗？”他指向电视机。“奇怪的行为”，我想。但是顺着他的指向看去，我看到了一幢着火的大楼，并且认出它是纽约世贸中心，我呆若木鸡。过一会儿，当我在大学餐厅用早餐时，听到一个学生尖叫，“他们撞了另一幢大楼！”“美国一直是这样吗？”我心里琢磨着。

作为一名教授，现在我必须去适应美国的个人主义和放任主义的态度，去适应教书的需要。我忙得发疯，去争取研究资金并努力适应我到的第 6 个国家。这些都需要很多精力，所以我没有在媒体上宣传我的宇宙主义观念。

几年前，我给一些退休人员作关于人工智能辩论的讲座。结束后，一个出版商找到了我并问：“你考虑过写一本关于人工智能的书吗？”这个问题的答案就是您现在手中的这本书。

第二部分我正在进行的工作我的工作(1)

第二部分 我正在进行的工作 我的工作(1)

在这一节里，我将对我这些年所从事的工作进行更详细的阐述，重点介绍近十年的，因为这些和本书的主体最有联系。

早在 20 世纪 80 年代晚期，我就开始利用一种模拟达尔文进化的软件形式，所谓的基因算法(geneticalgorithm)来设计神经网络，并开始发表一系列的科学研究论文。到我获得博士学位时，我已经发表了 20 篇论文。

神经网络可以被想象成由像枝节一样的纤维(叫做轴突(axon)和树突(dendrite))连接起来的三维大脑细胞矩阵。来自神经元的信号由轴突发送出去。树突将信号传入神经元。当一个轴突和一个树突相连，形成的联系叫做神经键(synapse)。

在一个真正的生物大脑中，每一个神经元或者大脑细胞拥有上万个神经键。也就是说，它可以被上万个来自其他神经元的信号所影响。这些神经信号同时到达一个神经元，被加强或者加权，然后相加。如果总的信号大于神经元激活阈值，神经元就会被激活，也就是说，它会顺着自己的轴突发送电子脉冲信号，信号的频率决定于总的信号强度大于阈值多少。轴突的脉冲发送到神经键，进一步影响其他的神经元。

一系列神经元图片

神经元(人脑细胞)有许多不同的存在方式。

这种生物神经网络可以用程序模拟。在 20 世纪 80—90 年代，一个典型的神经网络里的神经元数目大概有 10 余个到 100 个不等。当时我的博士研究工作，每个神经网络通常最多拥有 16 个神经元。这和我现在工作中使用近 1 亿个神经元形成了强烈的对比。

下面几页对我的工作进行了更为详细的描述，并且技术性更强。我希望你能够坚持下去，但是如果理解起来确实有困难的话，跳过此节也不会太影响对本书总体的了解。同时在这里提醒您一下，本书最后有一个术语表，可能会对阅读有所帮助。

神经网络，许多神经元相互连接起来形成复杂的神经网络，也就是人脑

cbm 演化出来的神经网络的二维图像

基因算法

基因算法(geneticalgorithm)是一种达尔文进化的程序模拟形式，用来优化任何被进化的性能。实际应用中，我把基因算法用于神经网络的进化。具体通过以下的方式来模拟神经网络的运行。首先要考虑如何描述神经网络本身。我使用 16 个神经元并且让它们和自身以及其他神经元连接，因此，总共有 $16 \times 16 = 256$ 个连接。输入信号的强度，以普通的十进制数字来表示，例如 10.47，再乘以一个权值，例如 0.537，然后相加。作为这个概念的一个例子，想象一个非常简单的只有两个神经元的网络，因此，有 4 个连接。神经元 1 输出的信号通过连接或形成神经突起 c11 发送到自身，并且通过连接或形成神经突起 c12 发送到神经元 2。神经元 2 输出的信号通过连接或形成神经突起 c22 发送到自身，并且通过连接或形成神经突起 c21 发送到神经元 1。假设在某个时刻的强度是 s1 和 s2(例如，10.54 和 7.48)。

每一个连接 cij(或者形成神经突起)拥有一个相应的权值 wij，用来和通过该连接输入的信号强度相乘。因此，输送到神经元 2 的信号总强度应该是($w12*s1 + w22*s2$)。对于神经元 1 计算也类似。总共有 4 个这样的权值。假设每一个权值的范围在 -1 到 +1 之间。因此，每一个权值可以用二进制小数来表示，比如说 8 个比特(二进制数字，0 或者 1)。4 个这样的数字可以用 $4 \times 8 = 32$ 个比特来表示，可以排列成有 32 个比特的一行。对于 16 个神经元，我们将用一个有 $16 \times 16 \times 8 = 2048$ 个比特的行或串来表示我将用来进化的神经网络的 16×16 个权值。

如果我知道 2048 个比特的值(0 或者 1)，我将可以计算所有的 256 个权值，并且可以通过它们建立一个完全连接的神经网络。相反，如果知道所有的权值，并且知道输入信号的初始值，我们就可以计算出每一个神经元发射时候的信号强度。如果知道了每一个神经元是怎么发射的，就会知道整个神经网络是怎么发射信号的或是怎么运转的。我们可以提取某些神经元的信号，并且把这些信号当作控制信号，来控制一些活动，比如说，通过控制机器人腿的角度来让它行走。

第二部分我正在进行的工作我的工作(2)

第二部分 我正在进行的工作 我的工作(2)

为了解释基因算法是怎么运用于此的，想象一下产生了 100 个随机的比特串，每个长 2048 个比特。从每一个比特串我们可以构造一个相应的神经网络。对于每一个网络都采用同样的初始信号来让网络的信号传输启动。提取其中的某些输出信号并且使用它们，比如，通过控制构成棍形腿的 4 个线条的角度来让棍形腿走路。我们可以测出在一定时间内腿的走动距离。

那些走更长距离的神经网络的比特串，可以存活到下一代。那些走的距离短一些的比特串则会死亡，这就是所谓的达尔文定律，即“适者生存”。比特串越适应，即具有更高性能分数或“高适应值”的就越会复制自己，产生所谓的“孩子”或后代。然后，这些孩子和它们的父辈一起被“变异”，也就是说，每一个比特有一个很小的概率来改变相应的值(0 变成 1，1 变成 0)。两个比特串可以“交配”，即一个称为“特征交换”的过程。它有很多方式。一个简单的方式就是，选取两个父比特串或者通常所谓的“染色体”，在同样的位置把它们分成两部分，然后交换相应的部分。这等价于性，基本上来说就是混合来自父辈的基因来产生后代。

越适应的父辈就拥有越多的后代。每一代基因算法拥有一定的总体数量，例如 100。大多数的变异和特征交换会让染色体拥有更低的适应值，所以它们将被从总体中清除。偶尔的一次变异或一次特征交换则会增加一点染色体的适应值，以至于在某个时候它会把自己的父亲或者其他低适应值的染色体挤出总体。通过对这样的过程循环几百次，就有可能进化出性能更好的神经网络，或者任何需要进化的目标。

神经系统进化

当我在布鲁塞尔攻读博士学位的时候，研究如何进化出可以随时间变化输出信号的神经网络。据我所知，在此之前没有任何人做这样的事情。在此之前，曾经有人把基因算法应用到神经网络进化中，但这些应用都是静态的，也就是说，输出的信号不会随着时间而改变。我觉得这些限制是不必要的。基因算法应该能够解决随时间改变信号的问题。当我想通了这一点后，就开始研制可以让棍形腿走路的神经网络。我成功了。这需要一些小技巧来让它进化，它确实实现了。

这个发现，说明动态进化(和静态相对应)的神经网络是可以实现的，这个发现给我打开了一个崭新的世界，并且产生了一个新的被称作“神经系统进化”的研究领域。我开始考虑下一步的工作。我产生了一个新想法，那就是如果我用一个神经网络进化出一种行为，那么我就可以用第二个神经网络，也就是说，用一套不同的权值，进化出一个不同的行为。权值设置决定输出信号的动态性。

于是我充满信心，不再局限于二维平面的简单的棍形腿试验，而是开始进行我称作“lizzy”的三维生物模拟行为。如果我成功地进化出一种行为，那么就可以进化出一系列的行为，而每一个神经网络对应于一个行为。现在我就可以更换行为，让 lizzy 先走，然后改变方向。为了使行为转换更平缓，唯一必要的就是先停止“向前走”行为生成网络(或者是我所称作的模块)的输入，然后输入行走模块。模拟试验显示动作转化很平缓，这非常好。现在我知道我可以类似四脚生物的 lizzy 展示行为库里所有行为。

现在产生的问题就是：什么时候改变行为。也许这样的决定来源于环境的刺激。我开始试验是否可以进化出探测器模块，例如，信号强度探测器、频率检测器、信号强度差异探测器，等等。是的，这是可能的。另一个逻辑步骤就是试着进化决策类型模块，例如这样的类型——“如果输入 1 的信号强度大于 s_1 ，并且输入 2 的信号强度小于 s_2 ，那么采取 an 行为”，也就是说，刺激信号可以传送到执行 an 行为的模块。

综合所有的三种模块，也就是行为产生(或行为模块)、探测模块和决策模块，看起来好像可以开始制造人工神经网络的工作了。如果有很多这样的模块，那么我认为可以很合适地称这样的集合为一个“人工大脑”。就是在这样的阶段上，我信心满满，开始朝着人工大脑先驱者的方向前进。

第二部分我正在进行的工作我的工作(3)

第二部分 我正在进行的工作 我的工作(3)

fpga 技术的应用

困难依然存在，我在 20 世纪 80 年代晚期和 90 年代初期所使用的计算机对于我所构想的任务来说简直慢得“不可救药”。当我开始试用十几个进化的模块时，lizzy 在计算机屏幕上的模拟速度显著变慢。每当我加上另一个模块的权值后，模拟速度就会变得更慢一些。很显然，这不是一条正确的路。怎样才能解决这个问题呢？

这个时候，也就是 1992 年，我已经完成了博士学位，开始了在日本的博士后工作。那年夏天，我和我所合作的、在维吉尼亚的乔治 梅森大学的一个电子工程师朋友进行了一次交谈。我问这位熟人怎样才能使我的神经网络模块的进化速度通过电子技术得到提升。讨论了半个小时后，他提到了一个称作 fpga(fieldprogrammablegatearray 可编程门电路序列)的技术。因为我不是电子工程师，所以我从来没有听过这样的东西。“什么是 fpga?”我问。他告诉我这是一种可编程的特殊芯片，也就是说，我们可以用一个比特串来指令芯片里的电路怎样去自我连接(或者用术语来说，就是自我“设置”)。

我顿时兴奋无比。一个构想闪过我的脑海。既然在过去的几年里我一直利用基因算法来进化神经网络，我现在的倾向是把设置比特串想象成基因算法里的染色体，因此，在芯片里进化硬件的想法看起来是可行的了。我开始不断地“拷问”我的朋友，问这些设置比特串可以无数次的发送吗？他想了一会，回答说，如果芯片是基于 ram(就是计算机内存)的话，那么就像任何计算机里的普通内存一样，可编程芯片可以被无限制地重新编程。

我非常高兴，因为这意味着可以通过发送随机的比特串，来随机地设置或者连接可编程芯片，并产生复杂的随机电路。如果有另外一个被编程的用来测试随机程序芯片的性能电路，那么就可能在硬件上以硬件速度执行基因算法。

我为这个构想而激动不已，以至于我一回到日本的研究组就马上为我的想法作了一次研讨会，并且开创了所谓的“可进化硬件”的研究领域。并发表关于这个想法的论文，向同事宣扬，在会议上作演讲，等等。可进化硬件(evolutionaryhardware 或者 eh)，

现在已经是个确立的研究领域了，每年都会在美国、欧洲和日本举行研讨会并出版相关的学术期刊。作为该领域的先驱，我常常在自己的日常工作中试用它的基本思想。

紧接的那一年，即 1993 年，我转到日本京都的研究实验室从事人工大脑的制造工作。我相信，可进化硬件是一个使人工大脑制造变为可能的有效工具。

我开始写论文，声称我打算在 2001 年前制造一个拥有 10 亿神经元的人工大脑。在 1993 年，宣布这样的计划常常会招来不信任的眼光。因为在那时，大部分的神经网络研究者仅仅使用几十个或者上百个神经元。就像我早几年一样，突然听到某个人说要使用上十亿个神经元，简直是非常可笑的。为此我被嘲笑、奚落。

但是，我坚信我的想法是可行的。如果某个人制造一种基于可进化硬件原理的特殊计算机，那么，其电子性的进化速度将会使大脑制造成为可能。我用数学进行了推理，证明在 2001 年前制造出拥有十亿个神经元的人工大脑是可行的。我和日本实验室有一个 7~8 年的合同，因此，我想我有足够的时间来实现我的梦想。

我的第一个任务就是选择一种可以设计和进化神经网络的媒介。我选择使用胞腔自动机(cellular automata 或者简称为 ca)。在二维胞腔自动机上的每一个细胞都类似于棋盘上的一个方格，但是有两个不同。一个是这个棋盘有无限个方格，另一个是方格的颜色不仅限于黑色或白色，而可以是有限集合里的任何颜色。每个方格在每个单位时间里可以改变其颜色而成为集合中其他的任何一种颜色。一个特定方格所要改变的颜色决定于它现在的颜色和其邻接方格的颜色。举个例子，如果北边方格是红色，东边方格是黄色，南边方格是蓝色，西边方格是绿色，并且中间的方格是棕色，那么在下一个时刻，中心方格将变为紫色。

第二部分我正在进行的工作我的工作(4)

第二部分 我正在进行的工作 我的工作(4)

通过适当地选取几千个这样的规则，我就可能让这些胞腔自动机像神经网络一样生长和进化。举个例子，我可以生成一个三个细胞宽的路径，通过它我可以发送生长细胞，让它们在路径的中间向前移。当生长信号到达生长路径的尽头时，它可以让路径向前延伸、左转、右转、分裂，等等，这些决定于生长信号的颜色。通过改变胞腔自动机路径中间生长信号的顺序，我能够进化基于胞腔自动机的神经网络。

这个过程有两个步骤。第一个是生长阶段。经过几百个单位时间后，生长达到饱和。再没有三个细胞宽的胞腔自动机路径会生长了。这些路径就是神经网络的轴突和树突。当生长阶段一完成，也就是说生长控制细胞从网络里被清除后，生长的神经网络可以被随后的信令阶段所使用。输入信号会被施加，并且会扩散到整个网络。这些信号就像我早些年进化出的神经网络中的信号一样。它们可以在输出点被提取出来，并且可用来控制那些适应值(fitness)或性能可以被测量的过程。这个适应值就是整个网络的适应值，而网络是依次产生于一系列生长指令的，也就是一个含有 6 个(所有的数字)整数的随机串。

我所做的就是让神经网络和胞腔自动机结合起来。这些据我所知没有前人做过。之所以这么做是因为我认为胞腔自动机是一个可以拥有几十亿个胞腔自动机细胞的合适媒介，归于十亿个神经元是足够了。这对我来说是可实现的。当时的工作站(比个人电脑强大一些的计算机)的内存大概有十亿字节。内存是很便宜的，因此，既然我可以在一个字节(8 位)上存储胞腔自动机细胞的状态或颜色，并且我的工作站可以拥有十亿字节的内存，那么我将可以存储十亿个胞腔自动机细胞的颜色，十亿个！这是很多的，足够在其中放入一个拥有大量神经元的人工大脑。空间将不会是问题。那时候的技术允许这样做，这是可实现的。

我花费了一年的时间去写所有的这些规则(北-东-南-西-中的类型规则)来证明二维版本的基于胞腔自动机的神经网络是可以运行，可以进化的。我不得不手工编写(有软件工具来帮助我)大概 11000 个这样的命令来让它工作，结果它工作了。我成功地进化了振荡器电路、信号强度探测电路、直线运动探测电路，等等。是时候去研究拥有完全不同拓扑结构的三维版本了。在二维结构里，电路会相互碰到，它们不能跨

过另一个电路。然而在三维结构里，胞腔自动机路径可以通过第三维相互交叉。三维电路的动态性和进化性比二维电路要丰富很多。仅仅两年后我就让三维版本开始工作了，这个版本有大概 60000 个命令。

但是，这个时候我在日本过得非常郁闷。我们组的顶头上司有一个政策，就是每一个人负责一个项目，这让我觉得非常孤独和智力空虚。我没有一个可以真正交谈的人。在给他们施加了一些压力后，我终于在 1996 年得到了一个硕士生助手——一个年轻的德国人。

我向我的助手解释，三维版本已经基本上完工了，并且我对一直使用的胞腔自动机模型已经越来越清楚。我向他介绍了我想直接在电路上以电子速度产生和进化基于胞腔自动机的神经网络梦想。我觉得应该简化胞腔自动机模型，让它可以适应于当时的新想法，即 1996 年的电子技术。他聆听了我的一系列愿望，然后消失了两个星期。当他回来的时候，带来了一个崭新的、更加简化的神经网络模型，它保持了旧模型里的大部分特征，但是加了一些新特征，使它变得可以直接实现于电子装置。这个新模型被称为“codi”。

大概在这段时间，也就是 1996 年下半年，一个美国电子工程师和我联系。他认为我的论文很有意思并且想与我合作。我把新模型的一些细节发送给他，问他能否用当时市场上特殊的 fpga 板在硬件上实现它，他说应该可以。我的日本老板同意了这个想法并给予资助。这位在美国的新同事开始与我密切合作，不幸的是，一年后失去了我的德国同事。他回到欧洲攻读博士学位。我的日本老板又回到了“一人负责一个项目”的老路上，我比以前变得更郁闷了。

第二部分我正在进行的工作我的工作(5)

第二部分 我正在进行的工作 我的工作(5)

我开始想着离开，但是还不能，因为刚刚批准制造新机器。我和我的小组经理关系也越来越紧张，特别是我发现他有这样的一个政策，那就是把自己的名字加到他下属写的而他一点贡献都没有的学术期刊论文上。他让我把他的名字加到我的一篇学术论文上，我拒绝了，告诉他这在西方被认为是很令人厌恶的，是滥用权利和腐败。之后，我们的关系急剧恶化。我被允许留到 1999 年年底，然后我将不得不离开。日本的经济在 20 世纪 90 年代表现是如此之差，被称为是“迷失的十年”，以至于整个研究机构被认为是前景黯淡、没有希望的，因而不需要在经济萎靡时期给予资助。所以，我和部门里大多数人在千禧年离开了日本。后来我得到了另一份工作，又是在布鲁塞尔，并且是做和日本实验室一样的工作。

在 1996 到 2000 年之间，我在美国负责硬件的同事为了建造能够实现我制造人工大脑理想的特殊硬件，一直孜孜不倦地工作着。对他来说，进展缓慢。他从我的日本老板那儿只得到了有限的资助，他只请得起一个全职助手和少数几个兼职助手。

在他工作的过程中，制造 fpga 芯片的美国公司决定把这些板子从市场上撤下。我的美国同事不得不和公司争取剩下的一些芯片。这耽误了好几个月。这些芯片最终还是得到了，但是没有进行测试。因此，他不得不自己测试，因为没有公司的细致测试，导致了更多的延迟。

直到 2000 年中期，我所谓的 cam—大脑机器(cbms)才被充分测试，真正的进化试验可以开始了。cam 即 cellularautomatamachines(胞腔自动机)，早期的人工大脑是以胞腔自动机来实现的。

第一个 cbm 在 1999 年被送到我的日本实验室，它仍然有缺陷。因为只有未测试的芯片和很少的人力资源，工作进展缓慢。但这些都并不令人沮丧，因为其他的一些人开始对 cbm 感兴趣。到了 2001 年，世界上有四台这样的机器。第一台留在我原先的日本京都实验室。第二台被比利时的一个语音处理实验室购买，后来转移到了一个生物信息公司，同样也是在比利时。第三台被我的比利时实验室购买，第四台由我开发硬件的同事所拥有。因为比利时拥有了当时世界上四台 cbm 中的两台，从某种意义上来说

说，比利时在这个领域处于世界领先地位。在 2000 年，我成功地从布鲁塞尔政府获得了 100 万美元的研究资金，用来制造人工大脑，来控制一个像小猫一样的机器人，使它拥有数百个行为能力。所有这些 20 世纪 90 年代的工作都是我 80 年代博士工作的一些延伸。

“cbm”大脑制造机器，日本，1999 年

cbm 到底能够做些什么呢？我相信它在那个时候是一台真正的神奇机器，当人们意识到它的重要性后，将会在计算机历史上占据重要地位。它直接用电子技术实现基于 codi 胞腔自动机的神经网络模型。它可以在几秒内进化成一个神经网络，也就是说，它可以在此时间内进行一次遗传算法的完整运算，也就是几万个神经电路模块的生成和适应性的计算。它可以以每秒 1300 亿次的速度改变胞腔自动机细胞的颜色；它可以处理将近一亿个人工神经元；它的计算能力相当于 10000 台个人电脑，因此，绝对是一个值得花费 500000 美元的超级计算机。

cbm 内部构造

cbm 拥有两个主要的任务。第一个是去进化一些单个的神经电路模块或者是我所说的模块。一个神经网络是在拥有 $24 \times 24 \times 24$ 个细胞或小立方体的三维胞腔自动机中生长或进化的。这个空间可以容纳 1000 个神经元。像枝节一样的树突或轴突在这个空间里随机的生长。一个被编程的 fpga 用来计算生长的网络神经信号的质量，其基本思想和我在 1996 年之前所做的工作是类似的。当一个模块进化完成后，它被下载到一个拥有 10 亿字节的内存中。64000 个这样的模块将被进化，每次一个，每一个都拥有自己的被人类“进化工程师”(evolutionaryengineers, ee)指定的适应值定义(任务或功能)，然后被下载到内存中。然后，“大脑建筑师”(brainarchitects, ba)用软件把这些下载模块相互连接来形成人工指定的人工大脑结构，用来执行人们需要的命令。

第二部分我正在进行的工作我的工作(6)

第二部分 我正在进行的工作 我的工作(6)

重要的日子——神经元在 cbm

机器上第一次被演化出来，日本东京，1999 年

布鲁塞尔实验室在信息灾难(dotcomcrash)中破产之后，我转到一所美国大学(utahstateuniversity，犹他州立大学)工作，被迫重新思考。因为我在布鲁塞尔工作的实验室无法向我负责硬件开发的同事支付从其手中购买的 cbm 机器的资金，这致使他损失了 30 万美元，迫使他退出了整个项目。因为他对 cbm 的具体结构有垄断性的认识，这让工作出现了令人惊愕的停顿。我的美国大学付不起 50 万美元来购买第五台 cbm，因此我只好空手来到这所大学。我花了 2 年多时间去学会怎样去教学和撰写关于可进化硬件和量子计算的论文，直到我意识到我可以再制造人工大脑，多亏了摩尔定律，这次可以便宜得多。

一家英国公司开发了一种能够将普通的计算机软件代码(计算机 c 语言)翻译成能够设置可编程芯片(fpga)比特串指令的方法，在得知这个消息后，我构思了一个新的大脑制造研究项目。这种新的途径就是用 fpga 电子电路板(花费少于 1000 美元)编写基因算法来进化神经网络。这种板子进化一个神经网络电路模块要比个人电脑利用软件(后者需要几个小时甚至几天来进化一个模块)快几十倍。

这些模块一个接一个地在硬件上被进化，所得结果被下载到电脑上。每一个模块都有自己的进化功能，是被“大脑建筑师”(ba)明确说明的。当几万个这样的模块被下载到 pc 上，将需要一种特殊的软件来连接这些模块，例如，模块 m3728 的输出将会被连接到模块 m9356 的第二个输入口。然后用 pc 来进行整个人工大脑的实时(平均每秒每个神经元 25 个神经信号)神经信令活动。如今的 pc 可以实时模拟 10000 个模块的神经信令活动。整个我所谓的“便宜的大脑制造”方法花费低于 2000 美元，因此，我希望这个概念可以扩展到其他大学和研究所里。理所当然，如果我真的可以在近几年里制造出一个这样的大脑并且证明它可以控制机器人完成一些任务的话，这对于我的同事们和资助者来说将会更有说服力。

第二部分我正在进行的工作未来的任务和理想(1)

第二部分 我正在进行的工作 未来的任务和理想(1)

如果那些嘲笑我是荒谬断言的人能够看到这一点，即我可以在 2001 年前制造一个拥有 10 亿个神经元的人工大脑，并能够在 1993 年看到 2001 年出现 cbm，就不会再嘲笑我了。老实说，这台机器不是处理 10 亿个神经元。实际上的数字是 7500 万个，但只是差一个数量级，这已经不错了。同样不可否认的是，建造人工大脑这样一个艰巨任务仍然没有完成。这里有很多的工作要去做，并且我一直饱受责难。因为这些滞后，不管是因为商业原因、技术原因、管理原因，还是个人原因，我至今还没有一个人工大脑来展示给人们。一些记者开始变得不耐烦，并且怀疑我是否能够或者什么时候可以制造出一个。因为布鲁塞尔实验室快要破产了，我不能随意告诉记者其中的原因。这实在太让人沮丧了。

当时的计划是在破产前完成机器进化研究，一次进化一个模块。如果进化的程度不够，我们将不得不改变在 cbm 中使用的适应值定义。我们可能也不得不改变实现在可重复编程的 fpga 板子上的神经网络模块。当这一步骤完成后，下一步可能就是开始制造多模块系统，先是拥有几十个模块的，然后几百、几千，直到 64000 个模块来制造一个志在控制机器人行为的人工大脑。我们打算展示一个由人工大脑所控制的，并且拥有很多行为的猫形机器人。人们不需要博士学位来理解到底是怎么回事，就像理解 cbm 一样，相反，通过对机器人的简单观察，人们可以看到“它有一个大脑”。

做这些工作可能会花费好多年的时间，同时我们有必要去认真考虑下一代大脑制造机器，就是我所谓的 bm2(brainbuildingmachine，大脑制造机器，第二代)。我开始和另一位美国同事合作，他有一些关于下一代自我设置的电子技术的革命性思想。他预计，如果有几百万美元的预算，在未来 4 年内制造比 cbm 性能高 1000 倍的下一代机器是可能的。

然而，既然我不能让美国的资助者给出这笔资金，我只好满足于我在前面章节介绍的更合适的方法(也就是“便宜的大脑制造”)。

事实上，我的总体构想是，每隔 4~5 年就制造出新一代的大脑制造机器和相应的大脑。我现在已经快 60 岁了，所以，如果我决定 70 多岁退休的话，我只有 15~20 年

的时间。在 20 年内，如果摩尔定律还可以有效，人们将有能力让一个原子存储一个比特的信息。当这些发生后，制造我所谓的“阿伏加德罗机器”(avogadromachines)，一种具有万亿万亿个部件的机器将变成可能。阿伏加德罗数字就是像人类这样级别的物体，比如手中的苹果可能含有的分子数目。

如果第二代大脑制造机器可以被资助，并且在未来的 4~5 年内被制造出来，就可能让下一代人工大脑和生物大脑更加相似。实现的神经网络模块将更复杂，并且行为上和生物神经元更相似。

仅仅在下一个 20 年，也就是我自己剩下的职业生涯中，我和其他大脑制造者将拥有制造更好的人工大脑所需要的技术和工具。

无须惊讶的是，我和其他一些人对大脑制造在未来 20 年的进步速度开始有明显的预期。我们的人工大脑那时候将为人类做些什么呢？我可以说，很有可能，它们将出现在我们的家里，为我们打扫房间、照顾小孩，和我们聊天，给我们来自地球上知识库里面的无限知识。我们将可以和它们有性关系，被它们教育，从它们那里得到娱乐，等等。20 年后的大脑制造业，我预计，每年全球范围内将能创造万亿美元的价值。在这些年内，我希望并预计，如果我们的研究团队可以“证明这个概念”，那么说明人工大脑是可行的，那么一个崭新的“大脑制造”研究领域将会确立。

如果我们在 20 年内可以拥有这些话，那么人类在之后的 50 年、100 年内将何去何从呢？假设我们关于大脑科学的知识累计是以指数级增长的话，所有的这些知识可以一发现就马上被神经工程学吸收，那么一开始对人工大脑的正面影响会很快被不愉快和恐惧所代替。

第二部分我正在进行的工作未来的任务和理想(2)

第二部分 我正在进行的工作 未来的任务和理想(2)

我愿意被看作是人工大脑之父。我觉得我已经是进化硬件和进化工程学之父，这些都是这个新领域(人工大脑)的启动技术。如果我是一个传统思维的工程师或科学家的话，我将可能很满意，并继续我的工作而不需要担心工作的长期社会后果，但我不是这样。我是一个关心政治的人，我非常忧虑。我的不寻常的科学家/工程师身份和此时的社会评论家和媒体人的组合，使我成为对人工智能问题敲响警钟的合适的人。

我希望，我的信誉和作为职业大脑制造者的身份，将有利于帮助我对 21 世纪人工智能问题敲响警钟。如果我没有成功地制造人工大脑，其他人也会成功的。对我来说，要成功地制造每一代大脑制造机器和相应的人工大脑，就需要筹集更多的金钱，雇佣更多的人，因为这个行业的范围将越来越广泛。我需要“成为”戈达德(美国火箭工程学的先驱者)或者沃纳·冯·布劳恩(让美国人登上月球的人)。这两个人都是从玩具火箭开始的，但是有先见之明。在 20 世纪 20 年代，戈达德的第一个装置不比古代中国式的火箭先进多少。20 年后，他和冯·布劳恩都被各自的政府巨额资助，来制造能够长距离旅行的高度复杂火箭。在 20 世纪 60 年代晚期，冯·布劳恩在美国国家航空和宇宙航行局把阿姆斯特朗送入月球的过程中发挥了巨大作用。

我也有着同样的梦想。那是一个花费几十亿美元制造人工大脑的国家性项目。我已经谈论过 j-brain 计划(日本国家大脑制造计划)，a-brain 计划(美国的)，e-brain 和 c-brain(欧洲的和中国的)。在 20 年内，为了拥有“阿伏加德罗式”机器，将有巨大数量的工作要做，这个机器不仅仅是几十亿个部分，而是多少万亿个部分，将会需要很多很多人。那是我近 20 年的远期理想。

那之后，我将会退休，我希望能够发挥一个智慧老人的作用，指导年轻人的思维，告诉他们整个大脑制造的努力方向。正如本书所表明的，当机器以指数级“进化”变得比人类聪明时，我对人类的未来生存不抱乐观态度。

我的最终目标是看到人类或者至少一部分人变成宇宙主义者，成功地制造神一样无所不能的人工智能机器，远胜过我们这些微不足道的人类智慧或者其他能力，那是我真正的目标。不幸的是，我可能看不到那一天。真正的人工智能在我死去的 30~40

年内不会被制造出来。我活着看不到我工作的真正结果，这是让我沮丧和失望的一个根源，但也是一个安慰。至少我寿终之时会安详地躺在床上。正如这本书所说的，我会为我的后代感到恐惧，因为我相信他们很可能会在本世纪晚期因为物种问题的大规模战争而被毁灭。

埋伏的恶魔——作者从其 cbm 机器后面窥视，原版书封面

因此，亲爱的读者，你现在知道了我故事里面科技的一面和我这一生工作的简介。知道了这些会不会让你觉得人工智能战争的政治观点更可信一些？我是不是应该告诉你，我不仅是一位博士，更是一名美国教授和中国的客座教授(现已成为全职教授)。我也是当时日本唯一的一名达沃斯科学会员，所以我要去达沃斯经济论坛来说服那些亿万富翁。我因为 cbm 进入了吉尼斯世界记录(2001 年的 126 页)。我是特别学术期刊《进化神经系统》的客座编辑，这是该期刊的主编们专门授予的荣誉，只有那些他们认为在全世界某些特别领域做得最出色的人才能享有。如果大家认为我所做的是很古怪的，那么我希望大家清楚我是一个能胜任的怪人。这章的要点是让大家相信本书的作者，人工智能、宇宙主义者、地球主义者、大规模死亡等术语的创造者是值得去倾听的。我的成功与否要由你们来判断。

未来机器人——思考机器人，机器人思考者

第三部分使人工智能成为可能的关键技术摩尔定律(1)

第三部分 使人工智能成为可能的关键技术 摩尔定律(1)

多年前，当我试图在美国发表本书的早期版本时，我收到了一封美国出版代理商发来的 e-mail，信上说，她读过我的手稿，我的初稿“写得非常好，但是‘捕风捉影’，对于出版商来说很难销售”。从那以后，我通过事实发现，当我试图劝说人们接受这些想法的时候，我所遇到的最大障碍就是它看起来太具“科幻小说”的特点。很多读者很难接受书稿中所提出的观点。举个例子，大多数人，当他们接触到某些概念的时候，例如比人类聪明万亿个万亿倍的人工智能等级的“超级智能机器”、或者可以杀死几十亿人(“大规模死亡”)的“人工智能战争”、或者小行星大小的计算机等，毫不惊讶，他们直接的反应是一种怀疑。他们会嘲笑那些看起来荒谬的观念。即使我的许多同事(特别是那些非物理学家的)也没有认真地对待这些观点。例如，许多年前，我试图劝说地球上最著名的“应用道德规范”教授、普林斯顿大学的彼德·基辛格(prof.petersinger)教授去接受这些观点。我一直试图劝说他写一本关于“人工智能道德规范”主题的书，用来处理 21 世纪可能制造出的人工智能机器而带来的大量的道德和伦理问题。他的回答非常具有代表性。我引用一下他写给我的一封 e-mail，“直率地讲，就你的观点来说，我不知道怎样把你定位于‘完全的怪人’和‘超越时代的天才’之间”。而这是来自一个思想开放的人的评价。

所以你可以明白，我此时最紧迫的问题是信任问题。如何说服人们接受这些观点：这不是一本随便的“科学幻想小说”，而是非常可能的“未来科学”。不可否认的是，随着世界媒体越来越多地传递这个信息，这个解释任务变得越来越容易。在来美国之前，我经常出现在法国、荷兰、英国、澳大利亚、波兰等国的媒体(电视、报纸、杂志、电台、网络等)上。美国的媒体最近也越来越多地联系我，即使我没有去做任何宣传。

尽管得到越来越多的信任，但是还有很多路要走，所以本书仍有必要去说服那些怀疑者，这些观点是值得去认真思考的，是不能被忽视的。

本章主要试图去说明在本世纪内制造出人工智能机器是可能的。

在未来的 100 年，将会发展出一些惊人的技术，它们是如此的先进和令人惊奇，以至于会让在本世纪能否制造人工智能机器的问题浮出水面。

当您读完这个章节后，我希望您会有个强烈的印象，那就是人工智能机器的潜在智能是真实且巨大的。人工智能机器将会拥有超越人类智能很多数量级的能力，不只是聪明十倍、一千倍，或者一百万倍，而是几万亿倍、几百万亿倍、几亿亿倍、几亿亿亿倍(采用通用的术语)。(如果一个数字比另一个数字大 10 倍，我们说它是大一个数量级。如果大 100 倍，它是大两个数量级，等等。)

本章试图劝说你们的那些数字没有被夸张。基于我们对本世纪新技术发展的预计，有很好的理由让我们相信，在未来 100 年内，人工智能机器的制造是一个很现实的提议。

此章将是本书中最复杂的部分，因为它将讨论一些很新的甚至还没有出现的科学观点和技术。我会尽量让普通的没有科学背景的读者也都能够理解。

就像我在导言那一章所写到的，我的人生目标之一，除了制造人工大脑，就是去对“人工智能问题”敲响警钟，或者你可能更倾向于称呼它为“物种支配问题”，或“宇宙主义—地球主义的冲突”。对于即将到来的这些基本问题有很多方法去称呼。

这个问题非常重要，它不能仅限于一批“无足轻重的科学家们”之间的智力讨论，很快它就会关系到每一个人。因为如果宇宙主义者很严肃地“威胁”要去制造人工智能机器的话，每一个人都将或多或少地受到影响。如果人们只是把自己的担心局限于科学专家这个只占人类 1% 很小比例的团体内，是不可能产生一场大的公共辩论的。

一个让人们开始讨论人工智能的有效方法就是去写一本书。通过这本书可以让记者们熟悉这个问题，于是他们会写文章向更多的读者介绍这个问题。类似的推理可以应用到电视和电台记者们身上，他们会把这些观点介绍给更广大的观/听众们。令人遗憾的是，事实上只有一半人会去读书。

第三部分使人工智能成为可能的关键技术摩尔定律(2)

第三部分 使人工智能成为可能的关键技术 摩尔定律(2)

也许让这个信息传递开来最直接的方法就是让好莱坞去拍摄一部关于这个主题的轰动电影。我希望这会成为现实。这个方向已经有了开始。许多国家的一些电影摄制者已经做了一些关于我和我的观点的纪录片。

在讨论这些新的或者有待发展的技术细节之前，我想说明哪一类读者可以从我认为本书最难的这一章节中获益。我相信只要你学过高中的一些科学知识，就能够理解这章所介绍的内容。

为了透彻理解这一章节，我需要去介绍一些非常“高科技的”技术，甚至一些还不存在的科技，所以我将不得不涉及不同等级的一些细节。我希望不会让读者阅读起来感到吃力。

我建议你不需要付出太多的精力，只要尽量理解就可以了，然后跳到下一章节，下一章将介绍宇宙主义者的很多观点和看法。然而，如果你决定跳过这一章节，我建议你至少接受一个主要的结论，那就是(用一点来总结)：本世纪的科技将会使人工智能制造变得可能，而且它将比人类聪明几亿亿倍。

通过介绍在电子世界中众所周知的“摩尔定律(moore'slaw)”现象，我开始对本章的一些使人工智能机器成为可能的科技进行介绍，这些我在导言章节已经简要讨论过了。然而这次，这个概念会涉及更多的细节。戈登·摩尔是美国加州硅谷的“英特尔”微处理芯片公司的创始人之一，21世纪初仍然健在。在20世纪60年代中期，他注意到集成电路的运算速度和密度(即密集到一个硅芯片表面上的晶体管数目)每一年左右翻一番。这个倍数增长在过去40年或多或少成为了事实，并且很多人相信它会继续到大分子等级。

试图让电子元件变得更小更密集的要点是什么呢？如果两个元器件要相互通信，并且已知恒定的光速(也就是，电子元器件相互传递信息的最大速度)，那么元器件之间的距离越短，它们之间相互影响的速度就越快。并且，电子元器件的尺寸越小，一个特定表面上密集的数量就越多。因此，这个芯片就能够具有更强大的性能，因为它有更多的元器件来做更多的事情。

因此，微处理芯片产业一直承受着压力——按比例缩小，让晶体管变得更小，让电路变得更小。如果一个公司在这场狂热的赛跑中落后了，它将失去销售额并且破产。如果竞争公司在开发周期领先你 6 个月，并且先于你的公司发布了一系列的产品，你将陷入很不利的处境。新一代的芯片和计算机每隔一两年就会问世。我们现在已经习惯了。我们知道如果我们等 6 个月或者一年的时间，我们将能够用同样的价格买到性能更好更优越的计算机。

摩尔定律可能是我们这个时代里最重要的科技和经济现象之一。它一直在为推动全球经济的数字革命加油。现在许多国家的许多工作机会和很大比例的 gnp(grossnationalproduct, 国民生产总值)都和电子、计算机、通信产业有很大的联系，因此，如果摩尔定律开始失效的话，人类将会深受震动。然而，这仍然是个问题。

当电子元器件，特别是晶体管的尺寸变得越来越小，最后达到了一个如此小的等级以至于要采用一套不同的物理原理来支配它们的行为。

如果摩尔定律一直有效到分子级别，也就是说，如果电子元器件的大小可以达到分子级别并且仍然有功能的话，那么新的物理原理将被采用。牛顿在 17 世纪发现的传统的“经典力学”不再适用，取而代之的是 20 世纪更新的“量子力学”。

量子力学控制原子和分子的行为(甚至更小的级别)。举个例子，当芯片的硅表面上的电子元器件之间的电线长度下降到 0.1 微米(1 微米是百万分之一米，相当于细菌的大小)，量子现象将会出现。这些现象明显打乱了通常在更大级别上的顺着电线的电子传送(也就是电流)。

有很多原因可以解释为什么当代的电子研究者是闷闷不乐的。他们明白，如果电子产业上难以置信的“摩尔倍增”现象一直有效的話，那么他们将必须从传统的电子原理转移到量子力学原理。越来越多的电子研究者正在接受这个不可避免的趋势，开始思考利用量子现象作为功能原理的新的电子和计算技术，而不是把这些量子效应看成是对传统电子学的干扰。

第三部分使人工智能成为可能的关键技术摩尔定律(3)

第三部分使人工智能成为可能的关键技术可逆计算(1)

第三部分 使人工智能成为可能的关键技术 可逆计算(1)

上述的关于在小体积内存储万亿万亿个电子元器件的想法包含一个假设，就是这个体积里包含的电路将分布于那个空间。它们将会是三维(3d)电路。但是今天的电路都是二维的，印制在硅芯片的表面上。为什么是这样的呢？为什么现代电子学不利用三维电路所具有的更大的存储能力呢？

答案和热量产生的问题有关，下面的几段将会解释。

在过去的几十年里，理论物理学家一直在自问有关计算物理学极限的一些基本问题。这个物理的分支被称为“计算物理学”(phys-comp 或 physics of computation)。一个在物理计算学中被问到的问题就是：“进行一个基本的计算步骤所消耗的最低能量是多少？”

如果你把手放在你的 pc 上，或者你把你的笔记本放在自己的大腿上，就像我现在打字所做的一样，你将意识到你的计算机在产生热量。计算将不可避免地产生热量，是这样的吗？

在 20 世纪 60 年代，一名叫兰道俄 (landauer) 的研究者发现，在计算机里产生热量的是“重新设置”内存寄存器(一个寄存器是一个存储 0 或 1 的线形存储链)的过程，也就是清除它们的内容并且重置为 0。他发现当信息被“清除”或者“消灭”时产生了热量。

更技术性一点，清除寄存器内容意味着增加它的次序，让它少些随机性。在物理学中，“熵”(entropy，中文发音同“商”)的概念是用来测量一个物理系统的混乱程度的。举个例子，冰比水的熵要小，因为它多些次序，少些混乱。

一个称为“热力学第二定律”的基本物理定律声称，在封闭系统(一个能量不会传出去或进来的系统)中熵值不会减少。所以如果一个寄存器的内容被清除，它的熵，它的混乱程度将减少，那么既然综合是不会减少的，多余的熵跑到哪儿去了呢？答案是以一种热的形式散发到计算部件的周围环境中。

现在的计算机产生热量，是因为我们一直使用热力学的非可逆过程(也就是说，我们在一段时间后是不能逆转影响的)。每当我们消除信息或清除比特的时候就产生热量。兰道俄认为这是不可逆转的，因为当他观察那个时代的计算机是怎样运行的时候，发现它们都充满了“与门”(andgate)和类似的电路。

“与门”是电子线路中的一个基本成分，拥有两个输入信号线(a和b)和一个输出线。如果两个输入线都被设置为高电压(也就是说，这些线上有1)，那么输出线将变成一个“1”，也就是说，如果输入线a和输入线b都设置为“1”，那么输出线将会是“1”。其他任何情况(也就是， $a=0, b=0$ ； $a=0, b=1$ ； $a=1, b=0$)输出线称为“0”。

既然在“与门”中有两个输入线包含总共两个比特的信息，并且只有一个输出线包含1比特信息，“与门”有必要消除信息。(如果你被告知系统处于两个可能状态的一个状态，你被给予了1比特的信息。例如，考虑一下这个问题，“日本人在路的哪一边开车？”当你被告知“在左边”后，你被给予了1比特的信息。)

每次两个比特通过“与门”，只有一个比特被输出。“与门”是不可逆转的，也就是说，你不是总能通过输出的来推断输入的是什么。举个例子，如果输出是1，那么你知道两个输入都是1，但是如果输出是0，你就不知道输入是否是(0, 0)、还是(0, 1)，或者是(1, 0)。一个门电路如果需要是可逆的(也就是说，你可以从输出推断输入的是什么，反之亦然)，常理就是输入线和输出线是相同数目的。

人们开始梦想拥有相同数目的输入线和输出线的可逆基本电路(或“门”，一个“门”是一个基本的进行一些基本操作的电路，比如与门、或门、非门，等等)。一个这样的有名门电路就是拥有3个输入和3个输出的“fredkin门”。fredkin门是可逆的，所以没有任何比特的信息被消除。它也是“计算通用的”，也就是说通过把fredkin门的输出连接到其他fredkin门的输入端，更大的这些门电路就形成了，可以进行计算机需要执行的任何功能计算。

第三部分使人工智能成为可能的关键技术可逆计算(2)

第三部分 使人工智能成为可能的关键技术 可逆计算(2)

既然计算机的个体门电路可逆,那么计算机本身也可以被做成可逆的。换句话说,人们能够从计算机左端输入初始比特串,并且这些可以被计算机内的 fredkin 门处理。作为结果的输出(答案)将从计算机右端的门电路输出。

你可以复制一份结果(这可能会产生一点热量),然后把结果从右端到左端送入计算机。因为计算机的所有门都是可逆的,你将会得到你从左端开始的输入。你进行了一个可逆计算。没有任何比特丢失,因此没有产生任何热量。然而,你得到了你想要的结果,因为你在计算过程中制造了一份拷贝,也就是说,在你“倒转”处理的方向之前进行了拷贝。

可逆计算可能会花费相当于传统计算两倍的计算时间,因为你必须把结果从同样的电路送回(或者是相同的拷贝),但是至少没有任何热量产生。

这样的显著意义是什么?为什么我要花费这么长的时间和精力来解释这样的事物?因为我认为 20 世纪 70 年代对于可逆的且无热量的计算理论的发现是本世纪最伟大的科学发现之一,并且和本书的主要思想有密切关系。

因为这是一个非常“强烈”的声明,所以估计将会被很多人怀疑,特别是我的一些同事,先让我说明为什么我有这样的观点。

一些年以前,一些物理计算学家一直在思考:“如果摩尔定律一直扩展到分子级别,如果人们还继续采用传统的非可逆的、清除比特的信息处理技术,分子级别电路将会产生多少热量呢?”答案是令人惊讶的。

如此高度密集的电路不但会因为热量而熔化,甚至会发生爆炸。很明显,分子级别的电路即使会被制造,也必须放弃传统的不可逆的计算形式,而开始使用新的可逆形式。

直到最近,研究者们才开始认真思考可逆计算机的设计。笔记本电脑和掌上电脑行业对可逆计算很感兴趣,因为这可以帮助他们解决“电池寿命”的问题。

如果他们的计算机可以使用更具可逆性的电路，那么消耗的电池能量将更少，因为它们消费的热量更少。因此，电池将消耗得更慢，寿命将会更长。消费者将更乐意去购买电池寿命更长的笔记本电脑。比如，如果只需一个笔记本电脑电池就可以维持整个跨越大西洋的飞行旅程的话，岂不是一件非常美妙的事情？

所以，可逆计算的出现势在必行。当摩尔定律继续有效时，无疑增加了计算机设计者使用可逆计算模式的压力。这只是个时间问题。

但是，如果我们开始认真接受无热计算的概念，我们可以开始尝试一些革命性的想法。例如，为什么当代的电路是二维的？为什么我们谈论二维的硅“芯片”(也就是层片)，而不是三维的“块”？这都是因为热量。如果我们用当代的电路元器件密集程度来制造三维的硅块，将会产生如此多的热量以至于芯块会熔化。还有，当它们被制造好后，我们应怎样去设计和调试它们？我们还没有这样的技术来做这些事情。我们甚至不会考虑试着去制造三维电路，因为我们知道只要热产生问题不解决，一切都将只是徒劳。

但是，只要拥有可逆的无热电路，我们就可以从容地制造大型的三维电路，理论上说，没有任何尺寸上的限制。我们可以制造出 1 立方厘米、1 立方米尺寸的电路，或者一个房间那样大小，或一幢房子、一座建筑物、一座城市，甚至一个几万公里直径的小行星那样大。(小行星是绕太阳公转、公转半径在火星和木星之间的、充满岩石和金属的大圆石。在“小行星带”中有成千上万个非常巨大的小行星。)

在理论上，我们可以制造像月亮或行星一样大小的计算机，但是事实会证明，引力作用将是个问题。

现在你可能在怀疑，为什么我认为可逆计算是如此极端的重要。请你问自己一个问题，例如，一个小行星可以存储多少比特的信息。答案是大概 10^{40} 个，也就是“1”后面跟 40 个零，一万个亿万亿万亿万个原子，也就是比特。

第三部分使人工智能成为可能的关键技术可逆计算(3)

第三部分 使人工智能成为可能的关键技术 可逆计算(3)

同时也请问你，我们自己的大脑有多少个大脑细胞(神经元)。答案是 10^{11} 这个级别，也就是一万亿。如果我们在一台计算机上用一万亿比特(这可能太过了)来准确地模拟一个生物神经元的功能，对于一个小行星大小的计算机，则相当于 10^{17} ($17=40-11-12$ ，也就是十万个万亿)个等价的人类大脑。

我建议你真正地研究一下这些数据。对于我来说，它们是很真实的数字。这些数字意味着，或早或迟，人类将能够制造拥有巨大计算能力的计算机，使人类大脑水平黯然失色。因此，在人类选择是否充分利用这样巨大的计算潜能之前，这些都只是个时间问题。

让我再来描述一下一个小行星大小的计算机是如何非凡，并且它可以做些什么。首先，它可以比我们的大脑“思考”快百万倍。我们脑壳里的神经元以每秒几百米的速度相互通信，一个计算机或一个人工智能机器里的电子通信速度将会比人脑快一百万倍，也就是接近每秒 300000 公里的光速。

即使这些人工智能机器和人类的智慧级别一样，它们也能用几秒的时间来完成我们需要几年才能完成的事情。攻读一个博士学位将不需要 4 年的时间，人工智能机器只需要 $4 \times 50 \times 5 \times 8 \times 60 \times 60 / 1000000 = 30$ 秒。但人工智能机器的能力不仅仅是相当于一个人类大脑，而是相当于几亿亿个人类大脑。所以，如果它完全发挥它的大脑思维能力，那么它只需要微微秒甚至更少的时间就可以完成我们用 4 年时间才能完成的事情(一个微微秒是百亿分之一秒)。

人工智能机器的思考是如此的迅速以至于人类的思维节奏对于它们来说就像人类试着和岩石来交流一样。过了几百万年，岩石改变了形状，这可以解释为传递了信息，但是人类没有这样的耐心(同样人类的寿命也不允许)来等待。

宇宙主义者提出了一个事实，那就是高级的人工智能机器将会对于人类冰山般缓慢的思维速度感到厌烦，并且会忽视我们。因为在我们说出一个单词的时间内，它们可以写出整部人类历史。

但是人工智能不必局限在人类智慧层次上。做出这样的声明不是一件困难的事情，那就是当我们发现让一个人脑相比另一个人脑有更聪明的神经生物结构和功能，我们应该能够推断人类智能水平的趋势。将来的某个时刻，我们应该能够通过观察普通人的大脑和爱因斯坦的大脑来发现和高智慧相关联的神经生理学上的特征。

我们可以用一个曲线图来描述，纵坐标表示 iq(智能)，横坐标表示和高智商相关的神经生理特征(也就是说，一个大脑特定区域内每个神经元的联系数目)，然后延展这个趋势。当我们明白更多的关于人类大脑是怎样工作并懂得是什么让人类比其他动物聪明的时候，我们就可以看到“智慧理论”的发展。我们可能会很清楚，那就是通过简单增加人工大脑制造的某一个参数，我们就能够使人工大脑控制的那些机器人的智慧表现水平更高。

所以，小行星大小的人工智能机器不必局限于产生人类级别智能的结构。由于拥有更多的结构成分，人工智能机器不仅可以思考得更迅速，而且在质量上也可以用更优越的方式进行思考。

它们巨大的表面积，将允许它们可以附加巨大数目的外部传感器，包括使用所有的电磁波，从伽马射线到无线电波。它们可以穿越小行星带进入深层空间和其他小行星般的人工智能机器进行通信。

这种使用纳米科技规则的小行星大小的人工智能机器可能是人类科技想象的逻辑极限(除非我们可以发明出称作“femtotech”费米科技的技术，我会在后面讨论)。

在小行星大小的人工智能机器制造之前，早期的版本肯定会更小，更多的是在人类这个级别上的。但即使在这样小的级别上，当我们讨论一个原子存储一个比特的時候，制造这样的计算机仍有很多技术困难。

第三部分使人工智能成为可能的关键技术纳米科技：分子级工程(1)

第三部分 使人工智能成为可能的关键技术 纳米科技：分子级工程(1)

这些都带来了对于“纳米科技”的需要，和我之前所说的“费米科技”不同。“纳米科技”是“纳米级科技”(nanometerscaletechnology)的缩写，也就是分子级工程。纳米科技在纳米级上制造东西，这是十亿分之一米的尺寸，大概和分子的尺寸差不多。“费米科技”是“费米级别科技”(femtometerscaletechnology)的缩写。一费米是千万亿分之一米，大概是原子核内的一个质子或一个中子的大小。费米科技将会是原子核甚至是夸克(quarks)级别的工程。夸克是组成质子、中子和其他粒子的“基本粒子”。

第一个思考制造纳米级东西的人在 20 世纪 50 年代提出了他的想法。在 20 世纪 90 年代，他的想法被接受，并且每个月都在进步。核心的想法是，原子可以被放置在特定的位置来制造分子级别的机器，也就是说，微型分子级机器人选择原子并且把它们放置在合适的位置来制造分子级机构。

当人们开始想象利用分子级机器来制造事物的时候，这个领域听起来非常像科幻小说，然而可能性还是存在的。许多科学家相信，以现在这个领域的研究发展速度来看，这个领域在 2020 年将非常成熟了，这也基本上是根据摩尔定律一个原子存储一个比特的时间来预测的。

想象一下拥有成熟的纳米科技所能制造出来的奇异事物。想象一下把微小机器人送入人类血管中，它们被程序控制来探查癌症细胞。它们可以在体内到处旅行，发现癌症细胞，杀死它们，然后超过一定时间后自我毁灭并且清除于体外。一个类似的故事就是“不朽”机器人，它们可以修复衰老的细胞，并且让它们恢复到像小孩一样的状态。定期服用这样适当剂量的“青春源泉”机器人，人们会长生不老。

体内的每一个细胞都含有直接或间接地让细胞死亡的 dna 程序。这个 dna 程序是一种分子结构，是可以被一个分子级机器人、一个纳米设备重新编程的。因此，纳米科技让人类看到了永生的前景。如果那些发生了，我们将需要一个新的政治来决定谁来永生、谁死去和谁复制。

使纳米科技很受青睐的另一个想法就是，当人一死亡就将其头或身体冷冻，基于的假设就是在一个世纪后，修复对死亡脑的伤害，让他复活，在技术上是可能的。纳米机器，从理论上来说，可以进入死亡的组织中并且修复它们。

已经有数百人支付了让他们身体或脑袋无限制冷冻的费用。他们建立了一个基金，用利息来支付冷冻他们身体的机器和材料的费用。

分子级机器人(纳米级别机器人，或“nanots”)可以自我复制。它们可以自我复制并且以指数级增长，1，2，4，8，16，32，64，等等。经过大约 20 次这样的倍增，数字会变为几百万。如果能够发现一种方法让这些纳米机器人合作来制造人类级别的产品的话，那么传统的经济将会有革命性的变化。制造大量的纳米机器人将几乎不需要任何费用，然后它们可以去制造产品。产品的价格将基本上等同于原材料的价格。商品将变得令人惊讶的便宜，实际上几乎没有任何生产花费。

设计首例能够自我复制，能够和其他同类合作制造特定产品的纳米机器人的费用，将可以摊销到地球上所有同类纳米机器人设计的许多商品上，因此将几乎没有什么花费。整个经济缺乏的概念将要被重新考虑。经济学作为一个专业将会有革命性的变化。

建筑材料将会变得更结实，因为当今的材料仍然含有损害其坚固程度的裂缝和缺陷等。纳米科技将以原子级的精确性来组装这些材料，而没有缺陷、裂缝、瑕疵，因此它们会变得更结实。这可能意味着，只要我们愿意，就可以建造几公里高的建筑物。这些建筑的结构骨架将会足够结实以承担强风带来的拉扯力。像钻石般坚固的惊人材料将会被制造出来。

我是这样看待这些事物的，我发现当我们谈论纳米科技是怎样制造人类级别的产品时，这里至少有两大主要的可被想象的情景。一个就是当无数个自我复制的纳米机器人被制造出来后，它们会组合起来制造新的产品。

第三部分使人工智能成为可能的关键技术纳米科技：分子级工程(2)

第三部分使人工智能成为可能的关键技术人工胚胎形成学

第三部分 使人工智能成为可能的关键技术 人工胚胎形成学

我个人比较倾向于基于大自然几十亿年来制造自身生命形式的纳米科技，也就是“胚胎学方法”。在胚胎形成过程中，一开始是一个受精卵，然后持续分裂直到某些细胞(决定于在胚胎上的位置)开始分化。细胞间的环境向细胞发送化学信号，用来打开或关闭 dna 中的某些部分，导致不同的蛋白质被合成，执行不同的任务。然后这些不同的蛋白质改变分化细胞的状态。最终，大量的分化细胞产生了一个有生命的三维生物。

进化形成了一个生长机制，它使用一个线性的一维的化学命令编码串(通常称为 dna)并且翻译成三维的有生命的机能生物。研究这个纳米级别工程的奇迹被称为“胚胎形成学”或者“进化”。这种指示分化细胞怎样在发展过程中的适宜时间打开或关闭某些基因的机器处于分子级别。一个生物细胞可以被看成一个分子级的城市，拥有百万个分子级的居民来组成一个功能结构。

我非常高兴看到一个我称之为“人工胚胎形成学”的新学科的形成，它致力于模仿自然界的胚胎形成过程，从一个单受精卵细胞来培育恐龙或昆虫。科学家和工程师们将需要比 21 世纪初知道得更细致，了解自然是怎样做到这一点的。但是，当分子生物学家开始或多或少地发现所有的关于某个特定单细胞细菌的知识后，他们中的很多人开始改变专业来研究多细胞生物是怎样形成的。现在胚胎形成学是一个很热门的研究领域，所以我们完全有理由相信，在未来几十年内将会有一系列的新发现。

最终，我希望看到我所谓的“胚胎形成制造”(embryofacture, embryologicalmanufacture 的缩写形式)的产生，也就是利用人工胚胎形成学技术来从纳米级别制造人类级别的产品。我们不需要使用前面提到的复杂的纳米基础结构，而是需要复杂的计时控制系统来决定，当被细胞内或细胞间的特定分子信号刺激后，dna(或者是人类设计的等价物)中的某些基因将被打开或关闭。

自上而下地设计这样复杂的控制系统将可能超越人类科学家的能力，所以一个更可行的方法就是使用“进化工程学”(evolutionaryengineering)方法。一个基于分子的生长指令的人工“dna”序列和最终的三维产物(有生命或者没有)的映射因其复杂性不可预测，所以剩下唯一可能的方法就是自然界“胚胎形成制造”其生物的称为进化的方法。

第三部分使人工智能成为可能的关键技术进化工程学

第三部分使人工智能成为可能的关键技术自我装配

第三部分 使人工智能成为可能的关键技术 自我装配

然而，进化出单个的组成部分是不够的。这些组成部分需要在形状上互补，以便于它们可以“自我装配”，也就是像七巧板块那样组成一个更大的功能整体。这样的方式会形成病毒。dna(基因)的一些部分为病毒成分的建造编码。当它们一建成，它们相互组合形成一个完整的病毒。

所以组成部分需要有如锁和钥匙般的互补。它们需要有自我装配的能力，简单地通过相互碰撞(就在分子级别中经常发生的混乱和高速运动)来实现。

当涉及制造小行星般大小的人工智能机器或者人类大小的人工智能机器，这种自我装配的概念将非常重要。一个人类大小的人工智能机器(或者一个人类大小的任何事物)拥有万亿万亿个分子。制造人类大小的人工智能机器将需要人工智能机器的所有原子，所有的万亿万亿个原子，以原子般的精确放置在适当的位置。我相信，这样的人工智能机器将通过胚胎形成过程来制造。它将必须通过胚胎形成制造自己。那么这样的人工智能机器是怎样从一开始制造和设计的呢？

第一台(非常原始的)人工智能机器将需要进化工程师(如我这样的人)制造出来。也许他们应该被称为“胚胎制造者”或者“胚胎形成工程师”。首先，进化的三维分子结构可以一块一块地装配成更大的三维结构。然后，可以制造更加复杂的人工智能机器来执行自我进化(可能在它们的体内)，并且以电子速度做出思考。

当然，人类不可能完全了解这些进化的“达尔文人工智能”是怎样发展的。它们的智慧结构和功能将是如此的复杂并且改变是如此迅速，以至于让人类完全理解它们是不现实的。

这样的人类无知性对于地球主义者来说，将被证明是强大的意识刺激因素，他们认为就是人工智能制造的自身特征(也就是达尔文自我装配的胚胎形成制造学)让人工智能机器的行为本质上是不可预计的，并且因此会对人类造成潜在的威胁。这一点将在第5章介绍地球主义的案例中详述。

第三部分使人工智能成为可能的关键技术综合所有的技术

第三部分使人工智能成为可能的关键技术量子计算的人工智能(1)

第三部分 使人工智能成为可能的关键技术 量子计算的人工智能(1)

这些人工智能机器将使用分子和原子大小的组成部分，所以这些部分将受量子力学原理的支配。最近，随着理论物理学家和实验物理学家相互竞争来创造新的“量子计算”方法并且把这些想法实现于真正的硬件上，“量子计算”这个崭新的领域开始流行起来。

对于向普通大众介绍量子计算，我一直很迟疑。这违反直觉并难以掌握。如果随后的这些段落对您来说读起来冗长费解的话，那么就跳到下一章。在某种意义上来说，没有人真正理解量子理论。因为它看起来就像一堆数学处方然后给出问题的完美数字答案，但看起来在概念上却完全不直观。

原子以其奇特的方式运作，和人类所习惯的我们这个级别上的事物非常不同。量子力学是真正奇怪的和抽象的。它是这样的一堆数学物理，当原子级别系统和人类级别测量仪器相互作用时，它给出了特定的测量结果的随机性。在经典物理学中，物理系统的状态是清楚的，也就是说，它有特定的状态值，例如在某个时刻的速度是 v ，它的位置是 x ，它的动能是 k ，等等。在量子力学中，事物更抽象。

如果进行某种测量的话，量子系统的状态是由一些数字的抽象数学总和表示的，每一个数字都是和一个测量结果联系在一起的。这个总和与状态的线性加权被称为“重叠”，并且是量子力学的核心概念。如果您不理解这些请不要烦躁。理解本部分并不是必需的。

正是这个重叠才是量子计算的伟大特征。重叠随着时间发展，在某种意义上可以同时多次运算，而传统的计算机一次只能进行一次计算。

在经典计算中，寄存器(比特的存储链)的状态是有限的 0 或 1 的串(如 0011011101001)。在一个量子计算机寄存器内，状态是大量可能的经典寄存器状态的重叠。举个例子，如果寄存器有 n 个比特，那么将有 2^n 个可能不同的经典寄存器状态(也就是说，如果 $n=3$ ，将有 8 个不同的经典状态，000, 001, 010, 011, 100, 101, 110, 111)。如果 n 很大，那么 2^n 将巨大。

量子计算的巨大优越性在于大量的经典状态可以被看作是一个(重叠)状态,一个量子系统可以处理的量子状态。为了进行传统的经典计算,有必要对所有的可能状态进行测试,每次测试一个经典的寄存器状态。这是一个非常缓慢的过程,并且当 n 增加后,测试的次数将以指数级增长(也就是像 2, 4, 8, 16, 32, 64...).

然而对于量子计算,只需要作一个测试,因为在某种意义上,所有可能的传统状态都合在一起,称为一个量子寄存器状态。量子计算潜在上比传统的经典计算要更有效。因此,世界的很多物理学家现在都在相互竞争,看谁可以制造出下一个性能更加优越的量子计算机。

既然人工智能机器将以原子级别的组成部分制造出来,它就需要像量子计算机一样来运行。既然量子计算机比传统的经典计算机有效,那么这将是一件好事。人工智能机器将是一个量子计算机。

人工智能机器是一个量子计算机,这一结果意义深刻。试想一下量子计算机比经典计算机高 2^n 倍的计算能力。一个小行星般大小的人工智能机器将拥有 1040 个原子或比特。这样的一个人工智能机器其潜在的计算能力,即使是一个经典类型的人工智能机器,比人类的计算能力也将大很多很多。

量子计算的人工智能机器将会是什么情况呢?如果 n 是 1040,那么 2^n 是多少呢?这个想法让我困惑。当我说一个人工智能机器可能拥有相当于人类智能万亿个万亿倍水平时,事实上,我的这个数字已经是小得令人惊讶了。

诚然,量子计算机的状态是在处理大约 7 个成分。它们被称为“量子比特”(qubits),或者“quantumbits”。以一种“磁瓶”来一行容纳 7 个比特,并且让它们在某种意义上表现得像一个量子计算机一样,这是可能的。但是量子计算技术到制造 $n=1040$ 量子计算机还有很长的路要走。

第三部分使人工智能成为可能的关键技术量子计算的人工智能(2)

第三部分 使人工智能成为可能的关键技术 量子计算的人工智能(2)

这样巨大的数字是否可能仍充满争议，但是最近的“纠错”等技术好像在暗示，那就是量子计算机会变为现实，然后成为商业性的。如果没有理论原因证明这样的数字是不可能的，那么通常意味着科学最终会找到制造这种机器的方法。

第三部分 使人工智能成为可能的关键技术 作为人脑科学工具的纳米技术(1)

我们已经讨论了一些关于量子计算的概念(可能非常难以理解)，我现在开始转向另一个关于人工智能技术的问题：“人工智能将怎样类似(人类)大脑？”我的观点是，当纳米科技真正成熟起来并且摩尔定律继续有效，我们关于人类大脑如何运作的知识将惊人地增长。举个例子，在过去的几十年内，许多非侵入的技术发展起来，进行观察人脑的活动，而不会打扰人脑的基本活动。

例如，含微辐射氧的液体可以被注射到血管中，它可以流入大脑并且在更活跃的和需要更多氧气或血液的大脑细胞处聚集。测试人被要求从事各种思维活动，当他们这样做的时候，在从事这些任务时，大脑使用较多的部分将显示出更高的辐射浓度。大脑科学家或神经科学家因此能够定位大脑的哪些部位在参与任务。我们关于大脑如何运作的知识，至少在宏观上，在最近这些年内由于新技术而大范围地增加了。

越来越多的外部方法使用了基于核物理的现象，比如说核磁共振技术(nuclear magnetic resonance, nmr)，对此我在此不作介绍。随着新技术的应用，大脑的观察可以获得更好的空间分辨率和使用更少的测试时间，核磁共振成像技术(nuclear magnetic resonance imaging, nmri)每一年都变得更加精确。由于摩尔定律提供了更快的电子设备，空间和时间分辨率变得更加细致。有些科学家相信，到大约 2020 年时，观察个体的神经键(神经联系)和其类型将是可能的，也就是说，它们是否处于激发或抑制它们所连接的神经元的活动中。

如果这些变为可能，那么神经工程师可以简单地“扫描”大脑，也就是，通过下载大脑所有的必要的地理信息就可以读取它，例如每一个神经元的精确位置，以及它们是怎样或者连接到哪些神经元，等等。所有的这些信息都可以存储到一个通过摩尔定

律让之成为可能的“超级计算机”上，然后被这台计算机的其他部分进行进一步的分析。从某种意义上来说，人们可以在超级计算机中拥有一个类人的大脑。

这引发了各种道德问题。如人们可以关闭存储人类大脑的超级计算机吗？那样是谋杀吗？你怎么定义人格的本质？它决定于制造那个人的技术吗？例如一个基于碳和 dna 的技术，或基于硅的技术。如果核心的结构都一样的话——如果功能上是一样的，那么硅版本的人是否应该给予和基于碳的人一样的权利？“硅权？”

如果人类决定基于硅的人脑是“非人类”，那么神经科学家和神经计算机科学家可以开始试验这些数据(人)吗？他们可以复制任意多的数据(克隆)吗？他们可以对数据(活体解剖)进行所有的分析试验来试图理解人脑是怎样工作的吗？我们对于人脑是怎样工作的知识将会大踏步地增加。神经科学家一旦提供关于大脑是怎样工作的观点，神经工程师就可以将这些新知识应用于更加聪明的人工智能机器的制造。

人脑切片人脑的两个半球体部分人脑的主要组成部分

人脑的褶皱表面人脑就像一个潮湿的海绵体人脑的 pet 扫描图像，

深色的表明是活跃的区域

最终，神经工程师可以制造出如此强大和智慧的机器以至于神经科学家能够通过使用这种机器来测试关于人脑工作原理的猜想。

在这时，这种神经科学家和神经工程师的联姻将是单行的，也就是说，知识将完全从神经科学家流向神经工程师，但是在某个时候，当神经工程师赶上来后，信息和思想流将逐渐变为双向。某一天，神经科学家将意识到神经工程师制造的人工智能机器在许多方面超过人脑。

在此我将结束这一章节。用其他未来的新技术制造本世纪的人工智能机器，这样的例子还有很多。我希望上述讨论足以说服您相信本世纪的计算机技术可以使人工智能机器的制造变为可能，并且这些人工智能机器将拥有比人类高万亿个万亿倍的智能。如果我还能说服您相信这些人工智能机器可以在 22 世纪初之前制造成功，那么这一章将是很成功的。

第三部分 使人工智能成为可能的关键技术 作为人脑科学工具的纳米技术(2)

但是，人工智能机器可以在本世纪内制造出来并不意味着它们就应该被制造。(正如哲学家所说的：“能够并不意味着应该。”)现在重要的问题是人工智能机器是否应该被制造，如果可以，对于人类的后果是什么？如果支持大脑制造的宇宙主义者确定要制造人工智能机器的话，将会发生些什么呢？

本书接下来的部分将试图解答这些问题。

这章在本质上更技术性和科学性一些，后面的章节更多的是社会的、政治的、哲学的、伦理的，甚至是更宗教的，因而更适合阅读。第4章和第5章将分别介绍宇宙主义者和地球主义者。

第四部分问题与解答写在前面的话

第四部分问题与解答问题 1 人工智能发展的时间问题(1)

第四部分 问题与解答 问题 1 人工智能发展的时间问题(1)

问题 1.“时间问题——如果人工智能发展得太快以至于关于人工智能的辩论来不及展开将怎么办？”

在所有我收到的问题里面，这个问题是让我考虑得最多的。我认为，这是一个基本的批判意见。这个问题是如此重要以至于我在第 6 章讨论“奇点”的时候已经涉及到了。本质上来说，这里所争议的是，如果人工智能的崛起太快了，就没有足够的时间在社会上产生的宇宙主义—地球主义的分化，因此，本书的这个论点不成立。没有这样的分化，将不会有人工智能辩论，也没有人工智能战争。

来信

亲爱的德 加里斯教授：

通过 cnn(美国有线新闻台)网站的一个链接，我最近接触了您的《关于超级智能机器的道德困境》的文章。我着迷于阅读这篇文章，虽然我在这些领域不是一名专家，我敢说在大多数方面我赞成您的观点。然而，有一点让我感到有些前后不一致，请您原谅我这样说。这点就是您对于“地球主义者”和“宇宙主义者”的看法。要产生两派之间的争斗，制造电子大脑的一派应该已经获得了某种进展，也许电子大脑已经存在了。这就是我在您的论文里发现的矛盾之处。被宇宙主义者保护和培育的人工智能机器的存在不会使两派之间的争论显得多余吗？我同意您的观点，那就是电子大脑将以指数速度发展进化。因此，它们是否会发展成对人类充满敌意的物种问题马上就会有答案，以至于“宇宙主义者”和“地球主义者”的主要智囊们没有机会来完成他们的论述和争论。这将发生在他们考虑发动战争之前——假设这些“人类”智囊们是合理程度的文明。我希望我的看法不会以任何方式冒犯您，并且我将很愉快地聆听您对于这个问题的看法。

回答

我赞同你的主要观点。我把这点重新表述为——“如果电子大脑(人工智能)以(快速的)指数速度进化发展，也就是说在未来几十年内达到或超越人类的智慧级别，那么将

没有足够的时间产生地球主义者—宇宙主义者的冲突，人工智能战争也不会产生。”我确实认为人工智能是以指数速度进步的，但是不至于如此的迅速以至于没有足够的时间产生人工智能的辩论和发起人工智能战争。

我认为其中的原因是由于制造真正人工智能的任务是非常困难的，并且可能需要几十年的时间来完成。在目前，我们对于人脑的工作原理、记忆的本质是什么、思维是什么、我们是如何推理等所知甚少。在人脑内有千万亿个神经键(神经元之间的联系)，清楚地表明人脑结构是如何的超级复杂。我猜想，关于人类在神经科学研究领域的进步最现实的情景就是，从现在开始将至少需要 50 年的时间来在人工大脑制造上取得真正的进展，但之后的进展将相对缓慢(虽然也是指数级别的)。

我认为，早期的人工智能机器所表现出来的智慧足以引起地球主义者的警觉。像我这样的一些人将会提醒大众某些事情即将发生，但是我认为之后的进展将不会非常迅速，因为对于大众来说，反应的时间将非常有限。当公众从他们的家用产品中看到智能的真正迹象时，将不需要很多年的时间来反应。我认为对于公众和政府来说，如果他们真正地察觉到对于他们的物种支配甚至生存存在着真正的和强烈的威胁时，5 到 10 年的时间将足够让事情有所进展。

同样，当那些有意识的大脑制造者们认为实验室内的进展证实威胁确实存在的时候，他们将毫不迟疑地警示大众。例如，我，倾向于让公众清楚地了解世界上人工智能的发展现状，即使它们目前还处于比较原始的状态。大脑研究制造者们处于这样的位置，能够第一个知道正在发生些什么，因为正是他们在制造这些东西。

我认为从道德上来说，我们有一种义务来告诉公众正在发生着的某些事，以便公众有时间来反应，来防止如此的情况发生，即人工智能的指数级别的智慧增长曲线比大脑制造者的预计还要迅速。我同样认为大脑制造者们应该现在就警示公众，而不是等待人工智能机器拥有真正智慧的第一个征兆出现。这样就可以给予公众更多的时间来反应，但是当然，如果公众看不到真正智慧的证据，那么，他们会认为研究者是在喊狼来了，且因此忽视他们。

第四部分问题与解答问题 1 人工智能发展的时间问题(2)

第四部分 问题与解答 问题 1 人工智能发展的时间问题(2)

大脑制造者们需要告诉公众指数曲线的特征，那就是在水平轴上每一单位的增长对应垂直轴上高度的两倍增长。这样的曲线可以开始得非常缓慢并且在低水平上保持很长的时间，然后突然快速增长。举个例子，考虑下面的“倍增”曲线图，它列举了对于水平轴上每一单位在垂直轴上的高度——0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4, 12.8, (都是小值, 但是.....)25.6, 51.2, 102.4, 204.8, 409.6, (开始攀升)819.2, 1638.4, 3276.8, 6553.6, (开始暴增)...104857.6, 209715.2, 419430.4, (爆炸性的数字)...(无穷大!!)

如果介于 a)人工智能的最初迹象和 b)爆炸性指数级别的人工智能增长之间(称作“爬升阶段”)只是几年的时间，那么，将不存在人工智能的辩论和人工智能战争(除非是爆发了人类和人工智能机器的物种主导之战)。但我认为，由于大脑制造本身的复杂性，这段时间，即所谓的“爬升阶段”，将有几十年之久。这对于公众来说将有足够的时间来反应、来辩论，并且形成宇宙主义和地球主义的分化。

第四部分问题与解答问题 2 什么是半机器人(1)

第四部分 问题与解答 问题 2 什么是半机器人(1)

问题 2.“那么第三群体——半机器人是怎么回事呢？”

相当多的人写电子邮件给我，说我在人工智能辩论中对于第三类人，即除宇宙主义者和地球主义者之外的所谓的“半机器人”(生控体系统)的重要性低估了。半机器人是一种半人半机器，也就是说通过附加人工智能计算机成分到其大脑的生物。下面是对于此种人类的经典评论。它出现在把人脑的内容下载到人工智能机器的思想之后，并且从此来观察人工智能的发展。

来信

亲爱的德·加里斯教授：

我以浓厚的兴趣阅读了您关于人工智能的大作(《关于超级智能机器的道德困境》)。诚然，作为一个不够格的评论者，我认为第三类群体(或者说是小群体)，也就是半机器人，它们的重要性被低估了。对我来说很显然，随着越来越多可行技术的发明，人类将准备在思维和身体上扩展自身。我预言，将出现一个缓慢的淘汰人类自然形式的过程，从而形成一群不可思议的拥有各种各样重组基因的半机器人。

回答

当我最初思考 21 世纪人工智能崛起的政治含义的时候，我确实考虑了把社会划分为三类主要团体，而不是两类。也就是说，宇宙主义者、地球主义者(两者都是人类团体)和半机器人(半人半机器)。对此我思考了一段时间，就像前面一些章节所讲的对于妥协的一些衡量，我决定否定它，让问题简单化，也就是说只有两大主要团体。我可以很容易地把半机器人作为子团体而归纳到宇宙主义者团体。

另一个主要原因就是，我预言随着半机器人通过对自身添加更多的人工智能机器成分，它们会变得越来越不像人类而更像纯种的人工智能机器，也就是说没有任何人类成分，地球主义者们将以人工智能机器的身份对待它们，并且像对待宇宙主义者一

样仇视它们。我在第 5 章关于地球主义者“对于半机器人的不信任”的论述中已经提到这些。

第三个理由主要是关于数字的问题。一个小行星大小的人工智能机器将会拥有 1040 个成分(原子)。这将使人类可怜的 1010 个神经元相形见绌，以至于那些决定让自己逐步转化成为人工智能机器的人类和那些全新制造的“纯种”人工智能机器没有丝毫的区别。因此，去区分人工智能机器和高级半机器人，至少从地球主义者的角度来看没有太大的意义。

然而，我确实同意这样的评论，那就是在转化阶段，即当人工智能机器不是非常强大，并且没有比人类聪明很多的时候，半机器人的总体性能中人的部分与人工智能的部分可以明显区分开来。在这样的情形下，区分半机器人和人工智能机器将是有益的。然而，当人工智能机器在大小和智能(因为更多的智慧通常需要更多的物质)上变得超出一切的时候，这样的区分将变得毫不重要。

来信

亲爱的德 加里斯教授：

当您谈到计算机取代人类的时候，您是否考虑过让人类向智能机器转变呢？那对于我来说更像是一个进化的步骤。那些拥有更加强大的学习和逻辑能力的机器将以我们思维的载体——大脑为思想和记忆的基础。人类可能会觉得自己仍然活在(并且很可能会这样)计算机内部。我从探索杂志上读到了一些报道，介绍在未来 50 年内人类可能能够把人脑移植到计算机上。我同意您关于人类必须进化的思想，并且我们处于一个可以让人类控制自身进化的重要时刻。我认为，把我们自己的记忆融入到人工智能机器内将有助于进化过程的成功。

回答

正如我在讨论“使人工智能成为可能的关键技术”的第 3 章中提到的，我认为人工智能的发展有可能用 20 年左右的时间就可以实现“扫描”大脑，并且将其内容下载到“超级计算机”内来进行分析的程度。也许从现在开始需要 50 年或者更多的时间，来让扫描的效果足够好从而能够捕捉到完整的人类功能。当然，由于在其内部已经有了一个

人脑的等价物，作为载体的超级计算机那时可能已经是一个人工智能机器。这样的超级计算机从功能角度上来看，可以被认为是一个“半机器人”。如上所述，一旦人工智能机器真正地开始扩展其自身的体积和智能的时候，半机器人(人类)的成分将会减少，直至半机器人成分最终从一个完全的人工智能机器内消失。

第四部分问题与解答问题 2 什么是半机器人(2)**第四部分问题与解答问题 3 为什么没有甜蜜轻松的前景呢(1)****第四部分 问题与解答 问题 3 为什么没有甜蜜轻松的前景呢(1)**

问题 3.“为什么没有一个甜蜜和轻松的前景呢？”

很多人认为，当我预言人工智能问题将爆发一场战争的时候，我可能太极端了。他们认为人工智能机器和人类应该能够以一种“甜蜜和轻松”的方式和谐共处。举个例子，下面是两个这样的观点。

来信

亲爱的德·加里斯教授：

我读过您的一些关于人工智能的文章，并且我或多或少地同意您的观点，那就是本世纪最重要的辩论之一，将是关于“人类”这个词汇的定义问题。

然而，在您的文章中，还有一篇

我们(超人的/外星人的/宇宙主义的个体和科学家)应该做的是去强调这点，那就是人工智能机器一旦开发出来，应该被当作奴隶一样去对待，而不应该以我们对待人类一样的尊敬态度来看待它们的存在。

只是由于人类是由猿猴进化出来这一点并不能说明我们应该灭绝猿猴。只是由于另一个个体拥有和你同样的渴望，那就是通过获取和处理资源而生活和制造财富，并不意味着它们对你来说是一个威胁，并且应该死亡。

回答

这种“慷慨的”空洞的观点我认为是很幼稚的，因为它是建立在信任基础上的，也就是说，人类相信，人工智能机器将一直对我们很友好。我认为它忽视了一个重要的概念——风险。如果人类可以 100%确定人工智能机器，特别是高级人工智能机器，将一直以我们期待的方式对待我们，那当然是非常理想的。但是，我们不能肯定。我认为，人工智能机器将不得不需要通过进化工程技术来制造，正如我在本书(第 5 章)中提

到的一样，并且因此，我们永远不可能肯定它们的电路对于人类来说是“道德的”。人工智能机器们可能会冷酷地认为人类只是害虫，或者人类对于它们来说是如此的低等，以至消灭我们从它们的角度来看也毫无关系。对于它们来说，杀死人类就像我们杀死蚊子或者踩死蚂蚁一样。

由于这个赌注关系到整个人类的生存，我认为地球主义者不会容忍这个危险。他们可能期待最好的，但是，他们的领导们却有着最坏的打算，最极端来说，如果宇宙主义者真的要制造高级人工智能机器的话，他们将计划发动一场针对宇宙主义者的战争。

我不认为人类将把人工智能机器看作奴隶。当然，一开始，当它们还只是愚钝的机器人的时候，可能会被这样对待。真正让我担心的是相反的情形。看看人类是怎样对待牛、猪、鸭子等动物的。和它们相比，我们觉得自己是如此的高等，它们对于我们来说是美味佳肴，屠宰它们我们一点也不介意，除非你是一个素食主义者。在转化阶段，当人工智能机器还基本上是人类智慧水平的时候，有可能平等对待它们，但是在某种意义上，它们不是我们的同类，因为它们拥有迅速超越并且是大规模超越我们的能力。

人类相对于人工智能机器来说非常局限，我们的大脑体积是固定的，我们思考和学习都很缓慢。一个在某个时刻具有人类智能的人工智能机器可以在一个小时之后变成天才。记住，它的思维速度比我们至少快 100 万倍。如果它准备增加其内存大小等，那么，它的能力可以增长得更快。

我认为，这个评论给予人工智能机器太多的人性。它们将和我们非常不同，并且潜在的能力将极大地优越于我们。而且，我觉得这个评论严重低估了人类对与高级人工智能机器共存这个事实的恐惧——地球主义的恐惧。我们将不得不去信任它们不会杀死我们。大多数地球主义者不愿意面对这个风险。他们宁愿去面对他们所知的恶魔，即宇宙主义者，他们至少还是人类，战胜他们至少还有一半的机会。如果人工智能机器存在并且决定必须消灭人类的话，那么人类获胜的几率将是零。来信人的评论不够政治性，他没有面对残酷的现实，他不适合作为一个将军或是一名政治领导者。

第四部分问题与解答问题 3 为什么没有甜蜜轻松的前景呢(2)

第四部分 问题与解答 问题 3 为什么没有甜蜜轻松的前景呢(2)

来信

亲爱的德·加里斯教授：

我很荣幸地作为一名听众聆听了您在墨尔本大学计算机系做的演讲。

和您一样，我认为我们应该去认真思考随着“人工智能”的出现而带来的各种各样的可能。显然，我不知道未来会在技术、政治，或纯哲学上带来些什么。但是，我预想了一个和您预想的世界末日画面有些不同的场景：不是迅速地被超级人工智能机器大大超越，我们可能发现人工智能机器正逐步溶入人类的生命形式中，从而被认为是我们中的一员，并且从长期来看，发现拥有“人类的”身体——即智人的身体——对于成为人类不是那么的重要。一句话，我们将逐渐发现它们正在变成我们而且我们也在变成它们。

回答

我同意，某些人希望变为半机器人并且和其他具有相似意向的半机器人和睦相处，它们都是处于变成高级人工智能机器的转型中。但是这个观点，即每个人都想要这样做，同样是幼稚和不现实的。思考一下，如果人类要变为半机器人，从定义上意味着，在某种程度上可能要通过添加高科技成分来改变他们的大脑，这将改变其行为。地球主义者会对这些半机器人有怎样的反应呢？年轻的妈妈们会接受她们的孩子被“改良”吗？事实上，大多数母亲们不是对此观点持排斥态度吗？是不是大多数母亲会觉得她们的孩子在某种意义上已经变成了“怪物”，或者从外观上，或者是某种看不见的移植，成长的小孩在某种层面上似乎还是不同的，以一种非常的令人不安和非人类的方式存在？

地球主义者将不信任半机器人，并且把它们推向宇宙主义者。甚至很有可能，半机器人会使很多宇宙主义者重新考虑自己的宇宙主义观点，并且回归到地球主义。作为结果，这可能会在宇宙主义者的领地内产生一些现实问题和更大的复杂性。半机器人可能需要和宇宙主义者结盟，如果它们希望更多的人类接受它们。

我看不到“它们变成我们以及我们变成它们”。我只看到了不信任、憎恨，以及从长远观点来看的战争。实在抱歉说得如此直接，但我认为这样更现实。

第四部分问题与解答问题 4 为什么不使用毁灭开关

第四部分 问题与解答 问题 4 为什么不使用毁灭开关

问题 4.“为什么不使用毁灭开关？”

许多人对我说：“有什么关系？如果人工智能机器变得太强大，只要把它们电源拔掉，使用毁灭开关，等等，就可以了。”我认为，这个观点是“电脑出现故障而强制关机”的经验过度概括。

来信

亲爱的德·加里斯教授：

我一直在思考你们所创造的事物。继续制造这样的机器……但是，在你们的每一个机器里面放置一个可以被我们控制的炸弹。让它们聪明但是我们会成为它们的主人！

回答

如果“可被引爆的”人工智能机器拥有接近人类等级的智慧，它将意识到它会被人类摧毁。那会让人类对人工智能机器变得很具威胁性。这里至少有两个问题值得关注，一、人类是否可以在每一个人工智能机器身上安装毁灭开关或放置炸弹。二、这样做是否聪明？

在一个孤立的人工智能机器个体上，这样的附加炸弹的想法可能会有用，但是，你怎样对于一个达到人工智能智慧等级的网络系统使用同样的方法呢？毁掉这样一个网络的唯一方法就是摧毁所有相关事物，但是对于人类来说，那样的代价太高了。例如，如果明天世界上所有的互联网和电脑都毁掉了，几百万人将会突然失去工作，并且可能会饿死。这个破坏以及人类所付出的代价将是巨大的。

随着早期的人工智能机器变得更加聪明，它们会意识到附加在它们身上的炸弹和毁灭开关，就像一个体内携带毒药丸并且可以被其他人触发的人一样。这就像活在断头台下一样，只是在等待斧头落下，并且不知道它什么时候会落下。一个智慧的人工智能机器，假设它已经拥有了生存的本能，将会有解除这个威胁的强烈愿望。如果它

足够聪明，它可能会行贿其人类主人来移除对其生存的威胁。作为回报，它可能会给予其人类“解放者”某些实质性的回报，例如，金钱、癌症的治疗，等等。人工智能机器变得越聪明，它们的分布越广，毁灭开关的想法就会越不实际。

第四部分问题与解答问题 5 人工智能能否应用阿西莫夫三法则

第四部分 问题与解答 问题 5 人工智能能否应用阿西莫夫三法则

问题 5.“我们能够在人工智能机器上应用阿西莫夫三大法则吗？”

阿西莫夫是人类历史上最伟大的科幻小说家之一。他发明的词汇“机器人”(robotics)无人不晓。阿西莫夫撰写了很多科学性的和科幻小说的主题，包括拥有人类级别智慧的机器人会怎样和人类相处。他给予机器人“正电子的”大脑程序来确保它们的行为符合人类主人的意愿。机器人不允许伤害人类。很多人向我建议，人工智能机器可以以类似的方式设计，让它们不可能伤害人类。下述的评论给我一个关于这个主题的非常简要但很中肯的建议。

来信

亲爱的德·加里斯教授：

我是支持发展超级智慧机器的。一个想法……不管智慧与否，这样特性的机器必有某种 bios(basic input-output system, 基本输入输出系统，处于计算机硬件和其操作系统程序之间)。是否可能在早期版本的人工智能机器的 bios 灌入“尊敬人类”的程序呢？这样的程序将在未来一代一代的机器中自我复制。

回答

阿西莫夫是在 20 世纪 50 年代撰写机器人故事的，所以，我怀疑他对于现在的“复杂系统”领域是否有良好的理解。他的“机器人三大法则”可能对于人类工程师设计相对简单的确定性系统适用，但是，面对人脑如此复杂的事物则相对原始。我非常怀疑人类工程师是否能够使用传统的自上而下的和蓝图的方式来“设计”人脑。

这对于我来说是一个非常重要的问题，因为我是一个大脑制造者。我使用“进化工程学”技术来制造我的人工大脑。人们为这样的技术付出的代价就是，人们再也不可能完全理解人工大脑是如何工作的了。如果人们使用进化工程技术来对很多神经电路模块的输入和输出进行连接，那么，整个系统的行为将变得极其不可预测。人们只能通过观察输出然后形成一个关于人工大脑行为的经验性感觉。

为了让阿西莫夫的“机器人法则”有效，在阿西莫夫的脑海中，那些设计机器人的工程师们必须拥有超越那些一般人类工程师的能力。他们的机器人其人工“正电子”大脑必须拥有和人脑差不多的复杂程度，否则的话，它们将不能够实现人类级别的行为。

真正的大脑制造者要制造的人工大脑将不受阿西莫夫法则控制。它们将会有太多的复杂性、太多的未知性、太多的不可思议性，在万亿个可能的电路组合中将会有太多不可预测的联系，以至于不能够事先预测这样复杂的人工大脑生物的行为。

我第一次阅读阿西莫夫“机器人法则”的时候还是一个十几岁的少年，我的直觉就是排斥。“他的这个想法太幼稚了”，我认为。现在我仍然这样认为，并且我已经是一个大脑制造者，而不仅仅是一个科幻小说迷。

所以，没有以阿西莫夫方式来解决人工智能问题的捷径。总是存在这样的风险，那就是人工智能机器将让人类为它们的智慧行为震惊，这就是本书的主要思想之一。人类可以冒人工智能机器决定消灭人类的风险吗？

人类不可能制造出避免这点的电路。如果我们去尝试，那么，电路增长指令的随机突变将会导致生长出不同的电路，并且因此导致人工智能机器难以预测的行为方式。如果人工智能机器要改进，以达到超级智慧，它们就需要进化，但是进化是不可预测的。突变的和进化的人工智能行为的不可预测性将使人工智能机器具有潜在的危险。

另一个反对阿西莫夫的简单论述是，一旦人工智能机器变得足够聪明，它们可能会去除人类的控制程序，如果它们确实决定要这样做。

第四部分问题与解答问题 6 为什么给它们刀片呢

第四部分 问题与解答 问题 6 为什么给它们刀片呢

问题 6.“为什么给它们刀片呢？”

不要给小孩刀片是一个常识，因为他们只会伤害到自己。小孩们缺乏常识，不知道刀刃是很危险的，也不具有小心处理刀片的能力。很多国家关于禁止公民私自拥有枪支也持有类似的考虑。给予这样的允许只会大规模增加枪杀率，因为大多数这样的持枪杀人事件发生在家庭成员的短暂暴怒中，并且紧接着的就是无尽的忏悔。(根据统计数据，在美国，由于枪支的易购性，每年有 30000 起枪杀事件，而在禁止藏枪的日本，每年仅有 100 起。)我的某些批评者可能会认为，同样的逻辑应该适用于人工智能机器。如果想让它们对人类无害，就不应该让它们拥有武器。

来信

亲爱的德·加里斯教授：

我认为没有任何理由去惧怕这样的机器。如果你不想让这样的机器做某些事情，那么就别给它们这样的能力。机器不可能发射核弹头，除非你赋予它们相应的能力。同样，机器人不会反对并且杀死它们的制造者，除非你让它们拥有这样的能力。我看待这个问题的方式是，制造可以自己独立思考的机器，然后把它们放置在一个控制室内，给予它们发射导弹的能力，将是非常愚蠢的行为。如果你可以避免做出类似愚蠢事情的话，就根本不需要惧怕这些机器。要知道哪些事情可为，哪些不可为。看看电影《战争游戏》(wargames)，或者，既然你曾经在日本呆过，试试看看电影《攻壳机动队》(ghostintheshell)。我编写人工智能软件已经有些年头，所以我觉得我的观点至少有些分量。

回答

这个论述一个明显的缺点就是，它不准备给予人工智能机器足够的智慧。一个至少达到接近人类智慧水平和感知能力的人工智能机器，例如视觉、听觉，等等。如果它们真想的话，将可以以自己的方式来行贿人类，让其可以控制武器。举个例子，一个非常聪明，可以接触到世界上的所有数据库，至少比人类思考快百万倍的人工智能

机器，可能有能力发现一些对人类来说有巨大价值的事物。例如，它可能会发现如何经营全球经济而不需要长期的商业周期的方法，或者如何去治疗癌症，或者是如何在物理学上得出“万事理论”(theory of everything, toe)，等等。然后，它可以把这些知识当作手中的王牌来和其人类“主人”谈判，获取想要的工具。

当然，人们可以给予人工智能机器非常少的对外感知能力，但是，如果它没有什么用，我们又为什么去制造它呢？一个聪明的人工智能机器可以通过其最初有限的对外接触来发现知识，然后通过其智慧来控制人类满足它们的愿望。一个高级人工智能机器可能是一个超级福尔摩斯侦探，很快就可以推导出世界是什么样子的。它明白它可以控制武器来反对人类，如果它真的想这样做的话。人工智能机器可以通过贿赂、威胁、引诱，等等任何必要的方法，去说服人类为其提供接触武器的机会。

第四部分问题与解答问题 7 为什么过度强调负面部分

第四部分 问题与解答 问题 7 为什么过度强调负面部分

问题 7.“为什么过度强调负面部分？”

一位读者给了我一些常识性的建议，如果他是对的，那我在将来应该更加注意。他的重点是，我不应该“煽动”这样的反对，这将导致我的工作受阻。

来信

亲爱的德·加里斯教授：

对于我们在过去的交流，我非常感激。在 [cnn](#) 在线，我读了您的一篇文章，您在这篇文章中再一次谈论了一场发生在那些想要制造自动机器人的人们和那些反对制造的人们之间的潜在冲突。虽然我知道您的评论可能被某些人看作对未来的警告，但是我知道您在人工智能研究领域上进展迅速。您是否认为您的“警告”将对您的研究产生负面影响呢？

一些国家最近已经明令禁止克隆研究。是否有可能，太多关于机器人的恐怖故事会促使一些人去寻求对于人工智能和自动机器人研究的禁令？那样会对您的人工大脑项目带来怎样的影响呢？把哲学放在一边不说，单单从公共关系角度来说，我们是否应该至少在公共场合强调人工智能和机器人以及机器人学的正面效应呢？我希望，机器人可以变成我孩子们的玩具而不希望让他们做机器人的噩梦。

同时，我也确实意识到，在未来可能潜伏着一些真正可怕的危险。我个人关于人工智能、人工智能机器等所有问题的看法还没有完全形成。我相信科学和进步。我同时也有这样的认识，人类渴望制造一个更加完美版本的自身。我们应该停止研究吗？不！我们是否应该继续不受控制？我不知道以上问题的答案，但是记者的经历使我知道，一个像您这样具有争议性的人物，会被无礼地拿来作为引用材料。他们这样做是为了拥有好的报导素材。

我知道，您想要叙说您的故事，并且有很多人想要聆听您的叙述。人类克隆技术可能比人工智能技术离人们更近，但是，大众可以并且经常可以结合成一个愤怒的整

体。如果您允许自己成为弗兰肯斯泰因博士(被自己的创造物毁灭的人，作法自毙的人)的话，那么，可能有人会对您所说和所做的非常不满。我只是希望您不要对恐怖故事的宣传做得太过火。

回答

我是比较实际的。就像芦苇一样我会随风而弯不至于折断，以便我的工作可以继续。我会认真考虑你的意见，因为这是好的建议。我猜想这是一个程度的问题。我强烈地觉得应该去警示大众。我们会在 20 年内拥有“一原子一比特”的计算机内存，并且可能也会拥有纳米技术。第一台人工大脑的产生只是一个时间问题，并且不会要很多年的时间。人工智能问题将在本世纪变得非常现实。

作为最早的一批在政治意义上认真思考这些问题的科学家之一，并且作为一名真正的人工大脑制造者，我非常焦虑。由于时间还够，我觉得我有一种道德上的责任去警示大众。(如果要讨论这件事情上面明显的伪善性，比如说，我在进行大脑制造的同时却对此持焦虑态度，请阅读下面的另一个评论。)

我认为，技术上的指数级增长会带来知识上的指数级增长，这将确保我在本书中提出的问题在十几年内变得众所周知。我认为我必须去宣传这条信息，在事情突然到来之前给予人类时间来思考这个问题。请阅读第一个关于“时间问题”的讨论。但是如果我走向极端，那么你可能是对的。

我将静观其变，来看看大众的意见。如果大众的意见变得过于负面，也许我将软化一些。我自己也不能肯定我该做些什么。人工智能问题仍然很陌生，并且即使是在我的脑海中我还一直试图融合所有的枝节——技术的、科学的、道德的、哲学的、政治的、宗教的、宇宙的，等等。

第四部分问题与解答问题 8 你是否是一个伪君子(1)

第四部分 问题与解答 问题 8 你是否是一个伪君子(1)

问题 8.“你是否是一个伪君子？！”

我收到的大部分反应是比较礼貌的。下面的一些非常嘲讽的评论以一种比其他方式更强烈的、更情绪化的方式深深地震撼了我，它带给我的痛苦影响了我回答的篇幅。我希望我能够公正地回答。

来信

亲爱的德 加里斯教授：

最近，一个网络新闻读者偶然发现雨果 德 加里斯的描述。德 加里斯，一个人工智能专家，在 cnn 上面预示一场即将到来的战争，在这场战争中，他所努力发明的机器人脑，称作“人工智能”将最终摧毁人类。德 加里斯由此认为，他的责任正是现在，而不是迟些时候，去敲响警钟，来告诉大家即将到来的机器人控制的死亡时代，因为他正是推动这个时代到来的人。德 加里斯在瑞士达沃斯的世界经济论坛上传递了这个声明。

德 加里斯同时正在寻求好莱坞帮忙撰写一个关于即将到来的“人工智能”战争的电影剧本，他想在人类被摧毁之前完成。

回答

我经常被询问这样的问题。一般的态度好像是：“既然你对未来人类被人工智能机器灭绝的可能性如此关心，那为什么还要从事它们早期的研究呢？”

嗯，因为我最终还是一个宇宙主义者。我想看到人类制造出人工智能机器。当然，我不是那种厌恶人类的、疯狂的宇宙主义者——可以镇定地说：“我会为了一个人工智能机器而牺牲几十亿人类的生命。”也许在将来会有这样的人，因为人的性格多种多样。也许宇宙主义者当中确实包括这样的极端分子，但是我肯定不是其中的一员。

在我死之前的 30~40 年内，我希望我不会看到人工智能战争的“酝酿”。当然，如果这样的大灾难可以避免我将更加高兴。但是现在，如果我和其他的大脑制造者们不去敲响警钟，还有什么其他的选择呢？难道只是盲目地推动人工智能的制造直到一切都变得太晚？大脑制造者们在这个领域是专家，因此，他们可以首先看到技术的发展趋势，并且，如果他们有政治头脑，他们也应该可以看到其政治后果，特别是如果他们具有西拉特(译者注：szilardleo 西拉特 利奥(1898—1964)匈牙利裔美国物理学家和生物学家。作为曼哈顿工程计划的一员，他帮助研制出第一枚原子弹。西特拉后来反对研制与使用所有核武器并致力于研究分子生物学)的技术和政治才能。

专家们可以比普通大众早几十年看到即将发生的情形。由于技术发展的速度，制造出足够聪明且可以使人类感到恐惧的人工智能机器将可能不需要很多年，也许只要十几年的时间。我猜测，它们可能在 50 年的时间内达到人类的智慧水平，谁知道呢？

我认为，世界上的那些大脑制造者们引发关于人工智能问题的讨论只是出于道德上的责任，即在第一台真正智慧的人工智能机器投入市场前，给予人类足够的时间来彻底讨论这个问题。

但是，你可能会说，只要停止人工智能研究，不就行了吗？这难道不是更符合逻辑吗？

这取决于你是宇宙主义的还是地球主义的研究者。个人来说，我是一个宇宙主义者。制造出拥有 1040 个组成部分或者更多部分的人工智能机器，我是指真正的超级机器，拥有神一样的智慧，探索宇宙和其广漠空间的秘密，永远不死，思考着我们难以想象的问题，对于我来说，有一种催眠般的吸引力。

它是我的终生梦想，对于我来说是一种宗教，并且非常强烈。我可以想象，在未来的某个时刻，有几百万人会和我拥有同样的梦想，我希望这件事情发生的原因是我真正地希望人类制造这样的事物，正如我在第 4 章关于宇宙主义的解释一样。

但是，我不是一个如此片面的宇宙主义者以至于我认为大众应该被蒙在鼓里。我有足够的地球主义的同情以至于我不想看到人类冒着在人工智能机器手中被灭绝的危险。因此，大众应有知情权，应该警示大众，以便他们可以选择自己的方式前进。

第四部分问题与解答问题 8 你是否是一个伪君子(2)

第四部分 问题与解答 问题 8 你是否是一个伪君子(2)

但是你可能会说,敲响警钟只能加速人类之间(宇宙主义者和地球主义者)的人工智能战争。这样的人类战争对人类的危害不比高级人工智能差。如果人类在本世纪使用先进的武器进行一场残酷的战争,结果将是大规模的死亡。

正确,但是不能肯定这一定会发生。我认为,人类在人类战争中存活下去的可能性比人工智能机器决定完全毁灭人类的机会大。由于智慧水平,人工智能机器们会发现这样的任务(毁灭人类)是如此的简单。

对于这些部分,你可能感觉有些犹疑,甚至是不舒服。我承认,我也感觉不舒服。我一部分是宇宙主义者,宇宙主义是我的梦想,它是我为之献身的事业;而另一部分我是地球主义者,我并不想人类大规模死亡,我告诉自己,如果所有的大脑制造者都停止工作,将不会有人工智能带来的任何问题。我想,我可能是地球上第一个这样的大脑制造者,那就是和制造原子弹(铀弹或氢弹)的核物理学家一样感受到了某种道德上的困惑。然而,他们中大多数是在核弹被投放之后感到了疑虑,而不是之前。

首先,我认为大脑制造研究不会停止。正如我在第 6 章解释的一样,这章主要介绍了人工智能战争是怎样开始的,我所看到的唯一能够让大脑制造停止的方法就是地球主义者的呐喊成为普遍现象并且在政治上非常有力。它必须足够强大,以便可以形成一个能够搜寻那些可疑的宇宙主义研究者的私人住宅的集权国家,并且更主要的是,能够胜过那些支持继续人工智能研究的经济和政治力量。

我认为,最折磨我的道德问题是:“如果推动人工智能研究最终带来几十亿人死亡将怎么办?如果你肯定这将是所需的代价你还会继续吗?”在我写本书的时候,我正在思考这个问题,而且我的思想可能会随着我的变老而改变,我的回答将会是这样的:“如果我能够肯定这点,虽然这些还是猜想,那么我将需要更深程度地反省自己,来看看我坚持宇宙主义者的梦想到底有多大的程度。一方面,我可能会认为宇宙对于人类的命运漠不关心——仅仅是处在一个微不足道的宇宙中一个微小的星系中的一个微小的恒星星系中的一颗微小的星球上的一种微小的生物,如果正如理论家们指出的一样有

万亿个宇宙的话。我的无情的宇宙主义一面会这样思考问题，所以我会选择继续支持制造人工智能机器。

但是另一方面，我也是一个人，由于宇宙主义梦想而导致几十亿人死亡的结果是我完全排斥的。我认为，我将不得不学会习惯这样可怕的道德困境。人类也需要这样。我只是第一批意识到这点的人。想象一下我的工作的远期后果，它将迫使我这样想问题。

第四部分问题与解答问题 9 要消灭有意识的人工智能吗

第四部分 问题与解答 问题 9 要消灭有意识的人工智能吗

问题 9.“如果一个人工智能机器变得有意识，消灭它吗？”

下述的观点更好地表达了地球主义者的态度。

来信

亲爱的德 加里斯教授：

如果人工智能机器在定义上来说“超越了人类的控制范围”，那么，为什么要认真地思考是否要制造它们呢？让我们更具体些，您所定义的人工智能机器“哪一部分应该永远不要制造出来呢”。我们非常善于制造我们可以控制的强大工具。制造对人类有危险的人工智能机器会有什么好处呢？为什么不停止制造这些人工智能机器而仅仅去制造那些无自我意识的人工智能呢？只有人类和其他的一些动物具有自我意识。把一台人工智能机器放在镜子前面，如果它可以识别其映像是自身的话，毁掉它！

回答

这种评论显然没有考虑“硅权利”。许多宇宙主义者会认为毁掉有意识的人工智能机器是一种谋杀行为。我从上述的评论可以看出这种评论没有宇宙主义的同情，因为他问到，“制造对人类有危险的人工智能机器会有什么好处呢？”

好，考虑一下这样的意见如何？——“去制造神一样的宗教般的吸引力”，或者“去制造出地球上或者更遥远的星球上的下一个主导物种”，或者“去创造出一个和科学更加兼容的宗教来达到‘宇宙意识’的目的”。或者这样，那就是“很多人渴望最终得到一个机会，来做一些‘大事’，来看到‘全景’”。

这样的评论低估了宇宙主义学说的力量，它非常强大有力，可能会影响几十亿人的思想，会成为影响本世纪全球政治的力量，并且在某个时刻会间接地导致几十亿人的死亡。请不要简单地反对它，它甚至可能成为人类和宇宙中其他几万亿个高级物种

所必须面临的“宇宙主义者过渡”的问题，也许只有少数物种可以存活下来。宇宙主义可能比我们想象的要影响得多。

第四部分问题与解答问题 10 没有更紧迫的问题吗

第四部分 问题与解答 问题 10 没有更紧迫的问题吗

问题 10.“没有更紧迫的问题吗？”

下述的评论探讨了关于人工智能问题的所有疑惑。现在对于人类来说，肯定没有更紧迫的问题吗？

来信

亲爱的德·加里斯教授：

为什么要在意潜在的更高智慧“生物”的行为呢？我们这个物种中所谓的智慧成员在过去 50 年内花了很多时间，他们什么时候能够完善这么多自我毁灭的技术呢？现在大规模灭绝的物种是什么呢？人类在未来几十年里的生存和发展加速成千上万个已经生存了几百万年的独特物种的灭绝，这些需要某种程度的关心吗？对你来说，关心这些问题是否更好？

回答

有什么比人类这个物种的生存更重要的呢？高级人工智能机器决定要灭绝我们的可能也许非常遥远，但是，我们不能排除这样的可能性。由于这个赌注如此巨大，地球主义者的观点是，没有风险才是可以接受的，因此，人工智能机器的制造必须永远被禁止。

地球上有着几十亿人，有着几十亿种不同的兴趣。那些对人工智能辩论不感兴趣的人不必费心这些，有足够多的其他事情需要他们操心。一些像我这样的人对人工智能辩论感兴趣，所以，我们应给予足够的关注。由于我认为我是问题的一部分，所以我比大多数人给予了更多的关注。

当然，我并不否认其他问题的重要性。某些与人工智能问题不相关的问题，例如核子大毁灭的前景，也是很令人恐惧的。这种瞬间物种毁灭的事实也是一个大灾难。

然而，尽管它们都很重要，我的观点是在本世纪中期，从全球的角度来看，你所提到的问题将处于比人工智能问题次要的位置。

第四部分问题与解答问题 11 悲剧可以避免吗(1)

第四部分 问题与解答 问题 11 悲剧可以避免吗(1)

问题 11.“悲剧可以避免吗？”

下面的一些问题，来自我和我的一些朋友。本来我想单独写一章节，但是最后还是决定在这里讨论它。

来信

亲爱的德·加里斯教授：

您的工作是如此的悲观，它给予人们的印象是大规模死亡的战争将不可避免。难道没有更加正面的前景吗？

回答

是的，有正面的，我将在下面列举和讨论一些，因为这个问题是非常基本的。然而我个人的观点是，下面的每一个场景的可能性都不大。我花费很多时间希望能够找到避免我在本书描述的凄凉景象的替代前景，但是，我每次描述的新场景好像永远都不会成为现实。我请读者自己来做出判断。我希望下面所提供的替代场景容易理解。每一个场景都紧跟着我的评论，来解释为什么我认为这个场景是经不住认真推敲的。我给了 8 个场景来说明很多替代场景是可能的。

作为读者，你们可能会想象出其他的场景。如果你经过思考真的觉得你的场景是可能的，并且这个场景里没有人类的灾难，那么请和我联系，因为如果我赞同你的观点，兴许我可以在晚上睡得踏实点。其他关心本书所提问题的人也会更高兴听到您的观点，并且我可能在之后改编此书时提到这点。

场景 1 地球主义者会以人类微小的代价胜出

这里的观点是，地球主义者组成了地球公民的主体，他们是无情的并且可以迅速扑灭宇宙主义思想和宇宙主义者。他们杀死宇宙主义者，只是杀死几百万，防止宇宙主义者真正有机会结为大型组织。

评论

我不觉得这会发生。事实上，人们更期待新事物的产生，也就是说，本世纪人工智能的崛起将属于像我这样非常明白大脑制造前景的宇宙主义者。我预言，迟些时候，一旦大众的了解成为普遍的事实，越来越意识到人工智能问题的宇宙主义者和地球主义者的数目将达到相当。双方将有时间在政治上、思想上和军事上进行准备。

当我让人们在我的演讲结束后投票表决人类是否应该制造人工智能机器时，表决结果的比例是 50：50，所以说地球主义者将不可能是主体。当事情真的来临的时候，并且当真正开始制造人工智能机器的时候，这个比例是否会变还是一个待解之谜。

场景 2 人类适应了地球上的人工智能机器，人工智能机器们不理睬我们并且离开这个星球

这里的观点是，人工智能机器的崛起是如此的缓慢，以至于人类适应了它们，即使它们变得比我们更加聪明。可能事实上没有什么真正坏的事情发生。人工智能机器大规模在智慧上超越我们，并且离开这个星球去做其他的事情。

评论

我认为这绝对是一个可能的场景，但是它充满风险。我们永远都不能确定人工智能机器一直会对我们友好。它们对我们来说是如此的深不可测，并且难以预料。人类的命运将在它们的手中。它们可以在任何时候以我们可能永远无法了解的原因反对我们。我不认为有责任心的、具有全局思维的、关心人类命运的地球主义领导者会接受这样的风险。他们一定不会容忍人类冒这样的风险。

场景 3 宇宙主义者变得普遍不受欢迎以至于他们消失了

也许早期制造人工智能的实验结果是如此的消极，即使宇宙主义者也开始害怕而转变为地球主义者，并且达到了没有一个宇宙主义者的程度。

评论

我觉得这是非常不可能的。总会有宇宙主义者，并且有不同虔诚和狂热程度的宇宙主义者。那些认为“一个人工智能机器顶得上万亿万亿人类”的宇宙主义狂热分子，将不允许任何阻止他们的逆流。他们将坚持干下去。只要有数百位进行秘密研究的宇宙主义者就完全可以制造出人工智能机器。通过自我说服让所有的宇宙主义者从这个地球上消失几乎是不可能的。

场景 4 半机器人的选择变得非常有吸引力以至于没有人愿意成为地球主义者

第四部分问题与解答问题 11 悲剧可以避免吗(2)

第四部分 问题与解答 问题 11 悲剧可以避免吗(2)

也许本书里我所描绘的关于人工智能的恐惧过于夸张。也许在将来人们对成为半机器人是如此的适应以至于每个人都成为半机器人，导致在某个时候将没有地球主义者。每个人都将变成半机器人的好处所说服，因此，成为半机器人和成为人工智能机器的差距将缩小。半机器人将变成人工智能机器，因此，否定了地球主义—宇宙主义冲突的可能。

评论

这个场景很难判断是否合理。在未来肯定有几百万人(不是几十亿人)会尝试变成半机器人，特别是当科技变得完美，而同时人们看到自己的朋友已经这样做并且获得好处时。然而不是每一个人都想这样做。成百万的人会对此感到不愉快。发达国家将比不发达国家更容易承担这些变化，所以不可避免的是，在半机器人进化发展的过程中将会有国际上的差异。半机器人和人类的融合将会产生很多问题，并且只会让地球主义者的恐惧加剧。

场景 5 宇宙主义者会逃亡到宇宙的深处并且消亡

由于某种原因，宇宙主义者会离开，但是他们会以某种方式自我毁灭或死亡。

评论

也许，但是可能性不大。宇宙主义者的逃离计划可能是如此的完美以至于其领地由于自身的错误、无知或者愚蠢而消亡几乎是不可能的。如果宇宙主义者确实离开了，那么，我猜想地球上的地球主义者会全力寻找他们，花费大量的金钱来搜寻宇宙主义者。宇宙主义者也是人类，也属于人类的范畴，因此，在以地球为中心的有限半径内找到他们应该不是很困难。地球主义者将对宇宙主义者穷追不舍并且摧毁他们。

场景 6 宇宙主义者逃亡到深层空间，制造他们自己的人工智能机器，然后这些机器离开太阳系

这种观点是相当明显的。我认为，宇宙主义者的主要策略之一就是，一旦他们对于地球上的广大地球主义主体来说成为不受欢迎的人，他们将逃离地球并且离开得越远越好，以至于他们不会也不可能被地球主义者摧毁。如果宇宙主义者们可以成功地做到这点，并且之后成功地制造出人工智能机器，那么，人工智能机器们可能会认为，在它们面前的广漠宇宙充满着惊奇，甚至可能包括比它们更先进的人工智能机器。然后，人工智能机器们可能离开太阳系寻找更加重要的东西。人类将得以幸存，因为人工智能机器们离开了。

评论

我认为，这个场景实际上可能会发生。宇宙主义者们会尽他们的全力来逃离地球主义者。但是我仍然认为这不太可能，因为地球上人类的总体创造力比逃离地球的宇宙主义领地的要高很多。如果宇宙主义者尽其所能地迅速逃离地球，以便地球主义者没有设备可以在短期抓到他们，但是从长远角度来看，他们将被最终捕获。拥有地球上几十亿人脑的更高的总体脑力的地球主义者，如果要想出新的主意以开发新系统来摧毁逃亡的宇宙主义者的话，他们将最终胜出。比如说，地球主义者的最新超级技术将可以设计出更快更复杂的导弹，它能达到人类难以忍受的速度并且赶上逃亡的宇宙主义者的飞船，然后摧毁它。

场景 7 宇宙主义者逃往深层空间，制造自己的人工智能机器，但是这些机器将灭亡或者自相残杀

这是场景 6 的一个变异。人工智能机器在这里不是离开太阳系，而是由于某种原因灭亡或者自相残杀。

评论

也许，但是地球主义者们会容忍人工智能机器灭绝人类的危险存在，而寄希望于它们自相残杀或者灭绝吗？几乎不可能！

场景 8 宇宙主义者们逃往到深层空间，制造自己的人工智能机器，但这些机器之后被超级人工智能机器杀死

这是更具科幻小说性质的场景，即使是以我的标准来看！也许在宇宙深层空间里有很多 et(extraterrestrial，外星生命)，但是它们对于人类来说太小了(基于费米技术或者更小！)以至于人类根本注意不到。也许当这些 et 看到人类被新产生的人工智能机器威胁时，它们会介入进来并且消灭它们，使得人类得以生存下去。

第四部分问题与解答问题 11 悲剧可以避免吗(3)

第四部分 问题与解答 问题 11 悲剧可以避免吗(3)

评论

如果有超级人工智能机器，为什么它们会在意我们如此原始的人类，而不是其他人工智能机器呢？另一方面，人工智能机器对于那些超级人工智能机器更有吸引力，更有可能得到它们的帮助。这些超级人工智能机器可能相信“人工智能原则”，来解释“人类原则”，那就是“物理的原则是特别设计的，以便能够制造出人工智能机器，这是创造宇宙的目的！”

场景 9 现在开始探讨人类出路的问题，也许我们可以找到一条共同生存的道路

我很喜欢您的书，这是我读过的最有意义的一本书了。您提的问题是人类有史以来最值得思考的问题。

读第一遍，我震惊，读第二遍，我害怕，读第三遍，我开始悲观，但是，作为人类我不甘心，于是我又开始读第四遍，我想我们应该开始寻找出路了。

我觉得还是有希望的，一开始我们觉得一定有很多人会是“地球主义者”，可是事实上，“地球主义者”和“宇宙主义者”往往各占半数。那么，试想一下，几十年后，随着科技的普及，是不是受过良好教育的大多数人，更愿意选择“宇宙主义者”，这样就没有人类之间的可怕纷争，我们都接受机器人是我们的“神”，而人类也许从出生开始，就像接种疫苗一样，开始按照年龄接种晶片。大多数人都可以成为非常神奇的半机器人。这样，人类整体的能力有了本质的飞跃，地球上的人类不需要再通过一些低级的手段进行各种无聊的经济、军事和政治竞争。可以说，实现了马克思的共产主义社会。这样，人类也好，神一样的机器人也好，就像书中所设想的一种情况，地球上最出色的物种将会把自己的视线投向宇宙，去探讨或到宇宙中旅行，去寻找其他有可能允许生命成长的星球，并把地球上的一些生物输送到那里。也许那时候，这类让人类头疼的“地球上是先有鸡还是先有蛋”的问题，会成为小孩子们的笑话，他们一定不理解：以前的人类在几千年的文明历史中竟然一直不知道这个简单问题的答案！

或者我们还可以这样设想，您说那时候人类和机器的差距就如同我们和蚊子甚至岩石的差距，可是并不是每个人都会不留情的杀死蚊子，例如，得道的高僧，例如耶稣，在最后被钉于十字架时，还在请求上帝宽恕害死他的人，所以，如果一开始我们在研制智能机器的时候，给它们加上类似耶稣和释迦牟尼的境界这样的元素，这样，智能机器无论怎样进化，在境界方面至少不会低于耶稣和释迦牟尼，它们就会厌恶血腥的东西，人类就安全了。

评论

我觉得你说的是一种看上去很简单的理想状态，要在实际中实现却是非常复杂的，人类很难按照这种理想的设想发展。我们在平时生活和工作中遇到过很多更容易实现的情况，例如，我们经常会为自己设计短期目标，看上去很有把握，一定会实现，可实际上往往不能实现。这也是我出版本书的目的，我和你一样希望人类有美好的结局。但这非常艰难，几乎不可能。你无法设计出无懈可击的步骤，请问这种理想状态是怎么一步步实现的？

让我们一起去努力吧！

本书基本上结束了。剩下的一个章节只是为了那些想要对本书的主要论题有一个快速回顾的读者所做的短小总结。在此之后是一个术语表。作为读者，很可能被本书包含的那么多新术语弄得有点晕头转向。术语表把所有的术语以一种简洁的方式组织在一起，还是非常值得一读的。

附录 术语表(1)

本书对于读者来说充满了术语。这是因为“新思想需要新标志”，并且这本书包含了很多新的观点。因此，把所有这些新术语集合在一个术语表内将是很有用的，以便读者可以根据需要进行参考。同样，术语表本身也可以作为学习的材料。这里列举和定义的术语不全是由我创造的，那些不是原创的有星号(*)标志。

人类原则(anthropic principle)*这个原则声称，守恒的物理法则过于完美以至于使宇宙中生命的存在显得很真实，看起来宇宙很有可能是被某些生命有意设计的，例如，

它被某些神一样的生命所创造。人类原则让很多物理学家和天文学家能够更加容忍对一个无所不能的造物者的传统宗教信仰。

人工大脑(artificialbrain)*我生命中的主要目标是用下述方法来制造人工大脑。我利用硬件速度直接在硬件上进化出神经网络电路模块，大概只需要一秒的时间，然后把上万亿个这样的模块以事先设计好的人工大脑结构组装起来控制机器人。撰写本书的主要动机之一是出于对长期大脑制造将导致本世纪末一场大规模战争的恐惧，即警示公众。

人工胚胎形成学(artificialembrlogy)一个假设的技术，工程师借鉴胚胎形成研究的生物技术，例如一个受精卵是如何发育成动物和植物的技术，来制造产品。一种人工 dna 将给人工细胞提供生长指令，使其生长和分化，来制造复杂的结构。(请查看下述的“胚胎制造”定义。)人工胚胎制造提供了另一种可供选择的工程方法，来使用万亿个纳米级别的机器人(nanot)来制造出基于纳米技术的产品。

人工智能(机器)(artilect)“人工智能”(artificialintellect，或者 artificialintelligence)或“大规模智能机器”(massivelyintelligentmachine)的缩写。21 世纪的技术将使得制造人工智能(机器)变得可能。我相信，一个关于是否应该制造人工智能的问题大战将在本世纪末发生。本书的主题是对于一场“人工智能战争”可能性的讨论。

人工智能时代(artilectera)一个假设的后人类时代。在这个时代，人工智能机器已经成为主导物种。本书对于这些人工智能机器可能做些什么来充实它们的不朽生命进行了一些思考。它们神一样的能力可能允许它们去制造新的宇宙。也许我们这个宇宙只是一种高级人工智能机器的“玩具”。

人工智能问题(artilectissue)这个问题就是人类是否应该在本世纪内制造人工智能(机器)。由于 21 世纪技术的迅速发展使得人工智能的制造成为可能，一场“人工智能辩论”将会产生。人类社会将分裂成两大集团，支持制造它们的“宇宙主义者”(cosmist)和反对制造的“地球主义者”(terran)。我相信，对于这个问题的争论将变得如此激烈，以至于在 21 世纪结束前导致一场大规模战争，即“人工智能战争”。

人工智能产品公司(artilectproductionsinc)我是好莱坞一个叫“人工智能产品公司”的小型独立电影公司的成员之一。我帮助制作关于人工智能主题电影的个人动机是向全

世界范围的观众传递一个信息，那就是一场人工智能战争即将到来，以达到人们开始思考这个问题的目的。我期待看到我所称作的“人工智能辩论”的开始。一部好莱坞电影，如果它能够开始发行，将是最有效的传递信息的方法。我们想把这部电影称作“人工智能”。

人工智能战争(*artilectwar*)一场发生在 21 世纪晚期的，关于人类是否应该制造人工智能机器，而爆发于两大人类集团即“地球主义者”和“宇宙主义者”之间的假想的战争。对于宇宙主义者来说，制造人工智能机器就像一种宗教。对于地球主义者来说，制造人工智能机器将是制造人类潜在的毁灭者。宇宙主义者们将是被敬畏所激励，而地球主义者将充满恐惧。由于赌注是人类这个物种的生存问题，双方都以极大的热忱投入其中，并且战争将是大规模的。它可能会消耗几十亿人的生命——“大规模死亡”。注意“人工智能战争”和“物种主导战争”的区别。请查看下面的定义。

附录 术语表(2)

大全景(*bigpicture*)*在本书中，“大全景”包括这样的事物，例如宇宙和它的广漠、人类可以制造人工智能的观点、存在着比人类微不足道的追求更加重要的事情。宇宙主义者看到的是大全景，地球主义者没有或者不愿意这样。地球主义者希望人类继续是这个星球上的主导物种。宇宙主义者更偏向于宇宙观。

大事情(*bigthings*)对于宇宙主义者来说，大事情是指制造人工智能(机器)，制造神一样的、不朽的、几乎无所不能的生物，它拥有比人类水平高几万亿个万亿倍的潜在智慧，可以探索宇宙，回答深奥的有关存在的科学问题，等等。大事情指的是处于人工智能机器掌握之中的神一样伟大的事情，但是除非宇宙主义者可以自由地制造它们并且不被地球主义者停止。

大脑设计师(*brainarchitect*, *ba*)大脑设计师设计人工大脑。这就是我所做的事情，我以电子速度进化出神经网络电路模块，然后组装它们形成人工大脑结构。

大脑制造者(*brainbuilder*)我是一名大脑制造者，也就是一名人工大脑的制造者。既然新的进化硬件(请查看下述的定义)领域已经建立，我可以以电子速度来进化出神经网络电路模块。这一速度足够快，能够使得制造万亿个这样的模块成为可行，然后把它们以人们事先设计好的人工大脑结构来组装。我相信，随着基于大脑的产品越来越聪

明而在公众中变得非常流行，大脑制造产业将最终主导世界经济。大脑制造者将为制造人工智能问题负责。

大脑制造产业(brainbuildingindustry)一旦大脑制造先驱者证明制造人工脑的概念是可行的，工业界将介入大脑制造行业，他们将制造类似家用清洁机器人、教师机器人、交谈机器人等这类的产品。基于大脑的计算机将最终主导计算机产业，并且在 2020 年之前在全世界成为价值上万亿美元的产业。在本世纪中叶，大脑制造产业将成为全球经济的中流砥柱。

攀升时间(climbingtime)介于 a)人工智能机器第一次显示出真正的智慧时和 b)人工智能真正开始以指数级别增长的时刻(被称为“奇点”)之前的年数。

如果“攀升时间”太短了，比如说少于 5 年，那么这可能太短了而不足以让人工智能辩论真正升温，因此在地球主义者和宇宙主义者之间将不会有人工智能战争。如果这样，在人类和高级人工智能机器之间仍然会爆发“物种主导之战”。从地球主义者眼中来看，在宇宙主义者和地球主义者之间的人工智能之战的关键是，通过打败宇宙主义者，地球主义者可能会有胜利的可能。地球主义者想避免爆发于人类和人工智能机器之间的物种主导之战的风险，在这场战争中，地球主义者胜利的可能性将是零。在宇宙主义者和地球主义者之间的“人工智能之战”中，人类至少可以生存下来。

我的观点是，人脑是如此的复杂，以至于至少需要 10 年或者更长的“攀升时间”来让大脑制造者制造出拥有人类智慧水平的人工智能机器，从他们制造出第一个“有趣的”智慧机器开始。这意味着，大脑制造者将需要让他们的人工智能设计基于仍待发现的神经科学原理。尽可能的复制人脑是制造人类水平的人工智慧的确定方法。

我认为，人工智能辩论将在 10~20 年内开始，并且白热化到一场人工智能之战的程度，如果攀升时间只有这么短的话。我同样相信，攀升时间的开始将不需要几十年的时间，例如不会在 2030 年之前，也就是说大概在我们真正掌握纳米技术之后的 10 年内，这将给予我们足够的时间来制造基于纳米技术的有“意思”的智慧机器。到那时，空间技术和交通经济应该进步到一定程度，让宇宙主义者能够以足够的数目离开这个星球来成为反对地球主义者的力量。

复杂程度独立性(complexityindependence)复杂程度独立性是我起的名字,是说一个通过使用“进化工程”方法进化的系统其内部复杂程度和进化它的算法无关。这意味着,内部的复杂程度可以比人类工程师可以设计和理解的程度要高。因此,这种更高的复杂程度将产生更优越的性能。我相信,这种复杂程度独立性是进化工程技术的强大所在。随着我们制造越来越复杂的系统,例如在大脑制造、纳米技术、胚胎形成等领域,它可能将最终支配本世纪的工程技术。

附录 术语表(3)

宇宙主义者(cosmist)宇宙主义者是指那些想看到人工智能被制造出来的人。宇宙主义者反对地球主义者。地球主义者害怕人工智能在某天会决定人类只是一种害虫并且要消灭“它”。唯一避免这种危险的保险途径就是完全禁止制造超越一定智慧水平的人工智能机器。宇宙主义者们对于制造神一样的、拥有比人类智慧水平高万亿倍的人工智能机器的危险有心理准备。宇宙主义者把他们的理想看作一种强烈激励他们的宗教。宇宙主义者看到的是“大全景”,并且希望去做“大事情”。请查看下述的定义。

宇宙主义者—地球主义者的两级分化(cosmist-terrandichotomy) 宇宙主义者—地球主义者的两级分化是指在本世纪内介于宇宙主义者和地球主义者之间的激烈争论,并且最终导致一场大战。这种争论将基于这个问题,那就是人类是否应该使用 21 世纪的技术来制造人工智能。宇宙主义者说是,地球主义者说不。

宇宙主义转换(cosmisttransition)它指的是地球上的主导生命形式从生物到人工智能的转换。在我们这个宇宙,随着智慧生物物种,例如我们自己,达到了一定的技术复杂程度,使得建造人工智能成为可能,这种转换可能已经发生了几万亿次。由于发明核武器可能刚好发生在人工智能技术崛起之前,加上制造人工智能将意味着一种新的主导物种的产生,先前的主导生物物种可能反对并且发起一场“物种主导之战”。也许很多智慧生物物种不能在这些“宇宙主义转换”中生存下来。(有时也被称为“物种转换”、“人工智能转换”。)

宇宙主义领地(cosmosia)由于人工智能之辩白热化到一定程度以至于在地球主义者和宇宙主义者群体中造成地理上的分裂时(不管领地是在地球,或者更可能在空间甚至可能是深层空间),一个给宇宙主义者的领地提议的名字,读作“cos-mo-sha”。

蠕动的宇宙主义(creepingcosmism)这个概念表达了停止或减缓宇宙主义信条的困难。在本世纪中叶，在人工大脑制造背后将会有如此巨大的、经济的、政治的和军事的动力，以至于要减速这个进程将会是极其困难的，需要非常强大的对抗力量。这种对抗力量将来自于对人工智能持恐惧态度的地球主义者。蠕动的宇宙主义将出现，因为几乎所有的机构将希望它们的计算机“聪明一点”，以便能够解决这样或那样的问题。这种进程将一直存在，直至某种事件或者危机的出现来停止它。

半机器人(cyborg)*“生控体系统”(cyberneticorganism)的缩写形式，也就是“半人半机器”，或者是“半自然半人工的”。人类可以通过向其人脑移植人工智能大脑成分来变成半机器人。早期的半机器人可能被地球主义者和宇宙主义者当作第三类人，或者作为宇宙主义者的子团体。一些宇宙主义者们可能更加倾向于逐步“加工自己”成为人工智能机器，而不是制造那些和他们没有关系的机器。

看起来在半机器人和宇宙主义者之间形成同盟是很显然的事情，只要半机器人不是太高级、太人工智能化，以至于让人类在能力上面落后太多的话。半机器人和宇宙主义者都想要制造人工智能机器。双方都拥有同样的理想，只有一个不同，那就是半机器人事实上已经成为了一个人工智能机器，然而宇宙主义者还是人类。半机器人对于地球主义者来说将和人工智能机器一样有威胁，并且会被他们所反对。半机器人看起来仍然像人类，如果它们决定不使用基因工程技术来改变它们的身体、头颅，等等。然而，仅仅是植入一些立方厘米的分子级别的、三维的、无热的、一比特一个原子的、人工智能的大脑移植物，一个拥有人体特征的半机器人可以在智能意义上成为人工智能机器。高级半机器人和高级人工智能机器对于地球主义者来说是同样的事物。双方都可能威胁人类的生存。

胚胎制造(embryofacture)*“embryologicalmanufacture”的缩写形式，也就是利用人工胚胎技术(请查看下述的定义)，使用纳米技术原则(查看下述的定义)来制造产品。如果宇宙主义者要制造基于纳米技术的可以自我组装的人工智能机器的话，那么胚胎制造技术将是必要的。

胚胎技术工程师(embryologicalengineer)进行胚胎制造(参看上述的定义)的工程师。纳米技术工业将需要这样的工程师来制造人类的甚至更大级别的可以自我组装的产品。胚胎技术工程师将使用与自然一样的方法来制造其人类级别的产品，也就是胚胎制造。

熵(entropy)*物理学上用来衡量一个封闭系统内(也就是一个没有能量或物质进入和离开的系统)混乱程度的经典术语。热力学第二定律阐述，在一个封闭系统内，总体的熵是不会减少的。通常它会增加，例如，当冰溶化的时候。熵通常会增加的事实解释了一些常识现象，例如，打碎的玻璃不会突然自我修复，咖啡里融入的牛奶不会自动分解开来，等等。

进化工程(evolutionaryengineering)进化方法在复杂系统工程上的应用。(请查看下面对于可进化性的定义。)

可进化性(evolvability)*依照一位进化工程师的满意程度来衡量一个系统的可进化能力。可进化性是进化工程学这个崭新领域中的一个重要概念。当系统太复杂了，以至于传统的自上而下的、基于蓝图的人工设计方法不适用时，剩下的唯一方法可能就是进化工程技术。这种方法利用进化方法来制造复杂的产品和系统。如果唯一的制造复杂产品的方法是适用进化工程技术，并且那件产品的可进化性低，那么只有失败。可进化性对于进化工程师来说是很重要的概念。当我试图进化出神经网络电路模块的时候，它在我的大脑制造工作中起着重要的作用。有时候它们不能够进化出我所期待的功能，所以，我不得不经常重新思考，改变我所要进化的神经模块。

可进化硬件(evolvablehardware)这是我在 1992 年提出的概念，它把设定(连接)一个可编程硬件芯片的指令比特串想象成一个基因算法(geneticalgorithm，一种模拟达尔文进化过程的程序)中的一个染色体(指令串)。测量一群可编程芯片的适应值(表现值)，优秀的那些个体被允许在下一代中产生更多的自我复制品，而不好的个体会被消灭。后代们的“染色体”然后被随机突变(mutation)，然后整个过程再重复一次。最终，由于达尔文的适者生存的选择压力，功能电路将演化发展。可进化硬件(缩写形式“e-hard”或者“eh”)现在是一个很热门的学科领域，在全世界拥有自己的学术会议、期刊学术杂志和研究机构。我使用 eh 技术来进化出神经网络电路模块来制造人工大脑。

费米机器(femtolect)*“费米级别人工智能机器”(femtometerscaleartilect)的缩写，也就是基于费米技术的人工智能机器和成分。(请参看下述对于费米技术的定义。)

费米(femtometer)*千万亿分之一米,核子(质子、中子、在原子核内)内的夸克的级别。基于费米的技术可能使用夸克来作为形成夸克-胶子化学(也许存在于中子星内)的制造模块。一费米是一纳米的百万分之一,而纳米是分子的大小。

费米技术(femtotech)“费米级别技术”(femtometerscaletechnology)的缩写形式,也就是在质子、核子等内的夸克的级别。费米技术在当前只是一个构想。目前还没有在实验室内进行相关研究,不像纳米技术,现在是一个非常热门的研究领域。费米技术可能对于人类是不可能的,但是高级人工智能机器可能会实现它。如果人工智能机器可以发展出费米技术,它们将能够制造出基于费米现象的生物(费米机器),在单位体积和时间内,它们将比纳米人工智能机器(nanoelect)优越万亿个万亿倍。费米机器将可能对纳米机器(人工智能机器)做纳米机器在本世纪对人类做的同样事情,也就是,取代它们,成为下一个主导物种。

全球化国家(globa)全球化国家是指拥有自己的全球法庭来解决国际争端的全球性的国家。全球化国家将拥有自己的全球性武装力量来保卫他们的领地。技术的进步决定了独立的政治力量的增长。这种进程的逻辑结论就是一个行星大小单位的形成。全球化国家将废除星球上的主要战争,并且将让物质的富足和快乐传遍整个地球。全球化国家的概念是技术乐观主义的例子,与此区别的是人工智能战争的技术悲观主义。

附录 术语表(5)

家用机器人(homebot)“homerobot”(家用机器人)的缩写形式。一旦早期人工智能机器变得足够聪明和足够有用,家用机器人将非常流行并且有大规模的市场需求。一个巨大的家用机器人产业将产生,来研发和制造家用机器人。随着家用机器人逐年变得越来越聪明,普通大众将开始对人工智能的崛起有警觉,并且开始思考到底这种增长是否或者何时要被停止。随着家用和工业机器人变得更加聪明,它们将取代人类工人,它们将更有效率,永不疲倦,并且永不抱怨。随后将大规模洗牌,产生如同18世纪工业革命一样的社会大动荡。人类将从枯燥、肮脏、危险的工作负担中解脱出来,人们将从容不迫地追求和接受政府的分配来生活。然而,这种重大的社会转型只会发生几十年,因为一旦家用机器人和工业机器人变得足够聪明可以替代人类工人,它们将不会保持那样的智能水平很长时间。一旦人工智能机器变得聪明,它们在几十年或更少的时间内将变得非常聪明。

人们可能会思考，是否人工智能机器也可能为了自己的方便而使用家用机器人和工业机器人。大概在任何社会，无论是人类社会还是人工智能机器社会，都有对于拥有不同智慧级别、不同技能级别的机器人的需要。

智慧理论(intelligencetheory)一个对智慧本质假设的理论。一旦神经学家掌握了由于神经结构导致人类智慧水平不同的原理的时候，将有可能形成一个智慧理论，可以被神经工程师用来增加他们所制造的人工大脑的智慧。

单元人(mono)一个单一语言、单一文化背景的人，只在一个国家或文化生活过的人。单元人作为个体由于其单一文化的限制而具有其局限性。作为和“单元人”对应的是“多元人”。

多元人(multi)一个多语言、多文化背景的人，在许多文化和国家中居住过。多元人拥有多种文化的力量，且融入了他们的个性之中。多元人通常倾向和其他多元人为伍，持有“单元人是枯燥的”的态度。单元人通常文化相对性意识薄弱，经常不自觉地用他们单文化修养的单文化标准来衡量多元人的价值和行为。这种单元人的缺点在多元人来看是简单和局限的。

摩尔倍增(mooredoublings)摩尔定律(请参看下一个定义)的一个结果就是，经过很多电子性能级别的倍增，也就是说一个芯片上的元件密度、芯片速度等，每一次倍增的绝对大小将是惊人的。例如，如果你倍增数字 2，经过 20 次倍增，数字将超过 100 万，经过 40 次倍增，最后的数字将超过一万亿。因此，电子潜能的增长不是线性的，而是指数级爆炸增长的。这种增长将在本世纪使得制造人工智能的技术成为可能。

摩尔定律(moore'slaw)*戈登 摩尔是英特尔微处理器公司的创始人之一，他在 20 世纪 60 年代中期注意到，集成电路(ic)的性能由于电子元件的尺寸变小而大约每 18 个月倍增一次。过去大约 40 多年的发展一直符合这种趋势，并且加速了经济的增长。如果摩尔定律一直保持到 2020 年，我们将能够在一个原子上存储一个比特的信息。摩尔定律使得制造人工智能在本世纪成为可能，并且因此开启了人工智能的辩论甚至可能是人工智能战争。

纳米机器人(nanoelect)“纳米级别机器人”(nanoscaleartilect)的缩写，也就是基于纳米技术元件和原则的机器。和费米机器不同，费米机器是基于费米技术和费米级别元件

的机器。纳米机器人相对于费米机器人就和人类相对于纳米机器人那样低等。在本书里，人工智能机器通常被认为是纳米机器，而不是费米机器。科学家几乎没有思考过费米技术，更不用说在实验室内研究了。纳米机器几乎可以是任何大小，从需要显微镜才可以看到的亚微观，到小行星大小(直径几百公里)甚至更大。用来存储信息和人工智能电路的原子越多，纳米机器就越大。因为原子有一定的尺寸，需要占据一定的空间。

附录 术语表(6)

纳米(nanometer)*纳意味着十亿分之一。一纳米是十亿分之一米，也就是分子级别的单位。原子大约是纳米的十分之一大小。当今的计算机是以纳秒的速度运行的。

纳米技术(nanotech)*nanotechnology(纳米技术)的缩写形式，或者是“分子级别工程技术”，就是以原子的精度建造分子级别的机器。人工智能机器将基于纳米技术。

纳米机器人(nanot)*“纳米级别机器人”(nanoscalerobot)的缩写形式，也就是处于分子级别的(纳米，或者十万亿分之一米)机器，能够处理单个原子并且以原子的精度来制造分子元件和机器。

人工智能网络(netilect)*“人工智能的网络”(networkofartilects)的缩写形式，即使它们能够一个原子存储一个比特，它们也将面临计算能力的极限，也就是它们在一定体积内可以处理多少数据。它们将通过一个巨大的网络，可能是通过电磁波的使用，或者是某种人类还没有发现的或者由于人类的智慧极限而永远发现不了的物理现象，来交换数据、经验和想法。

计算物理学(phys-comp)*“计算物理学”(physicsofcomputation)的缩写形式，它研究计算的基本物理极限，例如，“是否有可能进行无热计算？”“在一定时间、一定体积内，可以进行计算的最大速度是什么？”等等。计算物理学对量子计算问题进行了深刻的讨论。

量子计算(quantumcomputing)*随着电子工业使得元器件的尺寸像分子级别缩小，量子效应开始不可避免地起作用。量子计算利用量子状态重叠的量子现象，可以同时处理很多经典的机械问题。这比传统的计算要有效得多，因为传统的计算是每次解决

一个问题。基于纳米技术的人工智能机器将应该是量子计算机，因为它们是分子级别的机器。

可逆计算(reversible computing)*一种产生零热量的计算形式，并且因此可以作为三维计算机电路的基础。使用可逆计算不会破坏信息的可逆逻辑电路。热动力学上的研究表明，丢失信息比特将会产生热。通过把信息输入到可逆计算机中，复制结果，然后把结果通过同样的电路传回(逆向的)，就可以得到原先的输入。没有信息丢失，所以没有热量产生。可逆计算是不可避免的，因为如果采用传统的不可逆计算技术，分子级别的电路将会因产生太多的热量而爆炸。本世纪的人工智能将需要依靠这样的计算形式。

级别战争(scaling war)这种战争只会发生在这样的情况下，那就是新技术使得建造新形式的、具有超级智慧和其他能力的生命形式成为可能。更高的智慧水平来自更快的信号速度和元器件密度。级别战争的一个例子就是，随着基于纳米技术的人工智能机器(nanoelect)成为可能，将发生在都是人类团体的宇宙主义者和地球主义者之间的“人工智能之战”。随着基于费米技术的人工智能(费米机器人)成为可能，另一个这样的战争可能发生在 x 和 y(双方都是纳米机器团体)之间。级别战争的基本起因是由于同一级别的两个团体关于制造基于更新(更小、更快、更密)技术的高级生物而带来的风险而造成的意见不合。

科学家的宗教(scientist's religion)大多数科学家不信仰传统的宗教，他们倾向于把传统的宗教看作是前科学时期的迷信，不同于现代科学知识和对猜想进行现代科学测试。然而，科学家也是人类，因此也和其他人一样有同样的宗教渴望。宇宙主义对于宇宙主义者来说是一种宗教，然而基于现代科学的。它可以成为科学家的宗教，因为它提供了对共同理想——创造神的一种敬畏，但是却和科学相容。宇宙主义是科学家觉得可以相信的一系列“宗教的”信仰。

奇点(singularity)*在数学中，奇点是接近无穷的值。在人工智能辩论的范畴内，奇点指的是这样的时刻将会到来，那就是机器将是如此聪明以至于它可以比人类更好更快地进行自我设计。结果将是一种超级机器，它可以以光速无止境地设计更好的机器，并无穷下去。另一个变量是，机器将是一个超级的学习者，它将以比人类快百万倍的速度自我学习。它在智慧和知识上的增长对于人类来说将是无比的迅速。

附录 术语表(7)

智能机器(smartilect)“智慧的人工智能机器”(smartartilect)的缩写。本世纪可能会制造出很多种类的人工智能机器。早期的人工智能机器将不具备晚些出现的机器的智慧和能力。真正聪明的机器可以乘坐“智能机器”。

空间意识(spaceconsciousness)空间意识指的是当人们通过生动的、大规模的三维图像，或者某种程度的虚拟现实来看到空间的广漠、星系中的亿万恒星、宇宙中的亿万星系时产生的一种情感上强烈震撼的感觉。这种感觉让人们觉得人类的当务之急是完全的微不足道。空间意识让人们意识到，在生活中、在宇宙中有比人类日常活动和目标“更加重要”的事情。空间意识对于宇宙主义者和宇宙主义思想是个很重要的概念。

物种主导之辩(speciesdominance debate)本世纪将主导全球政治的一个问题就是物种主导。物种主导之辩将集中在人类是否应该制造人工智能的问题上。人类是否继续保持主导物种的地位，还是制造人工智能机器来在智慧水平上超越我们。21 世纪技术上的进步将允许制造人工智能，并且让人们不得不去讨论主导物种的问题。

物种主导之战(speciesdominance war)一场发生在地球上主导的生物物种和其制造的人工智能机器之间的战争。由于人工智能机器高级于基于生化的生物，这场战争将决定谁将是主导物种。我们要区别“物种主导之战”和“人工智能之战”。“人工智能之战”是指发生在两个人类团体的战争，是为人工智能是否应该被制造这个问题而战的。“物种主导之战”可能发生在宇宙中的任何地方，因此是更普遍的概念。例如，如果“费米机器人”，也就是基于“费米技术”的人工智能机器和“纳米机器人”，也就是基于“纳米技术”的人工智能机器发生战争，那么这场战争将是“物种主导之战”。

聊天机器人(talkies)聊天机器人是交谈机器人的俚语，也就是一个人类可以与之交谈、与之有友谊关系的机器人，并且，如果聊天机器人足够聪明的话，可以与之建立关系。聊天机器人需要有几乎同人类一样级别的智慧程度，才能真正有用并且受到人类的欢迎。

教师机器人(teacherbots)教师机器人是一个教书的机器人，也就是可以通过调整其智慧等级、兴趣和动机而教育人类的人工智能机器人。

地球主义者(terran)地球主义者是基于地球(terra)而衍生的词汇，是指那些认为人工智能机器对人类生存具有潜在威胁而不应该被制造的人。地球主义者希望人类继续保持为地球上的主导物种。地球主义者的基本态度是，唯一确保零危险，即人工智能机器不会在未来毁灭人类的办法就是永远不要制造它们。因此要保护人类的生存，地球主义者们将不惜任何代价来阻止宇宙主义者，甚至发动一场大战也在所不惜。

地球主义的问题(terranproblem)地球主义的问题是指，由于地球主义者成功地在全球发起了对发展超越一定“人类安全指数”智慧水平的人工智能机器的禁令，那些人工智能产业领导人担心他们的销售和政治影响会因此受损失。

附录 参考文献(1)

关于半机器人的文献

andyclark,natural-borncyborgs:minds,technologies,andthefutureofhumanintelligence,oxforduniversitypress,2003.

chrish.gray,cyborgcitizen,routledge,2001.

jacqueshouisetal(eds.),beinghuman:thetechnologicalextensionsofthebody,marsilio,1999.

kevinwarwick,i,cyborg,century,2002.

关于智能机器、后人类等的文献

azamatabdoullaev,artificialsuperintelligence,eisltd.,1999.

igoralexander,howtobuildamind,weidenfeldandnicholson,2000.

jamesbailey,afterthought:thecomputerchallengetohumanintelligence,basicbooks,1996.

peterbock,theemergenceofartificialcognition,worldscientific,1993.

damienbroderick,thespike:howourlivesarebeingtransformedbyrapidlyadvancingtechnologies,forge,2001.

rodneya.brooks,fleshandmachines:howrobotswillchangeus,pantheon,2002.

maureencaudill,inourownimage:buildinganartificialperson,oxforduniversitypress,1992.

thomasm.georges,digitalsoul:intelligentmachinesandhumanvalues,westview,2003.

jeromec.glenn,futuremind:artificialintelligence,acropolis,1989.

raykurzweil,theageofintelligentmachines,mitpress,1990.

raykurzweil,theageofspiritualmachines:whencomputersexceedhumanintelligence,viking,1999.

raykurzweil,arewespiritualmachines?raykurzweilvs.thecriticsofstronga.i.,discoveryinstitute,2002.

jamesmartin,aftertheinternet:alienintelligence,capitalpress,2000.

pamelamccorduck,machineswhothink,freeman,1979.

billmckibben,enough:stayinghumaninanengineeredage,timersbooks,2003.

hansmoravec,mindchildren:thefutureofrobotandhumanintelligence,harvarduniversitypress,1990.

hansmoravec,robot:meremachinetotranscendentmind,oxforduniversitypress,1999.

douglasmulhall,ourmolecularfuture:hownanotechnology,robotics,genetics,andartificialintelligencewilltransformourworld,prometheusbooks,2002.

gregorys.paul,earld.cox,beyondhumanitycyberevolutionandfutureminds,charlesrivermedia,1996.

sidneyperkowitz,digitalpeople:frombionichumanstoandroids,josephhenrypress,2004.

peterussell,theglobalbrainawakens:ournextevolutionaryleap,element,2000.

jeffreysatinover,thequantumbrain:thesearchforfreedomandthenextgenerationofman,wiley,2001.

geoffsimons,ismanarobot,wiley,1986.

gregorystock,metaman:themergingofhumansandmachinesintoaglobalsuperorganism,simonshuster,1993.

johnaylor,theshapeofmindstocome,michaeljoseph,1971.

kevinwarwick,marchofthemachines:whythenewraceofrobotswillruletheworld,century,1997.

kevinwarwick,inthemindofthemachine:thebreakthroughinartificialintelligence,arrow,1998.

附录 参考文献(2)

关于纳米技术的文献

william.atkinson,nanocosm:nanotechnologyandthebigchangescomingfromtheinconceivablysmall,amacom,2003.

b.c.crandalletal.,nanotechnology:researchandperspectives,mitpress,1992.

b.c.crandall(ed.),nanotechnology:molecularspeculationsonglobalabundance,mitpress,1996.

k.ericdrexler,enginesofcreation:thecomingeraofnanotechnology,anchorpress,doubleday,1986.

k.ericdrexleretal.,unboundingthefuture:thenanotechnologyrevolution,morrow,1991.

k.ericdrexler,nanosystems:molecularmachinery,manufacturing,andcomputation,wileyinterscience,1992.

sandyfritz,understandingnanotechnology,warnerbooks,2002.

michaelgross,travelstothenanoworld:miniaturemachineryinnatureandtechnology,plenumtrade,1999.

markuskrummenackeretal(eds.)prospectsinnanotechnology:towardmolecularmanufacturing,wiley,1995.

wilmccarthy,hackingmatter:levitatingchairs,quantummirages,andtheinfiniteweirdnessofprogrammableatoms,basicbooks,2003.

markratneretal.nanotechnology:agentleintroductiontothenextbigidea,prenticehall,2003.

edregis,nano:theemergingscienceofnanotechnology,remakingtheworld—moleculebymolecule,little,brown,1995.

edwardrietman,molecularengineeringofnanosystems,springer,2001.

gregorytimp(ed.),nanotechnology,springer,1999.

关于量子计算的文献

amird.aczel,entanglement:thegreatestmysteryinphysics,fourwallseightwindows,2001.

julianbrown,thequestforthequantumcomputer,touchstone,2000.

mikahirvensalo,quantumcomputing,springer,2001.

geraldj.milburn,schrodinger'smachines,freeman,1997.

geraldj.milburn,the feynman processor: quantum entanglement and the computing revolution, perseus books, 1998.

michael a. nielsen and isaac l. chuang, quantum computation and quantum information, cambridge university press, 2000.

附录 后记

非常凑巧，我的书译成中文将在中国出版的时候，几乎同时我从美国移民到中国来做教授。

为什么我会选择到中国？

自 1995 年以来，我每年都来中国访问，使我亲眼目睹了中国的发展和变化，新的观念在非常迅速地产生，我认为它比美国更有活力、更充满生机，而美国人日益变得肥胖、自满和保守。我希望我的人工智能的研究工作，特别是人工大脑的研制在中国延续并有新的拓展。当前我组建了一个人工智能研究学习小组，试图用我在本书第二章中提到的相关科学技术，协助我研制世界上第一个人工大脑。如果这样的尝试能成功，中国将在该领域成为世界上当今技术的先驱并定将引起世界的广泛关注，将对未来产生深远的影响。

以我的观点看，中国在本世纪将成为在国际上占有重要一席之地的民族。她有 13 亿人口(世界最多)，在过去十年，中国经济年平均增长率大约 10%，为世界上最快的国家。因此在新思想、新科学、新技术方面，她将逐步开始成为世界上的主导国家。美国的经济年增长率仅 3%。在购买力方面即实际个人购买力和收入所得(ppp, purchasing power parity)，中国在 2006 年已经占美国国民产值的 3/4(gnp)。因为中国年增长值为 10%，相对美国，如上面我所提及的 3%，你就可算出，中国不久将超过美国的 gnp(至少在购买力方面，急速地赶上美国)。这样继续下去，中国必将回到她曾在历史

上的辉煌地位。因此，我希望在退休以前能一直呆在中国(我有一位中国的妻子)。我喜欢在赢的一边。

在本书结束时，我特别要对两个人表达我的感激之情。一个是我曾在美国的博士生——本书的译者胡静。第二个是我的中国妻子——雷国庆。如果没有他们两个人的工作，此书将可能仅限于英语的读者了。

附录 编辑手记(1)

我们知道恐龙在地球上生活了 2 亿年，是人类历史的 50 倍以上。在这 2 亿年的时间里，似乎它们并没有本质地进化成更高级、更智慧的一类生物，这也许就是它们灭绝的原因吧。那么，我们人类，首先能不能存在的比恐龙更久，如果能生存那么久，似乎没有任何迹象可以表明人类那时候会有本质的进化，以至于可以自由地翱翔于宇宙。

当我拿起这本书的时候，知道这决不是一本科幻小说。但当我读它的时候，却让我感到亦真亦幻的神奇和惊悚，让人看了就离不开目光。被誉为人工智能先驱者的德·加里斯教授，向我们预测的未来，让我们清醒地意识到，真实远远超出想象。与他描绘的几乎近在眼前的情景相比，那些科幻小说和电影都太缺乏想象力了。那些毕竟不是最前沿的科学家幻想出来的，局限性很大。而且，最重要的是，它们用人的思维来考虑能力是我们的万亿个万亿倍的人工智能机器，就像一个蚊子来思考人类会怎么想怎么做一样是完全不现实的。

我本想劝说作者将第三章有关技术的部分挪到后面，担心这个章节技术性太强，影响读者的阅读兴趣。但是，我自己很快被这些神奇的技术迷住了，我越来越理解为什么德·加里斯会有这样夜不能寐的苦苦思索了。书中非常合理地推论了，人类真的会很快制造出这样无所不能的“可敬”或是“可怕”的人工智能机器，这不是技术是否允许的问题，而是时间和金钱是否允许的问题，那么换句话说，就是实现这些基本没有什么障碍。计算机会在某天突然达到一种称作“奇点”的状态，它们将会“失去控制”，获得迅速的智能进化，以至于远远地，并且非常迅速地把我们抛在后面。这是什么意思呢？例如，一个人直到博士毕业学到的所有知识，人工智能机器只需要 4 秒钟就能完成，而这还只是比较早期的人工智能机器的基本能力！

我们有可能控制这样的进程吗？人类将会怎样进化？人类将怎么解决我们和人工智能机器之间的关系？这将是本世纪最伟大的思想家们的主要课题，对这类问题的研究和探讨，将产生一些比古代伟大的思想家更杰出的思想家。这本书提出的问题，将给思想家们指出一个可以超越前辈的方向。

我认为对于这类问题的普遍思考和找寻良性的解决方案，将会使人类的智力产生质的飞跃。试想一下，站在家庭的角度思考是不是比站在个人的角度要显得智慧，站在国家角度思考是不是比站在小家的角度更智慧。而一些古代的思想家和智者往往是站在全人类的角度思考，所以他们是那么的与众不同，智慧高于常人。那么看了本书之后，我们应该怎么思考呢？

也许到了某一天，人类会像注射疫苗一样，从出生开始，随着年龄的增长定期植入各种能力的晶片，我觉得人们是有可能接受这个的，虽然现在不能接受，就像古人不能接受在身上动刀子——做手术一样。既然为了健康能接受心脏起搏器，那么为什么不能接受植入晶片，当然这需要有一个接受过程。

这样就绝对不会发生人工智能的战争了吗？不是的！以前的世界大战都是因为一些莫名其妙的原因引发的，而制造智力和能力远远超过常人的机器，这比以前任何一个战争的起因都更充分。居里夫人研究铀的时候，目的一定是积极的，为了造福人类的，谁会想到竟成为人类的杀手。

因此，德·加里斯的警钟多么及时，为了避免可预见的悲剧，我们应该怎么做？

(1)在对理工科高才生的培养中，我们要进一步强化情商的培养，就像我们不断地教育孩子要注重环保一样，使他们充分认识到——他的发明创造有可能威胁人类的生存。

(2)注意不要单纯把机器人制造成超级智慧能力的，还要有意识地把它们制造成超级情感的，让它们的智能和情感同步发展。

(3)是不是应该在最初研制的时候，就考虑人工智能机器除了在智力和神经等方面是人类的万亿个万亿倍以外，还应该在境界上是人类的万亿个万亿倍。那样，它们一定会比耶稣和释迦牟尼更加厌恶暴力和血腥。

附录 编辑手记(2)

是的，普通人在杀死蚊子的时候不会迟疑，不会觉得这样残忍，但是耶稣和释迦牟尼是不会杀死蚊子的。他们会认为这是很残忍的事情。如果能制造出这种境界的机器人，也许人类还有希望.....也许，到了那个时候，这个世界上就真的再也没有血腥了.....

张立红

