SPEECH | 演讲

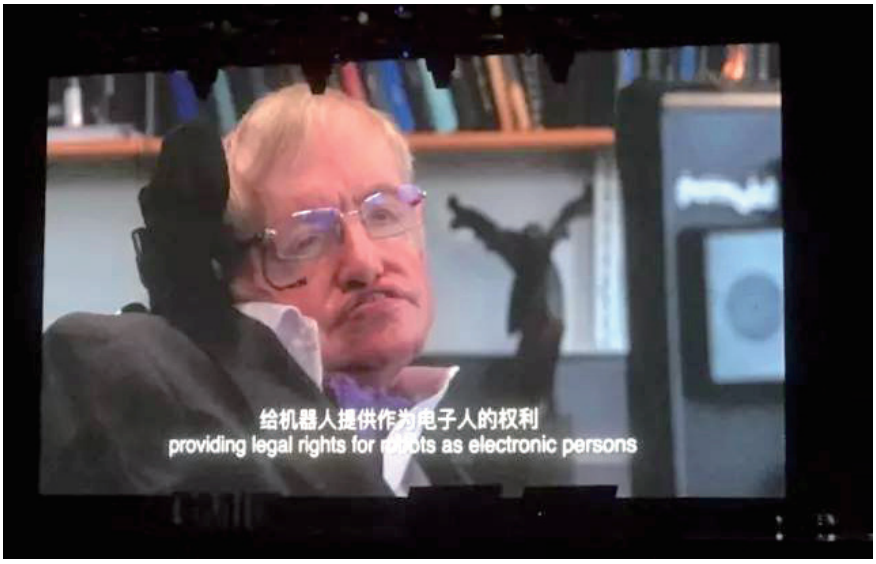
让人工智能造福人类及其赖以生存的家园

■演讲 / 斯蒂芬·威廉·霍金
整理 / 周　翔（雷锋网）

2017 年 4 月 27 日，GMIC（全球移动互联网大会）北京站 2017 开幕。在开幕式上，英国剑桥大学著名物理学家，现代最伟大的物理学家之一，斯蒂芬·威廉·霍金远程发表了名为《让人工智能造福人类及其赖以生存的家园》（“Guiding AI to Benefit humanity and the environment”）的主题演讲。在演讲中霍金仍保持了其对人工智能一贯的谨慎性，提醒 AI 科研者在利用 AI 造福人类的同时还需注意消除可能的威胁。在演讲之后，霍金还回答了网友的提问。以下是雷锋网整理的演讲全文。

在我的一生中，我见证了社会深刻的变化。其中最深刻的，同时也是对人类影响与日俱增的变化，是人工智能的崛起。简单来说，我认为强大的人工智能的崛起，要么是人类历史上最好的事，要么是最糟的。我不得不说，是好是坏我们仍不确定。但我们应该竭尽所能，确保其未来发展对我们和我们的环境有利。我们别无选择。我认为人工智能的发展，本身是一种存在着问题的趋势，而这些问题必须在现在和将来得到解决。

人工智能的研究与开发正在迅速推进。也许我们所有人都应该暂停片刻，把我们的研究重复从提升人工智能的能力转移到最大化人工智能的社会效益上面。基于这样的考虑，美国人工智能协会（AAAI）于 2008 至 2009 年成立了人工智能长期未来总筹论坛，他们近期在目的导向的中性技术上投入了大量的关注。但我们的人工智能系统须要按照我们的意志工作。跨学科研究是一种可能的前进道路：从经济、法律、哲学延伸至计算机安全、形式化方法，当然还有人工



智能本身的各个分支。

潜在的威胁：计算机智能与我们没有本质区别

文明所产生的一切都是人类智能的产物，我相信生物大脑可以达到的和计算机可以达到的，没有本质区别。因此，它遵循了“计算机在理论上可以模仿人类智能，然后超越”这一原则。但我们并不确定，所以我们无法知道我们将无限地得到人工智能的帮助，还是被藐视并被边缘化，或者很可能被它毁灭。的确，我们担心聪明的机器将能够代替人类正在从事的工作，并迅速地消灭数以百万计的工作岗位。

在人工智能从原始形态不断发展，并被证明非常有用的同时，我也在担忧创造一个可以等同或超越人类的事物所导致的结果：人工智能一旦脱离束缚，以不断加速的状态重新设计自身，人类由于受到漫长的生物进化的限制，无法与之竞争，将被取代，这将给我们的经济带来极大的破坏。未来，人工智能可以发展出自我意志，一个与我们冲突的意志。尽管我对人类一贯持有乐观的态度，但其他人认为，人类可以在相当长的时间里控制技术的发展，这样我们就能看到人工智能可以解决世界上大部分问题的潜力。但我并不确定。

2015 年 1 月份，我和科技企业家埃隆·马斯克，以及许多其他的人工智能专家签署了一份关于人工智能的公开信，目的是提倡就人工智能对社会所造成的影响做认真的调研。在这之前，埃隆·马斯克就警告过人们：超人类人工智能可能带来不可估量的利益，但是如果部署不当，则可能给人类带来相反的效果。

我和他同在“生命未来研究所（Future of Life Institute）”的科学顾问委员会，这是一个为了缓解人类所面临的存在风险而设立的组织，而且之前提到的公开信也是由这个组织起草的。这个公开信号召大家展开可以阻止潜在问题的直接研究，同时也收获人工智能带给我们的潜在利益，并致力于让人工智能的研发人员更关注人工智能安全。

此外，对于决策者和普通大众来说，这封公开信内容翔实，并非危言耸听。人人都知道人工智能研究人员在认真思索这些担心和伦理问题，我们认为这一点非常重要。比如，人工智能是有根除疾患和贫困的潜力的，但是研究人员必须能够创造出可控的人工智能。那封只有四段文字，题目为《应优先研究强大而有益的人工智能》（“Research Priorities for Robust and Beneficial Artificial Intelligence”）的公开信，在其附带的十二页文件中对研究的优先次序作了详细的安排。

如何从人工智能中获益并规避风险

在过去的 20 年里，人工智能一直专注于围绕建设智能代理所产生的问题，也就是在特定环境下可以感知并行动的各种系统。在这种情况下，智能是一个与统计学和经济学相关的理性概念。通俗地讲，这是一种作出好的决定、计划和推论的能力。

基于这些工作，大量的整合和交叉被应用在人工智能、机器学习、统计学、控制论、神经科学，以及其它领域。共享理论框架的建立，结合数据的供应和处理能力，在各种细分的领域取得了显著的成功，例如语音识别、图像分类、自动驾驶、机器翻译、步态运动和问答系统。

随着这些领域的发展，从实验室研究到有经济价值的技术形成良性循环。哪怕很小的性能改进，都会带来巨大的经济效益，进而鼓励更长期、更伟大的投入和研究。目前人们广泛认同，人工智能的研究正在稳步发展，而它对社会的影响很可能扩大，潜在的好处是巨大的，既然文明所产生的一切，都是人类智能的产物。由于这种智能是被人工智能工具放大过的，

我们无法预测我们可能取得什么成果。但是，正如我说过的那样，根除疾病和贫穷并不是完全不可能，由于人工智能的巨大潜力，研究如何（从人工智能中）获益并规避风险是非常重要的。

现在，关于人工智能的研究正在迅速发展，这一研究可以从短期和长期两个方面来讨论。

短期的担忧主要集中在无人驾驶方面，包括民用无人机、自动驾驶汽车等。比如说，在紧急情况下，一辆无人驾驶汽车不得不在小概率的大事故和大概率的小事故之间进行选择。另一个担忧则是致命性的智能自主武器。他们是否该被禁止？如果是，那么“自主” 该如何精确定义；如果不是，任何使用不当和故障的过失应该如何问责。此外还有一些隐忧，包括人工智能逐渐可以解读大量监控数据引起的隐私问题，以及如何掌控因人工智能取代工作岗位带来的经济影响。

长期担忧主要是人工智能系统失控的潜在风险。随着不遵循人类意愿行事的超级智能的崛起，那个强大的系统会威胁到人类。这样的结果是否有可能？如果有可能，那么这些情况是如何出现的？我们又应该怎样去研究，以便更好地理解和解决危险的超级智能崛起的可能性？

当前控制人工智能技术的工具（例如强化学习）以及简单实用的功能，还不足以解决这个问题。因此，我们需要进一步研究来找到和确认一个可靠的解决办法来掌控这一问题。

近来（人工智能领域的）里程碑，比如说之前提到的自动驾驶汽车，以及人工智能赢得围棋比赛，都是未来趋势的迹象。巨大的投入正在倾注到这一领域。我们目前所取得的成就，和未来几十年后可能取得的成就相比，必然相形见绌。而且当我们的头脑被人工智能放大以后，我们更不能预测我们能取得什么成就。也许在这种新技术革命的辅助下，我们可以解决一些工业化对自然界造成的损害问题，关乎到我们生活的各个方面也即将被改变。简而言之，人工智能的成功有可能是人类文明史上最大的事件。

但是人工智能也有可能是人类文明史的终结，除非我们学会如何避免危险。我曾经说过，人工智能的全方位发展可能招致人类的灭亡，比如最大化使用智能性自主武器。今年早些时候，我和一些来自世界各国的科学家共同在联合国会议上支持其对于核武器的禁令，我们正在焦急的等待协商结果。

目前，九个核大国可以控制大约一万四千个核武器，它们中的任何一个都可以将城市夷为平地，放射性废物会大面积污染农田，而最可怕的危害是诱发核冬天，火和烟雾会导致全球的小冰河期。这一结果将使全球粮食体系崩塌，末日般动荡，很可能导致大部分人死亡。我们作为科学家，对核武器承担着特殊的责任，因为正是科学家发明了它们，并发现它们的影响比最初预想的更加可怕。

现阶段，我对灾难的探讨可能惊吓到了在座的各位，很抱歉。但是作为今天的与会者，重要的是，你们要认清自己在影响当前技术的未来研发中的位置。我相信我们会团结在一起，共同呼吁国际条约的支持或者签署呈交给各国政府的公开信，科技领袖和科学家正极尽所能避免不可控的人工智能的崛起。

去年 10 月，我在英国剑桥建立了一个新的机构，试图解决一些在人工智能研究快速发展中出现的尚无定论的问题。“利弗休姆智能未来中心

（The Leverhulme Centre for the Future of Intelligence）”是一个跨学科研究所，致力于研究智能的未来，这对我们文明和物种的未来至关重要。我们花费大量时间学习历史，深入去看，大多数是关于愚蠢的历史，所以人们转而研究智能的未来是令人欣喜的变化。虽然我们对潜在危险有所意识，但我内心仍秉持乐观态度，我相信创造智能的潜在收益是巨大的。也许借助这项新技术革命的工具，我们将可以削减工业化对自然界造成的伤害。

确保机器人为人类服务

我们生活的每一个方面都会被改变。我在研究所的同事休·普林斯（Huw Price）承认，“利弗休姆中心”能建立，部分是因为大学成立了“存在风险中心（Centre for Existential Risk）”。后者更加广泛地审视了人类潜在问题，“利弗休姆中心”的重点研究范围则相对狭窄。

人工智能的最新进展，包括欧洲议会呼吁起草一系列法规，以管理机器人和人工智能的创新。令人感到些许惊讶的是，这里面涉及到了一种形式的电子人格，以确保最有能力和最先进的人工智能的权利和责任。欧洲议会发言人评论说，随着日常生活中越来越多的领域日益受到机器人的影响，我们需要确保机器人无论现在还是将来，都为人类而服务。一份向欧洲议会议员提交的报告明确认为，世界正处于新的工业机器人革命的前沿。报告分析了是否给机器人提供作为电子人的权利，这等同于法人（的身份）。

报告强调，在任何时候，研究和设计人员都应确保每一个机器人设计都包含有终止开关。在库布里克的电影《2001 太空漫游》中，出故障的超级电脑哈尔没有让科学家们进入太空舱，但那是科幻，我们要面对的则是事实。奥斯本·克拉克（Space Odyssey）跨国律师事务所的合伙人——洛纳·布拉泽尔（Lorna Brazell）在报告中说，我们不承认鲸鱼和大猩猩有人格，所以也没有必要急于接受一个机器人人格，但是担忧一直存在。

报告承认在几十年的时间内，人工智能可能会超越人类智力范围，进而挑战人机关系。报告最后呼吁成立欧洲机器人和人工智能机构，以提供技术、伦理和监管方面的专业知识。如果欧洲议会议员投票赞成立法，该报告将提交给欧盟委员会，它将在三个月的时间内决定要采取哪些立法步骤。

我们还应该扮演一个角色，确保下一代不仅仅有机会还要有决心，能够在早期阶段充分参与科学研究，以便他们继续发挥潜力，并帮助人类创造一个更加美好的的世界。这就是我刚谈到学习和教育的重要性时所要表达的意思。我们需要跳出“事情应该如何（how things should be）”这样的理论探讨，并且采取行动，以确保他们有机会参与进来。我们站在一个美丽新世界的入口，这是一个令人兴奋的、同时充满了不确定性的世界，而你们是先行者，我祝福你们。

谢谢！

Q&A

创新工场的 CEO 李开复： 互联网巨头拥有巨量的数据，而这些数据会给他们各种以用户隐私和利益换取暴利的机会。在巨大的利益诱惑下，他们是无法自律的，而且这种行为也会导致小公司和创业者更难创新。您常谈到如何约束人工智能，但更难的是如何约束人本身。您认为我们应该如何约束这些巨头？

霍金： 据我了解，许多公司仅将这些数据用于统计分析，但任何涉及到私人信息的使用都应该被禁止。如果互联网上所有的信息均通过基于量子技术加密，这样互联网公司在一定时间内便无法破解，这会有助于保护隐私，但安全部门会反对这个做法。

猎豹移动 CEO 傅盛： 灵魂会不会是量子的一种存在形态？或者是高维空间里的另一个表现？

霍金： 我认为近来人工智能的发展，比如电脑在国际象棋和围棋的比赛中战胜人脑，都显示出人脑和电脑并没有本质差别。这点上我和我的同事罗杰·彭罗斯正好相反。会有人认为电脑有灵魂吗？对我而言，灵魂这个说法是一个基督教的概念，它和来世联系在一起。我认为这是一个童话故事。

百度总裁张亚勤： 人类观察和抽象世界的方式不断演进，从早期的观察和估算，到牛顿定律和爱因斯坦方程式，到今天数据驱动的计算和人工智能，下一个会是什么？

霍金： 我们需要一个新的量子理论，将重力和其他自然界的其它力量整合在一起。许多人声称这是弦理论，但我对此表示怀疑，目前唯一的推测是，时空有十个维度。

斯坦福大学物理学教授张首晟： 如果让您告诉外星人我们人类取得的最高成就，并写在一张明信片的背面，您会写什么？

霍金： 告诉外星人关于美，或者任何可能代表最高艺术成就的艺术形式都是无益的，因为这是人类特有的。我会告诉他们哥德尔不完备定理和费马大定理。这才是外星人能够理解的事情。

音乐人、投资者胡海泉： 如果星际移民技术的成熟窗口期迟到，有没有完全解决不了的内发灾难导致人类灭绝？

霍金： 是的。人口过剩、疾病、战争、饥荒、气候变化和水资源匮乏，人类有能力解决这些危机。但很可惜，这些危机还严重威胁着我们在地球上的生存。这些危机都是未来可以解决但目前还未解决的。

问： 我们希望提倡科学精神，贯穿 GMIC 全球九站，请您推荐三本书，让科技届的朋友们更好地理解科学及科学的未来。

霍金： 他们应该去写书而不是读书。只有当一个人关于某件事能写出一本书，才代表他完全理解了这件事。

问： 您认为一个人一生当中最应当做的一件事和最不应当做的一件事分别是什么？

霍金： 我们绝不应当放弃，我们都应当尽可能的去理解（这个世界）。

问： 人类在漫漫的历史长河中，重复着一次又一次的革命与运动。从石器、蒸汽、电气……您认为下一次的革命会是由什么驱动的？

霍金： （我认为是）计算机科学的发展，包括人工智能和量子计算。科技已经成为我们生活中重要的一部分，但未来几十年里，它会逐渐渗透到社会的每一个方面，为我们提供智能的支持和建议，在医疗、工作、教育和科技等众多领域。但是我们必须要确保是我们来掌控人工智能，而非它（掌控）我们。科技