

后奇点时代的人工智能与人类的尊严

李建会 夏永红

摘要：人工智能的加速发展，使人们开始担心奇点时代的到来，因为越过奇点，机器就可能超越人类，甚至会取代人类。通过对人工智能的分析，文章认为，在人们还没有搞清楚人的大脑的本质和智能的本质的时候，人工智能发展再快，机器要想全面超越人类还很困难。即便其进一步的发展超越了人类，但只要自创生的人工智能没有出现，或者拥有鲜活的身体的人工智能没有出现，人工智能也不会取代人类。也许进一步的发展，未来的人工智能掌握了人类的所有能力，但我们也不用担心人类的尊严受到了威胁，因为人工智能作为人类的延展系统，它的智能越高，就越证明了人类的可完善性，人类也就越有尊严。

关键词：奇点；人工智能；人类尊严

中国分类号：TP18 **文献标识码：**A **文章编号：**1673-5706（2019）02-0055-05

早在20世纪90年代，人工智能专家库兹韦尔（Ray Kurzweil）就表达了这样的想法：当代科学技术正以加速的方式发展，到未来某一个时候会以指数的方式爆炸式地增长，尤其是人工智能技术更会爆炸式地增长，这时人工智能将大大超越人类智能，机器人也将成为人类“进化的继承者”和“思想的继承者”。^①这一时刻，计算机专家冯·诺伊曼称之为“奇点”。库兹韦尔认为，这一时刻要不了多少年就会来临。奇点真的会来临？人工智能真的会超越人类智能，并最终会取代人类吗？本文拟回答这些问题。

一、奇点临近？

自近代科学产生开始，尤其是进入20世纪，科

学技术就开始加速发展。早在20世纪50年代，冯·诺伊曼就说过：“技术正以其前所未有的速度增长……我们将朝着某种类似奇点的方向发展，一旦超越了这个奇点，我们现在熟知的人类社会将变得大不相同。”^②这里，冯·诺伊曼提到了两个概念：“加速”和“奇点”。“加速”的含义我们都比较好理解，就是越来越快。科学技术的加速发展对于今天的人来说已经司空见惯。我们的计算机、手机每隔一两年就要换新的；计算机软件和手机的APP不停地提醒你更新到新的版本；高铁越来越快。科学技术的加速发展推动着社会的日新月异。我们的生产方式、生活方式不停地在变化。“奇点”可能对我们很多人来说比较陌生，但也很好理解。在指数增长的曲

① 库兹韦尔：《奇点临近》，北京：机械工业出版社，2011年，第11-15页。

② 库兹韦尔：《奇点临近》，北京：机械工业出版社，2011年，第2-3页。

线中,我们可以看到,开始的时候增长很缓慢,几乎看不到有什么特异的地方,但是一旦超越曲线的拐点,它便以爆炸式的速度增长。我们很多人理解历史发展可能都是一种直觉的线性发展观,即缓慢而平稳的发展。20世纪以来,随着科学技术走在生产的前面,科学技术应用于生产的周期也越来越快。以至于我们今天的时代与过去的时代非常不同,我们一生会生活在一个日新月异的由技术推动的不断变革的社会之中。当这种加速到了一定的时刻,技术的发展就会出现爆炸性的增长,其带来的影响如此深远,人类的生活会彻底发生改变。

按着这种思路,库兹韦尔把宇宙进化的历史划分为六个纪元,并认为“奇点将随着第五纪元的到来而开始,并于第六纪元从地球扩展到全宇宙”。^①这六大纪元分别如下所述:

在第一纪元,即物理与化学纪元,宇宙大爆炸,产生时间和空间、原子和分子,如果说时间和空间、物质和能量是物理现象,那么,随着分子的产生,化学现象开始出现。在第二纪元,即生物与DNA纪元,出现了越来越复杂的由碳元素构成的化合物,随后产生了可以自我复制的大分子聚合物,最早的生命也由此形成,最终产生了可以保存生命信息的DNA。在第三纪元,即大脑纪元,以DNA为导向的生物进化产生了可以运用神经系统感知信息的生物,并最终产生了人类的高级大脑。在第四纪元,即技术纪元,人类的认知能力使人类迈进了人造技术的进化层次。开始是简单的机械技术,随着科学的发展,逐步发展到精妙的自动化技术,最终发展到以计算机为代表的通用信息处理技术。在第五纪元,即人类大脑与技术结合的纪元,人类知识与人类技术的结合,让强大的知识分享能力成为可能。奇点将会来临,人类与机器将深度融合,文明将超越人脑的限制,完全解放人类的创造力。在第六纪元,即宇宙觉醒时代,奇点之后基于生物智能与技术智能融合的人机智能,将会在物质和能量上受限,为了让智能进一步发展,就需要重新组织物质和能量以创造一种最优的计算,使更高级的智能在宇宙产生,

并将这种智能由地球推广至整个宇宙。

在这个历史过程中,人工智能被视为逼近奇点的关键技术。与其他技术的发展不同,人工智能加速发展之后,其不断提升的智能又会推进人工智能的加速发展,这样,到某一时刻,即所谓的技术奇点,人工智能的智能水平大大提高,在很多方面都会超越人类。然而,整体上,奇点之后,人工智能会超越人类智能,并最终取代人类吗?

二、人工智能能否超越人类智能

尽管库兹韦尔的奇点说非常吸引人,但也有很多学者认为,人工智能如何发展也不能超越人类,因为,人类有一些特殊的品质机器永远也达不到。实际上,自近代开始,随着机器的发展,人们一直就有机器是否会超越人类的担忧,但也有一些学者竭力论证,人类有一些品质,机器不可能获得。从近代的笛卡尔到当代的塞尔,不断地有哲学家在论证人类的优越性。

笛卡尔曾设想一种自动机器,它在外观上与我们的身体没有什么不同,可以模仿我们的任何动作,但笛卡尔认为它仍然不能算作真正的人类。他的第一个理由是,机器无法像人类一样使用语言以表达自己的思想。笛卡尔设想了一个聊天机器人,它可以根据我们的要求说出某些词语,但却无法以有序的句法规则将这些词语组合成适当的句子,最终在不同情境下表达不同的思想。第二个理由是,机器即便在某种类型的任务上比肩人类甚至超越人类,但却不能完成别的事情,因为它的结构是专用的,而人类的理性却是万能的,可以在很多情境下应付自如。^②可以看出,笛卡尔在第一个理由中,实际上把人类自由使用句法能力视为一种独有的技能,而他的第二个理由,类似于福多所说的中枢系统的各项同性或非封装性,也就是说,它是非模块化的,具有全局性的、整体性的认知加工能力。因此,笛卡尔最终把自然语言操演和全局认知能力视为人类的标志,而机器无法胜任这两种能力。

笛卡尔之后,二进制的发明人,计算机的早期先驱莱布尼茨也认为,机器不能像人那样具有真正

^① 库兹韦尔:《奇点临近》,北京:机械工业出版社,2011年第5页。

^② [法]笛卡尔:《谈谈方法》,北京:商务印书馆,2000年,第45页。

的智能。在《单子论》第十七节，莱布尼兹提出了一个可以称为“磨坊论证”的论证反对机器智能。他说：“可是我们不得不承认知觉和与之相联系的一切是不能通过机械的理由即性状和运动得到解释的。假定说有一架机器，构造得能够思想，产生感觉，有了知觉，还可以设想把它按原样的比例放大，人们可以像走进一座磨坊一样在里面一边观看，一边发现相互推动着的部件，但是都无法解释知觉来自何处。所以一定是在单纯实体内部而不是在复合物中或者说在机器中，去寻找知觉。也就是说，只有在单纯实体里面才能发现知觉及其变化。单纯实体的全部内在活动只寓于自身之中。”^①这里的所谓知觉，在莱布尼茨的哲学中实际上指的是单子的精神状态。可以看出，莱布尼茨认为，机器的运动并不能说明精神的产生，相反，精神只能在一个具有隐德莱希的单子内产生。所以，机器作为一个复合的物体，即便它表现得很智能，甚至与人的智能完全吻合，但依然不能算作真有智能。

当代很多反对人工智能可能真正具有智能的，其思想大都可追溯到笛卡尔和莱布尼兹的论证。塞尔的“中文屋”论证以及后来的“无心人”论证沿袭了莱布尼兹论证的总体思路，他们都认为机器无法具有人类的意向性和现象意识等；而德雷福斯基于框架问题对人工智能的现象学批判，其出发点则与笛卡尔的论证相似，他们都主张机器最终是模块化的专用系统，无法模拟人类智能，但德雷福斯并不像笛卡尔那样主张人类理性具有全局理解能力，相反，他把这种能力归结为非理性的熟练技能，也正是由此人类才不像机器那样会陷入由模块化系统的信息封装性而导致的框架问题。当然，当代也提出了其他很多的新理论，比如，从哥德尔定理出发反对强人工智能，认为机器无法获得人类的直觉和洞察力，此外，还有从自主性、创造性、自由意志等等人类特有的特征的不可代替性来说明机器智能的局限。尽管这些论证可能都存在争议，但在可预见的未来，人类对机器的优势仍然难以打破。因此，在人类智能的秘密还没有完全清楚之前，现在就说奇点马上来临，可能还不令人信服。

除了以上基于人工智能无法具有人类的心智能力的种种论证之外，还有一些论证主要出于伦理或意识形态的目的。人工智能对近代人文主义构成了这样一种挑战：如果机器能够思考，人类在理性上没有任何优势，那么人与机器之间的界限将被抹去，人的尊严将受到威胁。康德之所以把尊严赋予人类，一个重要的原因就是因有人有理性。如果机器也有了理性，机器就获得了和人一样的尊严。因此，也有很多人本主义者竭力反对发展强人工智能。但实际上，正如本文以上所指出的，虽然今天人工智能的发展异常迅速，但到目前为止，人类的优越性并未受到威胁，人工智能暂时无法超越人类的思维。当前的人工智能是专用系统，例如下棋程序不能用于语音识别，反之亦然。当然，我们还不能排除这种可能性，未来的智能机器将取得突破性的进展，并将在智能领域引领人类智慧，但这种可能性的前提是，我们对人类智能的秘密有更充分的了解。

三、人工智能能取代人类智能吗？

暂且假定人工智能可以全面超越人类，届时人们就可能担心人工智能是否会取代人类。事实上，关于这个问题，已有很多论证。比如，阿西莫夫的机器人三定律就是为了防止机器人对人类的伤害。博斯特罗姆（Nick Bostrom）在他的《超级人工智能》中认为，超级智能实际上可以与任何终极价值体系相结合，它们并不必然具有人类共享的道德观念，因此它很可能让人类面临毁灭性的生存灾难；此外，超级智能可能与人类争夺物质资源，这就会和人类存在生存冲突，从而破坏整个人类的发展潜力。自2015年以来，马斯克、比尔·盖茨和霍金都表达了对不受控制的超级人工智能可能导致人类灭亡的担忧。他们还与人工智能领域的专家签署了一封公开信，呼吁各界人士关注人工智能的安全性。

人工智能到底会否取代人类，需要我们对未来的人工智能的形态及其与人的关系进行认真的分析。生物学中关于生物之间的生存竞争理论可以给我们提供很多启示。假设在后奇点时代，人工智能在各种能力上都超越了人类，但它可能只存在于电脑的虚拟空间中，并不具有像我们一样的身体，从而也

^① 莱布尼茨：《莱布尼茨读本》，陈乐民编，南京：江苏教育出版社，2006年，第37页。

不会在现实世界中与人类争夺资源，这样的人工智能即便超越了人类，也不会与人类存在生存冲突，从而也不会取代人类。但如果超级人工智能像我们一样不仅具有自己的身体，而且具有生命的本质特征如自我复制和保存等，并且具有远高于人类的生存适应性，那么就会与人类在现实世界中争夺物质资源，在这种情况下，人类很可能被取代。

在生态学上，物种与物种之间存在多种关系：有竞争关系，有捕食或寄生关系，还有共生关系以及共栖关系等。其中，竞争关系对两个物种都不利，强者可能胜出；捕食者关系和寄生关系对其中一个有利，对另外一个是有害的；而共栖关系对其中一个有益，对另一个没有什么影响；互利共生关系让两个物种都从中受益。在这四种关系中，只有前两种存在着对一种物种有害的情况。因此，只有当人工智能与人类之间的关系属于捕食或寄生关系，并假设人类相对于人工智能相对较弱时，人类才会遇到被取代的风险。

但我们与人工智能的关系真的会是竞争关系吗？有相当多的人认为，我们并不会与人工智能构成竞争关系。因为，根据生态学原理，每种物种为了存活都必须占据一个生态位，以便从环境中获得物质和能量。如果两个物种占据了相同的生态位，那么因为资源的有限，两个物种不可避免会发生竞争排斥，导致一个物种必然趋于消亡。这就是生态学上的两个具有相同生态位的“完全竞争者不能共存”的原理。根据这一原理，如果人工智能取代人类，那么这种人工智能就要跟人类占有相同的生态位，从而发生实质的竞争。然而，人工智能会和人类占据相同的生态位吗？人工智能需要占据一定的生态位，需要深深植根于环境中，可以直接从环境中获取生存资源，以维持其自身的生存需求并培育后代。这需要人工智能具备这样的特征：自创生的能力和生殖能力。现有的人工智能都不具有这样的功能，未来是否具有这样的功能，现在还不清楚。

人工智能跟人类是否会形成捕食或寄生关系？这也是不大可能的。因为无论是捕食还是寄生关系，生物体都试图从其他生物获得物质和能量来维持和繁殖自己，然而，现有的人工智能的载体都是无机的机器，维持其工作的能量是电能，因此，没有必

要通过捕食和寄生获得能量。人工智能如果要通过捕食人类来维持和再生产自身，它必须发展到一种以有机材料为其身体的人工生命形态。也许未来的机器人是用纳米技术和生物技术制造的，身体结构和有机生命的身体的构成材料没有差别。但这样的机器人已经跟人类在身体结构的物质组成上没有差别。这样的机器人已经不是我们平常说的机器人，而是人类本身了。

综合以上分析，即便人工智能的智能远远高于人类，但由于人工智能与人类在可预见的将来还不会构成竞争和捕食关系，因此，除非人工智能具有人类式的鲜活身体，并与人类处于相同的生态位，否则人工智能就不会取代人类。

四、后人类尊严

我们之所以不断地在思考人工智能会否超越人类，一个非常重要的原因是迄今为止，人类是万物之灵，人类比其他生物都优越。如果人工智能超越了人类，人类的优越性就会受到挑战，人类的尊严也会受到伤害。人类尊严的传统观点基于某种固定的人性（如语言，理性等），但自20世纪90年代以来，我们对人性的看法已经被后人类主义（Posthumanism）话语深刻地纠正了。在《我们如何成为后人类》一书中，凯瑟琳·海尔斯（Katherine Hayles）认为人类已经与技术共同发展了数千年，并以各种方式深刻而微妙地塑造了人类。克拉克（Andy Clark）也认为，人类和技术构成了共生关系，心灵和自我分布在大脑和非生物系统中，而非局限在人类的头颅之内。人类与其他物种不同的原因是我们总是在寻找可以合作的非生物资源。各种技术和人类的相互作用证明了人类大脑的惊人可塑性。我们的思想和身体在本质上不断重新配置，外部技术设备不断融入我们的思想和个性系统。因此，人性是在漫长的进化历史中被技术逐渐建构出来的。

博斯特罗姆（Nick Bostrom）提倡了一种更具包容性的后人类尊严（posthuman dignity）概念，并用它来取代传统的尊严概念。后人类的尊严不再是基于人类对于非人类比如动物和机器的各种先天的优越性，而是基于我们现在是什么以及可能成为什么。这一切不仅依赖于我们的生物功能，更依赖于我们所处的技术和社会背景。这种尊严的概念一方面将

人类的卓越和优越性融入到人类的可完善性中, 另一方面, 它也指出了这种可完善性依赖于我们和世界的关系。也就是说, 人与技术形成了一个动态互动、相互构成和相互塑造的延展系统, 它体现出了人类面向未来的无限可能性。我们可以用可完善性来取代近代人文主义的优越性概念, 以其定义后人类尊严概念的本质。

基于这种后人类尊严概念, 即使未来某些人工智能具有超越人类的某种能力或大部分能力, 我们也不必担心人类的尊严受到威胁, 因为现在人工智能是一种人类的延展认知系统。即使人工智能在所有领域都击败了单个的人类, 它也只是证明了人类和人工智能构成的延展认知系统所具有的巨大潜能。人工智能和历史上的所有技术设备一样, 是人类智能的延伸, 它所具有的智能的越高, 就越能证明人类的可完善性, 人类也就更有尊严。

五、结语

人工智能的加速发展, 使人们意识到奇点时代的到来。越过奇点, 人们担心机器是否会超越人类以及机器是否会取代人类。在人们还没有搞清楚人的智能的本质的时候, 机器发展再快, 人工智能要想全面超越人类还很困难。即便进一步的发展超越了人类, 但只要自创生的人工智能没有出现, 或者拥有鲜活的身体的人工智能没有出现, 人工智能还不会取代人类。也许进一步的发展, 未来的人工智能掌握了人类的所有能力, 我们也不必担心人类的尊严受到了威胁, 因为, 人工智能是人类的创造, 它的智能越高, 人类也越有尊严。

参考文献:

- [1] 库兹韦尔. 奇点临近 [M]. 北京: 机械工业出版社, 2011.
- [2] [法] 笛卡尔. 谈谈方法 [M]. 北京: 商务印书馆, 2000.
- [3] [德] 莱布尼茨. 莱布尼茨读本 [M]. 陈乐民编. 南京: 江苏教育出版社, 2006.
- [4] Future of Life Institute. Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter [EB/OL]. <https://futureoflife.org/ai-open-letter/>, 2018-01-10.

[5] Hubert L. Dreyfus. What Computers Still Can't Do: A Critique of Artificial Reason [M]. MIT press, 1992.

[6] N. Katherine Hayles. How we became posthuman: Virtual bodies in cybernetics, literature, and informatics [M]. University of Chicago Press, 2008.

[7] Andy Clark. Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence [M]. Oxford University Press, 2003.

[8] Nick Bostrom. In Defense of Posthuman Dignity [J]. Bioethics, 2005, 19(3).

[9] Stevan Harnad. The Symbol Grounding Problem [J]. Physica D: Nonlinear Phenomena, 1990, 42(1).

[10] Tom Froese, Tom Ziemke. Enactive Artificial Intelligence: Investigating the Systemic Organization of Life and Mind [J]. Artificial Intelligence, 2009, 173(3).

[11] [英] 波斯特洛姆. 超级智能: 路径、危险性与我们的战略 [M]. 北京: 中信出版社, 2015.

[12] Larry L. Rockwood. Introduction to Population Ecology [M]. Wiley-Blackwell, 2006.

[13] 徐英瑾. 无用阶层论的谬误——关于人工智能与人类未来的对话 [J]. 文化纵横, 2017, (5).

[14] Patrick J. Hayes. The Frame Problem and Related Problems in Artificial Intelligence [R]. USA: Stanford University, 1971.

[15] [美] 丹尼尔·丹尼特. 认知之轮: 人工智能的框架问题 [A]. 玛格丽特·博登. 人工智能哲学 [C]. 上海: 上海译文出版社, 2006.

作者: 李建会, 北京师范大学哲学学院, 北京师范大学—香港浸会大学联合国际学院教授
夏永红, 华南师范大学公共管理学院特聘研究员

责任编辑: 周修琦