

# 论“奇点”对人类生存的挑战

陈仕伟

(江西财经大学 马克思主义学院,江西 南昌 330013)

**摘要:**“奇点”来临,人工智能在智能上将超越人类,甚至爆发“智能爆炸”。这既对人类生存产生颠覆性变革,也将带来诸多挑战。人类将越来越人工智能化,进而失去生存的意义,并导致人类将失去自主决策的能力和制定规则的权力而被动地生存。人机共存的环境资源也必将受到威胁。我们很难真正实现人工智能的伦理化,也无法在它们身上“嵌入”伦理规则。但是我们还是希望人工智能越来越友好。因此,我们有必要现在构建人工智能的责任伦理体系。

**关键词:**“奇点”;人工智能;人类生存;挑战;责任伦理

中图分类号:N031

文献标识码:A

文章编号:1008-7699(2020)01-0009-09

据库兹韦尔(Ray Kurzweil)考证,约翰·冯·诺伊曼(John von Neumann)首次提出“奇点(Singularity)”<sup>[1] 10-11</sup>真正进行全面阐述的应该是库兹韦尔本人,并预言将在2045年到达奇点。“21世纪的前半叶将描绘成三种重叠进行的革命——基因技术(G)、纳米技术(N)和机器人技术(R)”,<sup>[1]123</sup>这“三种重叠进行的革命”将为“奇点”来临奠定坚实的技术基础。一旦“机器人在智能方面超过人类的奇点”<sup>[2]231</sup>来临,人类的生存范式将发生根本性变革,人类将生存得难以想象的美好。国内学者黄欣荣教授甚至认为,大数据、互联网和人工智能将有助于共产主义的实现;<sup>[3]</sup>并且新人工智能将让人类彻底得到身心的解放,全面解放之后的人类可以尽情享受智能机器给我们带来的丰裕财富,在人工智能的帮助下阔步迈向更高境界。<sup>[4]</sup>可见,如果“奇点”确实存在的话,必将给我们的生产生活学习与工作带来颠覆性变革。同时库兹韦尔在论述GNR之后认为“会引进新的风险”。因此,“奇点”不仅给我们带来全新的生存方式,而且也必将带来诸多挑战,甚至直接威胁到人类的存继。面对这些未来的挑战,我们最好是从现在开始构建起人工智能的责任伦理体系,既是对我们自己负责,也是对人工智能负责,更是对构建良好人机关系负责。

## 一、是人工智能人类化,还是人类人工智能化?

“奇点”来临之后,意味着人工智能在智能方面超越人类,如何实现人机友好共存将是我们必须深入思考的问题。按照马克思主义的观点,“这些人把自己和动物区别开来的第一个历史行动不在于他们有思想,而在于他们开始生产自己的生活资料”,<sup>[5]</sup>即人类越来越能够脱离动物界而进行着真正人的生存:人类能够不断地制造和使用工具,不断确立起人的主体地位。因此,整个世界就出现主客二分的状况:主体就是我们人类,客体就是纳入到我们生存范围的客观世界(当然人类也能够成为自己的客体)。但是如果人工智能的智能远超越人类,就意味着具有超级智能的它们也同样能成为主体,整个世界就呈现出双主体。双主体的出现就意味着我们必须解决好如何与人工智能实现共存这一根本性问题,即到底是我们应该听命于人工智能还是人工智能应该听命于我们。

人工智能的强大要让我们来命令它们显然有点无能为力,反而是让我们越来越依赖于人工智能,特

收稿日期:2019-07-13

基金项目:国家社会科学基金重点项目“大数据技术革命的哲学问题研究”(14AZX006)

作者简介:陈仕伟(1979—),男,江西于都人,哲学博士,江西财经大学马克思主义学院副教授、硕士研究生导师。

别是在我们面对困境而不能自拔的时候,更应该听听人工智能的洞见。这样才能让我们生存得越来越好。当我们随时都要听命人工智能的时候,我们的生存就会越来越人工智能化。当然,我们可能会自豪地认为,毕竟人工智能是我们的作品。既然是我们的作品,它们就应该按照我们的要求来进行设计与制造。当我们能够利用人工智能来不断制造出新人工智能的时候,我们的意志就完全“嵌入”到它们身上,它们的行为必将最大程度地展现出人类的意志。因此,应该是人工智能越来越像我们。但是古德(I.J. Good)强调:“既然机器设计也是一项智能活动,那么超级智能机器应该可以设计出更好的机器。毫无疑问随后会出现一个‘智能爆炸’,人类的智能可能被远远甩在后面。因此,第一台超级智能机器是人类需要完成的最后一项发明”。<sup>[2]229</sup> 很难想象,在这个“智能爆炸”到来之后我们的意志在人工智能身上到底还能存留多少!特别是随着“智能爆炸”的不断发展且我们的生存已经习惯了人工智能带来的便利而越来越离不开它们的时候,人类只能越来越人工智能化,自觉不自觉地跟随着人工智能的发展而生存。因此,很有可能超级人工智能就是我们的最后一项发明。

表面上看,“我们的机器将变得越来越像我们,而我们会越来越像我们的机器”。<sup>[1]182</sup> 而实际上,应该是我们会越来越像人工智能。因为让人工智能越来越人类化会由于人工智能的过于强大而失去控制。毕竟“智能爆炸”的出现就意味着人工智能完全可以离开人类而独立完成自身的繁衍,并且每一代人工智能都将在智能方面超越上一代。因此,“人工智能计算机技术正不断取得成功,并迅速成为我们生活中不可或缺的一部分。……AI将会改变我们的生活方式、我们的社会,甚至是终结我们的族类”。<sup>[6]</sup> 我们的生存只能是越来越人工智能化!

因此,之所以人类越来越人工智能化就在于人工智能不仅在实践和意识方面和人类一样,而且在智能方面远远超越人类,以致于人类特有主体地位受到了前所未有的冲击,不得不与人工智能分享这个主体地位甚至让位于人工智能。

## 二、人类生存是有意义,还是无意义?

现在我们可以要求人工智能越来越人类化,但是“奇点”来临之后就只能是人类越来越人工智能化。既然我们将越来越人工智能化,那么我们就不得不思考生存的意义问题。意义是因为人的独有生存才得以产生,否则就无所谓意义的问题。毫无疑问,当我们完全依赖人工智能而生存的时候,人类独有的生存方式必将退却,所谓真正人的生存将难以存继,最终走向无意义化。

如果将人工智能仅仅界定为技术,那么人工智能越发展就意味着我们越能够更好认识、改造和控制世界,让世界不断向符合人类需要的方向变化。这样,我们的生存就必然越来越好、越来越有意义。比如,随着人工智能技术的发展,我们现在所进行的工作基本上都可以交于人工智能来完成,因为人工智能比我们完成得更好。如果所有的工作都可以交与人工智能来完成,我们就完全可以按照自己的个性来发展自己。可能就实现了人自由而全面发展的共产主义社会。因此,我们的生存必将按照自己的个性来不断发展而越来越有意义。

但是,人工智能不仅仅是一项先进的技术,更应该是一类具有人类意识且超越人类智能的新生命体。它们能够完成的工作远远不仅限于生产领域,而是超越人类所能涉及的所有领域,并且在任何一个领域都比我们要完成得更好。这个更好足够让我们望而生畏。比如现在的人工智能能够创作出非常优美的音乐和写出非常惬意的诗词,相信欣赏到这些音乐与诗词的任何想要通过音乐与诗词来寻找所谓“远方”的人们望而却步。因为当你经过勤学苦练达到与人工智能相媲美的境况时,人工智能瞬息之间又实现了更新换代,再次让你望而生畏。那么到底是你去寻找“远方”还是让人工智能帮你寻找“远方”呢?因此,“世上第一个超级智能一旦诞生,可能很快就会表现出强大的能力,它在任何领域的决策能力都比人类更强,不论哲学、艺术、科学、政治、社会还是人际关系。如果我们发现自己做任何事都比不过超级智

能,甚至远远不如它时,该有什么反应呢?从实际效果上来说,我们做的任何工作可能都会显得毫无意义。”<sup>[7]143</sup>

总之,人工智能的过于强大表面上会让我们的生存越来越有意义,实际上可能正在不断地消灭其意义。当生存失去意义的时候,“我们会不会所有人都为之折服,陷入绝望,任由自己纵情于沉浸式虚拟现实的灯红酒绿之中,或者干脆用电刺激大脑的方式自取灭亡呢?”<sup>[7]144</sup>赫拉利(Yuval Noah Harari)甚至认为:“由于接下来的科技发展潜力极其巨大,很有可能就算这些无用的大众什么事情不做,整个社会也有能力供养这些人,让他们活下去。然而,什么事能让他们打发时间,获得满足感?人总得做些什么,否则肯定会无聊到发疯。到时候,要怎么过完一天?答案之一可能是靠药物和电脑游戏”。<sup>[8]294</sup>如果真是如此,“奇点”不是使我们生存得越来越有意义,而是会消除我们寻找“远方”的最后避难所,生存彻底无意义。

因此,人类生存的无意义实质上就是人工智能的智能远远超越人类,导致我们赋予世界以意义的能力受到了人工智能的冲击,导致这种能力远逊于人工智能,不得不将此能力让位于人工智能。

### 三、是人类自主决策,还是被决策?

人类生存无意义的一个重要表现就是我们将逐步失去自主决策的能力而越来越被动地生存。本来,“奇点”来临之后的人工智能将为我们的决策提供更多的选择与便利,毕竟通过人工智能能够让我们更加理解整个世界,我们肯定就能够更好地作出决策而越来越自由。

恩格斯指出:“自由不在于幻想中摆脱自然规律而独立,而在于认识这些规律,从而能够有计划地使自然规律为一定的目的服务”。<sup>[9]</sup>因此,要求我们必须充分认识到“自然规律”的基础上做出正确的决策,进而寻求到真正的自由。“所谓自由,如所周知,就是能够按照自己的意志进行行为。……一个人自由与否,与他实行自己意志的自身的、内在障碍无关,而只与他自身之外的外障碍有关:自由也即不存在实行自己意志的外障碍”。<sup>[10]</sup>因此,要寻找到自由必须能够自主地作出正确决策。如果不能自主地作出正确决策,我们会完全受制于外在的障碍而不自由。而这个最大的外在障碍就是不能正确的认识和利用规律,导致陷入决定论的泥淖,因为“如果我们完全按照自然规则来生存和发展,那么我们的决策与行为就是不可控的,必须完全按照因果关系来施展”。<sup>[11]</sup>如果完全陷入决定论的泥淖,我们就完全没有自由可言。因此,我们要获得生存自由就必须自主地作出决策。如果对外在障碍的认识不够,就会在规律面前无能为力,我们的生存就完全受制于决定论而没有自由可言。如果能够借助人工智能,我们就更能够认识和利用规律,更能够根据自己的生存状况自主地作出决策,更能够获得我们所需要的自由。

但是人工智能的过于强大导致我们作出的一切决策都显得非常幼稚,还不如干脆让人工智能来帮我们作出。我们只需要被动地接受人工智能作出的决策就能够获得自由。人工智能具有的强大算法功能能够洞察世间的一切奥秘,并且它们的“计算”能力远超人间一切天才。因此,与我们比较,它作出的决策绝对是“最优”的。既然人工智能能够为我们作出“最优”的决策,那么我们就没有必要为作出自己的决策而苦苦煎熬,倒不如干脆放弃作出决策的权力,将这个权力转让或者异化给人工智能。这样,虽然我们放弃了自主作出决策的权力,但是也并没有因此而丧失自由。

问题的关键就在于,这样的生存方式到底是不是获得了我们所需要的自由。因为我们自主作出决策就意味着我们必须为此承担相应的伦理责任,在承担伦理责任的基础上继续作出更佳决策进而获得更大的自由。如果完全依赖于人工智能为我们作出决策,姑且不论人工智能作出的决策是否会导致我们朝着机器化的方向发展,关键是我们无需承担任何伦理责任。丧失责任的自由是不是真正的自由?显然这是一种动物式的自由,是被动地适应客观世界的生存方式;而人的自由是通过我们的决策及其形成的行为积极主动争取来的自由,是必须要承担相应伦理责任的自由。因此,“奇点”来临之后可能会导致我

们完全陷入人工智能造就出的新决定论境地,完全失去自主决策的权力而越来越被动地生存。

因此,从根本上说人类丧失决策的权力是我们的主体地位受到前所未有冲击的结果。由于主体地位的动摇,人类的决策权力必然要让位于享有超过人类主体地位的新主体即人工智能。如果人类还完全享有决策权力的话,那么人工智能就不可能享有相应的主体地位。

#### 四、是人类制定规则,还是接受规则?

正因为“奇点”的到来将导致人类丧失决策的权力,因此人类只能被动地接受人工智能制定的规则。这是人类生存无意义的另一个重要表现,即我们将在逐步失去自主决策能力的基础上丧失制定规则的权力。

人天生就是社会性动物,而社会性动物就意味着必须守规则,而规则就是人类为了实现共同目标而协商制定出来的。随着生存环境的变化,人类不断通过协商、废除、修改和制定规则。只要规则能够不断适应人类的生存现实,我们就应该严格遵守规则以便寻求到更好的生存。这就要求,要使我们的规则越来越符合人类的生存现实就必须不断地认识它。如果认识不清楚,我们的规则就不可能是好的规则,我们的生存就不可能越来越好。因此,我们要制定出好的规则,关键在于一方面能够认识清楚我们的生存现实,另一方面根据此前的认识进行协商。

显然“奇点”来临之后,人类肯定会利用人工智能来帮助我们认识清楚这个生存现实,并希望为我们制定规则提供重要参考。但是,完全依赖于人工智能制定出来的规则是否就完全适合于我们呢?更何况,人工智能作为新的生命体,肯定也需要遵守规则。这就涉及到这样的问题:以往制定出来的规则都是人与人相互协商的产物,但是现在有了人工智能的参与,制定出来的规则就不是人与人相互协商的产物,而是涉及到人与人工智能、人工智能与人工智能和人与人的相互协商。但是基于智能水平的差异,人类的认识与人工智能的认识肯定存有差异,我们肯定无法与人工智能进行平等地协商进而制定出适合人类与人工智能的规则。因为“人类技能的增长速度越来越缓慢,而计算机功能的增速却越来越快”,<sup>[12]</sup>智商极限在150的我们“或许已经能够接受智商为250的机器人,因为这接近于人类生物智能的智商范围。但是,对于智商1000的机器人或1万的AI来说,我们仍然心存不安”。<sup>[13]60</sup>因此,我们只能被动地接受智商比人类高出数个数量级的人工智能制定出的规则。

即使我们能够表面上平等地坐下来与人工智能进行协商,但是如果在协商的过程中出现争执,该如何处理呢?可以肯定,我们只能听命于人工智能。另外,既然人工智能是具有生命的新主体,它们在制定规则的时候也肯定优先考虑自己。长此以往,人类必将交出制定规则的权力而异化给人工智能。如果如此,人类将永远只能被动地接受规则而不能主动地废除、修改和制定规则。这样的结果就是人工智能开始设计起人类,而不是人类设计人工智能。我们的生存也许就是这样:“地球可能会变成一座没有围墙的动物园,一个实实在在的陆地动物饲养所,哪里只有阳光和孤独,我们的机械看管者为了维护正常的运转偶尔会推动我们一下,而我们会为了自身的幸福高举双手欢迎这样的帮助”。<sup>[14]</sup>

因此,由于智能水平的差异,我们必将逐渐失去制定规则的权力,进而完全依赖于人工智能来帮助我们制定规则。长此以往,我们就成了一个被动的执行者,更甚者,我们会逐渐丧失对我们的生存现实的判断能力。因为我们无需制定规则,根本不需要对我们的生存现实进行任何认识与判断。这样,我们反而成了被动执行指令的机器,并逐步被淘汰为动物甚至机器人的宠物。实质上,人类失去制定规则的权力就是人类已经无法控制人工智能,导致人工智能异化,即原来的主体(人类)转变为客体,原来的客体(人工智能)转变为主体。简言之,人工智能已经不是人类的创造与作品,反而成了控制和规定人类的主体。

#### 五、是环境资源优越化,还是危机化?

由于人工智能的过于强大,不仅人类生存的内在条件正在不断发生着不利于人类生存的变化,而且

人类生存的外部条件同样也正在不断发生不利于人类生存的变化,环境资源不是越来越优越,而是越来越危机,即可能会对人机共存的环境带来巨大挑战,直接威胁到我们生存所需的环境资源。现有的研究往往忽视这方面的挑战。

众所周知,人类生存环境的恶化从根本上说是由于环境资源的有限性,并且人的数量在不断增长以及需求在不断提升,难免会超出这个星球的承受能力。因此,要使我们能够一劳永逸地生存下去,首先就需要提高环境资源的利用效率以减少浪费;其次是将我们的需求控制在地球可以承受的范围;再次是大力发展可再生资源;最后是努力发展可替代资源。但是在现有的技术条件下,任何一个方面都显得举步维艰。不过库兹韦尔认为,“奇点”来临之后,特别是通过发展纳米技术,就会“使发展清洁、可再生、分布式和安全的能源技术成为可能”。<sup>[1]147</sup> 不仅资源的利用效率极大提高而几乎不产生浪费,而且我们将不再像现在一样依赖不可再生资源。因此,环境资源必将越来越优化,生存环境必将越来越好。

其实,库兹韦尔是根据现在的状况来预测未来,并没有意识到未来的人工智能和人类一样也需要消耗大量的环境资源。如果仅仅是我们需要消耗环境资源,当然可以通过发展高新技术来缓解甚至解决这个问题。但是人工智能本身所要消耗的资源又应该如何来计算呢?特别是,“智能爆炸”的到来完全有可能导致环境资源更加恶化。“智能爆炸”意味着人工智能数量的发展完全超出我们的控制。随着人工智能的数量呈现指数增长,制造人工智能所需要的特殊资源也必将呈现出指数增长。现在我们所能预测的是:“所有这些数字设备将需要更多的‘稀土’,更多的钨钽铁矿石,更多的锂以及许多其他日益稀缺的资源”;<sup>[15]196</sup>“为了查收电子邮件和短信,你需要诸如笔记本电脑、平板电脑和智能手机等设备,导致对锂和钨钽铁矿石等能源资源的需求也成倍增长”。<sup>[15]197</sup>而这些制造人工智能的特殊资源基本上都是不可再生资源。当然你可以说,人工智能所需要的资源不是人类所需的,因此,对我们的生存不产生任何实质性的威胁。但是不要忘记了,“奇点”来临之后,我们的生存已经与人工智能紧密结合在一起,如果人工智能不能得到持续发展,最终的结果就是我们也同样不能持续发展。

其实,不仅“智能爆炸”会带来新的环境资源危机,人工智能的不断发展导致效率的提高同样会带来环境资源的新危机。因为人工智能的发展仅仅是提高了效率进而减少浪费。但是从历史视角分析,以往的任何一次科技革命都提供了新技术,能够不断地提高环境资源的利用效率,但是效率提高了,我们的需要在质与量两方面也增长了。为了满足我们的需要,科技的进步必然得到进一步发展的动力。长此以往,人类所需的环境资源只能随着效率的提高而不断扩大,以至于超过地球的承受能力而不断产生新的危机。就目前而言,我们大谈特谈人工智能,也仅仅是因为提高了效率,并没有产生根治危机的效果。

因此,“奇点”的到来不仅使人类生存的内在条件(主体地位、赋予世界以意义的地位、决策权力和制定规则权力)受到前所未有的冲击,而且人类生存的外在条件即环境资源也受到了重大冲击,意味着人工智能对人类生存所产生的颠覆性影响是全方位的。

## 六、是促使人工智能理性化,还是理性化摧毁一切?

正因为人工智能对人类生存所产生的影响是全方位的,即“奇点”来临之后我们越来越像人工智能、我们的生存越来越无意义化、我们的自主决策权力正在被剥夺、甚至逐渐失去制定规则的权力、并且连我们生存所需的外部条件即环境资源也受到了威胁,那么我们就必须未雨绸缪地采取措施以避免这种状况的发生。现在看来,我们最容易想到的是赋予人工智能以理性。如果人工智能和我们一样具有人类的理性,那么人工智能的所作所为就会按人类的理性来展开,即可以借用理性来实现人工智能不断服务于人类!但是这可行吗?

从历史分析,“世界在不断进步,我们很幸运,因为当今社会比人类历史上任何时候都更加和平、有序,当代人培育、提高自己智力、艺术和情感能力的机会空前增加”。<sup>[7]130</sup>而整个人类历史能够呈现出进步

的状态关键在于人是有理性的。对于技术乐观主义者而言,人工智能的飞速发展并没有必要让我们保持高度的警惕与担忧,毕竟人工智能是人类发展的技术产物,并且“超级人工智能将会把理性作为自己的行动指南,并因此变得仁慈”。<sup>[7]131</sup>因此,人工智能必将在我们的作用下发展成为对人类越来越友好的人工物,毕竟它们会拥有和我们一样的理性。如果这样,我们的生存会随着人工智能的发展而不断进步。

确实,如果人工智能能够越来越理性化,对人类生存而言肯定是积极的。但是,如果人工智能越来越具有人的理性,那么人的理性是否存在着不可克服的矛盾?如果人的理性存在着不可克服的矛盾,那么拥有人的理性的人工智能是否也同样存在着不可克服的矛盾?如果人工智能因为具有人的理性而产生冲突,我们是否还有机会生存下去呢?

可以肯定,人类历史上曾经出现的各种危机都是理性的产物。至今让我们记忆犹新的两次世界大战都是理性的产物,还有冷战时期的美苏争霸导致随时都有可能发生毁灭地球的核战争阴云同样也是理性的产物,相信现在出现的各种生态危机也是理性的产物。因此,人类的理性并不完美,特别是理性冲突不可避免。如果人工智能将我们相互冲突的理性作为自己的行动指南,那么人工智能之间的冲突同样不可避免。一旦人工智能之间产生冲突,这样的冲突所产生的破坏力肯定要比两次世界大战超出数十倍甚至更多个数量级,必将危害到我们赖以生存的整个星球,世界末日就不是仅仅出现在宗教神话和科幻小说中。

可能更为恐怖的是,“奇点”来临之后的人工智能就像我们一样能够自我繁衍,它们的理性必然同样能够“自我繁衍”下去而产生出新的理性体系,那么人类的理性如果一旦和人工智能的理性产生了冲突,那么我们应该怎么办?是完全屈服于人工智能的理性还是奋起反抗而迫使人工智能继续接受我们的理性?显然,“完全屈服”可能是最好的选择,因为奋起反抗的结果只能是自取灭亡。有人可能会认为,我们还可以继续发挥理性狡计,比如利用拔掉插头以切断电源、贿赂、说服甚至威胁等方式,让人工智能不断服务于人类。但是现在我们就可以预见:“超人工智能将会变得超级精明,要靠拔插头解决问题绝不可能”,<sup>[16]177</sup>且“超人工智能将足够了解人类心理,可能知道什么样的贿赂或威胁可能起作用,也可能知道哪些人最容易被哪些说服方式说服”。<sup>[16]178</sup>显然在人工智能面前想通过理性狡计这样的雕虫小技来“迷惑”它们简直就是天方夜谭。而且我们更应该担心:这样的雕虫小技一旦不能(也根本不能)取得成功而被它们识破且报复我们的悲惨境况。

因此,促使人工智能越来越理性化的结果可能会违背我们的初衷,最终的结果很有可能就是摧毁一切。这就意味着促进人工智能朝着有利于人类生存的方向发展不能寄希望于将人的理性“贯穿”到人工智能身上。

## 七、是促使人工智能伦理化,还是限于新矛盾中?

可见,借用人类的理性来实现人工智能不断为人类服务的目标是不可实现的,当然我们不能就此而放弃促使人工智能不断服务于人类这一目标。除了赋予人工智能以理性之外,我们还可以通过“嵌入”伦理规则的方式来促使人工智能不断服务于人类。

早在1940年12月,刚刚20岁的阿西莫夫(Isaac Asimov)就曾站在人类中心主义的立场阐述了他的机器人三定律。<sup>[13]63</sup>其实早期的人工智能研究者都认识到一旦人工智能超越人类将会产生可怕的后果,特别是“聊天机器人”的发明者魏森鲍姆(Joseph Weizenbaum),在意识到人工智能带来的可怕后果后就开始研究人工智能带来的社会责任和道德问题。<sup>[17]</sup>因此,要使这种可怕后果不要发生的可能最佳方式就是为人工智能立法。比如,未来生命研究院提出“确保AI有益于人类社会”,美国计算机协会发布算法透明和可责性七项原则,德国提出全球首个自动驾驶车伦理原则,韩国国会提出《机器人基本法案》和IEEE发布《人工智能设计的伦理准则(第二版)》,等等。当然,为人工智能立法的最好办法就是在设计过程中“嵌

入”伦理规则,保证人工智能不断遵守伦理规则而为人类服务。段伟文认为,应该进行“探讨伦理嵌入的机器伦理研究”,“不仅是人工智能的设计者、控制者和使用者应该遵守特定的伦理规范,而且还需要将特定伦理规范编程到机器人自身的系统中去。”<sup>[18]</sup>这样,就能够通过“嵌入”伦理规则促进人工智能成长而培育出具有道德判断能力的可能,使人工智能成为道德智能体。

如果能够通过“嵌入”伦理规则而不断实现人工智能的伦理化,那么必将对我们的生存产生积极效果。但是问题的关键是我们应该为人工智能“嵌入”怎样的伦理规则。伦理规则和理性一样具有相对性,伦理规则之间的冲突不可避免,伦理规则与现实之间的冲突同样不可避免。就伦理规则之间的冲突而言,我们对伦理规则背后的基本伦理理念的理解仍然存在激烈的冲突,不同的理解必然会产生不同的甚至相互矛盾的伦理规则。人工智能在遵循这些伦理规则时必然会处于两难的悖论境地。比如,如果“嵌入”以人为中心的功利主义规则,那么首先就必须回答什么功利、什么是功利最大化和伤害最小化、功利到底如何计算,等等。如果在这些问题还没有搞清楚之前就将其中的伦理规则“嵌入”人工智能,那么它们能一以贯之吗?

另外,伦理规则一旦制定出来就意味着具有一定的普遍性,能够在任何情况下适用。那么将伦理规则“嵌入”人工智能之后,它们是否也会像我们一样,会出现一旦产生具体冲突该如何遵循的困境?比如现在人们普遍思考的无人驾驶。当无人驾驶汽车在面对不可避免的车祸的时候,到底是优先保护车上的人还是车外的人?即人工智能是否也像人类一样在普遍与特殊之间的矛盾中不能自拔?如果出现这样的状况,估计人工智能除了“死机”之外没有其他更好的办法。这就表明,人工智能很难将我们“嵌入”的伦理规则真正贯彻到底。

因此,一个具有人类意识的人工智能如果不能像我们一样严格遵守相应的伦理规则,肯定是可怕的;但是当我们在思考如何将伦理规则“嵌入”人工智能而使其伦理化的时候,却发现我们不知道应该“嵌入”什么样的具体伦理规则;如果真的“嵌入”,人工智能必然会像我们一样陷入普遍与特殊之间的矛盾中。

## 八、是人工智能越来越友好,还是越来越邪恶?

既然“奇点”的到来给人类的生存带来了前所未有的冲击,并且我们通过赋予人工智能以人类特有的理性和“嵌入”伦理规则都难以实现人工智能不断服务于人类,那么我们就不能不思考一个十分紧迫的问题,人工智能到底是越来越友好还是越来越邪恶?当然,如果从技术中立的角度来思考,人工智能肯定是既不好也不坏。但是技术中立性的观点已经不再适用于人工智能。因为人工智能不再像过去的技术一样呈现出死的技术产品,而是有意识且能够自我繁衍的新生命体。它的好与坏直接决定了在“奇点”来临之后,我们的生存是越来越好还是越来越坏。我们的生存与人工智能的好与坏产生了共进退的局面。但是,我们还是希望人工智能越来越友好。从人类产生至今的历史,我们完全有理由保持乐观的态度。世界只能越来越好而不是越来越坏,就像历史上曾经出现的历次科技革命一样,在革命的前期与中期人类总是会产生各式各样的悲观情绪,而在革命之后甚至更长的一段时间,突然发现每次科技革命都给我们带来了发展与进步。因此,可能现在我们对人工智能充满着各种担忧,相信一旦真正实现了“奇点”,完全有可能是另外一番欣欣向荣。

但是历史的乐观并不意味着我们对于“奇点”就可以保持十足的乐观,毕竟以往的科技革命都是我们控制下的技术进步,而“奇点”来临之后,人工智能将成为与我们并列站在客观世界面前的一个新物种,特别是“智能爆炸”,意味着我们的地位将发生根本性的改变。人工智能的过于强大很难使我们与之进行难分高下的竞争,结果只能是我们的生存越来越人工智能化,而不是人工智能越来越人性化,我们可能要不断地依附于人工智能这个新物种。当我们的生存越来越人工智能化,我们的生存也会走向无意义化,显得一切都不如委托给人工智能。因为这样,我们听命于人工智能的决策而形成行为就不用承担任何伦理

上的责任,“无为”地生存下去可能是更好的生存。

当然,面对强大的人工智能,如果我们能够“无为”地生存下去,也不失为一种较好的生存,反正我们可以按照自己想要的方式好好地生存下去。但是想要与人工智能保持相安无事地和平相处也有一定的困难。关键是人工智能可能具有我们的意识与理性,可能也正在践行着一整套独特伦理规则。它们的意识、理性和伦理规则能够和人类的和平相处吗?如果不能,我们就不要希望获得任何一方净土。其实我们现在思考人机共存环境的时候,很大程度上是鉴于技术具有中立性的原因。因此,人工智能不会主动出击。但是人工智能不会是完全中立的,毕竟它们是具有意识和理性的新生命体,更为重要的是,我们与人工智能在环境资源方面不可避免地会发生冲突。只要有冲突,哪怕是很小的冲突,谁都不能保证人工智能不会主动出击。赫拉利就曾悲观地预言:“一旦人工智能超越人类智能,可能就会直接消灭人类。人工智能这么做的理由,一是可能担心人类反扑、拔掉它的插头,二是要追求某种我们现在还难以想象的目标。”<sup>[8]294</sup>如果这样的情况一旦发生,我们人类的生存就难以为继。因此,谁能保证人工智能不会越来越坏呢?

因此,与以往的科技革命相比,“奇点”的到来只能导致人工智能在智能方面越来越超越人类,越来越失去人类的控制,甚至最终取代人类。这样的结果就是人类特有的主体地位必将完全为人工智能所取代。如果人工智能一旦与人类发生冲突,人类的命运就难以想象,人工智能就是邪恶的。

### 结束语:构建人工智能的责任伦理体系

综上所述,为了保证我们能够在2045年之后生存得越来越好,有必要现在未雨绸缪地采取措施,即担负起未来的责任。库兹韦尔指出:“科技是一把双刃剑,在奔向奇点的道路上,与走向奇点的可能一样,我们也很可能会走向岔路,造成令人担忧的后果。”<sup>[1]225</sup>现在看来,能否使未来的人工智能越来越友好,关键在于我们自己。因此,我们有责任、有义务促使人工智能越来越友好、越来越能够促进我们的生存。从责任伦理角度思考,这个责任与义务可能就是要求我们,从现在开始要逐渐构建起人工智能的责任伦理体系,负责任地研究、开发和运用。对人工智能担负起全面的责任,这既是我们对人工智能担负起全面责任,实质上也是我们对自己担负起全面的责任,从而实现人工智能也能够担负起全面的责任(如果人工智能真的能够成为继人类之后的新的有意识的生命体)。唯有如此,人工智能才能真正不断地服务于我们。正如现在我们面临的诸种生态危机一样,由于一开始没有对地球上的有生命和没有生命的存在物承担起应有的责任,导致我们现在处于各式各样的危机之中,直接威胁着人类的正常生存。如果我们一开始就承担起了应有的责任,诸种生态危机或许根本就不存在。也正因为一开始没有承担起相应的责任,导致我们现在不得不承担更大、更多、更沉重的责任。我们现在研究、开发和利用人工智能又何尝不是如此呢?因此,我们现在构建人工智能的责任伦理体系已势在必行,只有在研究、开发和利用人工智能的过程中不断承担起相应的伦理责任,才能在未来的人工智能发展过程中实现责权利的统一,最终促使人工智能不断为人类服务。不仅要负责任地研究、开发和利用人工智能,还必须使人工智能不断地成为我们的创造和作品;不仅要负责任地构建良好的人机关系,还必须实现在人类主导下人类与人工智能的和平共处;不仅要现在承担相应的伦理责任,还必须未雨绸缪地承担起未来的伦理责任;不仅要负责任地不断促进人工智能为人类服务,还必须在此基础上为人工智能的发展创造良好条件。总之,我们必须负责任地研究、开发和利用人工智能,不仅要承担好现在的责任,更应该面向未来而勇于承担未来的责任。

### 参考文献:

- [1]库兹韦尔.奇点临近[M].李庆诚,董振华,田源,译.北京:机械工业出版社,2016.
- [2]多梅尔.人工智能[M].赛迪研究院专家组,译.北京:中信出版社,2016.
- [3]黄欣荣.大数据、人工智能与共产主义[J].贵州省党校学报,2017(5):115-122.

- [4]黄欣荣.人工智能与人类未来[J].新疆师范大学学报(哲学社会科学版),2018(4):101-107.
- [5]马克思恩格斯选集:第1卷[M].北京:人民出版社,2012:146.
- [6]亨德勒,穆维西尔.社会机器[M].王晓,王帅,王佼,译.北京:机械工业出版社,2018:13.
- [7]蔡斯.人工智能革命[M].张尧然,译.北京:机械工业出版社,2017.
- [8]赫拉利.未来简史[M].林俊宏,译.北京:中信出版社,2017.
- [9]马克思恩格斯选集:第4卷[M].北京:人民出版社,2012:491.
- [10]王海明.伦理学原理[M].北京:北京大学出版社,2009:241-242.
- [11]陈仕伟.大数据时代的决定论与自由意志及其伦理问题[J].山东科技大学学报(社会科学版),2017(6):18-23.
- [12]科尔文.不会被机器人替代的人[M].瑜婷,译.北京:中信出版社,2017:9.
- [13]托斯.人工智能时代[M].赵俐,译.北京:人民邮电出版社,2017.
- [14]卡普兰.人工智能时代[M].李盼,译.杭州:浙江人民出版社,2016:200.
- [15]斯加鲁菲.智能的本质[M].任莉,张建宁,译.北京:人民邮电出版社,2017.
- [16]博登.AI:人工智能的本质与未来[M].孙诗惠,译.北京:中国人民大学出版社,2017.
- [17]尼克.人工智能简史[M].北京:人民邮电出版社,2017:143.
- [18]段伟文.人工智能时代的价值审度和伦理调适[J].中国人民大学学报,2017(6):98-108.

## On the Challenges of “Singularity” to Human Existence

CHEN Shiwei

(School of Marxism Studies, Jiangxi University of Finance and Economics, Nanchang 330013, China)

**Abstract:** With the advent of “singularity”, artificial intelligence(AI) will transcend human beings in the intelligence, and even “intelligence explosion” will emerge, which will not only produce the sweeping changes, but also bring lots of challenges to human existence. Human beings will become more and more artificial, and then lose the meaning of existence; what’s more, human beings will lose the ability to make decisions autonomously and the powers to make rules, so human beings will have to live passively. Environment resources of man-machine symbiosis will also be threatened. It is difficult for us to truly realize AI’s ethics and to “embed” ethical rules on them. We still hope that AI is becoming more and more friendly, therefore, it is necessary for us to construct the ethical system of responsibility for AI now.

**Key words:** “singularity”; artificial intelligence; human existence; challenges; ethics of responsibility

(责任编辑:黄仕军)