

未来的机器人将取代人类吗?

□ 韩东屏

摘要: 未来的智能机器人是否将取代人类? 对此问题的不同回答, 将决定我们该如何对待人工智能(AI)技术的发展, 而学界给出的回答是“取代论”和“否取代论”两种。取代论的 AI 技术发展目标理论和 AI 技术奇点理论都认为, 将来可以造出能超越和取代人类的超级人工智能, 并为实现超级人工智能提供了六条实现路径。否取代论则从三个方面论证未来的机器人不能取代人类: 一是技术性论证; 二是差异性论证; 三是“超而不代”的论证。纵观双方讨论, 尽管取代论的理论含有假说成分, 可否取代论的所有否定仍无效, 因为这些论证基本上都是以弱人工智能而不是以超级人工智能为前提的。所以, 不管取代论立论究竟如何, 否取代论都是落败的状况。但这并不意味着取代论就是正确的, 因为只要否取代论还能提供新的论证, 或者取代论还没确证未来机器人将取代人类的必然性, 取代论就还存在被否定的可能。可以肯定, 机器人要想取代人, 必须具有比人类更强的创造力, 但恰恰在这一点上取代论语焉不详。由于不同种类的实体必有不同的性能, 因而机器人及其机脑只要在质料、结构、形态上与人及其人脑存在一定差异, 就不可能具有人或人脑的所有性能, 尤其是不可能具有人之所以为人的最高能力——创造力, 并且即便假设机器人也能具有创造力, 机器人还是不可能取代人类。因为没有肉身的机器人由于没有需要, 就不会有自主意识, 也就不能成为主体, 于是只能是听从人的指令并服务于人的工具。

关键词: 机器人; 取代论; 否取代论; 创造力; 需要

中图分类号: B82-057

文献标识码: A

文章编号: 1671-7023(2020) 05-0008-09

美国人乔治·德沃尔在 1954 年造出世界上第一台可编程的机器人之后, 就时有人出来发布未来机器人会在能力上超越人类并最终取代人类的警告, 只是此类“取代警告”或者出自科幻作家的想象, 或者属于个别学者的单纯预言, 缺乏说理。事实也是如此, 那时计算机论域中与人相关的讨论, 还聚焦在电脑或人工智能是否会有类似于人的意识或智能的问题上, 而绝大多数学者对此问题的回答持否定性的观点。

可是近些年的情况明显不同。一个不同是, 发出“取代警告”的人已经不是科幻作家与个别学者, 而变成许多专家, 其中就有当今世界著名的科学家和技术家。如著名的物理学家霍金曾警告“人类由于受到缓慢的生物进化的限制, 无法与机器竞争, 并会被取代, 全人工智能的发展可能导致人类的终结。”^[1] 马斯克则是著名的技术家, 他也将人工智能说成人类“最大的生存威胁”, 并将发展人工智能比喻成“召唤恶魔”^[2]。另一个不同是, 此时的“取代警告”已经有了理论论证以及某种程度的事实佐证和可感受性。尤其是当人们发现战胜人类围棋世界冠军的智能机器阿尔法狗居然会自己学习时, 就不得不在惊叹之余, 相信人工智能还会有越来越快速的发展和越来越令人惊诧的表现。正是在这样的背景下, 关于未来的机器人是否将取代人类的讨论就空前火热地在学界展开了, 众多学者纷纷参与, 各抒己见。

这个“取代警告”之所以能成为令众人瞩目的热点问题, 在于对此问题的不同回答将决定人类社会对人工智能技术的发展会采取不同的态度和制度安排: 要么是积极推动, 要么是有所限制。此外, 不同的回答还意味着我们能据此判断, 现在除了“取代警告”之外的那些关于人工智能问题的讨论, 哪些问题是有意义的? 哪些问题由于不会出现而没有意义? 比如, 如果未来不存在机器人超越和取代人类的可能, 我们就没有必要继续研究以这种“可能性”为前提的人工智能问题。因此, 关于“取代警告”的讨论不仅必要和重要, 而且还应尽早形成共识性的明确结论。

作者简介: 韩东屏, 华中科技大学哲学系教授、博士生导师

收稿日期: 2020-04-10

那么,这个明确结论会是什么样的观点,这就需要对讨论中已有的各种观点进行全面的分析评价,再做最后判定。

目前学界讨论“取代警告”的学者,主要是科学家和哲学家,他们对此问题形成两种截然相反的回答,一种是给出肯定回答的“取代论”,即人工智能将来会超越人类并取代人类,成为地球新主宰;另一种是给出否定回答的“否取代论”,即人工智能永远也不可能超越人类和取代人类。

以下先对这两种回答分别进行考察,再逐步形成最终结论。

一、取代论

霍金、马斯克等发出“取代警告”的理论依据,是这些年盛行于人工智能学界的 AI(人工智能 artificial intelligence) 技术发展目标理论和 AI 技术奇点理论。

(一) AI 技术发展目标理论

AI 技术发展目标理论是经 20 世纪中叶的 L.G. 古德到现在的尼克·博斯特罗姆等一众科学家的论述而形成的理论。他们将 AI 技术的发展分为弱人工智能、强人工智能和超级人工智能三个阶段,并将超级人工智能作为 AI 技术发展的终极目标。时下已有的智能机,不论是非人形的智能机,还是人形智能机即智能机器人,如美国的阿尔法狗、埃及的索菲亚、日本的 Shibuya Mirai,均属于第一阶段的弱人工智能,是基本上只能完成一个单纯任务的专用智能机。第二阶段的通用人工智能属于强人工智能,能拥有像人一样的能力,可以通过学习胜任人的任何工作,只是没有创造能力和自主意识。第三阶段的超级人工智能更胜一筹,不仅具备人类的创造能力和自主意识,而且还比人类的所有能力都更为强大,可谓超人人工智能^[3]。古德在 1956 年首次提出超级人工智能设想的论文中说:“一台超级智能机器可以定义为是一台在所有智能活动上都远超人类——不管人有多聪明——的机器。由于机器设计属于这些智能活动的一种,那么一台超级智能机器当然能够设计更出色的机器,那么毫无疑问会出现一场‘智能爆炸’,把人的智力远远抛在后面。因此,第一台超级智能机器也就成为人类做出的最后的发明了——前提是这台机器足够听话且愿意告诉我们怎样控制它。”^[4]而用博斯特罗姆的话说就是“完全成熟的超级智能能够出色完成智能升级、战略策划、社会操纵、黑客技术、技术研发、经济生产等任务,这也是为什么科学家会认为超级智能将控制人类。”^[5]

与当年古德对超级人工智能只有设想不同,现在已经有了多种实现路径,它们来自诸多专家的不同构思和论述,大体可概括为如下六条。

第一条路径是全脑仿真或人脑复制。它直接通过扫描人的大脑,将扫描得到的原始数据输入计算机,然后在一个足够强大的计算机系统中输出神经计算结构,造出仿人脑的电脑^[4]。

第二条路径是提升复杂度。博斯特罗姆在《超级智能》中指出,“我们之所以能控制地球,是因为我们的大脑比即使最聪明的动物的大脑都要复杂得多”^[5]。据此,只要设法让电脑的“神经元”也足够多,电脑的系统也足够复杂时,电脑就能达到乃至超越人脑的意识。

第三条路径是功能模仿。此路径的提出者认为,人体有 11 种系统功能,每个系统的最底层都是微处理器,因而只要模仿这 11 种系统功能来建造人工智能系统并将它们合在一起,就可以使智能机器人达到人类智能的水平^[6],继而,就可以实现超越。对人工智能发展有重大贡献的马文·明斯基说:“如果我们能制造一个和人一样聪明的机器人,那我们也可以制造一个比人更聪明的机器人。”^[7]

第四条路径是机器与生物相结合。其要旨是将人体基因或仿人体基因植入机器人,使机器人不仅有比人更强大的逻辑思维运算能力,而且也有欲望、情感和非逻辑思维的智慧。比如有位韩国科学家就正在研究如何将人造染色体赋予机器人,使之也有性欲^[8]。

第五条路径属于本质性建构。这就是“学习胡塞尔的‘想象力自由变更’的办法,对‘智能’的本质进行直观剖析。……该办法的具体操作步骤是:对各种可能的智能类型进行展列,并由此为出发点对各种可能的智能形式进行想象,最终剔除关于智能的偶然性成分,找到智能的本质性要素”,从而用它制造超级人工智能^[9]。

第六条路径是制造以量子计算为基础的人工智能。随着量子力学的发展及其对量子特性了解的增

多,许多著名科学家都“提出了人类意识的量子假设”,猜测“人类智能的底层机理就是量子效应”。于是有人认为,“以量子计算为基础的人工智能”可以成为超越人类能力的人工智能,这种超级人工智能会成为取代人类的“后人类”^[10]。

(二) AI 技术奇点理论

AI 技术奇点理论是享誉世界的美国发明家和未来学家雷·库兹韦尔提出的,他在 2005 年出版的《奇点临近》一书中,预言人工智能将在 2045 年超过人类智能,而这一刻,就是 AI 技术的“奇点”。奇点预言的根据,是库兹韦尔本人受启发于摩尔定律所归纳出的“加速回报定律”。摩尔定律是指半导体芯片上集成的晶体管 and 电阻数量将每两年增加一倍,加速回报定律是指“对技术史的分析表明,技术变革是指数性的,与常识性的‘直觉的线性观’相反。所以我们在 21 世纪将不会经历 100 年的进步——它将更像是 2 万年的进步(以今天的速度)”。正因如此,“几十年内,机器智能将超越人类智能,并导致技术奇点的来临”^[11]。由于库兹韦尔以往的多种科技预言都足够灵验,尤其是他依据“加速回报定律”对人类基因组计划的完成时间为 7 年的预测,虽然与当时所有人预计的时间至少都要少一半以上(该研究项目计划用 15 年完成,对此很多人还是认为根本不可能),但最终的完成时间证明他的预言是准确的,所以他的 AI 技术奇点理论一经提出,也得到学界很多人的广泛认同,乃至可以开办奇点大学,广招天下学员^{[1][12][13]}。

“取代论”除了从正面用超级人工智能理论和 AI 技术奇点理论为自己立论外,有时也从反面通过对已经存在或可能的挑战观点进行预防性反驳。一个反驳针对的是“机器人作为无机物构成,不可能超过有机物”的观点。其驳辞是“然而没有人能够证明思想一定和身体有关。就信息处理和数据输出获取而言,机器人比人脑有更多的优越性”^[14]。一个反驳针对的观点是:模拟者怎么会超越被模拟者?人工智能怎么会超越被模拟的人类?其驳辞是“没有任何一条定理证明被创造的东西不能超过造物主,现实生活中反而有很多相反的例子,比如孩子比父母聪明,就非常典型”^[15]。

二、否取代论

否认未来的机器人将取代人类的学者似乎更多,他们中既有科学家,也有哲学家,偶有作家。由他们所分别给出的否定性论述也相当多,最终均可概括为以下三个方面的论述。

第一方面的否认可谓技术性论证。这就是从技术的维度出发,证明人工智能不可能超越人类智能。这类证明又有四种,基本上都是由人工智能专家做出的。

其一是对 AI 技术奇点论的否定。对其否定的依据有三个理由。第一个理由是奇点论者“忽略了十分重要的一点:计算速度的提升并不等同于智能的提升”。因为智能的提升要靠深度学习的算法的改进来实现,而计算速度的提升则“只是硬件提升和数据量增大带来的规模效应”。虽然“基于深度学习算法的人工智能系统在近些年取得了令人瞩目的成就”,但这只是技术应用的成就,并不是深度学习算法本身的提升,更不是什么指数级的提升^[11]。第二个理由是经验表明,“各种类型的人工智能系统在数十年来的实际发展中都经历着收益递减的过程。在研究初期,人工智能系统通常可以快速提升,甚至在某些时刻超越技术奇点理论所设想的指数增长速度,但随着完善度和复杂度的增加,人工智能系统往往会遭遇各类难以改进或跨越的瓶颈,导致无法维持固定的改进速率”^[11]。简单说,人工智能越是往上发展,需要解决的问题就越多、越复杂、越困难,这就导致研究投入的收益率也越来越低。第三个理由是奇点论者认为,“递归自我改进系统是实现技术奇点的有效途径”,这里不仅有硬件方面的困难,而且“递归自我改进系统在自我指涉方面也存在着严峻的挑战”。这就是,系统复杂程度的不断提升,将导致理解自身所需要的智能也不断提升。于是,要想使智能机器有一点儿智能提升,就得先有最高级的自我理解能力系统来适应,这说明人工智能的递归自我改进效率受到自我理解能力的限制,不会有高速度的改进^[11]。

其二是对超级人工智能的否定。该观点认为人工智能将来无论怎么发展,都不可能实现超级人工智能。首先,智能机器总是在人所严格规定的范围工作,不会做人没要它做的事。著名的机器人设计师亚瑟·塞缪尔就说“机器不能输出任何未经输入的东西”,“所谓‘结论’只不过是输入程序和输入数据

的逻辑结果。”其次,人工智能的升级能力也在人设定的程序之内。智能机器的升级能力只能在软件即人给定的范围之内。美国辛辛那提大学智能维护中心主任李杰也指出,人工智能的确有学习能力,但那属于程式里的学习,不会跳到程式之外,人工智能只能在指定领域升级。“不要忘了,人工智能里都有一个核心东西叫软件,软件不会自己思考,一定是人编程的。”^[15]

其三是对人工智能基础理论的质疑。根据学界公认的“哥德尔不完全性定理”,“证明任何无矛盾的公理体系,只要包含初等算术的陈述,就必定存在一个不可判定的命题,即一个系统漏洞,一颗永远有效的定时炸弹”。因而以二进制为公理的电脑逻辑运算系统也必然存在这样的漏洞。正是基于这一点,牛津大学的哲学家卢卡斯(Colin Lucas)确信“根据哥德尔不完全性定理,机器人不可能具有人类心智”^[16]。

其四是提出人类以脑机融合的方式防止被取代。这个想法的思路是,如果将来的超级人工智能真能全面超越人类智能,那就不妨让人脑与人工智能结合,比如在人脑中植入芯片,使人脑也能接受无线传输的大数据,拥有云脑,也能进行云计算和深度学习,等等,这就至少可以做到让人和超级人工智能一样强大,于是也就不会被机器人取代。现在这个设想已经被一些科技专家付诸研究,其中就包括发出“取代警告”的技术大家马斯克。

第二方面的否认可谓差异性论证。就是通过分析比较智能机器人与人之间的不同来说明机器人不能超越和取代人类。这些被论及的不同,可以分别命名为意识性差异、价值性差异、社会性差异和结构性差异。

意识性差异是指智能机器人在意识方面与人类相比,缺少了许多东西。因为智能只是意识的一种表现形式,其本质是逻辑思维的运算,但人的意识无比丰富。从思维方式说,人的意识不仅有逻辑思维,还有抽象思维、形象思维、直觉思维和灵感思维。这些思维形式与逻辑思维截然不同,是非逻辑的,根本不能通过编程或逻辑运算实现^[14]。从自主性说,人有自主意识,机器人没有,甚至连自我意识都没有,它是被人造出来为人服务的,机器人“永远无法摆脱作为人类工具的命运”^{[17][18]}。“人工智能可以模拟人类心灵的很多重要特征,比如自主性,但这仅仅是模仿,因为自主性概念本质上是排斥模仿的”^{[17][18]}。从能力上说,机器人只有计算能力,而人还有想象力和创造力,它们是人类智能的最大优势,“人类利用这类能力能够想象并且创造出自然界中不存在的东西”^[19]。从基本分类来看,人既有理性意识,也有非理性意识,如潜意识、欲望、情感等,而机器人则没有非理性意识。除以上差异之外,还有一类被作家韩少功注意到的意识性差异,这就是“人必有健忘,但电脑没法健忘;人经常糊涂,但电脑没法糊涂;人可以不讲理,但电脑没法不讲理——即不能非逻辑、非程式、非确定性的工作”。而其中的“可以不讲理”,意味着人有“同步利用‘错误’和兼容‘悖谬’的能力,把各种矛盾信息不由分说一锅煮的能力”。这种能力“有‘大智若愚’之效,还是只能让机器人蒙圈”^[16]。综上可知,人工智能与人的意识的差异,就是前者局部超越后者,前者整体不及后者的状况。

价值性差异是指人机在价值知识上存在差异,它包含两个层面的意思。首先,机器人存在做什么和不做什么的情况,这一情况意味着它的行动总要遵循某种价值原则。而这种价值原则“是要由人来设定的。而人给机器人设定怎样的价值体系,实际来自人对人类自己价值体系的理解和贯彻”^[20]。这个事实表明,机器人没有自主性,人机仍是主从关系。其次,无论是设计发展人工智能,还是对人工智能的实际运用,都面临价值难题。“一旦涉及价值观,机器人其实一直力不从心”。比如当事故难以避免,需要两害相权取其轻的时候,“肯定要让机器人蒙圈”。而这种价值难题之所以能困扰智能机器人的发展和应用,就是因为“人类这个大林子里什么鸟都有,什么鸟都形迹多端,很难有一定之规”^[16]。易言之,人类的价值观是多元的,没有统一性,也就不能确定该给机器人灌输什么样的价值观。

社会性差异是指从社会的角度考察人机差异,它也有两个层面的含义。一层含义是说,尽管机器人的硬件系统可以“相当于人的肉体”,软件系统可以“相当于人的思维体系”,但是人在社会历史中形成的“人的社会现实性,即人的历史文化本性”,“是不可能以机械的方式重复和再现的”,它“将成为根本的、很可能是无法超越的技术瓶颈和价值悖论。这是‘机器人难题’的核心和实质”^[20]。所以“人工智能不具有人的思维的社会性”^[21]。另一层含义是说,“人是社会性动物,个人的智能是有限的,个人的生

命也是有限的,但人类群体和整体的智能有巨大的加和效应,这种效应正常发挥出来是任何个人的智能无法比拟的。而且人类智能以及作为其基础的知识可以一代又一代的延续、传承下去。与之不同,机器人作为机器,不具有社会性、群体性,它总是单独的、孤立的个体,即使使用程序将它们联结起来,它们不可能在程序之外随意与同伴合作。它们之间的简单合作只会简单相加效应,似乎不可能产生加和效应。机器人不具有生殖功能,其智能当然也不可能一代一代传下去,更无所谓不断地进化发展。没有社会性,这也是机器人不可能取代人类、超越人类的根本性局限”^[22]。

结构性差异是指人与机器人的内在结构不一样,智能机器人没有人的结构。“人工智能虽然能够模拟出人的动作、人的表情、人的声音,甚至能够精确地展示出人的情感,但这些都表明人工智能能够成为人。核心原因只有一个,人工智能没有人的‘结构’,而结构决定着意识和本质属性”。具体说来,人体是一个极为复杂的性能完善的多级结构的巨大系统,“机器人无法拥有人类的系统结构、器官结构、细胞结构、基因结构,乃至终极的碳基分子结构。人工智能只能通过不断模拟逼近人类结构,或者演化成为一个全新的生命形态,但永远不可能成为人类”^[23]。

第三方面的否认可谓“超而不代”的论述。该论述虽然承认未来的机器人在能力上会超越人类,但不会取代人类。“超级智能时代的机器人一定会具有非常高的智能水平,在很多方面也一定会超过人类,但是,机器人智能发展的总体趋势是与人类愈加接近、亲近,而不是对抗,更不可能取代或淘汰人类。自然界之所以拥有200多万物种,正是因为大多数物种并不会取代老物种,它们宁愿与现有的生物体交织起来,挤进小生境之间,以其他物种的成就为基础”。因此,“人类不会被颠覆,反而会变成类似基础设施的存在”^[5]。

三、评析

经过梳理取代论与否取代论的论述可知,在取代问题上,这两种相互反对的观点都有各自的理由,这似乎意味着争鸣双方旗鼓相当、胜负难分的。但据笔者评析,否取代论明显处于下风。因为尽管取代论的理论中含有假说的成分,可否取代论对取代论的所有否认却仍是无效的,所以,不管取代论的立论究竟如何,否取代论都是落败的状况。

先看否取代论的技术性论证中的四种论述。

第一种论述是对AI技术奇点论的否定,但从该否定论述所依据的三个理由(即计算速度的提升不等于智能的提升、人工智能的研发存在收益递减的情况、人工智能的递归自我改进效率受到自我理解能力的限制)看,它们所否定的,实际上都只是奇点到来的速度绝不会像库兹韦尔设想的那么快,而不是它不可能出现。既然奇点未被否定,这就意味否取代论没有否认在未来的某个时刻,人工智能还是“会出现一场‘智能爆炸’,把人的智力远远抛在后面”。

在技术性论证中,第二种论述是对超级人工智能的否定,所依据的理由是智能机器人总是按照人的要求做事,人工智能的升级也只能在人的软件编程之内。这两个理由,在AI技术还没有发展出超级人工智能之前,都是可以成立的,可是一旦发展出超级人工智能就不再成立,因为超级人工智能已经被定义为“一台在所有智能活动上都远超人类”的机器的局限性去证明未来的超级人工智能也超越不了人类。

在技术性论证中,第三种论述是根据“哥德尔不完全性定理”,推论人工智能的理论基础即逻辑运算系统也会存在系统漏洞,成为“一颗永远有效的定时炸弹”,但它并没有具体说明,这个系统漏洞究竟会是怎样的漏洞?究竟会在何时爆炸?又会给人工智能技术的发展带来什么样的危害?更没说明,这个系统漏洞的存在,与否定取代论有什么必然性关联?如果漏洞就是机不如人的证据,那么,人类就是无比完美,没有任何软肋、短板和缺点的吗?

技术性论证的最后一种论述,是设想以脑机融合的方式防止人类被超级智能机器人取代。且不说脑机融合能否成真,也不说这种融合是否会对人脑造成损伤,只要承认未来智能机器人的机脑会实现对人脑的超越,那么,让一个不如机脑的人脑去和机脑结合的结果还是赶不上机脑,这就类似于任何负数加1,最后都是小于1而不是大于1。因此,实现脑机融合后的人脑,最多也只能是缩小人脑与机脑的差距,最终还是避免不了被机脑超越的命运。因为既然全面超越人脑的超级人工智能完全是以物理性物

质的电脑为基础实现的,那么以生物性物质为主构的脑机融合后的人脑就产生不了同样的功能。即便是以物理性物质为主构的脑机融合也仍会存在差距,而且这样的大脑应该也不再属于人脑。这就说明,无论怎样进行脑机融合都防止不了机脑(机器人之脑)对人脑的超越和超级智能机器人对人类的取代。

再看否取代论的四种差异性论证,也同样难称有效。

就意识性差异论证来说,它存在两个不足。其一是它基本上是根据现在的人工智能水平即弱人工智能的状况来说明智能机器人与人类在意识方面存在的各种差异,但有谁说过弱人工智能就能超越人类?取代论分明说的是超级人工智能。而超级人工智能之所以能超越人类意识,则在于它还被规定为不仅具有自主性,而且也拥有了以往只有人类智能才有的能体现人类“最大优势”的创造力,并且,由于还有深度学习、互联网、大数据和云运算的支持,超级人工智能的创造力还会比人类更强。其二是意识性差异论证虽然指出了一些人工智能在意识上有少于人类意识的东西,即情感、潜意识、健忘、犯糊涂和兼容悖论等,而且这种差异即便到超级人工智能时代也可能仍然存在,却没有说明为何有这些意识性差异的存在就意味超级人工智能不能超越人类?尤其是健忘、犯糊涂和兼容悖论,未必会比不健忘、不犯糊涂和不自相矛盾更厉害?人类顶级棋手赢不了阿尔法狗,不就是因为其记忆力、计算力和精准性都不如对手吗?

就价值性差异论证来说,也有两个问题:一是即便未来的超级智能机器人的行动价值原则还是由人来设定的,也并不意味着它就不能超越和取代人类,因为阿尔法狗的行棋价值原则也是由人给定的,而且与人类棋手的行棋价值原则是一样的,结果它还不是战胜了人类棋手?如此,以后在不作身份限制的世界顶级围棋比赛中,自然也不会再有人类棋手的事了。更何况有了自主性和创造力的智能机器人,难道就不可能再给自己创立价值原则吗?须知,人的价值观也不是一成不变的。二是把人工智能遇到价值难题会“蒙圈”说成是人工智能的“力不从心”是不对的,因为这分明是人类自己都没解决的问题,所以才不知该在这方面怎样设计人工智能。而人类之所以受困于此,来自于人类的价值观的多元化。

就社会性差异论证来说,同样不能成立。人的“社会现实性”或“历史文化本性”并不是生而有之的,而是在后天的社会实践中形成的。既然如此,有自主行动能力的未来机器人为何就不能拥有?难道它就不能在社会中进行自己的实践吗?显然不是。此外,由于人的历史文化本性不过是后天实践经验和文化学习的结晶,这就可以说,在这个事情上超级智能机器人只会比人更有优势。因为它是一诞生就有自主行动能力,而不需要一个成长的过程,并且也是一诞生就拥有了包含在大数据中的海量的历史文化知识。既然超级智能机器人有这样明显的优势,那它就势必比任何一个自然人甚至是自然人的最高天才都更聪明,于是所谓的“人类整体的智能加和效应”也赶不上它。这就是说,机器人不用联合起来就可以超越人类整体的智能,何况有自主行动能力并且更聪明的机器人为何就不能联合起来?至于机器人“不具有生殖功能,其智能当然也不可能一代一代传下去”,这也不是机器人的劣势所在,因为机器人不是肉身,是被认为可以永生不死的,这就比生命有限的肉身人类有优势。因此,无论是从以上哪种意义说社会性,都不能证明未来机器人不会超越人类。

就结构性差异论证来说,它的问题是,虽然其关于“人工智能没有人的‘结构’,而结构决定着意识和本质属性”的观点是对的,也符合现在的实情,并且由此思考问题的方向也是对的,但仅此还不够,它最多只说明了机器人目前的结构与人的不一样,还没有说明一种结构比另一种结构优越的因素是什么?这就无法从道理上保证未来的智能机器人也永远不可能有人的结构或比人更优越的结构,何况是目前的智能机器人也不是一切不如人,除了它的超强计算能力已经远超人类外,而且还有不死的身體。因而此差异论证最多只能说明机器人的结构和本质属性与人的结构和本质属性不一样,而不能用来作为机器人不可能超越人类的证据。因为有这种不一样,也可能形成的恰恰是机器人超越人类的结果,如非肉身的机器人相比肉身的人类,就具有永生不死的优势。

最后来看否取代论的“超而不代”论述。这个论述混淆了一个概念,即将“取代人类”等同于“淘汰人类”或“毁灭人类”,但实际上,前者是指使人类失去主宰地球的地位,后者是指使人类灭亡。有此区分之后,可以承认,如将来真有超越人类的机器人,它们是有可能不毁灭人类,但一定会取代人类,成为地球的新主宰。因为最聪明的存在者不会容忍比其笨的存在者来管理自己和地球。而人类一直以来能

成为地球的主宰,还不是因为人类比其他所有存在者都聪明?所以当机器人超越人类,变得比人更聪明之后,人类最好的结局也只能是充当世界的“基础设施”,而不会仍然是“上层建筑”。

或许正因为否取代论的三类论证都说理不力,完全不能让人信服,所以现在的状况是多数学者包括诸多知名学者,都相信了取代论的观点,并且其中一些学者以此为前提,开始思考此后的诸多问题。有的着手研究如何防范机器人反叛人类的问题,还成立了专门的研究机构^[24];有的开展了如何与超人机器人相处的问题研究^[25];有的则是想出了可以宽慰已知自己终结命运的人类的说辞,即“承认他们(指机器人)是我们自己进化了的后代,人类进入了一种改变了繁殖方式的发展新阶段。后代取代前辈继续生存,实现更好的发展,这不是人类的灾难,而是人类的进化飞跃到了一个新阶段”^[10];还有人提出,其实还不等机器人来淘汰人类,人类就可能因为太过安逸地无所事事而死于“人工智能创造的一切好事”^[26]。

四、我见

尽管否取代论的论证都是无效的,却并不意味着取代论就是正确的。因为只要否取代论还能提供新的论证,或者取代论还没确证未来机器人将取代人类的必然性,取代论就还是存在被否定的可能。取代论目前的情况正是这样,由于自身仍存在不少假说的成分,所以并未确证“取代”的必然性。

因而,未来机器人若想取代人类,必须先超越人类,而要超越人类,它不仅要成为具有自主意识和独立行动能力的主体,而且还必须具有比人类更强的创造力,至于是否具有人类的其他意识能力则都不是“超越”的决定性因素。能无中生有的创造力是最高的智慧,因而如果未来机器人的创造力比人类强大,就意味着未来机器人也比人类更智慧、更聪明。但恰恰在未来机器人何以也能拥有创造力这一点上,取代论语焉不详。

取代论宣称超级人工智能会拥有创造力,并提出了实现超级人工智能的多种路径。而这里可以确证,人工智能无论如何发展,无论是走什么路径,都永远创造不出有创造力的机器人。

人工智能要想发展出创造力,必须先搞清楚人脑如何会有创造力的机理,而目前是连人脑的许多基本情况都没弄清楚,又遑论能模仿人脑造出有创造力的机脑?并且,即便将来会有弄清楚的那一天,也还是无法使机器人具有创造力。

不同种类的实体必有不同的性能,这应该是一个普遍真理。因为在该真理的证伪性方面,不会有任何人能找到一个这样的反例:两个不同种类的实体,竟有完全相同的性能。不同的实体之所以会有不同的性能,在于它们或是质料不同,或是结构不同,或是形态不同,当然也可以是兼而有之。质料不同如矿石的化学成分不同,决定了有的具有可燃性,有的则没有;结构不同如建筑物因整体结构不同而有了不同的用途,有的成为民宅,有的成为厂房……;形态不同如钢筋和钢刀虽材质相同,仍因形态不同而用途不同。而机脑与人脑相比,就是这三种不同全部都有。从质料说,机脑的质料是物理性的电子元件,而人脑的质料是生物性的细胞。从结构说,电脑是平行结构,主要由电源、主板、CPU、硬盘、内存、调制解调器、软驱等构成,而人脑则是立体结构,人脑由脑干、小脑和前脑三大部分构成。脑干处于底层,又有延髓、脑桥、中脑和网络系统这四个部分的构造;前脑处于高层,也有四个部分的构造,即视丘、下视丘、边缘系统和大脑皮层;小脑处于中间层,由左右两个半球构成。从形态说,电脑是机箱中的板上楼盘状硬物体,人脑是颅骨中的球形玛瑙状软物体。正因机脑与人脑在质料、结构和形态上都存在差异,所以机脑再怎么设计、发展,都不可能具有人脑的所有性能。

而人脑的创造力就必然属于机脑所不可能具有的性能之一。因为机脑无论是现在的状况还是设想出的未来理想状况,都是以计算为基础功能的智能。计算是对已有信息或数据的处理,属于“有中推有”的逻辑思维,而创造则属于“无中生有”的非逻辑思维,是以想象为前提,先想象出世上尚未有的东西,再设法采用具体手段将它造出来。因此,机脑的计算能力再怎么强大,人工智能再怎么提升,都形成不了创造力,而且也形成不了人脑的所有非理性意识。

人工智能的提升,从方法上说无非两类:一类是系统模仿的方法,一类是还原构造的方法。系统模仿就是通过从整体上模仿人脑使人工智能具有人脑的所有功能,前述取代论实现超级人工智能的六条路径的前四条,即全脑仿真、提升复杂度、功能模仿和机器与生物相结合就属于这个方法;而后两条路径即本质性构造和量子计算则属于还原构造的方法,此法是意欲通过发现智能的本质来构造具有所有意识功能的人工智能。

可是,由于系统模仿方法的四条路径都是以计算的方式体现模仿,就不可能模仿出非计算的创造力,而且,不管今后还会有何种系统模仿人脑的方式,只要机脑所采用的建造质料与人脑的不一样,二者也就不可能有同样的结构和形态,于是也不可能模仿出人脑的所有功能。除非是机脑全部用人脑的质料来模仿人脑的结构和形态进行构造,才有这种可能,可是这样构造出的东西仍是人脑而不是机脑。

与之同理,还原构造的方法也不可能成功。因为我们即便是发现了智能的本质,但只要没用人脑的质料来构造机脑,机脑就还是不可能有人脑的结构和形态,因而也就不可能使按照智能的本质制造的机脑具有人类的所有智能,此其一。其二,所有昆虫、动物都有一般的智能而无特殊的智能即自我意识和创造力等,表明掌握了智能的本质不过是知道了各种具体智能的最一般的规定性而已,仅此并不足以造出智能的最高形态。量子计算路径实际上也属于本质还原。它认为在量子层面,物质与意识已经不分彼此,量子既是物质也是意识,因而人类智能的底层机理就是量子效应,于是以量子计算为基础的人工智能就可以具有人脑的所有智能。可是,这个想法忽略了一个最基本的事实。这就是,所有的实体性存在者都可以被还原到量子层面,或者说都有自己的量子层面,然而为什么却不是所有种类的实体,比如水、泥沙、岩石、石油、草木等自然物和诸多人造实体都有意识?而在所有有意识的生命活动者中,为何又仅仅是人类才具有创造力,而昆虫和动物则没有?这就说明,所谓“人类智能的底层机理就是量子效应”的观点是大错特错的。量子计算路径根本行不通,还原构造的方法不可取。另外,还可以从“整体大于部分之和”的古老命题得到说明。因为整体作为一个系统,不仅仅是所有构成因子的集合,还包括了这些构成因子之间的相互作用,而还原构造由于省略了这些相互作用,也就再造不出同样有这些相互作用的系统。由此推广,其他形式的还原构造方法,如分子计算、光子计算、纳米计算、生物计算和神经计算等,也都不可能使机脑具有人脑的所有功能,因为它们与量子计算一样,也是以计算为基础功能和以物理材料为主要质料。同时说明,还原论的方法不能滥用,如果超过了一定的界限就会沦为荒谬,这也正是它被后现代主义诟病的原因所在。至于还原论的界限,简单来说就是:高等级事物的问题,不能用低等级事物的道理来加以解释。

机脑不能具有人脑的创造力,除了以上解释之外,还可以从复杂性的维度得到说明。客观事实表明,创造力作为最高智慧,只是出现在最为复杂的人脑中,所以凡是达不到人脑复杂程度的脑,不管是物理性的机脑,还是生物性的动物之脑,包括灵长类动物的也已经非常复杂的脑,就都不可能具有创造力。

综上所述,机脑无论怎么构造,都不可能具有人脑的所有性能,尤其是不可能具有人脑的创造力。既然如此,未来的机器人无论怎么发展,也都永远超越不了人类,更谈不上对人类的取代。而且,这里还可以退一大步讲,就算未来的机器人可以具有创造力,甚至比人类的创造力更强,也不会取代人类。

机器人作为人造的机器,没有生物性的身体,而没有生物性的身体,就不会有任何自身的需要,也不用生长,没有自身的需求和生长,就不会有任何自发的行动。马克思说“任何人如果不同时为了自己的某种需要和为了这种需要的器官而做事,他就什么都不能做。”^[27]这话也适用于所有的动物和机器人。同时,机器人没有自身的需求和生长,也就没有所有生物都有的自利本性,并且也不会有属于自己的真正的情感,因为诸如喜、怒、哀、乐、忧、悦之类的所有的感情,最终都属于对需要满足与否的心理反应^{[28][29]},所以机器人即便被设计出了情感,也是虚拟情感。机器人既然没有自利本性和内涵善恶倾向的情感,也就不会像人一样具有自己的价值意识及价值观。因此,机器人为什么目的而出现、是否需要行动,在行动中需要遵循什么样的价值原则,就的确只能是听从人的安排。这就说明,未来的机器人其实也不会具有自主意识,更不会成为主体。

或许有人会说机器人有对能源,如电能的需要,但由于没有能源时机器人不仅不会“死”,而且也不会对自身产生任何危害,而有能源时也同样没有生长,因而能源并非机器人的需要。

或许有人设想可以通过制造有生物性身体的机器人来避免这些先天缺陷,可是一个物理性的机脑是不能长在生物性的身体上的,也无法与这个身体产生如同生物性之脑与生物性身体之间的相互作用,也就是说,前一种结合就算能实现,其结合效应也与后一种结合存在差距。如果为了解决这些问题,再将机脑换成生物性之脑来与生物性身体匹配,那么这时它就不再是机器人了,而是一个动物。因此,机器人的那些先天缺陷是注定不可避免的,机器人注定不可能有自主意识和属于自己的价值观。

由于人是为了满足自己的需求来创造机器人的,机器人就是人用来满足自己需求的工具,所以由人赋予机器人的本性就一定是“利人性”,人让它遵循的价值观就一定是服务于人的价值观,让它做的事

情包括假设它也能做到的创造,就一定都是有利于人的事情。这就是说,机器人是纯粹的利人主义者,又怎么可能会取代人或毁灭人?所以,所有关于未来机器人会对人类如何如何不利的各种担忧,统统都是毫无必要的杞人忧天。

至于有人可能会利用机器人来干损害他人甚至消灭他人之事的情况,也不会成为取代论的凭据。因为任何人造的工具,都可能被心术不正的人用来损害他人。由于人类早已明白相互恶斗的结果只能是谁也过不好,所以人类社会在任何时候都不会让这种心术不正的人得逞,因而这种人也就只能是偷着干坏事的极少数人,而由他们所造所用的机器人,也就一定战胜不了由其他绝大多数人所造所用的机器人。

因此,无论在什么情况下,哪怕机器人也具有创造力,机器人也不会取代人类和毁灭人类。既然如此,我们就尽管让人工智能快速发展吧,它给人类社会带来的只会是福音,这就是可以加速将人类从必要劳动中解放出来,从而进入自由王国。由于实际上机器人永远无法具有创造力,所以人类进入自由王国之后,也自然不会如某些人担忧的那样,因无所事事的安逸而亡,而是可以尽情发展自己的各种潜能天赋,进行各种发明、创造。

既然如此,如果不能颠覆本文的结论,我们就不必再浪费精力,去讨论任何以“机器人将取代人类”的观点为前提的各种问题。如果能颠覆本文论点,那就请多加指正。

参考文献:

- [1]陈如明.超级人工智能的务实发展策略[J].世界电信,2017(2):25-29.
- [2]叶子.比尔·盖茨:人工智能梦想即将实现,这是一切努力的终极目标[J].创新时代,2016(7):27-28.
- [3]陈自富.强人工智能和超级智能:技术合理性及其批判[J].科学与管理,2016,36(5):25-33.
- [4]何怀宏.奇点临近:福音还是噩耗——人工智能可能带来的最大挑战[J].探索与争鸣,2018(11):50-59.
- [5]刘英团.超级智能能取代人类吗?[J].商周刊,2015(8):10.
- [6]杨学山.走向通用人工智能[J].信息系统工程,2019(11):9-12.
- [7]邹琪编译.超级机器人计划——人工智能之父马文·明斯基访谈录[J].世界科学,2007(3):7-10.
- [8]Maggie Yu.未来机器人或取代人类智能?[J].中国科技财富,2009(3):116-119.
- [9]徐英瑾.人工智能技术的未来通途刍议[J].新疆师范大学学报(哲学社会科学版),2019,40(1):93-104.
- [10]翟振明,彭晓芸.“强人工智能”将如何改变世界——人工智能的技术飞跃与应用伦理前瞻[J].人民论坛·学术前沿,2016(7):22-33.
- [11]李恒威,王昊晟.人工智能威胁论溯源——技术奇点理论和对它的驳斥[J].浙江学刊,2019(2):53-62.
- [12]张梦然.2045年,现实版“黑客帝国”要上演?[J].今日科苑,2011(10):90-93.
- [13]褚波.库兹韦尔奇点[N].光明日报,2015-02-01(006).
- [14]杨溟.机器人已不止是人功能的延伸[J].商学院,2016(4):120.
- [15]刘力源.人工智能是人类的“终结者”吗?[N].文汇报,2015-04-17(T08).
- [16]韩少功.当机器人成立作家协会[J].读书,2017(6):3-15.
- [17]陶孝云.人工智能会超越人类心灵吗[N].光明日报,2016-05-25(014).
- [18]钱铁云.人工智能是否可以超越人类智能?——计算机和人脑、算法和思维的关系[J].科学技术与辩证法,2004(5):44-47.
- [19]龚怡宏.人工智能是否终将超越人类智能——基于机器学习与人脑认知基本原理的探讨[J].人民论坛·学术前沿,2016(7):12-21.
- [20]李德顺.人工智能对“人”的警示——从“机器人第四定律”谈起[J].东南学术,2018(5):67-74.
- [21]林可济.人类真的能够制造出“超级大脑”吗?——人工智能哲学论辩的历史回顾与现实意义[J].中共福建省委党校学报,2016(1):107-112.
- [22]江畅.高科技时代的人类自律及其构建[J].湖北大学学报(哲学社会科学版),2020,47(2):1-10.
- [23]崔瀚文.人类要如何适应人工智能时代[J].商学院,2017(10):88.
- [24]杜威.机器人取代人类[N].光明日报,2012-12-18(012).
- [25]程广云.从人机关系到跨主体间关系——人工智能的定义和策略[J].自然辩证法通讯,2019,41(1):9-14.
- [26]赵汀阳.人工智能提出了什么哲学问题?[J].文化纵横,2020(1):43-57.
- [27]中共中央马恩列斯著作编译局.马克思恩格斯全集[M].人民出版社,1963:342.
- [28]韩东屏.论价值意识[J].江汉论坛,2012(10):78-83.
- [29]韩东屏.人是元价值——人本价值哲学[M].华中科技大学出版社,2013:97.

(下转第59页)

- [35] Lee Y , Gordon R H. Tax Structure and Economic Growth[J]. Journal of Public Economics , 2005 , 89(5-6) : 1027-1043. -----
- [36] Agell J , Lindh T , Ohlsson H. Growth and the Public Sector: a Reply [J]. European Journal of Political Economy , 1999 , 15 -----
(2) : 359 -366. -----
- [37] 戴魁早. 要素市场扭曲如何影响出口技术复杂度? ——中国高技术产业的经验证据 [J]. 经济学(季刊) , 2018(1) : -----
337-366. -----
- [38] 甘行琼, 李玉姣, 蒋炳蔚. 财政分权、地方政府行为与产业结构转型升级 [J/OL]. 改革, 2020. <http://kns.cnki.net/kcms/detail/50.1012.F.20200706.1622.004.html>. -----

Does Fiscal Imbalance Hinder the Upgrading of Industrial Structure?

——An Empirical Research Based on Provincial Panel Data in China

LIN Chun , SUN Ying-jie , *Liaoning University*

Abstract: Deepening the reform of the fiscal system is an important and reliable path to achieve the upgrading of the industrial structure under the background of high-quality development in China. Based on this background , this paper makes use of system GMM and recursive model to study the impact of vertical fiscal imbalance on industrial structure upgrading based on 31 provinces' (municipalities and autonomous regions) panel data in China from 2000 to 2017. And the conclusion is that: (1) Vertical fiscal imbalances hinder the structural upgrading of inter-industry and intra-service industries , but promote the structural upgrading of the manufacturing industry; (2) There are regional heterogeneity and temporal heterogeneity in the upgrading effect of vertical fiscal imbalance on inter-industry and intra-industry; (3) The vertical fiscal imbalance can affect the upgrading of industrial structure through the market protection behavior , environmental protection behavior , investment preference behavior and technology investment behavior of local government. Finally , relevant policy enlightenment may be drawn based on the conclusions.

Key words: vertical fiscal imbalance; upgrade of industrial structure; local government behavior

责任编辑 胡章成

(上接第 16 页)

Will Robots Replace Human Beings in the Future?

HAN Dong-ping , *HUST*

Abstract: Will intelligent robots replace human beings in the future? Different answers to this question will determine how we treat the development of Artificial Intelligence(AI) technology , and the academic answers are “replaceable theory” and “irreplaceable theory”. Both the AI technology development goal theory and the AI technology singularity theory of replaceable theory believe that in the future , super artificial intelligence that can surpass and replace human beings may be created , and six realization paths are provided for the realization of super artificial intelligence. The irreplaceable theory demonstrates that robots in the future cannot replace human beings from three aspects: one is technical demonstration; second , the difference demonstration; the third is the argument of transcending but not replacing. Throughout these discussions , although the replaceable theory has a hypothesis component , all the disproof of the irreplaceable theory are invalid , because these arguments are basically premised on weak artificial intelligence rather than super artificial intelligence. Therefore , no matter what the replaceable theory is , the irreplaceable theory is in a defeated situation. However , this does not mean that the replaceable theory is correct , because as long as the replaceable theory can provide new arguments , or the replaceable theory does not prove the necessity that robots will replace human beings in the future , the replaceable theory is still possible to be denied. To be sure , a robot must be more creative than a human being in order to replace a human being , but the replaceable theory doesn't say clearly on this point. Due to the different types of entities and different performance , the robot and its machine as long as the brain with people on the material , structure , form and there exist certain differences of the human brain , it is impossible to have one or all of the human brain performance , especially to be creative in the highest sense of what makes a person human , and even assuming that the robots can be creative , they can't replace human. Since there is no need for a robot without a body , it will not have autonomous consciousness and cannot become a subject. Therefore , it can only be a tool to follow the instructions of human beings and serve human beings.

Key words: robot; replaceable theory; irreplaceable theory; creativity; need

责任编辑 吴兰丽