

论人工智能的伦理风险表征

李 旭, 苏东扬

(湖南师范大学 马克思主义学院, 湖南 长沙 410081)

[摘要]目前,对人工智能的伦理风险的关注已经达到了一定高度。人工智能的伦理风险有着不同表征维度,主要包括人工智能技术的社会伦理风险、人工智能技术的人文伦理风险和人工智能技术的生态伦理风险。随着人工智能的发展,这些风险还会不断出现新的变化,甚至可能产生新的风险类型。人工智能的出现不可避免,我们要做的就是采取合理有效的方式尽量削弱其负面效应。为此,可以让机器与人类各司其职、互相促进,且围绕人工智能的制度建设,尊重规律,顺势而为,才能更好地保障整个社会秩序和自由的秩序长期总体稳定与健康发展。

[关键词]人工智能;伦理风险;风险表征;社会秩序

[中图分类号]B82-057 [文献标志码]A [文章编号]1672-934X(2020)01-0013-05

DOI:10.16573/j.cnki.1672-934x.2020.01.003

On the Ethical Risk Representation of Artificial Intelligence

LI Xu, SU Dong-yang

(School of Marxism, Hunan Normal University, Changsha, Hunan 410081, China)

Abstract: At present, attention to the ethical risks of artificial intelligence has reached a certain level. The ethical risks of artificial intelligence have different representation dimensions, including the social ethical risks of artificial intelligence technology, the humanistic ethical risks of artificial intelligence technology, and the ecological ethical risks of artificial intelligence technology. With the development of artificial intelligence, these risks will continue to appear new changes, and may even generate new types of risks. The emergence of artificial intelligence is inevitable, and what we have to do is to take reasonable and effective ways to minimize its negative effects. For this purpose, machines and humans can perform their respective duties and promote each other, and around the construction of artificial intelligence systems, respecting the laws and following the trend can better ensure the long-term overall stability and healthy development of the entire social order and free order.

Key words: artificial intelligence; ethical risk; risk representation; social order

目前,人工智能技术的伦理问题,在学界已经受到了极大的关注。华南理工大学闫坤如教授指出,在人工智能发挥越来越重要作用的同时,需要高度重视其引发的潜在道德风险。智

能机器的道德主体地位问题、人工智能引发新的社会安全和公平正义问题及其他新的伦理问题都需要详细考察和谨慎决断^[1]。当代人工智能技术的伦理风险的主要表征可分为人工智能

收稿日期:2019-10-23

基金项目:湖南省教育厅创新平台开放基金项目(17K009)

作者简介:李 旭(1983-),男,湖南耒阳人,讲师,博士,主要从事马克思主义理论、技术伦理研究。

技术的社会伦理风险、人工智能技术的人文伦理风险、人工智能技术的生态伦理风险。这些风险的产生会随着人工智能技术的不断创新而发生变化,人类社会也将会因为人工智能技术所隐含的潜在风险而进入到一个“智能风险时代”。

一、人工智能技术的社会伦理风险

随着人工智能技术的应用与扩散,其对人类社会产生的多方面负向效应正逐渐显现,主要包括:大规模失业导致人类决策的边缘化、人工智能武器的安全隐忧、关于人工智能技术的相关权利与责任困境以及因人工智能技术的贫富差异所造成的社会分层和固化问题等。

第一,人工智能技术引发大规模失业,继而导致人类决策边缘化问题。以往机器被视为工具性或辅助性的生产工具,是为了提高或扩大人类物质生产力的“助力工具”。欧盟政治战略中心(EPSC)所发布的报告,指出人工智能技术(机器学习)对提高生产效率存有显著的助推作用。具体来说,预计到 2030 年,人工智能技术将为全球经济贡献 12.8 万亿欧元,这一经济贡献相当于推动现今全球总 GDP 增长 14%,而预计到 2035 年,这一数值则可提高到 40%。正因如此,全球各大企业目前正竞相收购以人工智能为重点的公司,据相关统计显示,自 2012 年以来,已有 250 多家 AI 公司被大企业所收购^[2]。至今,人工智能领域已成为世界各国竞争的主战场之一。尽管如此,因人工智能技术的发展引发大规模失业,继而导致人类决策边缘化问题已经成为了全社会关注的焦点之一。2000 年,高盛集团在纽约股票交易所的交易员有 600 人,十多年之后的今天,却只剩下了 2 人,其余的工作全部由人工智能包办。Kensho 公司 CEO 丹尼尔·纳德勒(Daniel Nadler)坦言:“在十年至二十年之内,高盛还会对内大量裁员。”除了高盛,每家主要金融公司的交易大厅也将如此^[3]。牛津大学诸多学者们一致认为,英国现存的工作种类,有三到四成将会在未

来的十多年内完全被人工智能取代。进一步说,随着机器人的“智能化”水平的迅速提升,从负面角度来看,人类将在决策制定的过程中被边缘化,这种“边缘化”的缘起正是由于现代智能技术的不断创新与发展。在某一具体的策略制定过程中,由于智能机器具备强大的数据收集与分析的自主能力,它们所做出的结论往往是“科学的”与“正确的”。有大量的数据表明,人类通常带有偏见,而且敏感,所以,依赖于直觉和理性的人类甚至比依靠数据统计的智能机器更可怕,毕竟“机器从不会说谎”。

第二,人工智能武器的潜在风险。早在上世纪末,最新人工智能技术被运用于某些军事领域,并形成了以智能机器人、无人舰、无人飞机为主的智能军用体系。在智能机器人领域,主要是以 iRobot 公司“费奇”机器人为代表,该智能机器人的作用主要为打扫战场、弹壳收集,除此之外,还有一部分智能机器人被用于生化武器和炸弹的检测。1999 年,美国的第一代用于鱼雷搜索的无人潜艇正式投入使用。作为无人机的原始模型——Q-2 训练靶机也只是为了训练相关飞行员而研发的。进入 21 世纪,随着深度学习技术的快速发展与应用,更是加快了智能军事武器的发展。美国作为引领现代军事技术发展的国家,所研发的一部分军用飞机系统已具备自动搜索与跟踪目标、自动辨别地形并选择最佳飞行路线等能力。例如,拥有几千万行代码被誉为“飞行计算机”的 F-35 第 5 代飞行战机和完全由电脑操控的海军 X-47B 无人轰炸机等。尽管在某种程度上,拥有强大杀伤能力的现代智能武器可以平衡不同国家间的军事实力,但是,当此类武器落入不法分子之手,后果着实令人堪忧,其所带来的危机将远超出战争引发的道德问题。需要加以补充的是,由于智能武器或智能武器系统不具备人类对战争的恐惧与战后所带来的各类痛苦的真实感受,战争的爆发变得更加的容易。当智能武器向强智能方向发展的过程中,人类可能会陷入“灭种的可怕漩涡之中”。

第三,智能技术的权利与责任认定之争。自机器诞生之日起,便被人类视为工具而服务于社会实践,成为人类的附属品。随着智能技术的不断发展,其智能性得到了前所未有的提升,这意味着作为智能化的各类技术的集合,将会对人类的主体性地位产生极大的动摇效应。换言之,智能技术的发展将会使得人工道德主体界定的模糊性增加,进而造成其权利与权益的混乱。具体而言,当人工智能仅仅被视为工具而服务于人类社会时,那么,经由此类机器所产生的一系列不良后果均属于人类的责任,因为“它们只是工具罢了”,犯错根源于人类。另外,如果人工智能以人类伙伴的身份参与人类的各项工作,那么它们的权利与义务又将如何界定?其中,如果人工智能被视为与人类一样,同样具备完全自主的道德抉择能力,那么,当它们“犯错”之时,智能机器的伦理设计者、创造者、使用者等是否也应承担一部分责任?反言之,如果人类为智能机器的错误行为埋单,那么,人工智能又是否可能成为一些别有用心之人推卸责任的借口?除此之外,当人工智能被视为能够完全为其危害人类社会的行为负责时,它们又将以什么样的形式予以承担呢?毕竟让人工道德主体意识到自己犯错并且接受道德约束,从方法论层面而言,迄今为止,并不存在具体可行的实际方案。如果人工道德主体需要承担刑事责任,那么,什么样的惩罚能让机器人恐惧且做到自主性自我约束呢?毕竟机器没有生命,只是盲目地遵循预设的算法和“目的”,遵循代码完成一系列随机试探、刺激—反应、加权评估、条件判断等过程,在求解空间中寻求一切满足目标方程式的可能性。这种计算对每种可能性的善或恶都漠不关心^[4]。

第四,技术的贫富差异造成社会分层和固化的风险。如前所述,随着人工智能技术被广泛应用到不同领域,各个领域的工作逐渐由人工智能取代,而这种取代并非纺织者被纺织机代替,或者人力车被汽车代替,而是对现有职业和工作版图进行大规模的颠覆。可以想象,当

滴滴公司完全由无人驾驶汽车运营;苹果公司可以不用去海外寻找廉价劳动力拼装,而是利用人工智能在本土进行机器组装,这其中的耗损,与传统生产方式相比,具有极大的优势。从某种角度而言,在这场颠覆性的技术变革中,受益最大者当属掌握大数据、智能研发技术的各类大型企业及其背后的国家。这是因为,数据优势对智能技术行业中的发展具有决定性作用。只有掌握大量原始真实数据的企业才能研发出更好的智能产品,这意味着其经济收益将会稳步增长,当资本积累达到一定程度之时,自然会吸引更多的人才加入其中并研发出更好的产品,如此往复,不断循环。但是,由此带来的影响在于有限的资源将会被那些少数的大型企业甚至国家所掌控,多数中小企业将会在这样的循环中被屠戮殆尽。可以预见,在未来的智能社会中,数据与技术将会与石油、天然气一样成为宝贵的资源。当技术的贫富差距被不断拉大,社会分层与固化的风险就越高。

二、人工智能技术的人文伦理风险

抛开人工智能技术对人类个体的正向作用不谈,就人工智能技术对人类个体的负面影响而言,至少存在两个方面的质疑:其一,人工智能体是否可以具备伦理道德?如若可以具备,那么,其二,人工智能体是否构成对人类主体性的威胁?

计算机伦理学家詹姆斯·摩尔(James H. Moor)曾根据人工智能对伦理的影响和参与程度对人工道德主体(artificial moral agents,简称AMAs)的类型进行了划分,它们分别是:有道德影响的主体;没有自主意图但是对伦理有影响的智能、隐式的道德主体;通过外在的硬件和内在的软件设置了伦理框架的智能、显式的道德主体;根据现实的情势和情况采取符合人类伦理的智能、完全的道德主体,拥有像人类的自主意识并且做出伦理应对。现阶段,尽管还未有一个完整的、通用的、置于人工智能技术内部的道德伦理框架对人工智能技术的开发、测

试、应用等方面进行价值引导,但事实上,当具备无自主意识的人工智能做出不同行为并产生了相应伦理影响之时,便可将其视为具备了伦理能力,由此转变成为伦理的关注对象。据此,一个问题得以出现,即自由与自主决策是否是人工智能被作为道德主体的前提?如果人工智能可以作为道德主体,人类的爱恨情仇等情绪是否为道德主体的基本组成要素?从某种角度而言,在自由与情感的加持下,人工智能与人类之间的区别会进一步缩小。但是,当给予人工智能以情感,将会产生的后果又是如何?人类目前无法得知。在面对错误时,人类虽已具备相当充分的经验,但是相同的错误仍会发生。同样的,拥有智能与情感的人工智能也会像人类一样,不断地犯同样的错误。这对于人类而言是经验的“真空区”。更为严重的是,那些出于恶意或是犯罪意图的人将不良“因素”输入于智能系统后,其给人类社会造成的破坏也将会是灾难性的。

除此以外,人类的主体地位将随着人工智能技术的日益成熟而受到威胁。众所周知,所有技术物都是人类智慧的产物,人作为使用技术的主体,通过各类技术的辅助,从而获得更加美好的生活。但与过往技术完全不同的是,人工智能技术的诞生,使人类的社会主体地位受到了前所未有的挑战。目前,已有不少科学家预测,在未来的 20 至 40 年间,人工智能将会突破“奇点”并进入一个全新的意识层面——强人工智能,“人工智能不同于历史上任何一次技术革命,它给予人类的认知、价值、关系、人的定义及其本质都带来颠覆性的影响。”^[5]人的主体性表现在其能力、社会地位及作为主体的作用等方面。在自然界中,人依托自然界的环境条件,是实施行为的主体因素,也是社会组建的主要要素,从马克思主义哲学的角度而言,人就是一切社会关系的总和,而这种关系将伴随人工智能技术的三大优势得以改变。第一,强大的学习能力。基于人类智慧成果产生的种族,人工智能体可以在极短时间内追赶上人类最高的智

能水平;第二,无种族差异性。没有个体差别,没有畏惧且全心全意为种群的发展而服务;第三,亲密的合作关系。与人类之间合作所考量的“回报”标准相比,人工智能体之间的合作是亲密无间的。在《我,机器人》中,出现了人工智能拥有自主意识并解放人类劳动负担,最终人类因此变得懈怠而不思进取的情节。随着人工智能自我意识的不断进化,它们一旦意识到人类本质的恶,由此对人类充满失望,继而为了它心中的完美世界而反抗人类,将会对人类社会造成巨大危害。据此,我们可以预测:既然人工智能可以拥有自主意识,并在各方面远超人类,那么它们又何必臣服于人类!

三、人工智能技术的生态伦理风险

人工智能技术是人类智慧的结晶,一门新兴科技带来的福与祸总是相伴而行^[6]。以人工智能技术为代表的“技术团体”,在为人类社会带来多种便利的同时,其相应的伦理问题也随之产生。就人工智能技术的生态伦理问题而言,其对自然生态的负面影响有二:一是人工智能技术发展的高能耗问题;二是人工智能技术发展与生物多样性的不对称性问题。

就人工智能技术发展的高能耗问题而言,我们可以清晰地了解到,人工智能技术的发展对自然能源是有着高于传统技术的能源需求的。据 2019 年 4 月 13 日人民网报道,咸阳市一个犯罪团伙,为了“挖矿”(即通过运行特定的计算机逻辑算法,与远程服务器通讯后得到的相应虚拟奖励)而盗用了 200 多万度电。又一案例,一位属于荷兰普华永道(PwC)分公司的研究员——亚历克斯·德·弗里斯(Alex de Vries),他曾通过大量的实验与搜集相关数据,对 2018 年比特币(Bitcoin)的全球能源消耗这一问题进行了研究。研究结论表明,全球现有的发电装置并不足以维持比特币的挖矿业务。通过实验得出的一组数据,亚历克斯·德·弗里斯认为,若要维持现有的挖矿业务,就必须满足高达 62.3 TWh 的电力要求,即每小时消耗

62.3 太瓦电量。更形象一点,这样规模的电量足以满足匈牙利或瑞士的电量需求。由此可见,人工智能技术的快速发展,需以高能耗的代价得以换取。从某种角度而言,人工智能技术的这种高能耗需求,会给生态环境的承载力带来巨大的“包袱”,并最终威胁到人类社会的可持续性发展。“环境的恶化很大程度上是由新的工业和农业生产技术的介入引起的,这些技术在逻辑上是错误的,因为它们被用于解决单一的彼此隔离的问题,没有考虑到那些必然的‘副作用’”^[7],康芒纳如是说。人工智能技术的发展,正是因为没有看到在生态自然方面的“副作用”,才会以前所未有的速度指数性发展。因此,其带给人类社会的影响之一,便是引起了全社会对人工智能技术的生态伦理风险的持续关注。正如卢风教授所言:“现代科技已渗透于文明(广义的文化)各个维度,科技的生态学转向不可能孤立地发生于科技领域,生态主义就是呼唤生态文明的崭新的哲学。”^[8]当然,人工智能技术的发展所伴随的高能耗伦理问题不仅局限于自身,也是当代高新科技发展所要面对的共同问题。

关于人工智能技术与生物多样性的不对称性问题,可以说,随着人工智能技术的发展,生物的多样性受到了挑战,而这种挑战源于生物与人工智能在生存与发展等方面的差异性需求。未来人工智能技术的发展具备大数据驱动、跨环境认知与推理、脑机协同、群体智能、高度自主化等特征,与生物以生存为目的向自然索取资源的多样性相比,人工智能技术发展的资源需求更加单一。正是由于二者存在这种差异性,从而导致为了发展人工智能技术而加大对生物多样性的破坏。举例而言,为了生存的目的,空气、水、树木、煤炭……,均成为了人类活下去的物质支撑,为了生活得更好,人类对自然资源的索取量将会变得更多、类型数目会逐渐增加。相反地,人工智能的发展更多地依赖于超体量的原始数据资源与各类驱动程序,尽管

在某种程度上,存储这些信息所需的物理存盘等也消耗了一部分物质资源,但其所占体量与人类所耗相比,是少数的那一部分。正是存在人工智能技术发展与生物多样性的不对称性问题,所以,人工智能为发展自身将会放弃选择保护自然环境以及其它物种。可以想象,当这一刻真正来临之时,便将是“至暗时刻”。

随着人工智能的进一步发展与应用,将会暴露出更多的伦理问题,人类面对的智能风险也将会更甚。为了将这些风险控制在一定的范围之内,就需要让机器与人类各司其职,互相促进……智能传播的法律一定要顺势而为、自然制定,才能更好地保障整个社会秩序和自由的秩序^[9]。在面临智能时代的“伦理大考”时,多措并举,共同应对是预防人工智能“转折背叛”的有效法宝。

[参考文献]

- [1] 李伦,刘梦迪.智能时代的数据伦理与算法伦理——第五届全国赛博伦理学暨数据伦理学研讨会综述[J].大连理工大学学报(社会科学版),2019(3):104—109.
- [2] The European Political Strategy Centre.The Age of Artificial Intelligence Towards a European Strategy for Human—Centric Machines[R]. EPSC Strategic Notes, 2018, Issue 29:1—14.
- [3] 谭燃,编译.人工智能正在颠覆华尔街 高盛已深受其害[EB/OL]. <http://tech.qq.com/a/20170228/037165.htm>.
- [4] 汤炜.对技术的现象学反思之进路——兼谈人工智能[J].哲学分析,2019(2):153—164.
- [5] 余乃忠.自我意识与对象意识:人工智能的类本质[J].学术界,2017(9):93—101.
- [6] 王云,刘亚平.人工智能技术发展中的生态伦理困境和启示[J].长沙理工大学学报(社会科学版),2018(4):7—14.
- [7] [美]巴里·康芒纳.封闭的循环[M].侯文蕙,译.长春:吉林人民出版社,1997:154.
- [8] 卢风.科技、自由与自然——科技伦理与环境伦理前沿问题研究[M].北京:中国环境科学出版社,2011:8.
- [9] 刘伟.智能传播时代的人机融合思考[J].人民论坛·学术前沿,2018(24):16—24.