

## 科技与伦理走向融合 ——论人工智能技术的人文化

○ 陈 静

(中国社会科学杂志社 编辑部, 北京 100026)

〔摘 要〕历史上,每一次技术大变革并未使人类走向灭亡,反而是文化的消解、人文精神的匮乏,更容易导致朝代更迭、文明消失。对人工智能过度悲观和盲目乐观均不是科学的态度,毕竟,人类制造工具,而工具让我们走得更远。应对人工智能带来的人文挑战,为人工智能体设置人类价值观十分必要。要强调科学家和科学共同体的道德义务,从源头上堵住不利于人类社会发展的的人工智能体出现。同时,以“机器人三大定律”为法则,在“不伤害”原则的前提下,培养人工智能体与人类的良好交互能力。此外,政府管控不能缺位,要保证研究结果获得分享,防止某个团体单独受益。人类仍要保持对大自然的敬畏之情,对理性抱有谨慎的信心。在以上几点的基础上,将以人为本、公正、和谐、可持续发展的价值观嵌入人工智能体,而这正是人工智能人文化的具体体现。

〔关键词〕人工智能;人工智能人文化;人机共存

DOI:10.3969/j.issn.1002-1698.2017.09.009

20 世纪以来,人们围绕可能带来巨大利益,同时又具有不可预料风险的尖端技术,展开一轮又一轮的激烈争论。争论的焦点主要不是技术,而更多地聚焦在伦理、政治以及决策层面。在这个过程中,应用伦理学的多个分支——科技伦理学、生态伦理学、生命伦理学、医学伦理学等学科迅猛发展,它们的使命就是规避科技进步所引发的伦理与道德风险,防止科技异化为奴役人的工具,使其必须为人类造福。

科幻小说家亚瑟·C·克拉克(Arthur C. Clarke)写过一句著名的话:“真正先进的技术,在人民看来就像魔术一样。”<sup>〔1〕</sup>人工智能(Artificial Intelligence)技

---

作者简介:陈静,中国社会科学杂志社编辑。

术(以下简称“人工智能”)具有这样的效果。从1956年概念的提出,到2016年“阿尔法狗”(AlphaGo)打败围棋高手李世石,60年光阴,人工智能以意想不到的方式深度参与人类生活,在交通、医疗、教育、家庭服务、公共安全、娱乐等领域,努力为改善人类健康、保障安全以及提升生产力等方面,发挥积极正向的作用。可以设想,未来几十年,其所带来的科技产品将推动社会各个领域的变革。

魔术是假的,但是人工智能技术带来的变革是真真切切的。变革总是充满着争议,何况是有可能“颠覆”人类自身的人工智能?就目前来看,尽管人工智能技术发展距离人类智能还很遥远,但它所影响的多个领域已面临着各种不同难题,比如:很难克服对人类被边缘化的恐惧(深度学习致使人类被某些就业领域淘汰)、难以创造安全的感知和作业硬件(无人驾驶和客服机器人)、得不到公众信任(低资源社区和公共安全)以及还存在减少人际交往的风险(娱乐领域)。

到底什么是人工智能?它将给人类带来怎样的世界?如何对其进行伦理规制?本文试图从人工智能技术的人文化,即强调科学家和科学共同体的道德义务,呼吁他们创造出有道德的人工智能体,并立足于坚持以人为本,促进人类社会公正、和谐以及可持续发展的角度,探索科技与伦理的深度融合。

## 一、人工智能的悖论

2016年,全球最大的公共关系咨询公司万博宣伟(Weber Shandwick)发布了一份与领英(KRC Research)联合调研的报告《AI—ready or Not: Artificial Intelligence Here We come!》<sup>[2]</sup>。该报告调查了5个全球市场(中国、美国、加拿大、英国和巴西)2100名消费者对于人工智能的看法和前景预测。报告显示,消费者对人工智能的理解比较粗浅,一提到人工智能,多数人的直接反应是机器人。导致公众这样的认识并不令人意外,因为科学界在如何界定人工智能的问题上至今尚未达成共识;谁也无法预料这一技术的终点是什么,何况目前人工智能技术各个领域的发展并不均衡。举一个例子:电子计算器比人脑计算快得多,而且几乎从不出错。然而,与智能手机相比,它所实现的功能只有后者的百万分之一。计算机的功能远远超过智能手机,而且它既不会偷懒,也没有情绪,在某些方面甚至比人更聪明,未来有可能会越来越聪明。计算器是昨天人们认为的人工智能,而当计算机生产出来后,计算机代替计算器成为人们眼中新的人工智能。

由此可见,随着技术的更新与发展,人工智能很容易推翻自定义的界限。因为“人工智能领域是一个不断努力推动机器智能向前发展的过程。具有讽刺意味的是,人工智能正在遭受失去话语权的长期灾难,最终不可避免地会被拉到边界内,即一个被称为‘人工智能效应’(AI effect)或‘奇怪悖论’(odd paradox)的重复模式——人工智能将一种新技术带到了普通大众中去,人们习惯了这种技术,它便不再被认为是人工智能,然后更新的技术出现了。同样的模式将在未来继续下去。人工智能并没有‘交付’一个惊雷般改变生活的产品。相反人工智能

技术以一个连续的、进步的方式正在继续更好的发展”<sup>[3]</sup>。

### (一) 人类智能与人工智能

要明晰人工智能的涵义,首先要搞清楚人工智能和人类智能的区别。

智能(Intelligence)是一种复杂现象,《辞源》第三版注释为“智谋和才能”,一般是指人在实践中对客观事物进行观察、记忆、分析、判断,以及有目的地行动和有效解决实际问题的综合能力。辩证唯物主义认为,智能的生理本质是以人脑和神经系统的生理活动为基础的心理活动,而它的社会本质是人在社会生活中对客观世界的反映。智能是人类最核心的特征,如果脱离了对心智甚至大脑的认识,就无法理解人。换句话说,智能特指人类智能,其主要特点是有着血肉之躯。

人工智能是一门综合计算机科学技术、神经生理学、心理学、哲学等学科的交叉学科。它主要利用计算机模拟人类的智力活动,是没有肉身的智能体。其产生可以省去有机物所需的很多活动,比如新陈代谢和繁殖。广义上讲,人工智能技术是关于人造物的智能行为,这种智能行为一般包括知觉、学习、推理、交流,以及在复杂环境中的行为。从工程角度来说,人工智能就是通过人工方法使机器具有与人类智慧有关的功能,如判断、推理、证明、感知、理解、思考、识别、规划、设计、学习和问题求解等思维活动。它是人类智慧在机器上的实现。<sup>[4]</sup>

综上,人工智能是对人的意识和思维信息过程的模拟,是人制造出来解决问题、增强人类能力的机器。它不是人类智能,但它能像人那样“思考”,在单个领域有可能超过人类智能。社会心理学巨擘马斯洛曾经说过,科学是人类的创造,而不是自主的、非人类的东西;科学产生于人类的动机,它的目标也就是人类的目标。<sup>[5]</sup>人工智能是人生产出来的机器,它为人类服务,从属于人,这是理解人工智能的基本立场;由此而引出的服从原则可谓人工智能伦理的首要原则。

### (二) 弱人工智能、强人工智能和超人工智能

伦理对科技的规制,使得伦理学家们仍然要先行思考科技可能为人类社会带来的巨大风险。针对人工智能,来自伦理的担忧主要还是随着技术发展,当强人工智能甚至超人工智能出现时,会否给人类带来灾难?

所谓强人工智能(Artificial General Intelligence),是指机器不仅是一种工具,而且它本身拥有思维,是各方面能力均可与人类比肩的人工智能。有学者认为,强人工智能有两种表现形式:一种是类人的人工智能,认为机器的思考和推理与人的思维一样;另一种是非类人的人工智能,主张机器产生和人不一样的知觉和意识,使用与人完全不同的推理方式。弱人工智能(Artificial Narrow Intelligence)是擅长某一方面的人工智能,比如“阿尔法狗”,但它只会下象棋,不会进行其他复杂的智力活动,目前我们看到的人工智能多属弱人工智能。当强人工智能发展到一定阶段,几乎所有领域的认知表现均远超人类时,就成为超人工智能(Artificial Superintelligence),它能够进行科学创造、具有通识,甚至掌握社交技巧。

超人工智能的出现,才是人类担忧人工智能威胁自身的根本。成功预言“机器 1998 年战胜人类棋手”的美国未来学家雷·库兹韦尔认为,在不久的将来,计算机将拥有类人脑新皮质功能并超越人类,人类将与机器结合成为全新物种,一如美国科幻作家弗诺·文奇在《深渊中的火》(A Fire Upon the Deep, 1992)中描述的超级智能。库兹韦尔在《人工智能的未来》一书中,甚至预言 2045 年是人类蜕变的奇点。不过,也有科学家表示了不同认识,“机器全面超过人类,保守时间为 2050 年”。

预言并非危言耸听,人工智能的发展前景要比人们想象的要远得多,毕竟科学家们最大的企图是让人工智能理解人类的情感和文化,而这个“冰山”依然沉潜在实验室里。就目前来说,人工智能技术还停留在弱人工智能阶段,多数科学家认为,创造强人工智能还有很长的路要走。他们期待有朝一日,人工智能像《人工智能》中被输入情感程序的机器男孩大卫一样,能够理解人类的情感和文化,这样的强人工智能才是他们意欲摘得的该研究领域桂冠。

## 二、解放人还是人类的“增强版”

在人类为什么需要人工智能这个问题上,有手段论和目的论两种观点。手段论主张,作为一种技术,它的诞生以服务人类为宗旨,更好地促进人类社会发展。目的论认为,人工智能技术是为了弥补人类的“先天不足”,再造一个人类的增强版。目的论和手段论的核心是孰为主体,孰为客体的问题。手段论以人类为主体,人工智能为客体;目的论则认为人工智能本身就是主体。

事实上,人类由于自身在宇宙中的局限性,对机器人的幻想由来已久。据史料记载,西周时期能工巧匠偃师制造出歌舞艺人,进献给周穆王,其误以为真人,后因机器人不懂君臣礼节,使得周穆王大怒;公元前 2 世纪,古希腊人发明了会动的青铜雕像太罗斯,它以水、空气和蒸汽压力为动力,能够自己开门,还可以借助蒸汽唱歌;1738 年,法国技师杰克·戴·瓦克逊发明了一只机器鸭,它会嘎嘎叫,会游泳和喝水,还会进食和排泄;1927 年,美国西屋公司工程师温兹利制造了电动机器人“电报箱”,它装有无线电发报机,可以回答一些简单问题。

亚里士多德曾经说过,如果机器能干很多活,岂不让人类解放出来。追溯人工智能产生的历史,它的出现就是为了解放人,实现人的自由全面发展。18 世纪英国发起工业革命,开启了机器代替手工工具的时代。第一次世界大战后,各国工业迅猛发展,人们迫切需要创造出一种机器,代替人进行体力劳动,这个想法成为机器人产生的客观需求。1936 年,被称为计算机科学之父的英国数学家阿兰·麦席森·图灵设计了图灵机。图灵从理论上证明了任何数学运算过程均可用图灵机进行模拟。20 世纪 40 年代电子计算机出现,为人工智能的实现奠定了技术基础。随着计算机技术的发展,在经历了孕育时期(1956 年前)、形成时期(1956—1970 年)、暗淡时期(1966—1974 年)、知识应用时期(1970—1988 年)和集成发展时期(1986 年至今)<sup>[6]</sup>后,人工智能近两年又成为炙手可热的名

词,学术界、产业界和媒体均对该领域表达了高度关注。因为,正如库兹韦尔所说,依据加速回报定律,“人类 21 世纪的进步将是 20 世纪 1000 倍”。而作为 21 世纪世界三大尖端技术之一(基因工程、纳米科学、人工智能),人工智能带来的变革一触即发,行业会被颠覆,旧的系统有可能面临崩溃。

### (一)生活中的人工智能

不管对人工智能持何种态度,它都已经潜移默化地渗入到人们的日常生活中去,并日益深刻地影响我们的思维习惯和生活方式。《Already or Not: Artificial Intelligence Here We come!》报告显示,全球消费者认为人工智能在完成危险任务、更易获取相关内容、解决复杂问题、节约资源、提升人类健康等方面给人类社会带来惊喜。未来,人工智能在很多领域大有可为。

#### 1. 医疗机器人:医生的超级助手

依据人类的构想,基于生物传感技术(可穿戴设备)和人工智能的新医疗选择,将会使我们每个人成为自己的“健康 CEO”。大规模的基因测序和机器学习会使人类探明癌症、心脏等疾病的病因,而机器人外科医生则能完美地执行每次外科手术。

变革悄然而生。2015 年,由 IBM 公司研究的人工智能计算机系统沃森(Watson)受到全世界关注。这是纽约州威斯特彻斯特县北部 IBM 研究中心一个研究小组耗时近十年的研究成果。工程师们原意是生产出一个在电视智力竞赛中击败人类高手的机器人,之后,它被广泛运用到医学等领域。现在的沃森可以学习大量文献,通过“假设自动生成”完成诊断。美国最大的医疗保险公司 Wellpoint 预测,沃森能够在很大程度上帮助缩短诊断癌症的时间。除了沃森,一个名为“达芬奇系统”的机器人去年以来也迅速走红,它被誉为当今世界最先进的微创外科技术平台,代表了世界外科手术机器人的最高水平。

同样被用于医疗领域的机器人,还有来自日本的“大白”机器人护士 Robear。而早在 2013 年,由 InTouch 和 Bedford 公司联合开发的机器人被美国食品药品监督管理局(FDA)认可,标志着机器人拥有给人看病的能力。除此之外,荷兰埃因霍温大学的 RoboEarth 项目同样引起世人关注。该项目的 4 个机器人通过与云端服务器的交互,进行信息共享和互相学习。也就是说,一个机器人学会的知识和技能可以“教”会其他机器人。这项技术一旦成熟,将会迅速复制一系列智能高效的护理机器人,帮助解决人口老龄化的难题。

#### 2. 智能交通:人机交互是关键

人工智能研究者痴迷的领域还包括智能汽车,目前,相关研究正在大步向前。当这套智能汽车技术成熟,被广泛用于城市时,不仅可以改变城市的整体面貌,解决交通堵塞和停车难问题,还会大幅度提高人类的工作效率。

商业巨头们看好这一行业的美好前景,纷纷将大量资本注入智能汽车的开发中来。虽然过去几年这一技术高歌猛进,但由此而认为无人驾驶投放市场的时机已经到来,还为时过早,安全问题仍然是其中的关键。2016 年 5 月,特斯拉

的自动驾驶汽车未能在明亮的天空下,识别出一辆白色的拖拉机,导致驾驶员死亡。由于事故的发生,这一技术再次受人质疑。业界认为,如何实现人与汽车的良好交互,仍是其中最大的障碍。

### 3. 智能家居:珍视意识共享和物理共享

家,以物理空间的形式再现居住者对文化的理解。它不仅是住宅产业化的商品,还是多个产业的交叉点,更是人类文化生活的基石。伴随时代发展和科技进步,人们的价值观和生活方式发生了巨大变化,老龄化、少子化现象以及远程工作者和“全球游牧族”的逐渐增多,迫使各国在崭新的时代背景下思考新的商业模式。“这里面蕴藏着把‘家’囊括其中的可称为‘生活产业’的巨大产业前景。无论是房地产、家电、住房设备、建材,还是建筑领域之外的诸如车辆、食物、能源、物流、医疗以及福利等,都将与生活融为一体,并作为全新的产业形式滋生出各种新的消费需求。”<sup>[7]</sup>

智能家居并不简单指通过智能设备控制照明和温度等,而是强调所有物品和服务能够以惊人的速度轻易到手。比如,近几年广受人们关注的一种新型居住模式——共享社区。一定数量的陌生人共有同一空间并生活在一起,在形式上,与大学学生宿舍、公司宿舍以及社会上的合租房相似,这些群居方式主要是以经济的合理性为目的,体现了一种共享的生活方式。而在体验角度,智能家居能够在—个私密空间内提供多样的个性化服务。

近几年来,在国际消费类电子产品展览会(CES)上展出的智能家居让人脑洞大开:帮主人减肥的冰箱、能展示新闻的桌子、远程遥控的儿童看护系列产品等。基于人工智能技术而实现的智能家居,使人们生活变得越来越便利的同时,也可能导致无法体验的社会。人工智能时代,如何增长实践经验,跨越无法体验的社会这一鸿沟,需要从文化和人文的角度去思考。

#### (二) 争议:机器之心迷人又可怕

尽管人工智能的前景令人向往,但数年来对它的批评也始终不绝于耳,最著名的代表人物是霍金。具有反讽意味的是,由于患有运动神经元疾病,霍金需要使用机器与外界交流。美英两家科技公司为他量身定制出一款打字软件,使其“说话”速度因此而提升一倍。这种可“猜测”使用者思维的软件,让霍金这位“人工智能悲观论者”成为人工智能的受益者。霍金担心的是人工智能技术最终会发展出与人类智慧相当,甚至超越人类的机器,“到时它将可能以不断加快的速度重新设计自己。而人类则受制于生物进化速度,无法与其竞争,最终被超越。”<sup>[8]</sup>

据外媒报道,美国当地时间2016年10月19日晚,霍金出席“未来智能中心”开幕仪式时,再次表达了他对“人工智能可能导致人类毁灭”的担忧。他说,强人工智能的崛起,可能是人类遇到的最好的事情,也可能是最坏的事情,人工智能的发展可能帮助人类消除疾病、消除贫困、防止气候变化,但是也可能产生人们不愿意看到的种种后果,包括产生自主武器、造成经济灾难以及发展出同人

类发生冲突时拥有自己意愿的机器。<sup>[9]</sup>

人工智能悲观论的另一个代表休伯特·德雷福斯,则运用海德格尔和维特根斯坦的观点,对认知模拟和人工智能的生物学假设、心理学假设、认识论假设和本体论假设进行反驳,最后得出的结论是:当前人类面临的风险,不是超智能机器的降临,而是低智能人的出现。人工智能悲观论者大多受牛津大学哲学家尼克·波斯特洛姆的影响,其著作《超级智能:路径、危险性与我们的战略》中的理念说服了很多人。与此同时,更多科学家表示,不必对人工智能的未来过于悲观。虽然人工智能正根据人类几十年前确立的逻辑及科技上的可能性取得进步,然而在实现真正的智能方面并没取得任何突破。就像“西尔勒中文屋子”所揭示的那样:计算机只有语法,没有语义,因此,机器永远代替不了人脑。

在我国人工智能研究领域,持乐观主义态度的科学家居多。例如,在2015年中国人工智能大会上,中国科学院院士、中国人工智能学会副理事长谭铁牛提出,“人的大脑是一个通用智能系统,可以举一反三、融会贯通。与之相比,当前人工智能的差距还比较大,比如有智能没智慧、有智商没情商、会计算不会‘算计’”。他认为,人工智能和人类智能各有所长,因此需要取长补短,“人机共存”将成人类社会结构的常态。

如果说科学家们看待人工智能更多地从全人类视角,那么普通大众关注的人工智能变革则更接地气,失业和隐私泄露的问题不容小觑。人们预测,律师、记者、上班族,甚至 CMO 均会因此加入人才迭代的大军。也有人认为,虽然人类一部分职业未来会被机器人代替,但是人工智能也会改善现有岗位,并催生一些新的行业。总之,站在人工智能的风口浪尖,人们唯有不断更新知识体系,增加新的技能,持续“长本领”,才不至于在新一轮技术革命的浪潮中被残酷淘汰,这也符合人类社会优胜劣汰的规律。

### 三、人工智能的人文化:做人类的合作伙伴

人工智能技术既会带来技术上的变革,也会带来文化上的变革。比如,随着“情感智能机器人”进入家庭,它们承担部分家教和陪伴的任务。相关的人文挑战也日趋显著:与机器的情感关系将怎样改变家庭关系和婚姻关系?谁来主导情感软件?人类有无操纵情感甚至放纵情感的隐患?这些伦理问题不容忽视,因为家庭关系是一切社会的人文基础。再如,手术机器人进入手术台后,医生的角色和功用是什么?如何界定医疗事故的责任?谁对濒临死亡病人的医疗方案做最后“判决”?除此之外,人工智能技术创造的财富在社会上如何分配,会不会加剧贫富差距或不平等现象?机器伤害或杀害人类的事故,到底是机器的责任还是人类自己的责任?这些机器的开发者、使用者是否要负相应责任,分别应负什么样的责任……诸如此类的伦理问题,均是这一技术所带来的人文挑战。

#### (一)人工智能的人文化

“人文”(humanitas)一词,最初与“天文”相对应而提出。《周易》中《贲

卦·彖传》称：“刚柔交错，天文也。文明以止，人文也。观乎天文，以察时变；观乎人文，以化成天下。”什么是人文？即文明以止。也就是说，人要规范自己的言行举止，根据自己的身份做该做的事情。人之所以是万物之灵，就在于它有人文，有独特的精神文化，知道止于何处。

科技的本质在于创造先进工具，使人类世界更美好。其精神实质是求真，获得真知，意即“打破砂锅问到底”。而伦理贵在求善，帮助人类在复杂多变的客观世界里寻求安身立命之道。近年来研究热度不减的应用伦理学既不是案例应用，也不是模型应用，而是一种批判性、反思性的应用。它批判的是先进的科学技术不能偏离人道主义的轨道，不能成为奴役人的工具；反思的是“枪不会杀人，只有使用枪的人才会杀人”的价值理念。

近代以来，科技被视为人类社会发展的决定性力量，并日益脱离人类的控制。始于叔本华、尼采的反对科学主义的思潮由是兴起。作为现象学代表的德国哲学家胡塞尔就曾表示：“我们的出发点在于对科学的估价的改变，这种改变是20世纪的末期出现的，这种估价的改变并不牵涉到各门科学的科学性，而只牵涉到各门科学、即绝对意义上的科学，对于人的实存所已经赋予的和可能赋予的意义。”<sup>[10]</sup>胡塞尔认为，科学的基础应该是以人作为主体而进行的探索。

所谓人工智能的人文化，意即强调科学家或科学共同体的道德义务，督促他们创造出有道德的人工智能体。这种人工智能体关怀人和人类社会的利益，以人为主体，重视人的价值，尊重人的尊严和权力，与人类一起迈向自由、平等和解放。

## （二）延展智能：将人类价值观嵌入人工智能体

“机器的自由化程度越高，就越需要道德标准”。<sup>[11]</sup>应对人工智能带来的人文挑战，为机器设置人类的价值观十分重要。在麻省理工学院的媒体实验室里，科学家用“延展智能”来表达这个问题。所谓延展智能，也就是说如何将人类的价值观嵌入人工智能。

首先，从源头上堵住不利于人类社会发展的的人工智能体出现。规范人工智能科学家的行为，在技术研发前引入伦理委员会机制，预测风险，并在程序设置中加以控制，重要的是找到使用人工智能造福人类的人。其次，人工智能体的设计，尤其是机器人的设计和应用必须遵循阿西莫夫的“机器人三大定律”，并以此为法则，在“不伤害”原则的前提下，培养人工智能体与人类的良好交互能力。“机器人三大定律”首要原则是不能伤害人类，或者保护人类不受伤害；其二是听命与人，但不得违反第一原则；其三是保障自身安全，但同时不得违背上述两点原则。三大定律定位了作为人工智能体的机器人，仅仅是人类在自身发展过程中创造的一个工具。不管机器人有无意识，有无自我决策能力，它都必须遵守“不伤害”的原则，这是应用伦理学秉承的核心原则，而其产生的最大诉求就是为了人类社会运转更经济、更有效率。再次，政府的管控不能缺位，要保证研究结果获得分享，不让某个团体单独受益。当一项能够给人类带来很大便利的技术



投放市场时,政府、私人企业和学术界谁来控制它,就成为一个必须明晰的课题。政府的管控是否缺位、及时,引导是否正确、到位,都是制约人工智能发展偏离轨道的重要因素。最后,人类要保持对大自然的敬畏之心,对理性抱有谨慎的信心。要知道,我们目前拥有的这些难以置信的感觉,相对于宇宙和自然界来说,还仅仅是沧海一粟,我们没理由狂妄自大,而应时刻提醒自己,科学探索无止境,技术使用有禁区。

### (三)人工智能的价值观

在有限制的基础上进行人工智能研究,从第一个关口“卡”过人工智能被滥用的可能。接下来的问题是将人类哪些价值观注入人工智能,使其更好地为人类服务。首先是以人为本的价值观。以人为本是最基本的人文素养,也是技术应用于人类的终极价值。其深层涵义在于:第一,珍视人的生命、保护人的尊严;第二,尊重人的自由和平等;第三,发扬民主精神和互爱精神;第四,促进人的全面发展。人工智能的产生是为了人,使用是为了人,它的消灭同样也是为了人。假设科学家或者伦理学家预测到人工智能体可能给人类带来灾难时,必须果断中止研发,对自己的行为负责。

其次是公正的价值观。技术进步最终受益者应是全人类、全社会,而不应成为加剧贫富差距,促进社会不公平的工具。以导航和无人驾驶为例,人工智能已经开始改变交通技术,但它绝不能止于让驾驶更安全、更经济、更便利。智能车的未来在于“消灭三大差别”:高档车和低档车之间的社会地位差别、有无停车位的使用成本差别、开车能力高低之间的安全差别。

再次是和谐的价值。在人工智能技术的实践应用中,其与人类和大自然之间的关系必须秉承和谐原则,这将有利于人工智能深化服务质量,满足人类需求。在人工智能的应用过程中,单纯以善为指导的伦理标准已越来越不适应社会发展以及人类当下的道德评判标准。随着科学的进步,人工智能的应用逐渐融入到人们的生产生活中,“不伤害”原则虽然是基本原则,但是也存在一些弊端。例如片面性,考虑并体现了一方的利益,必然就忽视了另一方的利益。而基于三大定律的最后一条原则,对于将来有可能出现的具有自主性,或半自主性的机器人而言,尤其是类人型、智能化机器人来说,为了实现人机之间的互相尊重,机器人与自然间的相互融合,这就需要有一个更有效的价值观,即和谐的价值。它不仅体现在机器与人之间,更为重要的是体现了机器、人与自然三者之间的和谐。

最后是可持续发展的价值观。科技的进步理应对现在和未来负责,对子孙后代负责。现代科技的不可逆性,警示我们不能随便拿人类命运冒险,更不能将技术的进步当作“赌注”,与人类未来和大自然较量。在现实社会,科技的进步方便了人类,但是也有可能使人的某些能力退化。例如,计算机时代的到来,特别是最近几年VR的兴起,有可能给人类的沟通能力带来障碍,导致无法体验的社会。再如,人工智能可能替代执行标准化的常规任务,但不是消灭工作岗位。人

类发明一项技术,并不只是要看它的短期效用,还要看它能否促进人类的进化,有利于人类社会的长期可持续发展。

#### 四、结 语

科技的发展显示了人类智慧的无限潜力。历史上任何一次技术大变革并未使人类走向灭亡,反而是文化的消解、人文精神的匮乏,容易导致朝代更迭、文明消失。我们没有理由武断地认为人工智能就是善的,或者就是恶的,因为简单以此判断,已经不适合现代文明,播种善因、规避恶意,才是更加理性的作为。毕竟,我们制造工具,而工具让我们走得更远。

伦理对技术彰显的是责任,而技术的发展需要伦理的介入和匡正。诚如马云所言,过去的机器是人类的工具,未来的机器是人类的合作伙伴。我们要明白机器不可能有智慧、使命、价值观和文化体系,人类最大的优势在于对文化的把握、对愿景的思考,以及巨大的想象力。未来,不是机器换人,而是机器扩人、机器升人、机器人。

#### 注释:

[1][英]默里·沙纳汉:《技术奇点》,霍斯亮译,北京:中信出版社,2016年,第95页。

[2]机器之心编译:《普通消费者怎么看待人工智能?新报告认为大部分是乐观的》,http://jiqizhixin.com/article/1691。

[3]参见斯坦福大学“人工智能百年研究”持续性项目发布首项成果——《2030年的人工智能与生活》,http://www.raincent.com/content-108-72。

[4]曹承志等编著:《人工智能技术》,北京:清华大学出版社,2010年,第2页。

[5]刘大椿:《在真与善之间:科技时代的伦理问题与道德抉择》,北京:中国社会科学出版社,2000年,第159页。

[6]蔡自兴等:《人工智能及其应用》,北京:清华大学出版社,2010年,第3-9页。

[7][日]原研哉主编:《理想家:2025》,上海:生活·读书·新知三联书店,2016年,第18页。

[8]霍金:《人工智能技术应谨慎发展》,新华网,http://news.xinhuanet.com/world/2014-12/03/c\_1113501243.htm。

[9]参见人民网科技频道:《霍金再次发警告:人工智能可能消灭人类》,http://ah.m.people.cn/n4/2016/1022/c155-7763434.html。

[10][德]胡塞尔:《欧洲科学危机和超验现象学》,上海:上海译文出版社,2005年,第127页。

[11]Picard, Rosalind, *Affective Computing*. Cambridge, MA: MIT Press, 1997, p.19.

〔责任编辑:汪家耀〕