社会科学进展

2019年6月第1卷第1期



人工智能的伦理规范问题研究

胡元蛟

中共合肥市委党校, 合肥, 230031

邮箱: husong12345@163.com

摘 要:本文旨在通过讨论人工智能所带来的焦点伦理问题,指出开展人工智能伦理规范问题研究的必要性,引起人们对人机伦理关系的关注,进而试图廓清人机伦理关系的应有之意,为现实社会中的人工智能发展建立更加完善、有效的伦理规范机制,以期能引起相关领域的从业人员及政府决策者的关注和重视。

关键词:人工智能;伦理;规范

收稿日期: 2019-04-12; 录用日期: 2019-05-22; 发表日期: 2019-06-06

Research on the Ethical Norms of Artificial Intelligence

Hu Yuanjiao

CPC Hefei Municipal Party School, Hefei 230031

Abstract: This paper aims to discuss the ethical issues brought by artificial intelligence, points out the necessity of carrying out research on artificial intelligence ethics, and

文章引用: 胡元蛟. 人工智能的伦理规范问题研究 [J]. 社会科学进展, 2019, 1 (1): 40-48. https://doi.org/10.35534/pss/0101005

call people to pay attention to the relationship between human and ethics, and then tries to clarify the meaning of human-computer ethical relationship. Finally it attempts to propose to establish a more complete and effective ethical normative mechanism for the development of artificial intelligence in the society, for purpose of attracting the attention of practitioners and government decision-makers in related fields.

Key words: Artificial intelligence; Ethics; Norms

Received: 2019-04-12; Accepted: 2019-05-22; Published: 2019-06-06

Copyright © 2019 by author(s) and SciScan Publishing Limited

This work is licensed under a Creative Commons Attribution 4.0 International License. https://creativecommons.org/licenses/by/4.0/



1 引言

人工智能起源于 1956 年的达特茅斯会议 [1],这一概念一经提出,就引起了众多学者的关注与研究。人们致力于让机器像人一样具有智能,能够与人进行自然的人机交互。从语音识别、机器翻译到医疗诊断、自动驾驶 [2],基于深度学习算法的人工智能在应用层面上正在不断成熟,甚至在很多方面已经开始超越人类,一个人工智能的美好时代似乎已然来临。然而与此同时,人工智能的迅速发展也不可避免地给人类生活带来了负面影响,类似机器事故引发的担忧提醒人们在乐观期待未来的同时,也应关注人工智能可能带来的风险,这就产生了人工智能伦理问题。

人工智能技术的突破及其高速发展,深刻改变了人们的生产、生活和工作方式,为更深层次促进人类社会发展提供了有力支撑和新的途径。然而,人工智能技术在服务人类社会的同时,产生了许多新的伦理问题并引起广泛关

注。从近年来出现的智能护理机器人代替人类工作、自动驾驶汽车撞人致死、AlphaGo打败人类围棋顶级高手等社会热点来看,这些极具前景的人工智能应用,在显著提升服务水平的同时,也引发了诸多伦理问题,带来了各种不容忽视的风险。如何应对这些迅速凸显的伦理问题,建立未来人工智能发展的伦理规范机制,成为一个非常必要且十分紧迫的课题,对于人工智能技术切实发挥作用从而真正服务人类社会具有重要意义。

2 相关概念与研究

2.1 人工智能的概念

人工智能分为以深度学习为代表的弱人工智能和目前难以完全实现的强人工智能。前者主要是指各种模拟人或动物智能解决各种问题的技术,包括问题求解、逻辑推理与定理证明、自然语言理解、专家系统、机器学习、人工神经网络、机器人学习、模式识别、机器视觉等[3];后者是指有自我意识、自主学习、自主决策能力的人工智能,是人工智能发展的终极目标[4]。

2.2 人工智能伦理的概念

"伦理"一词源于希腊文,意为"风俗""习惯"[5]。无论在西方还是我国,不同学派的伦理观差别很大。伦理学是哲学的分支,是研究社会道德现象及其规律的科学,其不仅可以建立起一种人与人之间的关系,而且可以通过一种潜在的价值观来对人的行为产生规范、制约和影响,使人们在实践中遵循社会普遍接受的准则。

有学者认为,人工智能伦理只是相应的人将人类伦理的概念强加在机器身上,从而才有了这一概念,反映了人类对于机器的高度期望,期望其能够像人类一样理解他人的想法,并与人类进行自然的交互。随着建立在现代科学原理上的一系列高新技术的发展,以及人类利用高技术控制自然能力的提高,人工智能技术似乎使得人类所创造的产物越来越具有理智并挑战现有的人类伦理,加上机器人可复制和再生产的优势,人工智能伦理已经成为一个不得不面对的

研究课题。

2.3 国内外对于人工智能伦理的相关研究

从国外研究来看,多数学者认为人工智能伦理与机器(人)伦理本质上没有太大区别,在人工智能一词诞生以前,就有学者对机器与人的关系进行过研究。有关学者认为:机器伦理关注于机器对人类使用者和其他机器带来的行为结果[6];随着机器越来越智能化,其应当也承担一些社会责任,并存在伦理观念,这样可以帮助人类更好地进行智能决策;人类应该尽快制定机器(人)相关方面的道德伦理准则。目前,国外对于人工智能伦理规范问题的探讨相对较多,主要集中在机器人技术与算法设计、法律、安全与隐私以及社会伦理问题方面的伦理难题讨论、伦理规则制定与面向公众的调查及实证研究。比如 2005 年欧洲机器人研究网络制定了《机器人伦理学路线图》、韩国工商能源部颁布了《机器人伦理宪章》、德国交通部下属伦理委员会发布了《自动和联网驾驶报告》、美国麻省理工学院媒体实验室上线"道德机器"网站进行了关于"道德算法"的大众调查等[7]。

相对而言,国内对人工智能伦理的研究起步较晚,但近年来,一些学者也开始重点研究人工智能的伦理规范方面。相关研究主要包括:①探讨机器人伦理和机器伦理的内涵与发展[6][7]。段伟文(2015)认为机器人伦理与机器伦理的不同之处在于机器人是否具有自主性以及如何与人类相处;于雪等(2016)认为机器伦理思想引导技术产品"负责任"地为人类服务,但也存在将道德行为数字化等局限性。②初步探讨人工智能与伦理道德、价值判断、法律责任、技术人文化等的关系[4][8][10][12]。孟伟等(2018)从现象学伦理学角度探讨了人工智能的一般性技术伦理问题和未来可能具有自主意识的特殊伦理问题;吴汉东(2017)提出,对于人工智能引发的负面风险,要形成以伦理为先导的社会规范调控系统和以技术、法律为主导的风险控制机制;周剑铭等(2018)认为人工智能技术应融合于"科学与人文"和"中、西文化"两种文化之中,这样才有可能促使人工智能更有效地为人类服务。总之,我国学者已经开始从单纯的技术伦理问题转向人机交互关系中的伦理研究。

3 人工智能伦理的焦点问题与典型案例

3.1 人工智能是否能从根本上代替人类工作

美国未来学家雷蒙德·库兹韦尔提出的奇点理论认为"纯人类文明将终止于 2045 年",就是说机器的智能很快就将超过人类,这使得人们更加担忧机器是否将会全面替代人类,但已有研究认为人工智能并不能从根本上代替人类的工作岗位。这是由于:①虽然机器与人类分别有自身的优势,但人类的优势是机器在短期内无法比拟与模仿的。人类具有思维能力,并可以在资源有限的情况下进行非理性决策。②目前人类对机器仍然占据主动,且产生不可逆的优势。机器的产生与发展是建立在人们对其环境的了解与改造上的,如果没有人类的指引与改造,机器只能停留在低端的机械重复的工作层次。③人工智能的高速发展给人类带来了机遇。人工智能的飞速发展整体拉动了相关产业(如服务业、金融业等)的发展,从而可以创造出更多的就业机会〔8〕〔16〕。

例如,以智能护理机器人为核心的智能老龄护理逐渐成为今后养老服务体系最具潜力的可行方案,但要解决其所涉及的诸多伦理问题,如权益保障、社会正义等伦理问题,必须通过人类利用自身独特的思维能力,对智能护理机器人进行指引与改造,构建出理性的智能老龄护理体系[9]。同时,智能护理虽然取代了传统家政服务的一部分工作,但其并不能替代家庭成员的心理慰藉以及尊严呵护等情感照料。此外,其未来的广泛应用也会拉动高新制造业、信息技术产业、服务业等相关产业的发展,从而带来更多的就业机会,而且新兴技术创造的工作机会还要高于所替代的传统工作机会。

3.2 人工智能事故存在责任分配困惑

如果人工智能出现事故致人伤亡、致财产毁损,法律责任应如何界定与分担? [10] 在关注事故本身之外,人们似乎更加担心机器作为执行主体,发生事故责任的承担机制。究竟是应该惩罚那些做出实际行为的机器,还是那些设计或下达命令的人,或者两者兼而有之? 如果机器应当受罚,该如何处置呢?目前也没有相关法律进行规范和制约。有关学者认为,如果将事故责任完全归

咎于机器智能系统设计方以及操作人员,却忽视机器一方的责任是有失偏颇的。 因为有时我们会发现系统发生故障时往往会出现未知甚至诡异的行为方式,并 不能够自行修复,操作人员也难以用经验察觉或下达指令进行恢复。当人工智 能开启机器完全自动模式时,系统是不稳定的,很容易造成不必要的事故。所以, 在一些复杂的人机环境系统中,人工智能事故仍然存在责任分配困惑[8][11]。

2018年3月,美国亚利桑那州一辆优步无人驾驶汽车在路测时致行人死亡,这是全球范围内曝光的首例无人驾驶汽车撞死路人的交通事故。这不得不让我们联想到2016年美国特斯拉无人驾驶汽车发生重大事故,导致司机丧生的事件。两年后的今天,从车主身亡到路人丧生,无人驾驶汽车所引发的交通事故依然面临着法律责任界定的困惑与困难。这其中包含了制造商、乘客、行人、监管部门甚至是汽车自身等多重主体之间的法律关系,以及对相关法律责任的分配等问题。[14]

3.3 在人机伦理关系中人类占据何种地位

人类创造人工智能的目的在于让机器变得更加智能,代替人类完成更多、更复杂甚至更危险的工作。虽然需要防范人工智能的"不确定性"所带来的巨大伦理风险,但可以预期的是,人工智能技术离我们设想的智能程度还相去甚远。面对机器越来越多地在某些领域超越了人的功能,以及个人如何与人工智能相处等人机伦理关系问题,有关学者认为,要正确理解机器、人工智能与人的智能之间本质的不同,明确人类在人机伦理关系中所占据的主导地位,调整现有的社会期望和行为,一要理性对待人工智能"不确定性"带来的希望和恐惧;二要始终将为人类服务、为人类进步做贡献作为人工智能的终极发展目标;三要充分理解人工智能是人类追求实现自身的意义、价值和地位的手段之一;四要人类必须承担相应的伦理责任,建立基于道德、法律、教育、社会制度等的伦理规范[8][9][16]。

例如, AlphaGo 战胜人类棋手后,似乎产生了一种恐惧,即"人工智能挑战了人的地位"。表面上看好像是"AlphaGo 战胜了人",但实际上,作为人工智能的"机器学习"只是因机器的模仿而表现了"学习功能"。而所有机器

模仿得到了类似人的智能在某方面的功能,都是基于人的模仿,正是这一点决定了人机关系。这说明了机器并不能像人一样成为天生的主体学习者;同时,AlphaGo的成功不是算法或逻辑的胜利,而是人的模仿的成功[9]。在这种意义上,AlphaGo在人、机对弈中的胜利最终证明了人的智能的胜利,体现了人类在人工智能发展中占据主动、发挥创造性潜能的自身价值。

4 建立人工智能伦理规范机制的对策

基于人工智能带来的伦理问题以及对人的价值和意义的重新认识,人类必须坚守自身的伦理信念,在人工智能的发展中承担起相应的伦理责任,建立起相应的伦理规范贯彻在自己所研究、制造、开发的人工智能中,使其服从甚至具有人信守的伦理准则「14」。为此,本文从以下几个方面提出对策建议。

4.1 嵌入符合人类社会规范的伦理道德

通过"自上而下"或"自下而上"的方式将符合人类规范的道德行为转化为数字符号,用道义逻辑、认知逻辑和行为逻辑等计算手段实现机器人伦理[6]。

已有研究认为,"自上而下"的伦理建构,即"运用道义逻辑将道义论的义务和规范转换为逻辑演算,同时通过净利益的计算实现功利的算计与权衡,机器人能够从一般的伦理原则出发对具体的行为做出伦理判断";"自下而上"的伦理建构则"通过机器学习和复杂适应系统的自组织发展与演化使机器人能够从具体的伦理情境生成普遍的伦理原则,通过道德冲突习得道德感知与伦理抉择的能力,最终使机器人的人工伦理能力得以从一般的人工智能中涌现出来[7]"。

4.2 贯彻与人工智能系统相匹配的伦理原则

首先,在人工智能研发人员中贯彻基本伦理准则,如有益性、包容性、透明性、隐私保护等[15];其次,建立人工智能伦理审查制度,规范人工智能伦理评估与审查流程,组织跨学科专家对人工智能技术和产品的伦理影响进行评估并提出建议[16]。只有将人类社会的伦理原则与人工智能系统紧密、有效地融合,才能让研发出的人工智能系统为人类所用,不致偏离人类伦理规范的轨道。

4.3 制定人工智能算法标准和规范

参考国际人工智能相关的算法标准和规范设计,建立符合国情的、切实可行的伦理标准和规范,从技术上对软件和算法设计进行规范和监管,防止恶意软件和算法产生,建立第三方检测、评估、认证和审批制度[16]。如编写《人工智能深度学习算法评估规范》,加快推进人工智能深度学习算法评估标准研制工作;进一步完善人工智能标准化体系等。

4.4 加强人工智能伦理风险防范教育

通过教育和宣传提高全体公民的人工智能素养,让公民知晓并防范人工智能技术可能带来的风险 [16]。借助互联网平台开展网络道德教育,树立正确的人工智能伦理观念;加强对人工智能技术人才的价值观培育;普通公民自身也应增强人工智能伦理自觉意识,传承优秀传统文化理念,提高伦理自律性和伦理责任心,最大限度地防范人工智能的伦理风险。

参考文献

- [1] 王冲鶄. 人工智能全球新浪潮 [J]. 中国信息化, 2017 (8): 9-11.
- [2] 薛风平. 中国人工智能产业发展路径研究[J]. 中共青岛市委党校. 青岛行政学院学报, 2018, 248(2): 20-24.
- [3] 熊英. 人工智能及其在计算机网络技术中的应用[J]. 技术与市场, 2011, 18(2): 20-20. https://doi.org/10.1365/s35141-011-0142-0
- [4] 孟伟,杨之林.人工智能技术的伦理问题———种基于现象学伦理学视角的审视「J].大连理工大学学报(社会科学版),2018,39(5):112-116.
- [5] 顾梅珑. 理性伦理的价值解构与审美伦理重构 [J]. 常熟理工学院学报, 2018 (3): 54-59.
- [6]于雪, 王前. "机器伦理"思想的价值与局限性[J]. 伦理学研究, 2016(4): 109-114.
- [7] 段伟文. 机器人伦理的进路及其内涵 [J]. 科学与社会, 2015 (2):

35-45.

- [8] 郑添元. 人工智能与伦理法律问题的思考 [J]. 商业经济, 2018 (4): 134-135.
- [9] 周剑铭,柳渝.两种"两种文化"交汇中的人工智能[J].科学与社会, 2018(1):59-71. https://doi.org/10.1111/jsap.12807
- [10] 吴汉东. 人工智能时代的制度安排与法律规制 [J]. 法律科学: 西北政 法学院学报, 2017 (5): 128-136.
- [11]刘伟, 赵路. 对人工智能若干伦理问题的思考[J]. 科学与社会, 2018(1): 40-48. https://doi.org/10.1097/01.NURSE.0000547798.23399.a2
- [12] 罗定生,吴玺宏. 浅谈智能护理机器人的伦理问题[J]. 科学与社会, 2018(1): 25-39. https://doi.org/10.1007/s12496-018-0024-3
- [13]白惠仁. 自动驾驶汽车的伦理、法律与社会问题研究述评[J]. 科学与社会, 2018(1): 72-87.
- [14] 苏令银. 能将伦理准则嵌入人工智能机器吗——以无人驾驶汽车为例 [J]. 理论探索, 2018(3): 38-45.
- [15] 张秀,田原,孙鹏. "大数据与社会公正"——上海市伦理学会第五届 青年论坛综述 [J].毛泽东邓小平理论研究,2017(10):102-104. https://doi.org/10.1080/1756073X.2017.1287043
- [16] 莫宏伟. 强人工智能与弱人工智能的伦理问题思考[J]. 科学与社会, 2018, 8(1): 14-24.