人工智能哲学研究述评

许 勇[1,2],黄福寿[1]
(1.上海师范大学马克思主义学院,上海 200234;
2.上海交通大学学生工作指导委员会,上海 200240.)

摘 要:人工智能哲学是人工智能技术与哲学联姻的产物,是系统研究人工智能技术的性质、演化规律、运行规律的一个哲学分支,其研究领域包括人工智能主体性研究、伦理规范研究、劳动分工研究、人的解放研究等领域。本文梳理了近五年人工智能哲学的研究成果,主要包括人工智能主体性研究、人工智能伦理问题研究、人工智能影响社会分工研究、人工智能推动人的解放研究四个方面研究内容。通过对这四个方面研究内容的分析,发现人工智能哲学领域还存在强人工智能时代人的主体性问题、伦理规范形成机制、情感伦理问题、人工智能立法、教育体系变革、人的解放前景、技术与哲学协同发展、构建人工智能哲学体系等课题需要进行更加深入思考和研究。

关键词:人工智能;哲学;伦理问题;劳动分工;主体性;人的解放;
DOI:10.13806/j.cnki.issn1008-7095.2020.01.015

人工智能哲学是人工智能技术与哲学联姻的产物,是系统研究人工智能技术的性质、演化规律、运行规律的一个哲学分支,其研究领域包括由人工智能产生的伦理问题、劳动分工变革、主体性挑战、人的解放等领域。以人工智能为主题词搜索,2014年1月1日至2018年12月31日期间,在北大中文核心和中文社会科学引文索引(CSSCI)发表的论文共2 053篇,其中包含哲学关键词的有114篇,本文选取其中有代表性的数十篇,并在图书馆查阅相关专著10余本,通过对这些文献的研究,笔者试图对当前人工智能哲学研究现状进行述评。

一、研究现状综述

(一)关于人工智能主体性的研究

人工智能主体性的问题将关系到人工智能伦理、法律等主体间关系的界定,是首先必须研究清楚的内容,其研究内容包括人工智能主体性内涵研究和人工智能主体性与人主体性关系研究。

1. 人工智能主体性内涵研究

围绕人工智能主体性内涵,学者们尝试通过主体性建构、意向性渊源、意向性建构和电子人格等方式对其进行解读。

关于人工智能自主体建构方面,代表性的文章有高新民的《自主体人工智能建模及其哲学思考》,其尝试了对自然自主体进行解剖,提出了几种不同角度的框架:"1)既能对所处环境进行

作者简介:许勇,上海师范大学马克思主义学院博士研究生,上海交通大学讲师;黄福寿,上海师范大学比较政党研究中心主任,教授,博士生导师。
基金项目:2019年上海市德育课程教学研究基地研究项目。

感知,又能根据自身的意愿、目标对环境作出反映的智能实体。2)具有感知能力、解决问题能力、与外界通信能力的全智能或半智能的实体。3)在没有外界干涉的环境下也能完成给定的任务的对象。"①作者讨论了慎思式自主体、反应式自主体、BDI 自主体(Belief 信念、Desire 愿望、Intention 意图)等自主体形式,认为现有的人工自主体还无法体现意识这一人类意向性的最根本特征。

关于人工智能意向性的渊源,余乃忠提出人工智能与自动化的根本区别在于人工智能拥有自我意识和对象意识。在马克思看来,人之所以具有自我意识和对象意识,其根源在于实践的对象性和目的性,马克思指出"人对自身的关系只有通过他对他人的关系,才成为对他来说是对象性的、现实的关系"。② 因此,余乃忠认为可通过人工智能对其对象的占有来确定人工智能的自我意识。作者认为人工智能可以在对其对象关系不断深化中逐步确认人工智能的自我意识。③

关于意向性存在的标识,维特根斯坦的《哲学研究》一书中曾这样解读:"只有对于人类和他们的类似者(行为上类似),方可断言:其有感觉;其能看到,还是失明;其能听到,还是失聪;其是有意识,还是无意识。"④根据此理论,常照强指出"如果智能机器(机器人)能在行为上与人类相类似,可认定这样的智能机器(机器人)是能够思维的",⑤从而可以推断智能机器(机器人)拥有一定的意向性。

关于人工智能意向性的建构,周昌乐提出机器意识研究的五个类别:"面向感知能力实现的、面向具体特定意识能力实现的、面向意识机制实现的、面向自我意识实现的、面向受蕴能力实现的,对于机器意识的开发,应搁置主观体验方面(身心感受)的实现研究,而研究意向性心识能力(环境感知、认知推理、语言交流、想象思维、情感发生、行为控制),采用脑和量子计算集合的方法,最终开发出有一定意向能力的机器人。"⑥

关于人工智能主体性的承载形式,蓝江提出"电子人格"的载体,讨论了人工智能的"电子人格"是否可以进入到与人类主体平等协商和对话的公共理性中,将人类与人工智能的关系类比于人与动物的关系,以基于对人自身的反射式关怀给予人工智能一定的权利。⑦

2. 人工智能的主体性与人的主体性关系研究

针对人工智能是否会挑战人的主体性,现存三种不同观点:一是认为人工智能拥有自主性并会挑战人的主体性。一旦人工智能拥有了主体性,是否会导致人类的存在危机:因为劳动的减少,人觉得生活无意义。邬桑尝试了四种自主理论:自由意志论、融贯论、理性回应论、推理回应论,指出理性回应论是唯一可能为人工智能自主性辩论的。根据理性回应论的定义:当一个人能够意识到理性,并有能力根据理性做出决定和行动的时候,他是自主的。在此理论下,人工智能在行动和道德理性上都是自主的。⑧

二是认为人工智能永远无法挑战人的主体性。持这样观点的学者认为人工智能仅仅是人的物化,机器没有私欲观念,不可能与人完全对立。人工智能在思维能力上无法超越人类思维

① 高新民,张文龙.自主体人工智能建模及其哲学思考[J].自然辩证法研究,2017(11):3-8.

② 马克思,恩格斯.马克思恩格斯文集(第一卷)[M].中央编译局,译.北京:人民出版社,2009:165.

③ 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.

④ [英]路德维希·维特根斯坦.哲学研究[M].李步楼,译.北京:商务印书馆,2000.

⑤ 常照强.中文屋思想实验的两个问题域——兼议维特根斯坦的人工智能哲学[J].洛阳师范学院学报,2015(12):30-33.

⑥ 周昌乐.机器意识能走多远未来的人工智能哲学[J].人民论坛(学术前沿),2016(13):81-95.

⑦ 蓝江.人工智能与伦理挑战[J].社会科学战线,2018(1):41-46.

⑧ 邬桑.人工智能的自主性与责任[J].哲学分析,2018(4):125-134.

和意识的整体性,不能产生人类主体性所依赖的社会关系和实践基础。人工智能领域存在的莫拉维克悖论指出:计算机在逻辑运算能力上无比优异,却在感知和行动能力上不如一岁的小孩。陈凡根据马克思的"人的主体性是在实践活动中形成并在实践中得以确认和强化"的观点,认为人工智能是人脑的延伸,其行为不具备实践性,因此人工智能不具备主体性。⑨ 王晓阳通过基于集体人格同一性的论证,证实了无法实现生产一种新的能以人类智能相似的方式做出反应的智能机器。⑩ 黄兆明认为尽管人工智能模拟了人的思维,但它没有也不可能穷尽人类的全部思维规律,不可能统治人类。⑪ 孙伟平从存在论、认知论、价值论的角度说明了人工智能无法真正获得主体地位,即使有人认可了人工智能产品的主体地位,这种地位也是人类赋予的。⑫ 蔡曙山从神经、心理、语言、思维、文化五个认知层级论证了人工智能都是在模仿人类智能,目前人工智能都远逊于人类智能,人工智能是在不断进步的,但在总体上并未超过人类智能。⑬

三是认为人工智能有主体性,但可以与人的主体性共存。叶妮通过对马克思的实践主体特质和乌托邦的实践模式的考察,得出了人工智能不会引起人的主体性崩塌,反而最终形成"人-物"和谐共存的结果。⑭ 潘坤认为人工智能可能带来的危机在根本上并不是人类被取代和统治的危机,而是人类自身的存在危机,人类应架构全新世界观,化解人的存在危机,最终实现人与人工智能的"同居"。⑮

(二)关于人工智能伦理问题的研究

人工智能伦理作为人工智能中的道德哲学,回答了人工智能与人之间的哲学关系和伦理规范,形成了伦理规范研究和具体伦理问题研究两大领域。

1. 伦理规范研究

一是伦理规范的形成机制研究。孙伟平主张成立人工智能伦理委员会,以人本原则、公正原则、责任原则为基本价值原则来建立人工智能伦理规范。⑯ 段伟文提出了构建更加透明、可理解和可追责的智能系统,并将伦理问题研究分为:凸显主体责任伦理研究、基于主体权利的权利伦理研究和探讨伦理嵌入的机器伦理研究,指出必须让伦理原则应用于实践,从具体问题入手强化人的控制,最终发展"基于负责任态度的可接受的人工智能"。⑰

二是人机伦理关系研究。人与人工智能机器未来将呈现怎样的关系?于雪提出了以"相互依赖""相互渗透"和"相互嵌入"为特征的人机关系递进结构模型,通过分析了人类将功能、意向和责任等机体特性赋予作为"人工机体"的机器的过程,展示了这一过程中意向转移、功能转移和责任转移的机制。⑱ 黄福寿提出了机器拟人、机器治人、人机共生三种人机关系的可能性并指出了人机关系的不确定性所在。⑲

---

⑨ 陈凡,程海东.人工智能的马克思主义审视[J].思想理论教育,2017(11):17-22.

⑩ 王晓阳.人工智能能否超越人类智能[J].自然辩证法研究,2015(7):104-110.

⑪ 黄兆明.人工智能的哲学思考[J].中学政治教学参考,2014(15):84-85.

⑫ 孙伟平,戴益斌.关于人工智能主体地位的哲学思考[J].社会科学战线,2018(7):16-22.

⑬ 蔡曙山,薛小迪.人工智能与人类智能:从认知科学五个层级的理论看人机大战[J].北京大学学报(哲学社会科学版),2016(4):145-154.

⑭ 叶妮,王宏波."乌托邦"与"实践性"-理解人工智能时代的物我关系[J].科学技术哲学研究,2017(6):113-119.

⑮ 潘坤.人工智能危机背景下的实践存在论重构趋势设想[J].云南社会科学,2018(4):21-24.

⑯ 孙伟平.关于人工智能的价值反思[J].哲学研究,2017(10):120-126.

⑰ 段伟文.人工智能时代的价值审度与伦理调适[J].中国人民大学学报,2017(6):98-108.

⑱ 于雪.人机关系的机体哲学探析[D].大连理工大学博士论文,2017.

⑲ 黄福寿.从"人是机器"到"机器是人"[J].团结,2017(6):37-41.

三是人工智能从业者的伦理教育问题。人工智能从业者作为人工智能的始作俑者,对好的或坏的人工智能的产生有重要的影响,因此多篇文献强调了对人工智能从业人员的伦理道德管理。王银春强调,科技人员应遵守科技伦理与职业道德,履行“通告和预防义务”,将其研究项目向有关当局或媒体进行通报,应加强对其进行常态化道德教育,培养道德自觉意识,一旦发现不当科研活动,应停止其研究进程。[20] 闫坤如也提出要规约智能机器设计者和使用者的行为,提倡人工智能从业人员的伦理道德,要研究人工智能的设计者、使用者、监督者、维护者及人工智能企业的伦理责任。[21]

2. 人工智能伦理的具体问题研究

一是侵犯人类隐私的伦理问题。魏屹东指出,当人工智能能够通过人脑思维的运行数据来探知人类的所思所想时,人类的思维隐私将不复存在,从而使人权得不到充分保障,最终会导致人人自危,缺乏信任。[22]

二是算法偏见问题。当人工智能大量应用,其算法是否能公正处理的问题,尤其深度学习后是否会进一步加强可能的歧视也是一个重要的伦理问题。周程提出有必要在原始的算法设计中,尽可能消除设计者的主观偏见,减少不公正的价值观嵌入,让人工智能成为社会公正的桥梁和尺度。[23]

三是责任伦理问题。邬桑在论证了人工智能具有道德自主性后,证实了其可以承担联合义务责任,强调了设计者、人工智能本体、使用者、非人工智能使用者都有承担一些义务的责任。[24]

四是战争伦理问题。人工智能是否可以应用于杀伤性武器一直是备受争议的话题,大部分专家对此都持反对态度,不过也有学者指出人工智能可以推动战争向着更有伦理关怀的方向发展。徐英瑾从人工智能本身就是为了“减少附带伤害“的角度论证了人工智能可以增加未来战争的伦理关怀。[25]

五是政治伦理问题。张爱军从乐观单向度、悲观单向度、平等双向度三个方向讨论人工智能的政治伦理话题,前两个向度都只是把人工智能作为工具来建构,而在平等双向度中,人工智能具有了成为政治人的可能性与现实性,进而要求人类要主动迎接人工智能的政治挑战,共同参与政治,共建新型的政治关系和政治伦理。[26]

(三) 关于人工智能影响社会分工的研究

劳动分工是马克思政治经济学的基本范畴,其不仅仅有经济学层面的意义,在社会和道德层面也有重要的价值,对智能时代社会分工的讨论也涉及一定的哲学意义。马克思指出“任何新的生产力,只要它不是迄今已知的生产力单纯的量的扩大(例如开垦土地),都会引起分工的进一步发展”。[27] 人工智能作为新的先进生产力的代表,必然影响分工形态,并最终对生产关系产生巨大影响。此方向的研究主要包括人工智能是否会导致社会分工调整、人工智能时代的“劳动价值”、人工智能时代的资本规制三个问题。

1. 人工智能是否会导致社会分工调整

一是在人工智能时代,劳动分工将发生变革。韩海雯提出人工智能将使社会分工从低级阶

---

[20] 王银春.人工智能的道德判断及其伦理建议[J].南京师范大学学报(社会科学版),2018(4):29-36.

[21] 闫坤如,马少卿.人工智能伦理问题及其规约之径[J].东北大学学报(社会科学版),2018(4):331-336.

[22] 魏屹东.人工智能发展对社会与伦理的影响[J].洛阳师范学院学报,2016(12):1-2.

[23] 周程,和鸿鹏.人工智能带来的伦理与社会挑战[J].人民论坛,2018(2):26-28.

[24] 邬桑.人工智能的自主性与责任[J].哲学分析,2018(4):125-134.

[25] 徐英瑾.人工智能将使未来战争更具伦理关怀——对马斯克先生的回应[J].探索与争鸣,2017(10):66-71.

[26] 张爱军,秦小琪.人工智能与政治伦理[J].自然辩证法研究,2018(4):47-52.

[27] 马克思,恩格斯.马克思恩格斯选集(第一卷)[M].中央编译局,译.北京:人民出版社,2009:520.

段(共同劳动产量提升)向高级阶段(共同劳动质量提升)转变,推动劳动工具从机器向智能机器发展,使劳动者回归劳动中心位置,劳动者不再处于被分配为某个或某组机器之附庸的被动地位,而是主动地根据需要组织和取用机器劳动或智能机器劳动,劳动者将实现自觉自主地生产。而社会生产的自主也会转变为面向需求的灵活、快速而精准地生产,使社会生产力直接服务于个人需求。[28] 张笑扬指出智能时代社会劳动呈现了三大特征:推动人的无机身体的延伸、加剧行业的普遍分化、重塑社会关系的总和,需从把握智能时代全球产业分工话语权和技术的自主权出发,从技术手段、社会关系、制度安排等方面探索分工合理化的路径。[29]

二是在人工智能时代,工作的内涵也会发生改变。人工智能时代的工作不仅将定位于有用,有意思也会成为工作的一个重要内涵。肖峰提出人工智能时代,人类的工作可能转变为"软工作",即消除了工作与休闲二元分割而同时并存的工作,它不以直接创造经济效益为主而以产生社会效益为主,它不是基于强制性的谋生需求,而以个体的兴趣爱好出发施展人的才华,最终将重塑工作哲学。[30]

2. 人工智能时代的"劳动价值"

面对人工智能的挑战,"劳动价值"也将发生一定的变化。何玉长分析了智能劳动价值,其体现在生产力诸要素中,在智能劳动中的决定因素是智能劳动者,但物质要素作用越来越大,甚至智能化工具代替了人力,需确定智能劳动与简单劳动的换算比例,在智能经济的视野下创新智能劳动的劳动价值论。[31] 王永章论证了人工智能不会推翻马克思的劳动价值论,因为人工智能不创造"以生命交换生命"的人与人的社会关系,智能机器人的"劳动"只创造使用价值,而不创造价值和剩余价值。[32]

3. 人工智能时代的资本规制

在人工智能时代,政府、企业和个人都需要充分重视资本逻辑在人工智能发展中的积极作用与消极作用。董志芯提出资本增值的逻辑是人工智能发展的深层动力,资本与科技的联姻推动了人工智能的快速发展。但需注意资本权利的过度介入造成人工智能的扭曲,尤其避免在军事和经济领域的错误使用,使掌握巨额资本的人成为利用人工智能去取得权威、话语权的掌舵人,最终加剧社会的不平等,引起国际局势的动荡不安。因此其倡导加强资本规制,力争使人工智能的成果由人民共享,为人的自由、解放和全面发展服务。[33]

(四) 关于人工智能是否会推动人的解放的研究

关于人工智能是否会推动人的解放的研究主要包括两个方向:一是人工智能是否会加速促进人的解放和人的自由全面发展,二是或者人工智能是否会导致无产阶级成为"无用阶级"。

1. 人工智能是否会加速促进人的解放

人的解放是马克思关于共产主义的重要呈现形式,在《共产党宣言》中马克思指出"每一个

---

[28] 韩海雯.人工智能产业建设与供给侧结构性改革-马克思分工理论视角[J].华南师范大学学报(社会科学版),2016(6):132-138.

[29] 张笑扬.论智能化时代马克思主义分工论的价值实践[J].理论导刊,2017(9):51-54.

[30] 肖峰.人工智能时代"工作"含义的哲学探析——兼论"软工作"的意义与"工作哲学"的兴起[J].中国人民大学报,2018(5):122-129.

[31] 何玉长,宗素娟.人工智能、智能经济与智能劳动价值——马克思劳动价值论的思考[J].毛泽东邓小平理论研究,2017(10):36-44.

[32] 王永章.马克思劳动价值在人工智能时代的指导意义[J].北方论丛,2018(1):114-118.

[33] 董志芯,杨俊.人工智能发展的资本逻辑及其规制——兼评《人类简史》与《未来简史》[J].经济学家,2018(8):20-26.

人的自由发展是一切人的自由发展的条件”。[34] 而人工智能作为先进生产力,如何推动人的解放是学界关注的问题。鲁品越提出智能时代本质上是物质世界人化的过程,通过以智能网络为基础,实现社会生产力的协调化、生态化、共享化、常态化创新和全球性开放,最终实现人类社会生产力的巨大飞跃。[35] 余乃忠认为,人工智能时代可以推动“人的异化”朝着积极的方向前进,并最终导致曾经作为异己、不可制服和无限威力的自然力量与人类最终形成一次重大和解,最终这次和解也带来人与人的关系的根本性和解。而社会关系的和解也将促进人的解放。[36] 何云峰提出,人工智能的发展虽然挑战人的劳动权利,打击了人类的自我优越性和自信心,但也为实现人的劳动解放创造了条件,因为人工智能的应用使劳动越来越复归到“自由的生命表现”上来,使劳动越来越充满快乐性,人成为越来越自由的劳动者,而威胁人类生命的因素也会因为人工智能的大量运用而不断降低。人工智能大量替代人类劳动将会促进人的全面发展。[37]

2. 人工智能是否导致无产阶级成为“无用阶级”

针对人的解放的讨论中,尤瓦尔·赫拉利在《未来简史》一书中提出的“无产阶级会成为无用阶级”[38]的观点引起了社会的广泛关注。蒋红群指出人工智能之所以排斥无产阶级,源于资本与劳动的对立,但从《共产党宣言》观之,无法改变“两个不可避免”的历史发展趋势,最终变为“无用阶级”的不会是无产阶级,而是资产阶级。[39]

二、对研究现状的述评

从前述的关于人工智能哲学的主题材料看,2014 年至 2016 年的文献主要讨论人工智能何以可能的问题,并且文献的数量仅为个位数。在 2017 年(含)以后,随着阿尔法狗与李世石的人机大战事件的发生,人工智能以颠覆性技术进入公众和学界视野,关于人工智能哲学的研究迎来了爆发性增长,对其主体性挑战、伦理问题、劳动分工、人的解放四个方面的研究也日益深入,取得了很多成果,回答了很多社会和学界关注的问题。但通过对文献的分析,笔者认为还有不少问题值得进一步研究和思考。

(一) 强人工智能视野下的人的主体性哲学研究

古希腊哲学家普罗泰戈提出“人是万物的尺度”,这句话定义了人作为主体性的依据,人类的自主性的自然合法性不容置疑。康德认为,自主不仅仅意味着自由意志,同时也意味着自由意志服从于自己制定的道德法则。[40] 由此看来,自主性包含两个方面的内容:自由意志和自主行动。“自主体”也经常与理性联系,“没有理性就没有自主体”。[41] 马克思指出“人的本质并不是单个人所固有的抽象物,在其现实性上,它是一切社会关系的总和”,[42]智能体之间联结将可能成为智能体自我意识产生的源泉,这也是智能体成为自主体的一个讨论话题。

---

[34] 马克思,恩格斯.共产党宣言[M].北京:人民出版社.2018:51.

[35] 鲁品越.智能时代与马克思生产力理论[J].思想理论教育,2017(11):10-16.

[36] 余乃忠.积极的“异化”人工智能时代的“人的本质力量”[J].南京社会科学,2018(5):53-57.

[37] 何云峰.挑战与机遇:人工智能对劳动的影响[J].探索与争鸣,2017(10):107-111.

[38] 尤瓦尔·赫拉利.未来简史:从智人到智神[M].林俊宏,译.北京:中信出版社,2017:293.

[39] 蒋红群.无产阶级会沦为无用阶级吗[J].马克思主义研究,2018(7):128-136.

[40] I. Kant. Groundwork of the Metaphysic of Morals [M], New York: Harper and Row, 1958.

[41] C. Cherniak. Rational Agency [M]// R.W. Wilson, F. Keil (eds.). The MIT Encyclopedia of the Cognitive Sciences. Cambridge. MA: MIT Press, 1999.

[42] 马克思,恩格斯.马克思恩格斯文集(第一卷)[M].中央编译局,译.北京:人民出版社,2009:501.

随着人工智能技术的不断更新迭代,拥有自主性的人工智能机器人极有可能不久会进入人类的视野,而人类作为唯一拥有自主性的物体的界限将被打破,此时将对人的心理产生重大挑战,就如同人类一直知道人必然会死亡,但一旦个人面临死亡,其内心的恐惧和压力会非常大,如何在前期做好强人工智能到来前的各种相关的预案,不断提升公众对强人工智能时代的理性认识,是政府和人工智能企业的任务之一。在拥有自主性的强人工智能时代,人类将处于怎样的地位,现有情况下很难预知,所有关于人的主体性哲学的研究都是基于各种可能性来开展的,现有的多种可能性似乎无法回答人类对自身主体性唯一性的期待,是否可以建构符合公众期待、哲学上合理、技术上可实现的强人工智能视野下的人的主体性哲学理论是摆在学者和技术人员面前的问题。

(二) 人工智能规范的形成机制研究

人工智能规范包含伦理规范和法律规章两个部分。在伦理规范方面:无论是人机关系、研究规范、隐私伦理、算法偏见、责任伦理、战争伦理、政治伦理等伦理问题,都涉及一个问题:人工智能伦理规范机制如何形成?以何种方式方法来处理人工智能伦理失范问题?虽然业界曾试图建立了阿西洛玛人工智能原则、阿西莫夫三原则等人工智能发展规范,学者们也通过发表各种研究报告提出了关于伦理问题的认识,但很少涉及伦理规范的形成机制问题,以及以何种方式方法来处理人工智能伦理失范问题。在法律规章方面:随着人工智能的广泛应用,其带来的社会风险也加速释放,如自动驾驶、智能安防、自然语言识别、人脸识别等都会引起一定的社会问题,例如当自动驾驶导致行人伤亡时,究竟应如何定罪,主体责任由谁承担,立法的原则是选择保护技术创新为主还是保护公众安全为主。国家的立法机构和法律专家可在伦理和主体性哲学研究的基础上,适时研究出台《人工智能法》的相关法律条款,以良法促进技术发展,减轻公众对人工智能安全的忧虑,最终形成人与人工智能良性互动的局面。

(三) 人机情感伦理的研究

人工智能的产生离不开技术的进步,技术的进步又与使用者的情感相伴而生,随着技术的进步,人与机器的情感可能会强化。人与强人工智能的情感问题首先呈现在各科幻电影和小说中,但随着人工智能产品的普及,人机之间的情感问题无法避免成为每个人都要面临的问题,如各种智能助手、机器人都会与人建立密切的情感沟通关系,人机情感问题背后的伦理问题值得关注。围绕人工智能意识与情感的发展,孙振杰提出将导致人工智能发展遭遇"五化"(退化、进化、蜕化、异化、黑化)问题,并指出需要增加对人工智能自身产生情感的控制。[43] 如何建立完整的人机情感规范,在强人工智能时代来临之前使公众知晓如何与其相处,把握"人机情感""人机共生"尺度、遵从人机情感伦理是一个有价值的课题。

(四) 人工智能影响劳动分工带来的教育体系变革研究

人工智能可能带来社会分工的重大调整已经成为广泛共识,由人工智能导致部分行业的失业可能无法避免,这些失业人群可能会像机器大生产导致的路德主义者一样对人工智能产生极大的愤怒,从而导致一系列社会问题。要解决这一问题,可以从以下三个方面开展研究:一是可提前开展学校教育体系的变革研究,改善学科专业布局,对未来可能失业的行业减少员额,提前改变专业类型,使专业教育更好地适应人工智能引起的变革。二是在大中小学教育中适时推出人工智能的教育课程,在学校教育中尽早融入人工智能方法、逻辑、技术等教育内容,让同学们从小适应人工智能带来的变革,从心理上增强适应性。三是针对失业人群的再就业教育也应提前做好计划,避免失业造成的社会问题扩大化。

---

[43] 孙振杰.关于人工智能发展的几点哲学思考[J].齐鲁学刊,2017(1):77-81.

(五) 人工智能赋能下的人的解放的前景研究

人工智能在促进生产力的极大进步后,人类将面临越来越多的自由时间,此时如何适应自由时间增多的闲适生活,如何设计丰富多彩、有价值的活动来让人保持自我的价值感是一个值得研究的问题。同时,应积极开展人工智能时代的社会福利政策研究,避免善政养懒汉。

(六) 人工智能哲学与人工智能技术实践的协同发展研究

实践与理论是辩证统一的,当人工智能的技术发展超越了人工智能哲学的范畴时,犹如脱缰的野马引起恐慌,而当人工智能哲学过快地超越了人工智能技术现实时,就会出现很多不切实际的空想理论。过冷或过热的技术和哲学都不利于社会健康发展,寻找人工智能哲学与技术实践的协同发展理论是摆在学者面前的一个课题。

(七) 如何构建完整的人工智能哲学体系

构建完整的人工智能哲学体系的意义在于预测人工智能的未来,规约人工智能的发展,完善人工智能时代的思想基础。目前人工智能哲学的各个分支已经形成了四个板块:人工智能主体性研究、人工智能伦理问题研究、人工智能影响社会分工研究、人工智能推动人的解放研究,如何找到四个板块间的逻辑关系是构建完整的人工智能哲学体系的前提。随后,如何持续发现人工智能哲学研究的新热点,如何建立一个基础理论完备、研究深入、框架清晰的人工智能哲学体系是值得深入研究的问题。

A Literature Review of Philosophy of Artificial Intelligence

XU Yong[1,2], *HUANG Fushou*[1]

(1. School of Marxism, Shanghai Normal University, Shanghai 200234, China;
2. Student Affairs Committee, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: The philosophy of artificial intelligence(AI), as a hybrid of AI and philosophy, is a branch of philosophy that systematically studies the rules of the nature, evolution, and operation of AI technology. It covers the subjectivity of human intelligence, code of ethics, division of labor, and emancipation of manpower. This paper outlines the fruits of the philosophy of AI over the last five years, mainly from the above fourfold aspects. After the analysis, some existing problems are discovered, such as human subjectivity in a strong AI era, the formation of ethics code, issues of emotional ethics, AI legislation, education system reform, prospects for emancipation of manpower, the synergetic development of technology and philosophy, and the construction of AI philosophy system, which require more in-depth insights and studies.

Key words: artificial intelligence(AI); philosophy; issues of ethics; division of labor; subjectivity; emancipation of manpower

(责任编辑:黄谷香)