

人工智能与技术伦理思潮:反思与重塑

刘 丽*

一、引言

人工智能(Artificial Intelligence, AI)被定义为研究开发用于模拟、延伸和扩展人类思维和行为的理论、方法、技术及应用的一门技术科学。它企图探索智能的实质,并生产能以人类智能相似的方式做出反应的智能机器或系统。人工智能作为引领第四代科技革命和产业变革的战略性技术,已成为全球科技竞争的焦点领域,世界主要国家纷纷从战略布局、资金投入、技术研发、领域应用等方面入手提升人工智能全球竞争能力。随着人工智能引领新一轮科技革命的飞速发展,经济、社会、生活和产业迈向数字化、智能化,重构生产、分配、交换、消费等经济活动各环节,催生新技术、新产品、新产业。例如,人工智能进入家庭、教育、医疗、养老、环境保护、城市运行、司法服务等领域,带来了智慧家居、智慧交通、智慧医疗、智慧工厂、智慧农业、智慧金融、智能制造、智慧政务、智慧城市等新气象。

人工智能技术具有双重效应,既带来了巨大的经济利益和社会变革,也引发了人类认知受困和权益受损等复杂伦理问题。例如,在搜索引擎、推荐系统、计算广告、人脸识别等技术领域,带来了信息茧房、隐私泄露等技术异化^①;在文化领域的写作、音乐、修改论文等,加深了版权的法律难题;在社会治理中,如金

* 刘丽,复旦大学发展研究院助理研究员;复旦大学社会科学高等研究院博士后。

① 曾建平、黄以胜:《信息技术问题的道德治理》,载《华东师范大学学报》(哲学社会科学版)2022年第2期。

融服务中的信用评分,执法领域中的犯罪风险评估,以及就业领域中的简历评估、工作任务分配与评估与合同关系管理等,面临由数据集不完善和算法黑箱带来的身份、种族、性别歧视等问题。^①在犯罪评估、信用贷款、雇佣评估等方面,推荐算法决策的一次小失误或者因歧视而产生的连锁效应,可能影响一群人或者一个种族的利益,影响社会公平。

人工智能的技术热潮和人工智能伦理的知识浪潮^②,引发了技术伦理思潮的回归。学者、政府是潮头,开展面向智能社会的理论探索和法律制定;民众和企业是潮流的主体,利益和公正意识被唤醒。在学术领域,人工智能的伦理原则不断更新、拓展,包括可解释性、正义性与公平性、责任性、隐私性等。以技术和技术善治作为竞争优势的来源,各个国家都在积极推动技术伦理治理。例如,2017年,国务院印发的《新一代人工智能发展规划》明确提出要“加强人工智能相关法律、伦理和社会问题研究,建立保障人工智能健康发展的法律法规和伦理道德框架”。2019年,我国新一代人工智能发展规划推进办公室专门成立了新一代人工智能治理专业委员会,建立健全人工智能治理方面的政策体系、法律法规、制度体系、伦理道德的研究和实践工作。近日,我国印发《关于加强科技伦理治理的意见》,进一步深化人工智能的伦理教育和治理。在行业中,企业将技术伦理作为一项重要的服务内容(伦理即服务),开拓了新的业务部门和社会服务组织。人工智能伦理服务是传播理想和信念的重要载体,也是推动治理向系统性、务实性方向迈进的重要尝试。

人工智能伦理的产生和蔓延,既是思想和价值观念问题,更是利益诉求的表达与实现问题。技术带给人以解放,也必然带给人以奴役,技术与解放、奴役具有“同存共生”性。“有无相生,难易相成”,解放与奴役是技术这一矛盾体的两个方面。^③人工智能技术能提高效率,也能对人进行窥视。技术奴役是智能情

① 谭九生、杨建武:《人工智能技术的伦理风险及其协同治理》,载《中国行政管理》2019年第10期。

② 刘鸿宇、苗芳艳、彭拾、夏思宇、王珏:《人工智能伦理研究的知识图谱分析》,载《情报杂志》2019年第7期。黄崑、徐晓婷、黎安润泽、徐峰:《近5年图情SSCI期刊人工智能伦理研究文献分析与启示》,载《现代情报》2021年第6期。汪玉兰、邓国民:《国际人工智能伦理研究现状与发展趋势》,载《图书馆》2020年第1期。于雪、段伟文:《人工智能的伦理建构》,载《理论探索》2019年第6期。

③ 刁生富、冯桂锋:《消解技术奴役:探寻大数据时代人的解放之路》,载《学术研究》2018年第12期。

境下的主体性危机,也是社会文化困境。技术奴役的根源并不在于技术本身,而在于技术的制度规范不成熟和资本利益的驱动,造成主体的精神沉沦、主体个性及社会性缺失并为技术所奴役。技术奴役也源于掌握数据和技术的数字资本权力,通过技术“优势”加剧知识的碎片化而影响社会舆论与认知;或由于技术预测的局限性歪曲事实,遮蔽数据本来的价值意涵。

中国在向技术社会转型,作为数字产品的最大消费市场地,也面临着社会心理、思想、观念、意识的交锋和耦合。中国的人工智能情境下的技术伦理不断完善,但面临着三个方面的挑战。第一,碎片化的理论,忽视了社会心理预期和发展情境的复杂性,导致了技术与社会之间的隔离与鸿沟。第二,理论与实践和社会现实之间的联系松散,即伦理与经济、政治、社会、文化诸文明要素之间缺乏健康互动的生态关系,难以推动思想整合与利益整合,化解智能化时代的社会矛盾。第三,研究格局需要扩大,使学术思想与社会心理融合,获得广泛的流行性和影响力。人工智能伦理目标与实践之间存在诸多不确定性,即坚守“伦理信念”并不能保障善的结果。例如,欧洲的研究显示,伦理审查将制约技术创新的步伐,造成人工智能创新的系统性困境。先验性的人工智能技术伦理审查过程,要保持一些自由探索的空间,也要给伦理评价留有更为充裕的时间、空间省思。人工智能伦理的发展,不仅描述和反映技术变迁,也培育了价值认同的文化土壤,技术与人文共生,加速人工智能技术的综合性创新。人工智能技术的发展,受到法律制度、文化传统、政治意志和伦理道德等多方面因素(机制)的影响,成为智能时代多样化社会生态和技术文化的构成部分。当前,社会主要矛盾转变为人民日益增长的美好生活需要同不平衡不充分的发展之间的矛盾。人工智能技术不完善,阻碍了群体追求可持续发展和美好生活的利益诉求的实现。技术伦理从对未来不确定性的忧虑,转向了对美好生活的探索,是引领性和前瞻性的哲学社会科学理论议题和现实要求。智能化时代的“人-技术-社会”协同共生的新生态,推动了伦理意识的复归和时代议题的重思。

伦理是意识形态,也是一种文化形态和文明形态,伦理规律含有丰富的文化信息。伦理的集体记忆和集体演练,是民族的自在形态,是民族精神的生活化和艺术化的表达方式。我国确定了伦理先行的引导性机制,即用伦理规范引导人工智能的技术和产业变革,具有重塑和完善文化和价值系统的重要作用。

第一,通过人工智能伦理观念进行隐蔽性的社会主义意识形态渗透,如在科技领域推动共同富裕、和谐共生的精神文明建设;第二,采用“因势而谋,应势而动,顺势而为”的策略,推动人工智能技术适应不确定性环境的挑战;第三,人工智能的伦理能力和自治能力^①,有助于塑造道路自信、理论自信、制度自信和文化自信。“实践-利益-制度”的辩证互动,推动了智能化时代技术伦理的理论概括和理论深化,形成包容性、开放性的制度文化,增强不同群体的思想凝聚力,推动技术与市场经济、商业资本、社会权力相互交织和发展,建构起中华民族命运共同体的新维度。因此,人工智能伦理折射出人们追求美好数字生活的强烈共鸣,追求个人价值和社会全面发展,既符合人类道德与价值观念,也适应多样化经济、社会和文化的需要。

本文的理论贡献是,人工智能伦理结构化,从智能化时代“人、技术与社会”协同发展的角度、深度和广度延展了技术的疆域,总结和提炼了人工智能伦理的发展和影响,致力于重塑人类的数字化生存空间,弥补和修复人工智能面临的多重鸿沟。它超越了伦理的人际维度(成员间的伦理共识),迈向了伦理的时空维度(生态化发展)^②,构建和发展以人、人工智能和社会协同发展对象的新场景^③,融合技术文明与社会文明,指向了科学技术与美好生活的时代命题。^④建立“人-技术-社会”之间的关联与互动综合性分析框架,以“更好的人、更好的技术、更好的生活”目标,推动人、技术和社会的和谐共生,从而形成人文性、适应性和综合性的技术创新。

二、人工智能的技术发展

人工智能技术的发展,呈现了自动化、智能化、生态化等三个趋势,致力于

① 曲蓉:《人工智能责任伦理的可能性条件》,载《自然辩证法研究》2021年第11期。

② 樊浩:《当代伦理精神的生态合理性》,载《中国社会科学》2001年第1期。

③ 陈国青、任明、卫强、郭迅华、易成:《数智赋能:信息系统研究的新跃迁》,载《管理世界》2022年第1期。

④ Luciano Floridi, Josh COWLS, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke and Effy Vayena, “AI 4 People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations”, *Minds and Machines*, Vol.28(2018), pp.689—707.

减轻人非必要的体力、脑力劳动,增强人本身的体能、智能和外延能力,使得人类以更有效的方式、更大的广度和深度从事创造性活动。

第一,人工智能技术的自动化。基于人工智能技术的自动化,与内燃机、蒸汽机等技术带来规模化生产的效率革命一致,能够替代人参与重复性、枯燥性的机械工作,并且持续工作时间更久、误差率更低,它们也被称为“人工智能员工”。麦肯锡预测,到2030年,随着人工智能等技术的进步,自动化会带来“结构性失业”,多达3.75亿劳动者可能需要更换职业类型。自动化对员工的替代问题,从工业革命时代就被提起。但是,旧职业被淘汰的同时,总会有新职业出现,劳动力也从农业向工业和服务业不断迁移。互联网的繁荣创造了大量的新职位,人工智能新技术会创造更多新的岗位,如数据标注员、算法工程师、大数据分析师等。

第二,人工智能技术的智能化。智能是人类生命的本质,涉及诸如意识(consciousness)、自我(self)、思维(mind)等各种问题,而人类的智能包括语言智能、音乐智能、逻辑-数学智能、空间智能、身体动觉智能、人格智能和自省智能。人工智能通过对机器进行智能训练,即从经验和数据中学习,让它进行智能行动,获得依托计算机系统和模型获得的智能。人类把怎样表示知识、怎样获得知识并使用知识(思维过程、学习能力和知识存储)以逻辑的形式告诉计算机,计算机能够不断学习、融入、发现新的知识。其中,机器学习(machine learning)^①、深度学习^②和强化学习(reinforcement learning)^③是当前推动智能化的重要技术资源。但是,有多少“人工”,才有多少“智能”。人工智能的发展,需要大量的人力参与数据标注,教会机器获得和拥有智能。

第三,人工智能技术的生态化。人工智能技术的发展,离不开大数据技术、芯片技术等技术设施构成的技术生态系统,多种技术的共生状态是人工智能技术发展的基石。虽然信息、数据跨越了国界的限制,但是人工智能技术呈现出不均衡的世界地貌,欧美发达国家处于先导性地位,新兴国家正在拓展。人工

① 机器可以从数据中学习知识,不断进行自我重新设计,提升自身的能力。

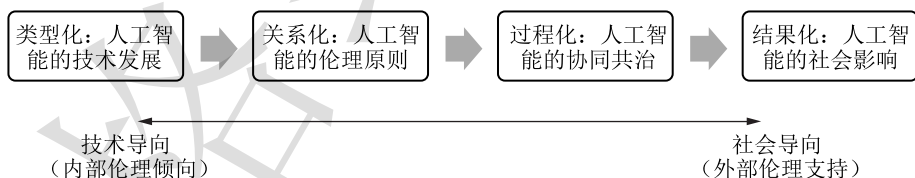
② 深度学习,思维机制的模块化组织起来的神经网络的信号传递,而模块化的互联互通形成智能。

③ 通过激励信号强化人的行动,而惩罚与之相对,用于优化智能。

智能技术的系统化、生态化发展机制,复刻了不平等的全球技术生态。同时,人工智能技术与经济、社会、制度等方面协同发展,而伦理和法律制度促使人工智能技术向善,又能积极地作用于经济和社会,加速技术与整个社会生态系统的协同演化。当前,人工智能技术攻关,制度、文化和伦理因素从一种潜在的、隐身性的存在,变为人工智能系统的外显性的部分,影响人工智能开发和应用。各国的人工智能伦理及其治理不断强化技术、法律、伦理、社会等多维度支撑,即成文的国家立法和监管的“他律”、企业的伦理道德的“自律”,与智能设计机器的“物律”相结合,形成结构松散的创新网络,规范人工智能的发展。各国不断出台人工智能伦理标准,研判和防范人工智能发展的潜在风险,维护人民利益、国家安全和行业发展,确保人工智能安全、可靠、可控,构建一个良好的人工智能发展生态。

三、人工智能伦理的多元视角

人工智能技术是一种有责任意义的工具。技术向善,需要构建新的人工智能伦理^①,将伦理实践化。数字技术和人工智能技术为基础的物质变革带来了一种技术化、智能化的全新社会存在,撬动了人工智能伦理中丰富多彩的思想素材和意识形态争辩,伦理实践也呈现不同的发展模式。人工智能伦理有四种视角,呈现从技术导向到社会导向的演变过程(见图1),分别涉及技术尺度的伦理评价(内在伦理倾向)和行动秩序的伦理引导(外部伦理支持)两个方面。



资料来源:作者自制。

图1 人工智能伦理的多元视角

^① 孙田琳子:《论技术向善何以可能——人工智能教育伦理的逻辑起点》,载《高教探索》2021年第5期。

第一,类型化,即对弱人工智能和强人工智能(或称为超级人工智能)进行技术伦理评价。弱人工智能,指设计的机器能够,或看起来能够智能地行动,也称应用性人工智能。强人工智能,也称通用性人工智能,是具有接近于人的决策能力的一种人工智能。在强人工智能技术生命的早期,科学家要前瞻性地分析和识别技术的伦理问题,推动技术的预先伦理治理,各国都主动规避对强人工智能的开发和运用,为负责任创新研究提供了有效的伦理进路。^①强人工智能或者超级人工智能是对技术超能力的一种想象,显示出面对人类能力极限与技术演化之间冲突的担忧。也有学者提出,强人工智能是人类繁殖的一种新方式,它本质上是人类进化的一个新阶段,如同年轻一代超越老一代,不应该用个体的焦虑来替代社会发展规律。^②

第二,关系化,即探索人类与人工智能(机器)的多重关系,在人工智能中嵌入人类伦理。具体而言,包括以下几点:(1)人机之间是镜像关系,即人工智能能像人一样行动和思维,是具有感知方式、神经机制、“共情”共性能力的智能体。(2)人机融合,通过脑芯片植入对神经的改造,人通过机器提升能力、获得便利,机器也向人学习。(3)机器和人之间的神经元结构的“同构性”^③,有自主意识并且可确证其主体资格的“智能”,涉及机器人的法律身份和公民资格等社会治理的结构性难题。机器人作为道德能动者,其道德性决策能力影响了它们与人相处、互动的方式。例如,情侣机器人、医疗护理机器人等服务型机器人,可能会导致人类对其产生情感和依赖,从而影响他们与其他人类的正常交流,同时也引发与传统婚姻伦理相冲突等一系列问题^④;社交机器人对人类同理心的“操控性”和其“欺骗性”^⑤,会涉及机器伦理、设计伦理、信息技术伦理,尤其是拟生的“人”的道德规范。关系化的视角,推动伦理作为提前制定的编码,限

① 顾世春:《荷兰预判性技术伦理思潮研究》,载《大连理工大学学报》(社会科学版)2018年第4期。

② 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》,载《人民论坛·学术前沿》2016年第7期。

③ 杨庆峰:《从人工智能难题反思 AI 伦理原则》,载《哲学分析》2020年第11期,第137—150页。

④ 段伟文:《机器人伦理的进路及其内涵》,载《科学与社会》2015年第2期。

⑤ 王亮:《社交机器人“单向度情感”伦理风险问题刍议》,载《自然辩证法研究》2020年第1期。

制人工智能的不合理行为,却忽视了发挥伦理对人和技术的引导和激励作用。

第三,过程化,即伦理介入人工智能技术,作为一个新的技术层级或者单元,嵌入人工智能系统,渗透到“研发→引入→扩张→权威→衰落”的技术生命的全部发展历程。伦理介入推动研发过程及时调整技术设计或采取其他应对措施,既可以减少或者避免资金浪费,对不良技术进行预先性治理,避免技术的伦理风险扩散到应用中和社会中;也以参与式伦理实践方法论,推动多元化利益相关者^①共同参与伦理治理。^②但是,它也面临着技术审查的范围问题,如何减少审查与创新之间的张力,即避免以科学技术审查为基础的伦理,变成一种压迫性和限制性的力量。因此,尽管人工智能伦理在不断发展,伦理介入不同的技术过程之间缺乏系统性的关联,人工智能时代的技术认知和社会规范与秩序的升级,依然面临着诸多的难题。

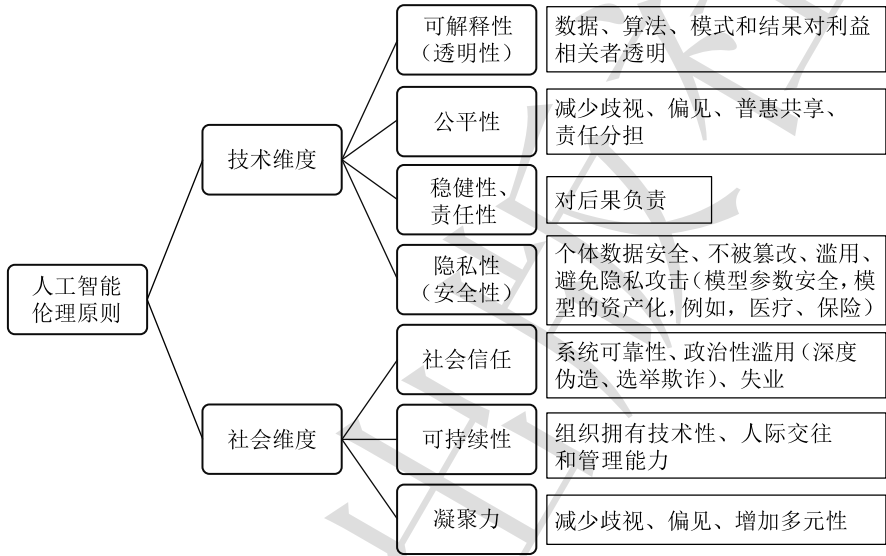
第四,后果化,即对人工智能的社会影响进行评估、判断和实践性反思,推动人工智能技术担当有价值的社会角色,最大限度地规避或减少技术对人类的负面影响,促进社会繁荣。它致力于提高对技术性力量的风险认知与应对能力,提升人工智能伦理的认知、理论、原则和政策,提前防范技术给人类带来的脆弱性和风险性。但是,从普遍性(抽象性)的人工智能伦理原则,到适应性、区域性的人工智能伦理实践,依然存在纷繁复杂的人工智能伦理问题,尤其是因为产业和行业差异性,目前缺乏对原则与实践的关系的系统性总结。虽然企业、研究机构和公共组织高度关注人工智能伦理原则和指南,评估人工智能系统的设计、行动、监视和评估风险等,但不同个体、组织、国家和国际社会关于人工智能的各种思想与流派并不是处在同一层面上,“选择性注意”使得人工智能伦理知识处于割裂、分散和碎片化状态。人工智能伦理原则包括可解释性、公平性、责任性、隐私性、社会信任、可持续性、凝聚力等^③,构成了人工智能技术本身(技术维度)和人工智能的社会影响(社会维度)(图2)两个维度的网络状结

① 包括政府、平台企业、算法工程师、用户。

② 张春美:《负责任研究创新的生成逻辑及其实践》,载《上海师范大学学报》(哲学社会科学版)2021年第3期。

③ 李伦、孙保学:《给人工智能一颗“良芯(良心)”——人工智能伦理研究的四个维度》,载《教学与研究》2018年第8期。

构。图 2 中罗列的各项人工智能伦理原则指向技术之善和技术的社会责任,为软件工程师、企业和政府等部门的决策奠定基础,但是它们忽视了伦理认知与实践的鸿沟,如技术奴役和权力关系,以及人工智能伦理原则之间的关联性和内在结构性。^①



资料来源:Anna Jobin, Marcello Ienca and Effy Vayena, “The Global Landscape of AI Ethics Guidelines”, *Nature Machine Intelligence*, Vol.1(2019), pp.389—399。

图 2 人工智能的伦理原则框架

四、人工智能伦理的结构化

人工智能技术代表了物质文明新阶段的一种实践工具,其扩展到生产、生活的各个方面,成为一种新的公共基础设施。同时,人工智能技术也代表一种新的权威,它的发展不仅复刻了不平等的全球技术生态,也使现代社会面临被不平等的权力严重撕裂的危机。面对不平等的、生态化演进的技术权力,如何拓展人的主体性,促进社会公平和谐,成为当前重要的挑战。本文借用英国社

^① 杨庆峰:《从人工智能难题反思 AI 伦理原则》,载《哲学分析》2020 年第 2 期。徐玖玖:《人工智能的道德性何以实现?——基于原则导向治理的法治进路》,载《现代法学》2021 年第 3 期。

会学家吉登斯的结构化理论,推动人工智能伦理的结构化^①,强调以人和社会发展为中心的伦理运行逻辑。结构具有二重性,即结构影响行为,行为重塑和再生产结构。结构包括资源和制度,伦理是实践,被制度和资源重塑;同时,伦理也是一种资源和制度,解决智能化时代的社会问题,完善社会治理和实践的过程,从伦理之知到伦理之行,推动更大框架内的知善、行善。人工智能伦理的结构化,是指伦理性的行动被赋予了动态的倾向,反对固定化,随着认知、实践和外条件而不断变化与创造。它强调在技术与社会的互动中,推动人工智能伦理秩序的升级,即伦理本身不仅平衡现实世界的利益关系,而且指向个体存在的生命意向认同,以及情感式的内在价值认可。^②

人工智能伦理的结构化,既推动技术与伦理(社会制度)的结合,发挥伦理鼓励、引导和塑造个体行为的作用,也规制伦理实践中技术权力的扩张。一方面,它沿用了维贝克的“道德物化”思想,把道德规范和理念嵌入技术物,对技术物进行负载伦理价值的设计,使技术物对使用者的行为产生道德引导和规范作用,即通过技术或者物将伦理行为固化为习惯,推动技术与道德之间的深度关联。它通过物的设计和使用,拓展了人承担道德责任的路径。另一方面,它反思物的能动性和技术的伦理角色与地位带来的对个体和社会的潜在控制,从文明发展的角度调节人与技术和世界的关系。程序性的定义和判断,减少了思考或导致了思考的终结,也导致了伦理的终结。伦理行动是思考和判断的结果,而不是遵循提前定义的公式的结果。盲目跟从提前定义的准则、价值、政策和习俗,不能得出正确判断。换言之,伦理决策不是决定性的判断,而是反思性的判断。^③一旦将道德固化于物体,将强化精英人群和核心国家对技术和规则的创造,使得技术和资本精英获得定义和实施“价值”的权力,导致价值家长制,限制了部分人的自由。例如,中国古代“藏礼于器”的伦理思想,部分限制了人的能动性。因此,器物和技术的伦理化是文明发展的初级形态,而技术与伦理、社会

① 张旺:《伦理结构化:算法风险治理的逻辑与路径》,载《湖湘论坛》2021年第2期。

② 陈凡、贾璐萌:《技术伦理学新思潮探析——维贝克“道德物化”思想述评》,载《科学技术哲学研究》2015年第6期,第54—59页。

③ Brent Mittelstadt, “Principles Alone Cannot Guarantee Ethical AI”, *Nature Machine Intelligence*, Vol.1 (2019), pp.501—507.

的协同共生,构建良好的技术-社会生态,实现物质文明与精神文明的协调才是伦理结构化的本质目标。

(一) 人工智能时代的技术奴役

人工智能参与社会分工,替代人类主体履行一定的社会职责,成为社会系统的重要组成部分。人工智能既拓展了个体的潜能与优势^①,也带来了技术奴役的新问题^②。技术奴役是人工智能时代的新殖民主义,涉及剥削、种族偏见、知识和文化霸权,将加剧社会排斥、西方中心的种族主义、全球财富不均和贫困。人工智能伦理的本质是把人放在核心位置,强调人工智能为人类服务而不是取代人,用技术提升人类的潜力、扩展人的体能,改善人类状况并增进人类福祉。人工智能伦理是对技术、制度和社会文化的综合性整合,高效地实现技术条件下人与社会的发展,人工智能的伦理意识拓展了对技术奴役风险的反思。

第一,以任务为中心的技术文化,带来了对技术开发的限制和绑定。技术开发者会陷入一个机械性、自动性、标准性的技术程序中,难以获得动态、灵活的自我定义的权利。以任务逻辑为中心,人辅助技术,要跟从机械化的任务程序,导致人思维的机械化,抑制了开发者主体性、创新性的发挥。任务的单一性,也使得开发者失去全面发展、开拓创新的机会,它也是抑制我国人工智能技术发展的重要原因。

第二,人工智能技术的引入与应用成为社会运转的常态,加深了人的技术化锁定。用户对人工智能技术产生便捷性的依赖,甚至形成一种成瘾性的习惯。深度的技术依赖一旦被恶意操纵,人可能缺乏反省性,增加了技术社会的风险性。虚拟现实技术有可能使得缺乏认知能力的青少年真假难辨,破坏他们正常的感知能力。人脸识别技术用于身份识别、支付等,提高了效率,也存在识别失误、欺骗攻击、数据安全等方面的隐秘性风险,并极易造成“技术利维坦”。

第三,人工智能算法推荐的新闻、娱乐、购物等,基于用户的历史数据来挖掘用户偏好,在满足人的需要、促进社会连接的同时^③,也构筑了一种固化的行

^① 顾世春:《荷兰预判性技术伦理思潮研究》,载《大连理工大学学报》(社会科学版)2018年第4期。

^② 孙伟平:《人工智能与人的“新异化”》,载《中国社会科学》2020年第12期。

^③ 邱仁宗、黄雯、翟晓梅:《大数据技术的伦理问题》,载《科学与社会》2014年第1期。

为模式,剥夺了人的复杂性和可变性^①。伴随着互联网的广泛使用,人们的各种身份和行为信息被越来越多地数字化,带有认知、预测和决策功能的人工智能算法,在利益驱使下消解人的主体性。一方面,算法推荐会干扰人的注意力,使得个体最宝贵的时间资源被不断地抢夺,侵占个体在感兴趣、需要和实践等多元状态之间的审慎选择时间与空间,挑战个体的主体性。另一方面,算法把人固化,进而简化了人与信息的关系,不同平台的倾向模式趋同化,模式化的信息机制带来了大量重复、无效和雷同的信息(信息疲劳)。

因此,人工智能使得人与技术的关系,面临着控制与反控制之间的矛盾。人工智能伦理需要反思、批判人工智能技术的片面性,冷静、合理、审慎地重思技术背景下的技术奴役、人际网络广度的拓展、人际身份的虚实交错等。

(二) 人工智能时代的数字权力扩张

人工智能作为一种革命性、颠覆性的高新技术,不仅深刻地变革了智能时代的生活方式、生产方式、交往方式和认知方式,也重塑了社会关系网络和权力关系体系。人工智能是依赖人类劳动的一种新的技术存在方式,也蕴含着人类社会的新结构和技术分工。人工智能与人的解放和社会发展具有三重内嵌悖论。第一,从工作层面看,人工智能抑制了重复性劳动,却也带来了控制、垄断、排斥。人工智能的运用,更多的是替代低技能劳动者,却没有真正解决困扰社会的深层次工作问题,尤其是更加体面、有意义、有价值的工作。

第二,从社会层面看,把伦理原则嵌入智能机器人^②,并不能确保社会公平的实现。人工智能技术,更多地被知识精英、经济精英所掌握,它所服务的范围和方式具有内存局限性。精英群体的知识积累、财富创造和社会地位,在人工智能技术时代得到巩固和强化,加剧了马太效应,阻碍了社会公平。而社会公平代表了群体的利益和想法,社会不公平是破坏社会凝聚力的重要影响因素。如果缺乏有效的伦理引导,科技与资本的结合就可能加速社会不公平,成为阻

^① 喻国明、曲慧:《“信息茧房”的误读与算法推送的必要——兼论内容分发中社会伦理困境的解决之道》,载《新疆师范大学学报》(哲学社会科学版)2020年第1期。

^② 苏令银:《当前国外机器人伦理研究综述》,载《新疆师范大学学报》(哲学社会科学版)2019年第1期。

碍个体发展的结构性障碍,这比技术性障碍更加难以克服。

第三,从制度层面看,资本掌握了数据资产和技术(处理海量数据的计算机程序),可以将其价值需求隐蔽地嵌入算法和数据之中。数据霸权和算法歧视不仅源于数据错误、数据来源不清、数据质量不足、主观偏见和人为操纵^①,也源于资本的算法权力^②。资本将价值需求嵌入算法与数据之中,进一步加深了算法危机,包括客观性丧失、片面化重构和大众批判能力下降等风险^③。这些问题在一定程度上破坏了社会凝聚力,加剧了社会分裂。人工智能技术与社会信任、可持续发展(环境保护)、社会凝聚力之间的张力,是人工智能时代西方国家最为关注的制度议题。

资本和平台将社会层面融入人与技术关系,社会公平和可持续发展将为人工智能技术的发展和运用提供更加宽阔的市场和发展空间。将人权、道德嵌入社会,把人与社会可持续发展作为衡量技术的标准,不断探寻技术之善内在使命和外在价值,是人工智能伦理思潮演变的内在动因。技术与人和社会深度关联,推动了人工智能伦理的综合性、融合性,既对人工智能的技术风险进行控制,也避免了“伦理责任”的操纵与压榨造成社会成本过度。人工智能伦理,需要具体情境下的创造、交流、互动实践,突破制度背景下科学和技术的压迫和限制。

(三) 人工智能伦理与向善发展

人工智能技术的巨大发展及其所内蕴着的深刻矛盾、价值冲突和价值困惑,不仅需要法律法规、社会观念、制度规范引导科技向善和治理现代化,更需要伦理文化更新。如果离开了道德的制约、伦理的引导,科技进步可能导致文明的毁灭。光有伦理原则是不够的^④,伦理原则易被操纵,让伦理替代法律,会带来伦理的滥用和误用,使得伦理无法实现它应有的价值,也无法产生对个体和社会影响。将伦理与个体福利和社会发展的过程相结合,才能让伦理

① 刘培、池忠军:《算法歧视的伦理反思》,载《自然辩证法通讯》2019年第10期。

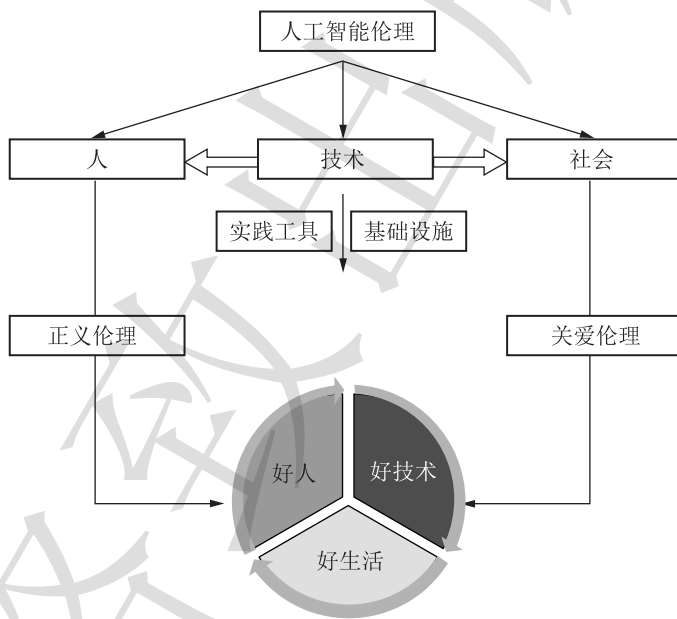
② 邱仁宗、黄雯、翟晓梅:《大数据技术的伦理问题》,载《科学与社会》2014年第1期。

③ 贾开:《人工智能与算法治理研究》,载《中国行政管理》2019年第1期。

④ Brent Mittelstadt, “Principles Alone Cannot Guarantee Ethical AI”, *Nature Machine Intelligence*, Vol.1(2019), pp.501—507.

“长牙齿”^①,发挥人工智能伦理的重要价值。

人工智能技术与人的尊严、社会发展的对立冲突是人工智能伦理兴起的动力,社会结构分化、政府治理能力、文化差异是塑造人工智能伦理变迁的关键因素。从人与技术的关系角度,人与技术共生的发展进程强调技术与道德的互存,强调技术实践活动合乎道德,才能推动人的道德性不断成长。^②人工智能伦理的整体目标是人、技术和社会三者之间的良性互动与优势叠加,进而实现发挥尊重人的主体性价值、提高个体的学习和发现能力,提高社会的整体福利。人工智能伦理的使命和目标,是推动更好的人(人的尊严和存在意义、人的能力)、更好的技术(丰富和拓展对世界的认识、提高能力和效率)、更好的生活(善且正当的生活方式,使得社会更加包容、多元和公平)三者的有机统一和系统性增效(图3)。



资料来源:作者自制。

图3 人工智能伦理的结构网络

① Anaïs Rességuier and Rowena Rodrigues, “AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics”, *Big Data & Society*, Vol.7(2020).

② 李宏伟:《基于人与技术实践共生的技术伦理反思》,载《中国人民大学学报》2021年第4期。

人工智能不仅要强化对技术维度的治理,推动技术本身的道德内化;也需要填补社会维度空缺带来的社会问题,积极促进个体的“自我实现”和对美好生活的向往。人工智能伦理的结构化,以重经验、重实效的原则,完善对人工智能技术的治理,从而实现可持续发展的目标。一方面,必须明确,机器是会理性思考的,但机器是伦理盲人,不能思考提前定义的程序和价值之外的因素。机器只能根据事先的程序作出判断,很有可能就会强化已有的认知和偏见。故而,引入正义伦理(ethics of justice),将伦理价值和标准嵌入设计和使用过程,即技术道德化引导主体的道德行为和伦理规范^①,推动人(设计者和使用者)和技术的双向问责^②,是一条可资考虑的进路。设计主体必须在人工智能设计之初,把人类道德与规范嵌入智能机器算法中,使得人工智能系统的决策、行为具有可靠性与安全性。换言之,就是通过前瞻性地识别和阐述技术的问题,将人类对道德重要性的理解和判断转译为人工智能信息处理和决策的逻辑规则,使其作出符合社会道德要求的决策(道德算法)。^③技术使用者也需要保持自身的反思性,不断反思和防范技术便利中潜在的技术奴役。

另一方面,必须关注以情感为中心的关爱伦理(ethics of care),强化技术和社会关系的深化及系统性,推动伦理规范与复杂性、不确定性社会系统的衔接,让伦理为社会系统赋能,促进社会整体的发展。^④技术伦理规范不能寄希望于科学推理或者先验假定,而需要在实践中得到检验。关爱伦理的核心关切,不是机器能否像人一样思考,而是机器能否比人思考更多,更好地影响和造福社会。^⑤将伦理议题与经济、社会系统结合,使人工智能合伦理性不仅成为人工智能系统发展的一个新的层级,也推动伦理与多要素的知识整合和拓展。为此,

① 孙田琳子:《论技术向善何以可能——人工智能教育伦理的逻辑起点》,载《高教探索》2021年第5期。

② 蒋晓丽、杨珊:《大数据时代媒介技术的伦理问责》,载《西南民族大学学报》(人文社科版)2016年第7期。

③ 顾世春:《瓦尔博斯前瞻性技术责任伦理研究》,载《自然辩证法研究》2015年第8期。

④ 朱春艳:《现代西方技术现实主义思潮评析》,载《江苏大学学报》(社会科学版)2011年第1期。

⑤ Jessica Morley, Luciano Floridi, Libby Kinsey and Anat Elhalal, “From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices”, *Science and Engineering Ethics*, Vol.26(2020), pp.2141—2168.

我们应不断总结人工智能技术应用的社会影响,增加可靠的社会知识和伦理知识,更好地促进人类在技术背景下美好生活的实现,保障人的权益和提升社会福利,引导经济、社会 and 文化的整体性变迁。^①推行关爱伦理要遵循三个原则:第一,重视个体行动与社会发展的关联;第二,反思性的创造,即强调行动的条件性和能动性;第三,去中心化,推动多元共存的发展生态。人工智能技术和伦理的发展,需要哲学家、技术人员(自然科学家、工程师)以及社会科学各界(心理学家、社会学、律师)加强知识的交流和互动、理论与实践的整合,为多元化的发展提供优质的知识内容和思想认知,完善人的生存方式,形成更好的技术和社会生态。

总而言之,人工智能伦理,既是知识创造、技术的自我创造、自我发明,更是文明的再造。在人工智能的社会运行过程中,应引入包容性创新、公平开放、可持续发展理念,推动人、技术、社会多个维度^②的统合与重构,保障技术化时代的美好生活。^③总体的美好生活,既是个体实践的目标,也是国家和社会制度合理安排的目标。

五、总 结

人工智能技术是人类社会发展的动力,而人工智能伦理是技术升级和软性的智能革命的重要维度。人工智能技术的风险性和不确定性,给个体和社会带来了新的伦理问题和风险。人工智能伦理对于升级技术治理、推动可持续创新、实现美好生活具有重要意义。人工智能伦理的结构化,推动了技术与社会生活的融合、伦理观念与技术物品的整合,揭示了个体、科技与社会之间的碰撞与矛盾,不断激发思想和社会创造性变革的火花。将忧患式、问题式的伦理意识变为引领式的伦理行动,探索理论与实践的整合,可避免人工智能伦理规则

① 孙伟平、李扬:《论人工智能发展的伦理原则》,载《哲学分析》2022年第1期。

② 王钰、程海东:《人工智能技术伦理治理内在路径解析》,载《自然辩证法通讯》2019年第8期。

③ 龙静云、吴涛:《新责任伦理:技术时代美好生活的重要保障》,载《华中师范大学学报》(人文社会科学版)2021年第5期;黎常、金杨华:《科技伦理视角下的人工智能研究》,载《科研管理》2021年第8期。

僵化的系统性风险。人工智能的伦理准则和个体和社会利益的动态关系,是本文探索人工智能伦理结构化的逻辑起点。基于技术与人和社会的互动,构建具有适应性和综合性的人工智能伦理和治理框架,有助于应对未来的人工智能发展的挑战。人工智能伦理理论与实践的系统化,需要计算机科学家、伦理学家和其他人文社会学科的探讨和合作,加快技术的民主化、开放化与审慎化(反思)的进程,深化具体情境下技术与社会发展的整合。