

人工智能视域下科技伦理

于奉鑫

华北电力大学 马克思主义学院, 北京昌平, 102206;

摘要: 新兴科技的迅猛发展不仅带来了前所未有的机遇, 也引发了深刻的伦理思考。在人工智能前沿技术快速突破的今天, 科技伦理不仅关乎技术发展的方向把控, 更是维系社会信任、保障人类福祉的关键支撑。这一领域的重要性已从学术讨论层面上升为全球治理的重要议题, 成为科技创新过程中不可或缺的规范性力量, 成为引导科技向善发展的关键因素。智能技术的深度应用不仅关乎个体权益的保护, 更涉及社会整体利益的平衡。面对这一全球性挑战, 中国在国际舞台上积极倡导“以人为本、智能向善”的发展理念, 致力于构建负责任的人工智能治理体系, 推动全球人工智能技术沿着伦理轨道健康发展。本文立足于人工智能伦理治理的现实需求, 通过系统探讨人工智能伦理的基本内涵, 深入分析伦理审查的道德正当性与政策规范的必要性, 旨在构建人工智能伦理审查的理论框架。

关键词: 科技伦理; 人工智能; 科技向善; 以人为本

DOI:10.69979/3041-0673.25.05.008

人工智能以惊人的速度渗透到现代社会的各个领域, 在带来颠覆性变革的同时, 也引发了深刻的伦理反思。其迅猛发展不仅为人类开辟了前所未有的机遇, 也对人类的主体性构成了严峻挑战^[2]。正是这些问题的紧迫性, 使得伦理治理成为人工智能治理中不可回避的核心议题。智能技术凭借其强大的变革力量, 正在重塑人类的世界观与价值观。然而, 无论技术如何日新月异, 以人为本的核心价值始终不可动摇^[3]。

1 人工智能伦理问题的表现形式

人工智能的广泛应用, 给各个领域带来了前所未有的机遇, 同时也引发了一系列复杂的伦理问题。这些伦理问题不仅关乎个体权益的保护, 还涉及社会整体利益的平衡。在某些情境下人工智能的自由裁量权已超越了日常事务的范畴, 甚至在涉及生死的关键决策中, 人工智能的影响力也不容忽视^[6]。人工智能伦理问题的关注度与其技术发展阶段密切相关。在强人工智能阶段, 伦理问题则进一步延伸到深度伪造、舆论操控、价值观干预等领域^[7]; 而当人工智能发展到超级智能阶段时, 其引发的伦理挑战将更加复杂和深远。这一阶段不仅需要应对智能水平差异带来的社会分化, 还要解决国家间技术发展的不平衡问题, 更将直面人类生存安全的根本性威胁。这些跨领域、跨层次的伦理议题构成了人工智能治理的核心挑战。接下来, 我们将针对若干具有代表性的人工智能伦理困境展开具体分析。

人工智能系统依赖于海量数据进行训练和优化, 但

在数据收集过程中, 可能侵犯个人隐私, 甚至滥用敏感信息。大规模数据的集中存储和处理增加了数据泄露的风险, 可能导致个人信息被非法利用。例如, 社交媒体平台利用用户数据进行精准广告投放, 虽然提高了广告效果, 但也引发了用户隐私泄露的担忧。此外, 医疗领域的人工智能应用需要处理大量敏感健康数据, 一旦泄露, 将对患者隐私造成严重侵害。此外, 当人工智能系统做出错误或有害决策时, 责任归属难以界定。是开发者、使用者还是算法本身应承担 responsibility? 这一问题尚未有明确的法律规定。在医疗诊断领域, 如果人工智能系统给出错误诊断, 导致患者健康受损, 责任应由医生、医院还是算法开发者承担? 这些问题的解决需要法律和伦理框架的进一步完善。

算法黑箱现象主要存在于深度学习等复杂人工智能系统中, 其特征表现为系统内部决策机制的高度不透明性^[8]。这种不透明性不仅使普通用户难以理解, 甚至对开发者而言也构成了认知障碍, 最终可能导致决策结果缺乏可解释性, 进而引发潜在的公平性问题。如果训练数据本身存在偏见, 人工智能系统可能会放大这些偏见, 导致对特定群体的歧视。例如, 在招聘系统中, 如果历史数据中存在性别或种族偏见, 人工智能可能会在筛选简历时延续甚至加剧这些偏见, 导致不公平的招聘结果。在刑事司法系统中, 人工智能。由于算法黑箱的存在, 即使受影响者遭受歧视性决策, 也难以了解其背后的依据, 这在一定程度上削弱了公众的合法权益。如何

确保公平性并消除偏见,已成为一个亟待解决的伦理问题。

2 人工智能伦理审查的理论逻辑

人工智能技术的迅猛发展使得相关伦理问题愈发突出,促使科技企业、研究机构和监管主体积极探索可行的治理路径。在此背景下,借鉴生命科学领域的伦理审查机制因其显著的特性和优势,正在演变为人工智能伦理治理体系中的重要组成部分。这一制度不仅能够促进多方主体协同共治,形成政府、企业、学术界和社会公众共同参与的治理格局,还能通过预先评估和防范潜在风险,降低人工智能技术应用可能引发的伦理问题。此外,伦理审查制度具有灵活性和适应性,能够根据人工智能技术的快速迭代和社会需求的变化,动态调整审查标准和流程,从而有效应对技术发展中的不确定性。这些优势使得伦理审查制度在人工智能伦理治理中展现出重要的实践价值,为构建负责任的人工智能发展框架提供了有力支撑。

2.1 人工智能伦理审查的道德逻辑

人工智能伦理审查的首要道德逻辑是以人为本,确保技术的发展和应用始终服务于人类的福祉。伦理审查的道德逻辑要求人工智能系统在设计和应用中必须体现公平正义,避免算法偏见和歧视。审查过程中需要评估算法是否对特定群体(如种族、性别、年龄等)存在不公平对待,并采取措施确保决策的公正性。强调责任归属的明确性,确保在技术应用过程中出现问题时能够追溯责任主体。这一逻辑要求开发者、使用者和监管者共同承担伦理责任,建立清晰的责任链条。此外,人工智能伦理审查强调风险预防,通过前瞻性评估识别潜在伦理风险,并采取相应措施加以防范。这一逻辑要求在设计阶段就考虑技术可能带来的负面影响,并制定应急预案。人工智能伦理审查的道德逻辑还涵盖可持续发展,要求人工智能技术的开发和应用必须兼顾环境效益和社会效益。审查过程中需要评估技术的能源消耗、碳排放以及对社会的长期影响。人工智能伦理审查的道德逻辑还强调全球协作,推动国际社会在人工智能伦理问题上达成共识。这一逻辑要求各国在伦理审查标准和实践上进行协调,避免因文化差异导致的伦理冲突。

在人工智能伦理治理中,亟需构建新型审查框架与标准体系。首要任务是科学界定审查的适用范围与边界,

其次应建立兼顾多方诉求的程序公开机制。更为关键的是,审查环节必须贯穿人工智能研发的全流程,包括数据获取、算法构建、模型优化等关键环节,从而实现风险防控的无缝覆盖^[9]。这种全周期、多维度的审查模式将有效提升人工智能伦理治理的系统性和实效性。

2.2 人工智能伦理审查的政策逻辑

人工智能技术的迅猛发展对现有伦理监管体系提出了新的挑战。传统的科学共同体自治模式,主要依托同行评议和学术交流来管控伦理风险,这种机制在技术影响相对有限的时期确实发挥了重要作用。然而,面对人工智能技术日益显著的跨学科特性和广泛社会影响,单纯依靠内部自治已难以应对复杂的伦理监管需求,建立系统化的外部监管框架势在必行^[10]。

建立规范的伦理审查机制具有多重积极意义。其一,这一制度能够促进社会对人工智能技术的认知与接纳。虽然技术创新是推动社会进步的重要动力,但其引发的价值冲突和伦理困境也不容忽视^[11]。人工智能在就业替代、算法公平性、决策透明度等方面带来的挑战,已经引发了广泛的社会担忧,甚至造成了公众信任危机,严重制约了技术的推广应用。通过伦理审查提升技术透明度,有助于消除公众疑虑,为人工智能的良性发展营造有利的社会环境。

从国际竞争视角来看,伦理审查制度的建立将显著提升我国在人工智能领域的话语权。随着国际社会日益重视科技伦理,伦理评估已成为人工智能研究和产品开发的重要环节。我国构建完善的伦理审查体系,不仅能够提高科研成果的国际认可度,还能为相关产品进入全球市场提供伦理保障,从而增强我国在该领域的国际竞争力^[12]。

此外,前瞻性的伦理审查制度因其动态适应性,能够更好地满足新兴科技的治理需求。人工智能技术的快速迭代带来了持续涌现的新型人机交互模式和伦理挑战,这对传统的静态治理模式构成了严峻考验。伦理审查制度通过持续优化审查标准和流程,能够及时响应技术发展中的新问题,为人工智能的可持续发展提供制度保障^[13]。这种兼具灵活性和前瞻性的治理模式,为新兴科技领域的伦理治理提供了创新性的解决方案。

3 人工智能伦理审查的原则建构

在明确人工智能伦理审查的理论依据后,构建相应

的原则规范至关重要。这一过程应以“以人为本”为基石，确保人工智能朝着“向善”的方向发展。科技活动应围绕人民需求展开，通过改善民生、促进社会进步来推动经济增长，提升人民的幸福感，支持社会的可持续发展。

“智能向善”应作为伦理审查的核心原则，其设计应以人本主义为基础，确保科技服务于人类福祉。科技的价值取决于使用者的引导，因此，人工智能的发展必须尊重和保障基本人权。在技术研发和应用中，需对其伦理风险进行严格评估，特别是对隐私权和平等权的潜在影响。通过建立严格的伦理审查机制，确保技术应用不损害人类核心价值，推动社会公平与进步。

人工智能伦理审查应着重评估技术对基本人权的可能影响，并建立相应的问责与救济机制，确保公平公正。在坚持“以人为本”原则的同时，还需关注因透明度不足导致的损害，防止技术发展侵害合法权益。

参考文献

[1] 习近平在省部级主要领导干部坚持底线思维着力防范化解重大风险专题研讨班开班式上发表重要讲话强调提高防控能力着力防范化解重大风险保持经济持续健康发展社会大局稳定李克强主持栗战书汪洋王沪宁赵乐际韩正出席[J]. 时事报告, 2019, (02): 9-11.

[2] 《关于加强科技伦理治理的意见》出台[J]. 机器人与应用, 2022, (02): 1.

[3] 张平. 人工智能伦理治理研究[J]. 科技与法律(中英文), 2024, (05): 1-12. DOI: 10.19685/j.cnki.cn11-2922/n. 2024. 05. 001.

[4] 于雪, 凌昀, 李伦. 新兴科技伦理治理的问题及其对策[J]. 科学与社会, 2021, 11(04): 51-65. DOI: 10.19524/j.cnki.10-1009/g3. 2021. 04. 051.

[5] 沈林兴, 刘英. 数字科普是数字化时代科普的主流[J]. 科普研究, 2011, 6(S1): 66-70. DOI: 10.19293/j.cnki.1673-8357. 2011. s1. 017.

[6] 江小涓. “十四五”时期数字经济发展趋势与治理重点[J]. 上海企业, 2020, (11): 56-57.

[7] 王硕, 李秋甫. 数字伦理: 数字化转型中科学普及的新使命与新规范[J]. 科普研究, 2023, 18(03): 57-64+72+113. DOI: 10.19293/j.cnki.1673-8357. 2023. 03. 007.

[8] 陈勇川. 回顾与展望: 我国生物医学研究伦理审查的发展趋势[J]. 医学与哲学, 2020, 41(15): 1-7.

[9] 孟令宇, 王迎春. 探索人工智能伦理审查新范式[J]. 科学与社会, 2023, 13(04): 97-113+165. DOI: 10.19524/j.cnki.10-1009/g3. 2023. 04. 097.

[10] 李建军, 唐冠男. 阿希洛马会议: 以预警性思考应对重组DNA技术潜在风险[J]. 科学社会, 2013, 3(02): 98-109. DOI: 10.19524/j.cnki.10-1009/g3. 2013. 02. 009.

[11] 王学川. 科技伦理价值冲突及其化解[M]. 浙江大学出版社: 2016: 12. 285.

[12] 杜严勇. 人工智能伦理审查: 现状、挑战与出路[J]. 东华大学学报(社会科学版), 2024, 24(02): 32-39. DOI: 10.19883/j.1009-9034. 2024. 0124.

[13] 吴兴华. 从“回应式”治理到“前瞻式”治理: 人工智能伦理治理模式的变革[J]. 理论导刊, 2024, (02): 62-68.

[14] 张永忠. 论人工智能透明度原则的法治化实现[J]. 政法论丛, 2024, (02): 124-137.