

人工智能决策中的道德伦理考量

郭丹丹

新疆大学马克思主义学院/铸牢中华民族共同体意识研究基地 新疆乌鲁木齐

【摘要】随着人工智能技术在各领域的广泛应用，其引发的道德伦理问题日益凸显。本文从马克思主义理论的角度出发，深入分析了人工智能决策中存在的算法偏见、算法不透明性和隐私侵犯等道德伦理问题，探讨了马克思主义伦理学中关于人的主体性尊严、社会公平和正义、责任和权力等理论在人工智能决策中的作用，剖析了道德伦理考量在实践中的挑战，并提出了相应的对策建议，旨在为人工智能决策的伦理治理提供理论支持。

【关键词】人工智能；算法伦理；科技伦理

【收稿日期】2025 年 5 月 15 日

【出刊日期】2025 年 6 月 17 日

【DOI】10.12208/j.sdr.20250037

Moral and ethical considerations in AI decision-making

Dandan Guo

School of Marxism, Xinjiang University/Research Base for Strengthening the Consciousness of the Chinese Nation's Community, Urumqi, Xinjiang

【Abstract】 With the widespread application of artificial intelligence(AI)technologies across various domains, the ethical issues they raise have become increasingly prominent. From the perspective of Marxist theory, this paper conducts an in-depth analysis of moral and ethical problems in AI decision-making—such as data bias, algorithmic opacity, and privacy infringement. It explores the role of Marxist ethical concepts concerning human subjectivity and dignity, social equity and justice, responsibility, and power in AI decision processes, dissects the practical challenges of ethical considerations, and proposes corresponding countermeasures and recommendations. The aim is to provide a theoretical foundation for the ethical governance of AI decision-making.

【Keywords】 Artificial Intelligence; Algorithmic ethics; Technology ethics

人工智能在各个领域中的快速发展和应用，使其在决策中的道德伦理问题日益突出。如何确保人工智能的决策过程符合道德伦理原则，避免可能出现的道德风险，成为一个亟待解决的问题。伦理问题的出现，既源于人工智能技术的特性，也源于其在应用过程中的复杂性。人工智能的决策过程涉及到大量的伦理原则和道德标准。这些原则和标准不仅来源于人类社会的道德规范，也来自于人工智能的设计和开发过程中所遵循的伦理原则。然而，这些原则和标准如何有效应用于人工智能的决策过程中，仍然存在许多挑战。

1 人工智能决策中的道德伦理问题

1.1 数据偏见

数据偏见是人工智能决策中的典型伦理问题，表现为算法因数据来源或处理方式而对某些群体产生不公平对待。系统性偏见往往源于训练数据集本身的不完整性或不公正的收集与处理方法，而结构性偏见则反映了数据在性别、种族、经济地位等社会因素下的固有不平衡^[1]。二者共同作用，使得算法在教育入学评估中可能对某些性别或背景的学生产生歧视，在医疗诊断中也可能高估或低估特定患者的风险，从而导致错误决策，侵害个体的健康和权利。

这种偏见不仅削弱了 AI 决策的公平性，还会在

作者简介：郭丹丹（1994-）女，汉族，山西省大同市人，硕士，新疆大学马克思主义学院硕士研究生，研究方向：马克思主义基本原理、马克思主义文化理论。

实践中加剧社会不平等。为了缓解其影响,首先需要训练数据和算法模型进行公正性评估,及时识别并修正潜在偏差;同时,通过公开模型原理与决策依据,提高算法透明度,使开发者和用户都能理解、监督其运行;此外,建立严格的行业监管和责任追究机制,防止算法被滥用或恶意操纵;最后,应优化数据收集方案,扩大数据来源的多样性,并在算法开发和应用全流程中持续关注公平原则,确保 AI 在各领域的决策更加合理与公正^[2]。

1.2 算法不透明性

算法不透明性是指在人工智能决策中,由于模型结构复杂、数据来源不明和处理流程缺乏可追溯性,使决策逻辑对开发者和用户都难以理解和检验。结构性不透明源于深度神经网络等黑盒模型本身的复杂性;数据源不透明则表现为训练数据的质量、来源和偏差难以被准确把握;过程性不透明指算法在数据预处理、参数调优和结果生成等环节缺乏清晰说明或可视化^[3]。

这种不透明性不仅提高了决策错误的风险,可能在教育、医疗等场景中造成歧视和偏见,侵害个体权利,还削弱了公众对 AI 技术的信任,阻碍其社会各领域的推广应用。

为化解这一伦理难题,应在技术、制度和社会层面协同发力。首先大力发展可解释 AI,设计更透明的模型并配套决策过程可视化工具,让人们能够直观了解算法依据;其次完善监管和审查机制,要求算法开发者公开关键文档和运行细节;同时鼓励公众参与和监督,及时收集反馈并纠正偏差;最后加强教育培训,提升行业监管者和社会公众对算法透明度重要性的认识。只有在多方协作下,才能构建公平、可信的人工智能生态。

1.3 隐私侵犯

隐私侵犯是指未经个体许可,擅自获取、使用、披露或干涉其私人信息或活动。在人工智能领域,这种侵犯常见于擅自采集用户的身份信息 and 行为轨迹,将数据用于广告投放或画像分析,因系统漏洞或管理疏忽导致的大规模泄露,以及通过恶意软件或社交平台进行的持续监控与追踪^[4]。隐私侵犯不仅会引发公众对 AI 系统的信任危机,还可能触及伦理与法律红线,当决策基于未经授权的数据时,更容易带来偏见或歧视性的结果,损害个体权利并加

剧社会不公。

要遏制隐私侵犯,需要从多个层面协同发力。首先,应增强用户和企业的隐私保护意识,使其在数据采集与使用时保持谨慎;同时,完善并严格执行相关法律法规,明确数据处理边界并加大违法成本。技术上,要推广加密和匿名化等手段,提升系统安全性,防范泄露与非法访问;在行业层面,应建立常态化的监管机制并鼓励自律,制定统一的隐私保护标准;最后,通过提高 AI 系统的透明度与可解释性,让用户清楚地知晓数据流向与决策逻辑,从而在源头上削弱未经授权使用个人信息的可能性。

2 马克思主义理论视角下的道德伦理考量

2.1 人的主体性和尊严

马克思主义认为人的主体性是社会历史发展的产物,是人的本质和尊严的体现。人能够依靠“自由”“自觉”的活动,即有目的、有意识的阶级斗争、生产劳动和科学实验去改造自然、改造社会和改造自身,同时,在改造对象过程中获得自身发展的自由。这种自由是受到社会历史条件的制约^[5]。

从伦理学的角度来看,人的主体性原则是整个伦理学的精神主旨。人不是作为被动的客体存在,人既是历史剧作者,又是历史的剧中人,他们在前人遗留的物质和精神活动前提下按照自己的目的开展物质实践活动。人的主体性原则在伦理学中的体现,主要表现在人对于自我和他人的道德责任。人应该对自己的行为负责,对自己的生活负责,在实践活动中不断反思和审视自己的行为,提升自身的道德素质。同时,人应该对他人的行为负责,对他人的生活负责,在实践中尊重他人的权利,关爱他人的利益,促进他人的全面发展。此外,人的主体性原则在伦理学中也体现在人对于社会和自然的关系上,人应该尊重自然,保护自然,与自然和谐相处,在实践中尊重自然的规律,合理利用自然资源,实现人与自然的和谐发展。

2.2 社会公平和正义

社会公平和正义是伦理学领域的重要议题,它涉及到社会资源的合理分配、利益的公平获取以及人民主体性的尊重等问题。社会公平和正义是指在社会生产和发展过程中,按照一定的标准和原则,合理分配社会资源和利益,使全体人民共享社会发展的成果^[6]。马克思主义认为,社会公平和正义是社

会发展的基本要求，是人的主体性和尊严的体现。社会公平和正义的实现需要建立在生产资料公有制的基础之上，通过大力发展生产力，大幅增加社会财富，消灭工农之间、城乡之间、体力与脑力劳动之间的差别。保证全体社会成员享有平等的政治地位和社会地位，消灭社会一切不公正现象，实行各尽所能、各取所需，实现人的自由和全面发展。

马克思主义认为政治正义是制度正义的前提和基础，是整个制度伦理的价值基础和底线。在人工智能决策中，社会公平和正义要求人工智能系统的设计和使用应该避免加剧社会不平等和歧视现象，应该确保所有群体和个体都能够平等地受益于人工智能技术的发展。

2.3 责任和权利

在马克思主义伦理学视域下，责任与权利是辩证统一的：责任要求各方在社会规范下承担义务，权利则保障个体或集体的自由与利益。在 AI 决策中，责任应由开发者、使用者和管理者共同承担^[7]。开发者需设计符合道德与法律的系统，确保安全可靠、透明可解释；使用者和管理者则应遵守法规与伦理，对运行过程和结果进行监督评估，及时纠正偏差与漏洞。

与此同时，用户作为直接受益者，拥有知情权、申诉权和隐私安全权，能够了解决策逻辑、质疑结果并保护个人数据；社会公众作为间接受益者，也应参与和监督 AI 的开发与运行，以维护公共利益和社会正义。只有在多方共同履责、切实保障各方权利的基础上，人工智能决策才能实现可持续发展，推动社会和谐与进步。

3 道德伦理考量在人工智能决策中的应用与实践

3.1 数据偏见与道德伦理考量

大数据时代，数据质量、多样性和价值密度直接决定算法效能。从道德伦理角度看，必须在数据获取、处理和分析的全过程中坚持真实性与完整性，尊重并维护数据多样性，杜绝因样本偏倚或标签偏差引发的算法歧视。尽管数据清洗与增强可提升代表性，却也可能带来新的偏见，因而需谨慎权衡并反复验证^[8]。

在算法设计与应用阶段，应当嵌入公平性约束，强化模型的可解释性与问责机制，对所有可能影响

决策公正的参数进行独立审查，避免对弱势群体造成二次伤害。同时，建立动态的伦理审查与评价体系，持续监测并纠正数据和算法中的不公现象，确保 AI 技术始终致力于促进而非阻碍社会正义。

3.2 算法不透明性与道德伦理考量

算法不透明性的伦理应对依赖于提升透明度和可解释性的道德手段。马克思主义伦理学批判“技术至上”观，强调技术必须服务于人并置于公共伦理和法治之下。在关键领域，尤其是公共决策系统，应部分公开决策流程、关键参数和模型结构，以增强社会监督；同时加强可解释性研究，让用户和监管者理解算法逻辑，为申诉和问责提供技术支持。推进技术与伦理并重，妥善平衡透明性与系统安全、创新与知识产权保护，防范过度披露带来的新风险。立法和政策层面则需细化算法问责机制，明确各环节责任，构建“可管、可控、可追责”的治理格局。

3.3 隐私侵犯与道德伦理考量

在大数据技术中，首先必须尊重人的主体性和尊严，在数据收集、处理和使用各环节保障用户的知情权、选择权与监督权，采取加密、脱敏等技术手段，确保个人信息不被泄露或滥用。只有让用户在每一步都能清楚地看到并控制自己的数据流向，才能真正维护其隐私安全与自主权。

与此同时，应公平对待所有社会成员，防止大数据分析对特定群体造成歧视或负面影响，将公平原则嵌入算法设计和业务流程，对关键决策变量进行独立审查并定期评估。通过技术方案中融入尊重、隐私保护与公平正义等核心价值，才能实现大数据应用的伦理规范和可持续发展，为用户提供更安全可靠的服务。

参考文献

- [1] 姜延强;. 人工智能的伦理困境与正解[J].道德与文明.2022.(01):133-141.
- [2] 岳大为;张亚迪;黄子睿;. 人工智能时代的伦理道德思考[J].教育教学论坛.2020.(03):278-279.
- [3] 钱燕娜;储召锋;. 人工智能的社会影响研究[J].重庆科技学院学报(社会科学版).2023.(06):69-79.
- [4] 唐凯麟;李诗悦;. 大数据隐私伦理问题研究[J].伦理学研究.2016.(06):108-112.

- [5] 王军;.人工智能的伦理问题:挑战与应对[J].伦理学研究.2018.(04):85-89.
- [6] 梁宪飞;.对人工智能时代算法歧视的思考[J].中国信息化.2020.(07):56-57.
- [7] 王艺霖;刘芸瑞;.马克思主义语境下的人文精神解读[J].中共青岛市委党校.青岛行政学院学报.2014.(01):34-37.
- [8] 陈思坤;.人的尊严伦理及价值[J].教育学术月刊.2010.(12):15-17+69.

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<https://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS