

历史唯物主义视域下机器道德的本质研究

姬文静, 崔玮琳

北京建筑大学马克思主义学院, 北京

收稿日期: 2026年5月9日; 录用日期: 2026年6月1日; 发布日期: 2026年6月11日

摘要

在历史唯物主义视域下, 机器道德的本质并非真正意义上的道德能力。从生产力与生产关系的辩证关系来看, 机器道德是人类基于自身生产生活需求, 通过技术手段赋予机器的特定功能集合, 是人类道德认知和实践在技术层面的映射。其局限显著, 一方面, 机器缺乏主观意识和情感体验, 无法真正理解道德情境中的复杂情感因素, 只能依据预设程序和算法运行。另一方面, 机器道德受技术发展水平和社会生产关系的制约, 不同社会阶层和利益群体对机器道德定义和应用存在差异。同时, 机器难以适应社会历史发展中不断变化的道德观念和价值标准, 无法进行自主的道德判断和选择。因此, 对机器道德需理性看待, 重视人类在道德领域的主体地位。

关键词

历史唯物主义, 机器道德, 情感, 本质

Research on the Essence of Machine Morality from the Perspective of Historical Materialism

Wenjing Ji, Weilin Cui

School of Marxism, Beijing University of Civil Engineering and Architecture, Beijing

Received: May 9, 2026; accepted: June 1, 2026; published: June 11, 2026

Abstract

From the perspective of historical materialism, the essence of machine morality is not moral ability in the true sense. From the perspective of the dialectical relationship between productive forces and production relations, machine morality is a set of specific functions given to machines by human

beings based on their own production and life needs, and is the mapping of human moral cognition and practice at the technical level. On the one hand, machines lack subjective consciousness and emotional experience, cannot truly understand complex emotional factors in moral situations, and can only operate according to preset programs and algorithms. On the other hand, machine morality is restricted by the level of technological development and social production relations, and there are differences in the definition and application of machine morality by different social classes and interest groups. At the same time, it is difficult for machines to adapt to the changing moral concepts and value standards in the development of social history, and cannot make independent moral judgments and choices. Therefore, machine morality needs to be viewed rationally and the main position of human beings in the field of morality should be valued.

Keywords

Historical Materialism, Machine Morality, Emotion, Essence

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

随着人工智能的迅猛发展, 机器道德成为学术界与社会各界关注的焦点。从自动驾驶汽车的伦理决策到服务型机器人的道德响应, 相关实践不断挑战传统道德理论的边界。在这一背景下, 以历史唯物主义为理论视角剖析机器道德的本质, 具有重要的现实意义与理论价值。历史唯物主义强调生产力与生产关系的辩证运动, 以及社会存在对社会意识的决定性作用, 这为我们揭示机器道德背后的社会根源与发展规律提供了科学方法论。本文将深入探讨机器道德在历史唯物主义框架下的本质属性, 剖析其产生的社会经济基础与技术动因, 以期为理解机器与人类道德关系提供更深刻的理论支撑。

2. 历史唯物主义视角下机器道德的产生与本质属性

历史唯物主义的基本原理为剖析机器道德的本质提供了科学且系统的理论框架。马克思在《〈政治经济学批判〉序言》中明确指出, “人们在自己生活的社会生产中发生一定的、必然的、不以他们的意志为转移的关系, 即同他们的物质生产力的一定发展阶段相适合的生产关系。这些生产关系的总和构成社会的经济结构, 即有法律的和政治的上层建筑竖立其上并有一定的社会意识形式与之相适应的现实基础” [1]。这一论断深刻揭示了生产力与生产关系, 社会存在与社会意识的辩证关系, 由此可见, 机器道德并非孤立的技术现象, 而是当代生产力与生产关系、社会存在与社会意识辩证运动的必然产物。它的出现与发展, 始终深深植根于人类社会的物质生产与社会关系之中, 只有从历史唯物主义的视角出发, 才能准确把握其本质。

生产力与生产关系的辩证运动从根本上制约着机器道德的发展进程。马克思在《资本论》中强调, “劳动资料不仅是人类劳动力发展的测量器, 而且是劳动借以进行的社会关系的指示器” [2]。由算法、算力和数据构成的技术体系, 虽能使机器模拟人类在特定场景下的决策行为, 但这种模拟本质上是人类将自身道德认知通过技术编码赋予机器。在当前大数据与深度学习技术的支持下, 机器能够处理海量信息并做出看似符合道德准则的决策, 然而这一切都源于人类预先设定的程序和规则。机器道德行为并非源于其自身对道德的理解, 而是人类改造自然、社会过程中道德观念的技术投射。随着生产力的持续进

步, 量子计算、脑机接口等前沿技术的发展, 也将为机器道德带来新的发展可能, 但这种发展始终会受到生产关系的影响和制约。

资本主义与社会主义两种不同生产关系, 塑造了机器道德截然不同的发展方向与价值取向。在《资本论》的注释中提到过: “如果有 10% 的利润, 资本就保证到处被使用; 有 20% 的利润, 资本就活跃起来; 有 50% 的利润, 资本就铤而走险; 为了 100% 的利润, 资本就敢践踏一切人间法律; 有 300% 的利润, 资本就敢犯任何罪行, 甚至冒绞首的危险。” [3] 在资本逐利本性的支配下, 企业对机器道德的研发和应用往往以成本控制与利润增长为首要目标。与之相对, 社会主义生产关系以“满足人民日益增长的美好生活需要”为根本宗旨, 使得机器道德的发展更契合社会公共利益。这正呼应了马克思设想的共产主义社会中“每个人的自由发展是一切人的自由发展的条件” [4], 社会主义制度下对人工智能技术的应用, 致力于打破技术垄断, 推动技术普惠, 让机器道德真正服务于全体人民的福祉。

社会存在决定社会意识, 决定了机器道德是人类道德认知在技术载体上的具象化呈现。马克思认为, “不是人们的意识决定人们的存在, 相反, 是人们的社会存在决定人们的意识”。不同社会形态、文化背景下, 人类对机器道德的期望与定义存在显著差异, 其根源在于社会存在的多样性。在个人主义文化盛行的西方社会, 对用户隐私保护等“道德”功能的重视, 折射出其对个体权利的强调。恩格斯在《路德维希·费尔巴哈和德国古典哲学的终结》中指出, “一切以往的道德论归根到底都是当时的社会经济状况的产物” [5], 随着社会历史的发展, 道德观念不断演变, 机器道德也必然需要随之调整更新。

从时间维度审视, 机器道德的产生与发展具有鲜明的历史性特征。马克思在分析机器大工业时代时提出, “手推磨产生的是封建主的社会, 蒸汽磨产生的是工业资本家的社会”, 这表明技术进步能够推动社会形态变革, 同时也重塑着人类对机器功能的认知。回顾历史, 工业革命初期, 机器主要服务于物质生产, 尚未被赋予“道德”属性; 信息技术革命后, 机器信息交互能力的提升促使人们开始关注其在信息传播、隐私保护等方面的责任; 到了人工智能时代, 机器道德成为研究热点。这一演进过程充分印证了恩格斯所说的“每一个时代的理论思维, 包括我们这个时代的理论思维, 都是一种历史的产物, 它在不同的时代具有完全不同的形式, 同时具有完全不同的内容”。当下, 随着人工智能技术的快速发展, 机器道德能力面临着新的挑战 and 机遇, 其发展也将继续遵循历史发展的规律。

从空间维度来看, 不同国家和地区的政治制度、经济发展水平、文化传统差异, 造就了机器道德发展的地域特征。以北欧福利国家和发展中国家为例, 前者基于对公平、人文关怀的重视, 在智能养老设备研发中强调情感陪伴功能; 后者出于经济发展与社会稳定需求, 侧重利用机器道德解决民生问题。这种地域差异本质上是不同社会存在条件下, 人类对机器道德进行差异化构建的结果。同时, 全球化背景下, 不同地区在机器道德发展方面也存在相互影响和借鉴, 共同推动着相关技术和理念的进步。

在历史唯物主义视域下, 机器道德是生产力与生产关系、社会存在与社会意识辩证运动的产物, 其本质是人类道德认知的技术外化, 具有鲜明的工具属性与社会历史性。马克思主义经典作家的理论为我们穿透技术表象、把握机器道德的本质提供了锐利的思想武器, 这有助于我们在人工智能时代正确处理人与机器的道德关系, 坚守人类在道德领域的主体地位。

3. 机器道德的现实实践与局限性

机器道德的现实实践, 深刻反映了人类道德认知与技术发展的紧密交织。“万物的主人理应是人类, 所有非人存在物, 无论是否有生命, 都应该为人所用, 它们的存在便是以人为目的的。” [6] 技术作为劳动的重要手段, 理应服务于人的自由发展与社会进步, 这为理解机器道德提供了关键视角。从交通出行到家庭生活, 从养老服务到教育辅助, 机器道德在多领域的实践, 既彰显了技术赋能伦理的巨大潜力, 也暴露出其作为工具性存在的固有局限。

1) 无人驾驶领域的安全与秩序维护

无人驾驶汽车通过传感器、算法与决策系统的协同运作, 将“保障生命安全”“维护交通秩序”的道德准则转化为可执行的技术程序。无人驾驶汽车的研发初衷, 在于利用智能技术降低因人为操作失误而导致的交通事故发生概率。实践中的道路测试也表明, 无人驾驶汽车在降低事故发生率方面的确具有显著优势[7]。然而, 当无人驾驶汽车面临类似“电车难题”的困境时, 其局限性便突显出来。在极端情况下, 无人驾驶系统需在撞向突然闯入道路的行人, 或是转向可能导致车内乘客受伤的障碍物之间做出选择。其决策完全依赖预先设定的算法规则, 缺乏人类道德判断中对情感、社会共识的考量。机器因缺乏情感体验与自主意识, 无法像人类那样基于对生命价值的权衡、社会道德舆论的预期等做出决策, 这清晰地暴露了其“道德”能力的机械性本质, 也凸显了技术在模拟人类复杂道德判断时的巨大挑战。在此背景下, 如何在法律层面对人工智能技术进行有效规制, 是社会必须予以回应的重要问题[8]。

2) 智能家居领域的关怀

智能家居系统借助物联网与人工智能技术, 将“保护家庭安全”“提升生活质量”的道德诉求转化为智能设备的自动化运行。智能家居通过实时监测、智能控制等功能, 构建起安全、便捷的生活环境, 成为人类生活实践的有效延伸。智能安防系统对入侵行为的自动报警, 健康监测设备对居住者身体状况的实时追踪, 均体现了技术对人类生存关怀的切实实现。

智能家居的“道德”能力受限于数据隐私与算法偏见问题。若设备采集的用户数据被非法利用, 如一些不法分子通过破解智能摄像头获取家庭隐私画面并用于非法交易, 这严重违背“尊重隐私”的道德原则。此外, 若算法在设计中存在性别、年龄等偏见, 可能导致智能客服对不同用户提供差异化服务, 违背“公平对待”原则[9]。马克思在《资本论》中深刻揭示了资本主义生产方式下对劳动者权益的侵犯, 在智能家居领域, 类似的问题也反映出技术在资本驱动下可能偏离服务人类的初衷, 沦为侵犯权利的工具, 这与马克思对技术异化现象的批判相呼应。

3) 养老陪护机器人领域的关爱

养老陪护机器人运用传感技术与自然语言处理算法, 承担起陪伴老年人、保障其健康安全的“道德”责任。养老陪护机器人通过协助老年人日常活动、提供情感陪伴, 在技术层面践行了对老年人关爱的伦理要求。其对老年人身体状态的实时监测与异常情况的及时反馈, 有效降低了意外风险, 显著提升了养老服务的质量与效率。

当机器人的行为导致老年人受伤或权益受损时, 开发者、制造商与使用者之间的责任界定存在法律空白[10]。例如, 若养老陪护机器人在协助老人起身时, 因程序故障导致老人摔倒受伤, 此时难以明确是算法设计缺陷、硬件质量问题, 还是使用者操作不当所致。在当前技术快速发展而制度规范相对滞后的情况下, 养老陪护机器人的责任困境凸显了社会规范未能及时适应技术变革的问题, 亟待通过制度创新与法律完善加以解决。

4) 教育辅助机器人领域的发展

教育辅助机器人通过大数据分析 with 个性化算法, 将“促进教育公平”“因材施教”的教育伦理转化为实际教学功能。教育辅助机器人依据学生学习数据智能推送学习内容, 为不同基础的学生提供定制化学习方案, 有助于缩小教育资源差距, 推动教育公平的实现。在偏远地区, 其弥补师资不足的作用, 更是直接促进了教育资源的均衡分配。

但教育辅助机器人的应用可能加剧“数字鸿沟”。若算法设计过度依赖学生的历史数据, 可能固化学习路径, 限制学生的创新思维。例如, 对于一些基础薄弱但潜力较大的学生, 算法可能持续推送简单内容, 阻碍其提升。同时, 经济条件的差异导致部分学生无法获得机器人辅助, 从而使教育资源分配不均。技术应用的现实效果深受社会经济条件的制约, 这要求我们在推动教育辅助机器人发展时, 必须充

分考虑社会公平因素, 避免技术进一步拉大教育差距。

机器“道德”能力在现实中的多领域实践, 是人类道德理想与技术能力结合的产物。其既展现了技术对道德实践的拓展潜力, 也暴露了技术工具性与人类道德复杂性之间的矛盾。

4. 科技向善与人机平衡的发展路径

1) 坚守人类价值主体地位

人类是社会历史的创造者与价值主体, 机器道德作为技术实践的产物, 其本质是人类道德认知的外化与延伸。这决定了机器道德行为无法脱离人类的价值设定与目标导向。从理论层面看, 机器的算法逻辑、决策规则均源于人类对道德规范的抽象提炼, 其“道德判断”本质是对预设程序的机械执行, 缺乏人类基于实践经验与情感认知的自主反思能力。因此, 必须确立人类在道德领域的核心地位, 将技术发展严格限定在服务人类整体利益与全面发展的框架内。在制度设计层面, 需要构建以人类价值为核心的伦理规范体系^[11]。参考康德“人是目的而非手段”的哲学理念, 制定具有普遍约束力的人工智能伦理原则, 强调技术应用要尊重人类的尊严与权利, 明确机器道德的应用边界。依据“权责一致”原则完善法律制度, 建立技术责任追溯机制, 将开发者、使用者的道德义务纳入法律规制范围, 确保技术应用符合人类社会的基本道德共识。同时, 对技术影响下人类道德主体性的潜在变化进行深入分析, 从根源上维护人类在道德实践中的主导权。

2) 推动技术创新与道德进步协同

生产力与生产关系、经济基础与上层建筑的辩证关系, 这一原理同样适用于理解技术创新与道德进步的互动逻辑。技术作为生产力的重要组成部分, 其发展必然引起生产方式的变革, 进而影响社会的道德观念; 反之, 道德作为上层建筑的构成要素, 通过价值导向对技术创新形成约束与引导。从理论构建角度, 应建立技术伦理的跨学科研究范式。结合“己所不欲, 勿施于人”等传统道德理念, 以及现代社会倡导的公平、正义等普适价值, 为机器“道德”能力的算法设计提供价值基础。运用系统分析方法, 将技术与道德演进视为相互作用的整体, 通过梳理两者的关联因素, 揭示协同发展的内在规律。同时, 从社会批判的角度出发, 审视技术创新中可能存在的不合理因素, 避免技术被不当利用, 确保技术进步与道德发展相互促进。

3) 构建人机共生的道德生态

历史唯物主义强调人与自然、人与社会的辩证统一, 在智能时代, 这种统一性体现为人机关系的重构。从沟通与协作的角度来看, 人机共生的道德生态本质上是人类与机器在道德实践领域的新型合作关系。这种关系既非人类对机器的单向控制, 也不是技术对人类的替代, 而是基于功能互补的协同发展模式。在理论层面, 需要构建人机伦理的元理论框架。借鉴“物尽其用, 人尽其才”的理念, 重新审视机器道德的本质, 明确其工具属性与价值限度。运用辩证思维, 分析人机在道德实践中的矛盾与统一, 探讨两者协同进化的可行路径。通过对人机交互过程的深入思考, 总结其中的规律与特点, 为构建新型道德规范提供理论依据。最终形成以人类价值为核心、以技术理性为支撑、以伦理规范为保障的人机共生道德理论体系, 推动人类社会向更高层次的道德文明演进。

5. 结语

在历史唯物主义视域下探究机器道德的本质, 它既不是技术自主演化的产物, 亦非独立的道德主体, 而是人类基于生产生活需求, 通过技术手段将道德准则编码为机器功能的实践结果, 其发展始终受制于生产力水平与生产关系的辩证运动。

从本质属性看, 机器道德具有鲜明的工具性特征。它是人类将“保障生命安全”“维护公平正义”等

道德共识转化为算法规则的技术投射, 例如无人驾驶系统的决策逻辑、智能家居的隐私保护机制, 均源于人类预设的价值框架, 而非机器自主意识的产物。马克思主义关于“劳动资料是人类劳动力延伸”的论断在此获得当代诠释: 算法、算力构成的技术体系, 成为数字时代人类道德实践的新型工具, 但其功能实现依赖于人类对道德规则的抽象提炼与程序输入。机器虽能模拟道德行为, 却无法理解道德情境中的情感张力与价值冲突——这种基于从历史维度看, 机器“道德”能力的演进呈现清晰的阶段性逻辑。工业革命时期, 机器聚焦物质生产, 尚无伦理赋能需求; 信息技术革命后, 人机交互深化催生隐私保护等初级“道德”功能; 人工智能时代, 算法决策对社会生活的深度介入, 使其成为技术伦理研究的核心议题[12]。这一过程契合恩格斯“理论思维是历史产物”的判断, 表明其内涵随生产力发展与社会道德观念变迁而动态调整。在实践中, 机器道德的局限性集中体现为技术工具理性与人类道德主体性的矛盾。一方面, 机器缺乏情感体验与价值反思能力, 无法在复杂道德情境中进行自主判断; 另一方面, 其应用受技术垄断, 可能加剧资源分配不均, 暴露“技术中立论”的虚幻性[13]。这些局限本质上表明, 机器道德是人类道德的功能性延伸, 而非替代性存在, 其价值实现依赖于人类对技术的合理规制。

展望未来, 随着量子计算、脑机接口等技术的发展, 机器道德可能呈现更复杂的形态, 但其本质仍需置于历史唯物主义框架中考察——技术形态的迭代不会改变其作为人类道德延伸的属性。机器可以拓展道德实践的边界, 却无法取代人类在价值创造与意义生成中的核心作用。唯有坚持马克思主义的方法论, 将技术进步纳入“人的自由全面发展”的轨道, 才能在人机共生中构建真正符合人类文明进步方向的道德生态。

参考文献

- [1] 《马克思恩格斯全集》第26(I)卷[M]. 北京: 人民出版社, 1972: 15.
- [2] 《马克思恩格斯全集》第23卷[M]. 北京: 人民出版社, 1972: 204.
- [3] 《马克思恩格斯全集》第5卷[M]. 北京: 人民出版社, 2009: 875.
- [4] 《马克思恩格斯全集》第1卷[M]. 北京: 人民出版社, 1972: 273.
- [5] 《马克思恩格斯文集》第9卷[M]. 北京: 人民出版社, 2009: 99-100.
- [6] Gartner, C.B. (1983) Rachel Carson. Frederick Unger Publishing, 120 p.
- [7] 刘云江, 李保元. 论无人驾驶汽车强制责任险[J]. 财会月刊, 2020(24): 130-138.
- [8] 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(6): 65-75.
- [9] 李猛, 谭九生. 智能家居的伦理风险、成因透析及治理建议[J]. 宁夏社会科学, 2024(1): 104-113.
- [10] 黄欣荣. 新一代人工智能研究的回顾与展望[J]. 新疆师范大学学报(哲学社会科学版), 2019, 40(4): 86-97.
- [11] 《学术前沿》编者. 算法治理与科技向善[J]. 人民论坛·学术前沿, 2022(10): 18-19.
- [12] 陈姿含. 人工智能算法决策中的敏感个人信息保护[J]. 法律科学(西北政法大学学报), 2024, 42(6): 63-75.
- [13] 郭冲辰, 陈凡, 樊春华. 论技术的价值形态与价值负荷[J]. 自然辩证法研究, 2002(5): 37-39, 57.