

人工智能与伦理问题的马克思主义分析： 技术理性与人的解放的辩证

孙悦

新疆大学马克思主义学院，新疆 乌鲁木齐

收稿日期：2026年3月9日；录用日期：2026年3月29日；发布日期：2026年4月9日

摘要

人工智能作为当代科技革命的前沿领域，正通过数据化、算法化与平台化重构人类社会的生产范式、社会关系与价值秩序。在推动生产力革命性跃升的同时，也引发了主体性遮蔽、数据权属失序、算法偏见固化与技术异化等一系列伦理危机，暴露了技术理性扩张与人类解放目标之间的价值断裂。本文基于马克思主义伦理学批判范式，运用历史唯物主义方法论与异化劳动理论，系统揭示人工智能伦理困境的生成逻辑：在资本主义生产方式下，技术理性被资本逻辑异化为增殖工具，导致技术发展与人本价值的结构性背离。在此基础上，主张重建以人为本的技术伦理观，探索具有社会主义制度优势的人工智能治理路径。研究旨在为新时代人工智能伦理治理提供理论支撑与实践方向，推动技术发展真正服务于人的全面发展与社会整体进步。

关键词

人工智能，伦理问题，技术理性，人的解放

A Marxist Analysis of Artificial Intelligence and Ethical Issues: The Dialectic of Technological Rationality and Human Liberation

Yue Sun

School of Marxism, Xinjiang University, Urumqi Xinjiang

Received: March 9, 2026; accepted: March 29, 2026; published: April 9, 2026

Abstract

As a frontier field of the contemporary technological revolution, artificial intelligence is reshaping human society's production paradigms, social relationships, and value systems through digitization, algorithmization, and platformization. While driving a revolutionary leap in productivity, this has also triggered a series of ethical crises—including the obscuring of agency, the disorder of data ownership, the entrenchment of algorithmic bias, and technological alienation—exposing a value disconnect between the expansion of technological rationality and the goal of human liberation. Drawing on the critical paradigm of Marxist ethics and employing the methodology of historical materialism and the theory of alienated labor, this paper systematically reveals the underlying logic behind the ethical dilemmas of artificial intelligence: under the capitalist mode of production, technological rationality is alienated by the logic of capital into a tool for accumulation, leading to a structural divergence between technological development and human-centered values. Building on this foundation, we advocate for the reconstruction of a people-centered ethical framework for technology and the exploration of AI governance pathways that leverage the strengths of the socialist system. This research aims to provide theoretical support and practical guidance for AI ethical governance in the new era, thereby ensuring that technological development truly serves the all-round development of individuals and the overall progress of society.

Keywords

Artificial Intelligence, Ethical Issues, Technological Rationality, Human Liberation

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

中国高度重视人工智能发展，积极推动互联网、大数据、人工智能和实体经济深度融合，培育壮大智能产业，加快发展新质生产力，为高质量发展提供新动能。人工智能技术正以前所未有的速度深度嵌入人类社会的各个维度，技术对社会结构和人类生活方式的影响早已突破工具层面，进入到价值秩序与伦理结构的重构阶段。在这一转型进程中，人工智能带来的不仅是效率提升与便捷体验，更伴随着日益复杂的伦理挑战和人文危机。数据安全、算法科学性、劳动异化、主体边界模糊等问题频繁出现，既暴露出现代技术理性的工具化局限，也揭示了资本逻辑在技术体系中的决定性作用。

尽管主流人工智能伦理讨论多以风险规避、规范制定与道德原则建构为核心议题，但这些路径大多停留于经验层面的调节，缺乏对问题根源的系统揭示与解决。伦理危机不断加深，表明人工智能的治理不能仅依赖于技术层面的突破或规范层面的制度建构，重要的是解决社会制度结构与意识形态机制的深层矛盾。

马克思主义伦理思想是以唯物史观为理论指导，并以人的自由全面发展为价值追求的伦理思想。“劳动者利用物的机械的、物理的和化学的属性，以便把这些物当作发挥力量的手段，依照自己的目的作用于其他的物。”^[1]从这一视角出发，人工智能所引发的伦理问题并非偶然，而是资本主义生产方式与技术理性相结合所导致的必然结果。人工智能不仅是一种生产力工具，更是一种社会治理机制，必须以马克思伦理思想的视野中加以深度分析与重构。

在此基础上，本文主张将马克思主义作为理解与改造人工智能伦理问题的根本理论，重申“人的解放”在智能社会构建中的核心地位。同时，结合中国社会主义制度优势，探索人工智能治理中的伦理路径。

2. 人工智能的技术逻辑与伦理挑战

人工智能的发展不仅是一场科技革命，更是一次深刻影响人类社会结构与价值系统的变革。它以超越以往技术的智能性、预测性和适应性嵌入社会运行的各个环节，使传统的社会规则、伦理规范与主体边界面临新挑战。与此同时，人工智能的发展路径并非纯粹技术性的演进，而是在特定社会经济结构特别是资本主义逻辑主导下生成的复杂结果。伦理困境由此不仅表现为局部道德风险，更体现为技术理性主导下的结构性张力。理解人工智能的伦理挑战，必须从其技术逻辑、现实表现与价值取向三个层面展开，进而为马克思主义视域下的批判与重构提供理论支撑。

2.1. 人工智能的技术特征与演进路径

人工智能作为集数学建模、算法计算与数据集成于一体的综合性技术体系，其发展轨迹不仅体现为算力提升与算法迭代，更深刻地改变了知识生成、社会协同与主体认知的基本机制。从专家系统的逻辑推理，到深度学习对非结构化数据的处理，再到当前大语言模型在自然语言生成与人机交互中的广泛应用，人工智能已经从初级工具演化为在特定领域具备代理能力的“准主体”。其核心特征包括：以庞大的数据驱动为前提，通过复杂的模型结构实现自主学习与动态反馈，同时具有决策过程高度不透明的“黑箱”属性。这些特征使得人工智能在社会生活中不仅执行任务，更在塑造规则、筛选信息与调节行为中发挥主动作用，逐步改变人类在社会中的位置与角色，从而埋下伦理结构被技术逻辑重构的深层种子。

2.2. 人工智能引发的主要伦理问题

随着人工智能在社会运行中深度渗透，其带来的伦理问题也逐步显现并日益复杂化，主要表现在主体性削弱、隐私权侵犯、算法歧视与技术异化四个层面。在智能推荐与个性化服务中，个体选择被系统预测与调控，人类的认知主动性与判断力面临被算法规训的风险，体现为自动化支配下的主体退场。同时，人工智能系统依赖对个体行为数据的长时间收集与深度分析，用户在被建模的过程中丧失了对自身数据的处置权与知情权，形成典型的数字不对称权力结构。数据的收集与使用过程中包含大量的数据流动与数据交易，用户对自身数据的控制与管理逐步减弱，隐私泄露逐步加剧为恶意攻击、信息骚扰等安全隐患。此外，AI系统由于训练数据的历史偏见而复制甚至放大现实社会中的性别、种族、阶层歧视，使不平等以技术中立的名义合法化并固化。而在劳动力市场层面，大量中低端岗位被智能系统替代，形成“高技术集中 - 低劳动贬值”的结构性失衡，人工智能不仅没有缓解异化劳动，反而将其进一步技术化与系统化。这些问题不仅指向个体层面的道德困扰，更显现出整个人类社会面临的系统性伦理危机。

2.3. 从技术理性到工具理性批判

人工智能伦理危机的深层根源在于技术理性在当代社会中逐步异化为资本工具理性，它以效率最优化、可计算性与控制性为唯一目标，在决策逻辑中排除了价值判断与人文考量，导致技术系统对社会价值的侵蚀性主导。正如韦伯所指出的现代工具理性会压缩目的理性的空间，人工智能的发展越来越呈现出手段主导目的的结构倾向。当“最优解”被限定为经济效益最大化或资源配置效率提升时，道德权

利、社会公正与人类尊严等价值维度被技术排除在系统之外。“要知道，机器之所以存在，在机器上之所以投入这样大量的资本，恰恰是为了通过这一资本来榨取工人的劳动。”^[2]在资本主导下，这种理性被进一步强化为利润逻辑的执行机制，技术成为资本增值与劳动力控制的利器，伦理问题也由此转化为政治经济问题。这种“价值冷漠”的技术建构方式不仅使伦理规范失效，更使社会发展失去方向感。因此，人工智能的伦理问题不能仅以规范建构与道德修补加以应对，而必须从批判现代技术理性的哲学基础入手，反思其社会建构过程中的权力结构，从而为技术的价值转向开辟可能性。

3. 马克思主义视域下人工智能伦理问题的根源分析

人工智能伦理问题的生成，并不仅仅是技术快速发展引发的制度滞后或道德空缺，更深层次地根植于特定社会关系和生产方式中的结构性矛盾。主流伦理治理框架往往着眼于制度改进或风险规避，忽视了技术发展的社会动力与历史根源。马克思主义作为揭示社会基本矛盾与人类解放路径的科学理论，为分析人工智能伦理困境提供了根本的视角。从生产力与生产关系的张力、技术异化与人的物化过程，到人的全面发展理念的遮蔽，人工智能的发展逻辑在资本主义条件下被深度嵌套进利润驱动与控制机制之中，导致技术理性系统性地背离了人的价值理性。本章将基于马克思主义基本原理，对人工智能伦理问题的根源进行理论剖析，揭示其表象背后的制度性诱因与意识形态遮蔽机制。

3.1. 生产力与生产关系矛盾中的 AI 技术

人工智能作为新质生产力的重要代表，其技术属性与应用路径深刻嵌套于当代资本主义社会的生产关系之中，呈现出典型的“工具嵌套结构性矛盾”现象。从马克思主义生产力与生产关系矛盾运动的理论出发，技术发展原本应是推动社会历史进步的重要力量，但在资本主义生产关系下，技术往往被锁定为资本增值的手段，其解放潜力被异化为压迫机制。在人工智能场域中，这一矛盾表现得尤为突出：一方面，AI 大幅提升生产效率，推动自动化和智能化转型；另一方面，资本将其掌控于平台巨头与跨国公司手中，使劳动力被边缘化、劳动价值进一步贬低，新的阶级分化与数据权力不平等愈加严重。AI 不仅没有打破人与人之间的不平等关系，反而在某些领域强化了生产关系的封闭性与等级性，导致技术与之间的矛盾日益显性，伦理问题由此根植于结构失衡之中。

3.2. 资本主义条件下的技术异化与人的物化

马克思在《1844 年经济学哲学手稿》中指出，资本主义制度下的劳动异化不仅剥夺了工人对劳动产品的控制权，也使劳动过程本身异化为对人的压迫。在人工智能时代，这种异化不仅没有消解，反而以更隐蔽、更系统化的方式重现。AI 系统通过算法控制与行为预测，将人的行为变得可量化、可测量、可分类，使人沦为数据提供者与算法对象，其主体性被剥夺。这是一种新的“数字异化”形态，人工智能将人从生产过程的中心驱逐，转化为被技术调度的资源或变量，其存在意义仅在于满足算法系统的最优化逻辑。在“人 - 人工智能 - 自己”的关系中，可能出现人与人工智能之间主客体关系的异化；在“人 - 人工智能 - 他人”的关系中，可能出现人与他人之间的信任危机，这些人际层面的伦理风险直接影响了人们社会生活的体验感。与此同时，平台企业通过技术封装将价值判断内嵌在代码与模型之中，使价值通过技术路径完成再生产，人在其中不再是伦理主体，而是被动接受的对象。这种物化过程削弱了人的主体性和独立想法，也构成了现代技术伦理危机的深层土壤。

3.3. 人的全面发展与人工智能的悖论

马克思主义始终以实现人的自由与全面发展为根本价值追求，技术理应作为解放人类、缩短劳动时

间、拓展创造空间的工具。“自主性即自由意志，是人之为人的规定，指向具有独立人格的个人遵循自己的内心展开一系列活动的自在自为的状态。人的自主性以自在自为的活动为载体，而自在自为的活动则是基于人的意识，人的意识又以大脑对物质世界所反馈的各种信息的处理为基础。”^[3]然而，在现实的人工智能应用中，这一目标被严重遮蔽甚至背离。AI的发展并未普遍促进人的发展条件，相反，在全球范围内制造了新的数字鸿沟、文化依赖与能力被动等问题。人工智能基于其庞大的数据库、精确的复杂算法和独特的预训练语言模型，可以为用户提供他们所需要的信息，且让人们认为人工智能更加的便捷高效，从而开始长期地依赖人工智能并让渡自己的决策能力。这不仅意味着用户信息来源单一化，同时意味着人工智能在逐渐脱离被动的工具向能动的“主体”转变。这种悖论根植于当代社会将发展简单化为经济增长与技术迭代的单向逻辑，忽视了发展的价值目标应当是增强人的自由、自主与合作能力。当技术发展脱离人的发展而成为自身目的时，技术理性便会压倒伦理性，导致发展的主体人类本身沦为工具性存在。这一悖论唯有确立以人的自由为目标的社会伦理秩序方可破解，也为马克思主义重申“人的解放”提供了新的理论。

4. 人工智能伦理治理的马克思主义路径探索

人工智能所引发的伦理危机，实质上是资本主导下技术理性膨胀对人的主体性与社会公共价值的侵蚀。马克思主义强调以人为本、注重整体性、追求社会公平，其伦理立场和制度构想为人工智能的治理提供了强有力的理论基础和现实路径。在中国，人工智能治理应坚持以人民为中心的伦理原则，强化公共权力对技术的引导与规制，推动技术理性向价值理性转化。同时，应充分发挥思想政治教育在价值塑造、责任意识培育与批判性思维训练中的独特功能，构建具有中国特色的智能社会伦理秩序与治理范式。

4.1. 重建以人为本的技术伦理观

我们要未雨绸缪，加强战略研判，确保人工智能安全、可靠、可控坚持以人为本、智能向善。面对人工智能在多个领域介入人类生活与社会治理的趋势，亟需从根本上重建以人为本的技术伦理观，确立“技术为人”的本体性价值立场。这种伦理观强调技术发展的出发点和归宿应是人的自由、尊严与全面发展，拒绝将效率、控制与收益作为评价AI发展的唯一标准。相较于传统以规则约束和风险预防为核心的“工具性伦理”，马克思主义所倡导的本体性伦理更关注人的现实处境与主体生成，强调技术在实践中必须回应人的现实需求与精神诉求，要实现从以物为本向以人为本的价值转向。

4.2. 构建社会主义制度优势下的人工智能治理体系

中国特色社会主义制度为应对人工智能伦理挑战提供了现实的体制保障与政策空间。任何伦理准则都无法自我执行，需要借助于一系列相互配合的运作机制才能得到落实。伦理准则、伦理审查、伦理教育、伦理监督和伦理问责相互配合，以确保技术应用符合道德规范，并对违法行为进行惩戒和追究我国在治理理念上呈现出了“促进-安全-促进与安全并重”的变化路径，在治理主体上则高度重视由政府单向治理转向共建共治共享的治理格局。在具体制度设计上，应从以下方面推进：

第一，探索数据产权的公有与共享机制。在数据产权制度设计上，应超越传统私有化或完全公有的二元思维，构建国家所有、分层授权、收益共享的制度框架。具体而言，可将涉及公共利益的高价值数据界定为公共数据资源，由国家设立数据信托机构统一管理，企业可在合规前提下通过数据使用许可获取开发权，其产生的社会效益反哺公共财政。同时，鼓励平台企业通过数据合作社等形式，赋予用户对其个人数据的集体处置权与收益分配权，探索数据价值的民主化共享路径。

第二，建立国家主导的算法审计机制。算法审计应包含以下核心标准与程序：一是合规性审查，重

点审查算法是否存在歧视性、操纵性与不透明性,评估其是否符合《生成式人工智能服务管理暂行办法》等法规要求;二是社会影响评估,引入第三方评估机构,对高风险 AI 系统进行模拟测试与动态监测,评估其在性别、种族、阶层等维度上的公平性;三是可解释性标准,要求算法开发者提供可理解的技术文档与决策逻辑说明,确保受影响的个体享有知情权与申诉权;四是审计结果公开,建立算法审计结果公示制度,接受社会监督。此外,充分发挥国有企业与公共科研机构在伦理创新中的示范作用,通过公共算法平台与开源机制促进技术大众化,推动人工智能治理从市场主导转向国家引导与人民参与相结合的协同治理格局。

第三,推动思想政治教育与计算机伦理工程教育的深度融合。在高校教育中,应构建“思政 + 技术”的交叉课程体系,将马克思主义伦理思想、技术社会学与算法设计、软件工程等专业课程有机结合。具体路径包括:在计算机专业中增设技术伦理与工程设计必修模块,采用案例教学、项目制学习等方式,引导学生在算法设计、数据处理等环节主动嵌入伦理考量;在思政课程中引入人工智能伦理议题,通过技术 - 社会互动视角,帮助学生理解技术发展中的权力关系与价值冲突;鼓励高校设立人工智能伦理与治理辅修专业,培养具备技术能力与伦理意识的复合型人才。

4.3. 加强思想政治教育在人工智能伦理中的作用

人工智能背景下,自然语言处理与人工智能的融合变革了人类知识生产的方式,并带动了跨学科融合与创新方法的变革,人机协同生产和传播知识的流程发生了变化。思想政治教育在人工智能伦理治理中不仅是价值传播的通道,更是价值生成与伦理自觉建构的重要机制。首先,思政教育能够有效提升公众特别是青年群体的技术伦理意识,帮助其理解人工智能所隐含的社会风险与权力关系,防止陷入技术崇拜与价值虚无。其次,思政课程可借助 VR、AR、数字孪生等智能技术创设沉浸式、立体化的数智化教学场景教学[4],强化学生对 AI 时代人的主体性地位、权利与责任的认知,引导其形成基于人的尊严和社会正义的价值判断能力。此外,高校应构建交叉学科范式:设立人工智能伦理交叉学科,培养既懂技术又通伦理、法律的复合型人才,推动人文学科与工科教育融合,培养具有道德责任感和批判性技术素养的学生。更重要的是,思想政治教育要发挥“立德树人”的根本任务,加强对青年大学生获取信息的价值目标引导,引导青年将技术能力与社会使命有机统一,在“技术 - 人 - 社会”的张力中形成自觉而坚定的价值立场。

4.4. 国家主导模式的反思与制衡机制

在强调社会主义制度优势的同时,也应清醒认识到国家主导模式下人工智能治理可能存在的潜在风险。主要表现在:一是行政主导的伦理审查体系若缺乏独立性与专业支撑,可能沦为形式化程序;二是监管机制若滞后于技术迭代,可能出现监管俘获现象,即被监管对象反向影响监管规则;三是过度的政府干预可能削弱市场活力与社会自治空间,影响技术发展的多样性与适应性。

为此,应在治理体系设计中嵌入以下制衡机制:其一,设立独立于行政系统的国家人工智能伦理委员会,由技术专家、伦理学者、法律专家与社会公众代表共同组成,对重大 AI 应用项目进行伦理审查与公共听证;其二,引入监管沙盒机制,在可控范围内允许企业进行技术试验,同时设置伦理底线与动态监测,实现创新激励与风险防控的平衡;其三,强化社会监督与公众参与,建立算法审计结果公开制度,畅通公民对 AI 应用的申诉与救济渠道,形成政府、市场、社会三方协同的多元治理格局。

4.5. 推动全球视野下的人工智能伦理话语权建构

世界各国聚焦人工智能发展前沿问题,深入探讨人工智能快速发展条件下教育发展创新的思路和举

措，凝聚共识、深化合作、扩大共享，携手推动构建人类命运共同体。人工智能伦理治理不仅是技术管理问题，更是全球意识形态博弈中的价值较量。当前，西方国家主导的 AI 伦理话语体系往往隐含新自由主义价值取向，将技术中性、市场优先与个人主义合法化，忽视对结构不公与资本控制的反思。在此背景下，中国应以马克思主义为理论根基，提出具有自主性与普遍性的人工智能伦理治理方案，构建体现社会主义核心价值观的全球技术伦理话语权。人工智能发展为了人民群众，人民是历史的创造者[5]。一方面，以多边主义为基石，深度参与国际伦理规则制定，推动形成包容多元文明底色的 AI 治理共识。中国应主动作为，在联合国教科文组织《人工智能伦理建议书》框架下，联合发展中国家提出“技术普惠性”“文化敏感性”等修订议案，反对将单一文明价值观强加于全球伦理标准。具体可借鉴 IEEE《道德化设计自主系统国际标准》制定经验，推动成立跨文明伦理工作组，针对算法歧视、数据主权等共性议题，建立包含儒家“仁爱”、非洲“乌班图”等多元文明要素的伦理评估指标体系。同时，依托 OECD 人工智能政策观察站，推动建立“伦理-技术-法律”三轨联动的国际审查机制，确保高风险 AI 应用接受多边伦理委员会的合规性审查。另一方面，以自主创新为杠杆，强化中国在全球伦理治理中的知识供给与规则引领能力。针对当前“技术领先国主导伦理话语权”的失衡格局，中国需从两个维度突破：一是科研平台赋能伦理工具开发：支持鹏城实验室、之江实验室等国家级平台研发“伦理影响评估沙盒”，通过模拟不同文明场景下的算法决策，量化文化适应性指标，破解“西方中心主义”伦理工具的适用性困境；二是人工智能伦理的政策应从生产源头企业角度率先出发，出台政策鼓励企业采用隐私保护技术，如差分隐私、联邦学习等，减少数据泄露风险。推动百度、deepseek 等头部企业将隐私计算、算法可解释性等中国技术方案转化为国际标准提案，实现伦理规则与技术方案的同步输出。

5. 结语

人工智能的快速发展，尤其是生成式 AI 的广泛应用，已推动伦理问题从技术边缘走向社会治理核心。当前主流伦理治理模式多聚焦风险管控与行为约束，虽能缓解短期矛盾，但难以触及技术理性背后的制度性异化根源，本质上仍是资本主义工具理性下人的异化在数字时代的再现。这一矛盾源于生产关系与科技发展之间的结构性张力，表现为技术逻辑对伦理价值的隐性侵蚀。本文基于马克思主义伦理学与历史唯物主义，系统剖析人工智能的技术逻辑、伦理困境与意识形态机制，提出以人的自由全面发展为核心的技术伦理重建路径。具体而言，需依托中国特色社会主义制度优势，构建国家主导、社会协同、公众参与的多元治理体系，推动技术正义与伦理正义的统一。同时，高校思想政治教育应主动介入 AI 伦理教育，通过提升青年群体的伦理意识与批判能力，培育其公共责任感与人类命运共同体意识。

未来对于人工智能与伦理的研究需聚焦三大方向：一是构建跨学科伦理研究框架，融合马克思主义伦理学与技术哲学、法学、教育学；二是推动 AI 伦理在公共政策中的制度化落地，促进伦理共识的社会化生成；三是深化思想政治教育与技术伦理教育的协同机制，打造中国特色智能社会的德育新范式。通过解决这些问题，才能引导人工智能技术服务于人的解放与社会公正，构建价值理性主导的人机共生未来。

参考文献

- [1] 马克思恩格斯文集(第 5 卷)[M]. 北京: 人民出版社, 2009: 209.
- [2] 马克思恩格斯文集(第 8 卷)[M]. 北京: 人民出版社, 2009: 293.
- [3] 杨礼银, 李海艺. 论人工智能对人的主体性的冲击及化解路径——基于马克思机器论视角的考察[J]. 云南大学学报(社会科学版), 2024, 23(4): 5-13.

-
- [4] 郑泉. 生成式人工智能的知识生产与传播范式变革及应对[J]. 自然辩证法研究, 2024, 40(3): 74-82.
- [5] 肖福赞. 人工智能驱动高校思政课教学改革的内在机理、风险挑战与应对之策[J]. 电化教育研究, 2025, 46(5): 103-107, 115.