

从“控制”到“共生”：行动者网络理论视角下人机关系的伦理反思

郭佳楠

(郑州工商学院马克思主义学院, 河南郑州 450000)

摘要: 随着人工智能、算法系统及智能代理的快速发展, 人机关系正从传统的单向控制逐步演进为多维共生。通过对行动者网络理论的概述及其在人机关系重塑中适用性的剖析, 进一步指出人机关系演进过程中存在着自主性与依赖性悖论、责任归属错位、可解释性与信任断裂, 以及效率与价值冲突等核心伦理问题。为此, 行动者网络理论为发展良性的人机关系提供了一条新的路径: 即保持人类行动在技术网络中的持续在场, 确保人机互动的生成轨迹在网络结构中得以持续显现, 维持人类经验与技术转译之间的循环关系, 积极将伦理原则嵌入进技术行动链条内部, 以期实现技术的能动性与社会价值在网络化互动中的同步协调, 进而明确人机共生不仅是技术优化的目标, 更是伦理与治理的必然要求的结论, 为未来社会中智能系统的可持续发展提供理论与实践参考。

关键词: 人机关系的伦理困境; 行动者网络理论; 控制与共生; 价值共识; 权力结构

1 引言

当代人工智能的迅猛发展正深刻重塑人类社会的认知结构与行为方式。从早期以算法驱动的“工具理性”阶段, 到如今以生成式大模型为核心的“智能共创”阶段, 人机关系的形态正经历从“单向控制”到“协同共生”的范式转移。在此背景下, 行动者网络理论 (Actor-Network Theory, ANT) 提供了一种重要的分析视角。该理论强调技术、制度、符号、算法等非人要素同样具备行动能力 (agency), 并与人类共同构成复杂的异质网络。人机关系不再被理解为“主体—客体”的

单线结构, 而是被重新界定为多行动者共构的动态网络关系。由此, 人工智能不只是“被使用”的对象, 而是与人类共同行动、共同生成意义与价值的“伙伴性存在”^[1]。

随着人工智能、物联网和大数据等技术的深度融合, 人机关系研究逐渐成为科技哲学与伦理学领域的重要议题。早期相关研究主要受勒内·笛卡尔 (René Descartes) 二元论与现代技术哲学影响, 马丁·海德格尔 (Martin Heidegger)、卡尔·海因里希·马克思 (Karl Heinrich Marx) 与雅克·埃吕尔 (Jacques Ellul) 等思想家均在

作者简介: 郭佳楠, 郑州工商学院马克思主义学院讲师, 研究方向为科学技术与公共政策。

项目来源: 郑州工商学院科研创新项目: 基于 ChatGPT 类生成式人工智能的红色文化传播现状与策略研究 (项目编号: 2023-KYZD-01); 2023 年度河南省社科联调研课题“当代新兴 Chat GPT 技术的伦理风险及其治理路径研究” (项目编号: SKL-2023-252)。

不同程度上揭示了技术工具化的逻辑及其异化风险，认为技术的发展虽可提升生产效率，却可能导致人的主体性被技术体系吞噬^[2]。进入21世纪，随着生成式人工智能和智能体技术的崛起，人类与机器之间的界限日益模糊。在此背景下，行动者网络理论成为理解这一转变的重要理论框架。该理论由布鲁诺·拉图尔（Bruno Latour）、米歇尔·卡隆（Michel Callon）与约翰·劳（John Law）等人提出，主张应打破传统人类中心主义，承认非人行动者（包括技术、制度、算法等）在社会系统中的能动作用^[3]。近年来，人工智能伦理研究者逐渐将ANT视角引入人机共生关系分析^[4]，尝试解释算法如何在网络中具备“道德位置”与“行动权能”^[5]。马克·科克尔伯格（Mark Coeckelbergh）则进一步提出“关系性伦理”框架，认为伦理应关注人机关系的互动过程，而非仅限于责任主体的归属划分^[6]。尽管研究成果丰富，但现有研究多数仍停留在“工具理性”或“控制风险”的层面，未能充分揭示人机共构关系中的动态伦理生成逻辑及应对路径^[7]。基于此，本文拟在对既有文献的系统梳理基础上，运用行动者网络理论重新审视人机关系的伦理结构，从“单向控制”向“人机共生”的转型中揭示技术能动性与人类责任的交互逻辑，进而为构建具有共治意识与共生精神的人工智能伦理体系提供新的理论支持与现实启迪。

2 控制的悖论：人机关系演进中的伦理困境

在人机关系的演进过程中，“控制”作为技术理性与伦理反思的交汇点，始终贯穿于人类与机器互动的全过程。然而，随着人工智能、算法学习与自主决策系统的不断演化，机器的“他者性”与“能动性”逐渐显现，人类对技术的绝对控制正在被动摇。控制不再是单向度的命令与执行关系，而成为一种多维的伦理博弈过程：人在试图掌控机器的同时，也在被机器的逻辑、算法与数

据规则所“反向规训”。这种控制关系的模糊化与动态化，折射出人机关系从支配到协作、从外在控制到内在耦合的深层转型。

2.1 自主性与依赖性的悖论：控制的边界模糊化

在人机关系的演进过程中，“自主性”与“依赖性”的张力构成了控制伦理的首要悖论。传统控制逻辑建立在人类中心主义基础上，将机器视为人类意志的延伸与工具性存在。然而，随着人工智能、大数据与具身计算等技术的发展，机器逐渐具备了感知环境、分析信息、生成策略与自我优化的能力，表现出一定程度的“行为能动性”。这种能动性虽然源于人类设定的算法规则，却在复杂的学习与反馈过程中生成了超越人类预期的行动模式。人类对机器的控制权因此不再稳固，技术从“被动执行者”演变为“参与性行动者”。行动者网络理论指出，社会系统由多种人类与非人行动者共同构成，权力与行动不是被垄断的，而是在网络互动中动态分配的^[8]。

与此同时，这种技术自主性的增强并未带来人类主体性的强化，反而加深了人类的“依赖性”。从算法推荐到自动决策系统，人类在追求便利与效率的过程中，逐步让渡了思考与判断的权力，将自身的认知活动“外包”给技术系统。例如，在智能导航、自动翻译或生成式人工智能创作中，人类可能接受算法输出而非自主检验，形成“技术信赖依附”的心理结构。依赖的加深削弱了个体的能动判断，使人类陷入“被控制的自由”，即表面上掌控技术，实则受制于技术逻辑的操控与引导。技术以“服务”的名义重塑了人的行为模式与价值取向，使控制关系从“主控”转向“互控”，再到“隐性规训”^[9]。

2.2 责任归属的悖论：控制权与责任权的错位

在人工智能深度融入社会运作的时代，人机关系中的控制权与责任权之间出现了前所未有的错位现象。传统伦理学与法理学假定技术仅是人类意志的延伸，因而“控制者即责任者”成为普遍认可的原则。然而，智能算法具备自我学习与

自我演化的能力，技术系统的决策过程逐渐超出人类的直接掌控，控制权呈现出多主体分布与动态转移的特征。人类设计者通过算法设定影响技术的初始方向，用户通过交互输入塑造系统的运行逻辑，而 AI 系统本身则在庞大的数据流中形成具有一定独立判断能力的反馈机制。于是，原本清晰的责任分界被打破——控制的主体不再单一，责任的归属也不再可被线性追溯。人机共构的行动网络在提升效率与智能水平的同时，也带来了伦理上的责任缺失，使得传统基于意志与行为的责任归属不明确。

随着控制权的多元化与技术系统的自主化，人类的道德判断能力在责任体系中被进一步边缘化。AI 决策过程使人类无法完全理解算法的运行逻辑，进而难以确定行为后果的意图来源与责任根基。当自动驾驶系统、医疗辅助诊断或内容生成算法造成损害时，社会往往陷入“追责无门”的困境——开发者辩称系统行为超出预期，操作者声称仅依赖算法建议，而算法本身又不具备法律人格。在这种情境下，责任的指向变得模糊且不断转移，伦理上的可追溯性被技术复杂性所遮蔽。控制与责任由此出现分离：掌控行为者不再是最最终的承担者，而被控制或间接影响的系统反而在伦理判断中失去了位置。

2.3 可解释性与信任的悖论：控制与理解的断裂

人工智能时代的人机关系，正在从可控的线性逻辑转向复杂的算法逻辑，而“可解释性”成为关键问题。传统技术系统的运行遵循透明的因果链条，人类能够通过逻辑推理或经验理解技术行为，从而保持认知上的“掌控感”。然而，随着深度学习与神经网络模型的广泛应用，AI 系统的决策过程越来越呈现出不透明的特征：算法在庞大的数据集中进行非线性运算，其判断依据和决策路径往往超出人类理解的边界。这意味着人类即便拥有技术上的控制权，也可能失去认知上的解释权。人类与机器之间由此形成一种新的伦理断裂——控制的形式仍在，理解的实质却已消失^[10]。

同时，可解释性的缺失直接引发了“信任的

悖论”。在人机互动关系中，信任是伦理稳定的基础，它建立在人类对技术行为的可预测性与透明度的认知之上。然而，当算法系统的运作逻辑超出人类理解时，这种信任便从“理性信任”转变为“盲目信任”。用户在无法解释算法行为的前提下，仍依赖系统作出决策，形成“信任替代理解”的心理机制。例如，医学影像 AI 在诊断中提供高精度结果，但医生难以说明其决策的具体依据；金融智能推荐系统为投资者提供最佳组合，但算法的偏好取向却不可知。在这些情境中，人类的信任不再基于知识与理解，而是基于对算法权威与数据准确性的信赖。这种“技术信任”表面上提高了人机协作效率，实则削弱了人类对技术的伦理约束与反思能力^[11]。

2.4 效率与价值的悖论：技术理性压制人文伦理

在人机关系的持续演进中，技术理性成为现代社会的重要驱动力，其核心诉求在于提升效率、优化流程与实现资源最大化配置。然而，当效率被奉为最高原则时，人类社会赖以维系的价值理性与伦理规范却被不断边缘化。人工智能、大数据与算法系统以“去价值化”的中立姿态渗透进决策体系之中，使“最优解”替代了“应然性”的价值判断^[12]。另外，效率与价值的冲突不只是生产逻辑的矛盾，更是人类对自身存在意义的再度拷问——当机器以更快、更准的方式完成任务时，人类是否仍能保持对目标与意义的主导解释权？

此外，更深层次的伦理困境在于，技术理性不仅改变了行动方式，也重塑了人类的思维范式。随着算法治理和自动化决策渗透至教育、医疗、公共管理等领域，人类逐渐习惯将复杂的伦理抉择交由系统完成，从而陷入“技术依赖型伦理麻痹”。这种对效率的盲目信任削弱了个体的道德判断力，使伦理反思让位于数据分析，也使公共讨论被算法推荐所主导。行动者网络理论提醒我们，技术并非价值中立的中介，而是具有能动性的“非人行动者”，它通过算法规则与数据逻辑深度参与社会价值的建构。若效率逻辑不受伦理约束，技

术系统的“自我强化”倾向将不断放大工具理性的支配力，使人类从控制者逐步转变为被系统逻辑规训的“数据行动者”。

3 行动者网络理论及其在人机关系重塑中的适用性

在人工智能加速融入人类社会的时代背景下，传统的人机关系理论逐渐暴露出解释力的局限。以“主体—客体”二元划分为基础的分析框架，难以充分揭示人机共生语境中复杂的行动逻辑与伦理关联。为突破这一认知瓶颈，行动者网络理论提供了一种去中心化、关系化的分析视角。正是在这一意义上，行动者网络理论为重新理解人工智能时代人机关系的互动机制与伦理重塑提供了重要的理论支撑。

3.1 行动者网络理论概述

行动者网络理论（ANT）是20世纪80年代兴起于法国科学技术社会学领域的重要理论取向，主要由布鲁诺·拉图尔（Bruno Latour）等学者提出。该理论最初旨在解释科学知识与技术系统的社会建构过程，但其核心思想逐渐突破学科边界，成为理解社会—技术关系与人机共生结构的重要分析工具。行动者网络理论作为后现代科学哲学与社会学的重要理论创新，源于对传统社会科学“人类中心主义”与“结构决定论”的反思。其思想核心在于将社会理解为由多种异质行动者（heterogeneous actants）共同编织的动态网络，这些行动者既包括人类个体，也包括非人要素，如技术装置、算法程序、自然物体、制度规范等。与传统社会理论不同，ANT并不预设主体与客体的本体区分，而强调所有行动者在网络中通过不断的“关联”（association）与“调适”（adjustment）形成社会事实。拉图尔认为，“社会并非一种独立存在的实体，而是一种不断生成的关系过程”^[13]。因此，ANT本质上是一种“生成论的社会观”和“关系本体论”，主张用动态、非中心化的方式理解社会结构的形成机制。

3.2 行动者网络理论在人机关系重塑中的适用性

在人工智能深度嵌入社会结构的背景下，人机关系已不再停留于“工具使用”的层面，而演变为多主体协同参与的复杂情境。ANT以“异质行动者共同构成网络”为基本立场，强调行动能力来源于关系编排而非主体本身，因而为分析人机关系中的控制悖论提供了契合现实结构的理论工具。它既不将技术简化为中性媒介，也不将人类理解为绝对主导，而是通过追踪行动如何在网络中被连接、转译与稳定，揭示控制、责任、理解与价值如何在关系结构中被重新分配。这种视角的引入，使人机伦理问题获得结构层面的解释基础。

就自主性与依赖性的悖论而言，ANT的适用性体现在对“能力来源”的重新阐释。表面看来，个体在智能系统中的决策依赖不断增强，似乎意味着自主性的削弱；但从网络视角观察，个体的行动并未消失，而是借助技术节点获得新的实现路径。理论强调，任何节点的能动性都依赖于网络支持，自主性因此表现为关系性存在而非孤立属性。控制边界之所以模糊，是因为人类与技术共同构成了行动的条件集合，传统主体—客体划分已难以描述这种状态^[14]。该理论正好提供了理解“依赖增强却仍具行动性”这一现象的解释框架，使自主性问题能够在关系维度而非实体维度上获得理论说明。

在责任归属悖论中，行动者网络理论同样展现出高度解释力。智能系统的决策结果往往经历长链条转译，涉及数据供给者、算法设计者、系统维护者、情境使用者等多重参与方，责任难以固定在单一主体上。传统归责模式以行为主体为中心，难以覆盖这种分布式行动结构，而行动者网络理论则认为，责任应当与行动路径相联系，沿着网络关系加以识别。技术装置并非被动执行命令，而是在规则设定与流程限定中实质参与了行动的形成，因此也是责任网络中的一环。这一

理论取向使责任错位现象不再被视为制度漏洞，而被理解为网络化行动的结构特征，进而为重新认识责任的分布形态提供理论支撑。

围绕可解释性与信任、效率与价值两类悖论，行动者网络理论的适用性进一步显现。对前者而言，理论指出理解并非仅来自对算法内部逻辑的掌握，而是在持续互动中形成的关系性认知。界面反馈、制度保障、他人经验与技术表现共同构成信任生成的环境，信任因此是一种网络稳定状态，而非单纯认知结果。这种观点有助于说明为何在算法不可完全透明的条件下，社会仍可能形成可运作的信任结构。对后者而言，技术理性之所以呈现效率优先倾向，是因为在网络建构过程中，量化指标、资本逻辑与管理规范等行动者获得了更强影响力，而人文价值未能同等嵌入关系结构。行动者网络理论通过揭示价值同样可以作为行动者进入网络，说明伦理并非外在约束，而是可以参与行动编排的要素。

4 走向人机共生：人机关系良性发展的路径

当人机关系从“工具理性”走向“智能共生”，传统以“人类中心主义”为基础的控制范式已难以应对智能技术的复杂能动性与伦理挑战。人工智能、算法系统、虚拟代理等“非人行动者”不再是被动的执行对象，而逐渐成为社会行动网络中的积极成员，参与知识生产、价值塑造与决策过程。由此，人机关系的伦理重构不应停留于“谁控制谁”的单向逻辑，而应转向“如何共生共治”的协同逻辑。行动者网络理论为这一转向提供了重要的分析框架——它强调人类与非人技术共同构成动态的关系网络，通过协作、转译与互构实现社会秩序的生成。

4.1 保持人类行动在技术网络中的持续在场

在智能系统深度嵌入社会运行的条件下，自主性问题并非表现为“人被技术替代”，而是体现为行动在网络中被层层委托后逐渐失去可辨识形

态。行动者网络理论认为，技术装置并不是中性工具，而是承载并延展行动的中介物。当算法模型、数据规则、界面逻辑等在系统稳定后转化为“理所当然”的背景存在时，人类行动的作用路径也随之隐形，个体对自身影响决策过程的能力产生疏离感。因此，人机共生关系的关键并不在于恢复单向控制，而在于维持人类行动在转译链条中的持续显现，使人始终作为网络中的有效行动者存在。围绕这一目标，需要在技术结构、运行流程与使用场景中嵌入多层次的人类参与环节，使行动不完全封闭于技术回路之中。

在系统设计与运行层面，可以通过若干方式保持人的参与。其一，在高风险或价值判断密集的决策场景中嵌入人工参与环节，使算法输出需要经过人的情境性判断与确认，避免形成完全自动闭合的决策链条。其二，在模型训练与迭代阶段引入多元主体协同评议机制，由开发者、领域专家及实际使用者共同参与模型调整，使技术演化本身成为持续协商的转译过程。其三，在交互界面提供可调节选项，如推荐偏好设置、排序逻辑选择或风险阈值控制，使用户行为能够反向影响系统结构，而不只是被动接受结果。

在组织与社会应用层面，同样需要维持人类行动与技术中介之间的关联。可以通过持续的使用者培训与情境演练，使操作者理解系统适用范围与局限，从而在关键情境下发挥判断作用；在系统评估与反馈环节收集实际使用经验，使实践中的问题与修正意见进入后续模型更新；在制度层面将人工复核、异常情境报告等环节纳入常态流程，使技术系统始终处于人类的持续关注之下。这些做法并非削弱技术能力，而是保证人类行动在网络中的位置不被替代或遮蔽。

4.2 确保人机互动的生成轨迹在网络结构中得以持续显现

行动者网络理论指出，任何结果都不是单点因果，而是连续转译的沉积形态。数据选择方式、模型设定逻辑、场景部署条件、组织使用习惯等

因素共同塑造行动方向，但当系统运行趋于稳定，这些形成过程被压缩为“系统输出”，转译历史被折叠进黑箱之中。责任困境由此产生，因为行动虽是分布生成的，却被认知为整体自动发生。因此，使行动生成轨迹持续显现，本质上是在网络层面恢复行动来源的可见性，使结果始终携带其生成结构。

在技术运行层面，需要让不同阶段的行动贡献保持结构性存在。数据的采集范围、筛选标准与标注方式应与模型运行持续关联，而不是在进入系统后失去来源标识；模型的版本演进、参数调整与训练环境变化应伴随系统输出保留记录，使每一阶段的转译活动能够被识别；系统部署情境，包括应用场景、使用规则与操作流程，也应与技术结果保持联系。这样，行动并非从“系统内部自然产生”，而是始终呈现为多环节协同生成的状态。转译链条因此不再隐匿，责任不再脱离行动结构而悬置于单一主体。

在组织与制度维度，同样需要维持行动轨迹的延续性。不同参与方在系统运行中的角色应在持续的记录与说明中得到确认，使其行动位置在网络中保持稳定存在；技术评估、伦理审查与风险复盘等活动应与系统更新周期相互衔接，使规范性判断成为转译历史的一部分；实际使用过程中的反馈与修正经验也应被纳入系统演化记录，使实践层的行动不断进入网络结构^[15]。通过这种方式，责任不再只是事后归因的结果，而是伴随行动生成始终存在的关系属性。

4.3 维持人类经验与技术转译之间的循环关系

当算法模型通过数据关联、模式识别与概率预测来“解释世界”时，其所呈现的并不是经验世界本身，而是经由数据抽象、特征压缩与规则编码后的“技术化现实”。若这一转译过程长期处于单向运行状态，人类经验将逐步退化为算法训练的原始材料，而不再作为独立的认知资源参与意义建构。由此产生的风险并非认知能力下降，而是理解权的转移——人类逐渐丧失对“为何如此”的诠释主导地位，只能被动接受技术系统输

出的“已解释结果”。因此，实现人机关系的良性发展，关键不在于彻底消除算法中介，而在于重建人类经验与技术转译之间的循环关系，使理解活动在网络中呈现出持续往复的双向流动，而非单向替代。

维持这一循环关系，首先需要在制度与技术架构层面嵌入“经验回流机制”，防止算法解释过程与现实情境发生结构性脱节。具体而言，在算法系统的训练、部署与迭代环节中，应制度化引入来自一线实践主体的情境反馈，使技术转译持续接受来自真实经验场域的修正。比如，在公共治理、教育评价、医疗辅助决策等场景中，不应仅依据数据指标评估算法有效性，而应将使用者的经验判断、情境感知与价值考量纳入模型更新依据之中，形成“数据—模型—实践经验—再建模”的循环链条。同时，在界面设计与系统交互层面，应为用户保留对算法输出进行解释补充、情境标注与异议表达的通道，使人类经验能够重新嵌入技术网络，而非被排除在决策链条之外^[16]。通过这种方式，技术转译不再是封闭的自动过程，而成为与经验世界持续对话的开放结构，从而避免算法认知与社会现实之间出现意义断裂。

进一步而言，还应在认知结构与文化层面重塑人类对技术转译的理解方式，使人类经验从“被动提供数据”转向“主动参与诠释”。这意味着需要强化公众与专业群体对算法运行逻辑、数据局限性与模型假设条件的认知，使技术解释重新被视为一种“可讨论的中介过程”，而非不可质疑的客观真理。在教育与培训体系中，可通过案例对照、情境模拟与反事实推演等方式，训练个体识别算法解释与现实经验之间的差异，从而提升对技术转译结果的批判性理解能力。当人类能够意识到算法只是网络中的一种行动者，其“理解”同样受到数据结构、模型设定与目标函数的限制时，人类经验便重新获得诠释主动权。

4.4 积极将伦理原则嵌入技术行动链条内部

在行动者网络理论视角下，价值困境的实质并不是“技术缺乏伦理”，而是伦理在网络中的位

置发生了位移。传统社会结构中，价值判断主要由人类行动者在具体情境中完成，而在智能技术深度嵌入社会运行之后，算法系统通过规则设定、参数权重、优化目标与分类标准，事实上承担了部分“价值分配”功能。例如，推荐系统决定何种内容被优先看见，风控模型界定何种行为被视为风险，教育与招聘算法参与筛选与排序，这些过程均包含隐含的价值取向。然而，这些价值并非以伦理语言显现，而是被编码为技术逻辑的一部分，从而造成“价值去显性化”的结构性后果。人类往往只能看到效率提升或结果输出，却难以识别其中所隐含的公平观、风险观与利益排序逻辑。由此，价值判断并未消失，而是被技术化、结构化并固化为系统运行的“默认前提”，这正是价值困境的核心所在。因此，实现人机关系的良性发展，关键不在于在技术外部进行道德呼吁，而在于推动伦理原则真正进入技术行动链条内部，成为网络运行的构成要素。

要实现伦理的内嵌化，首先需要在技术设计阶段建立“价值前置机制”，使伦理原则在模型构建之初即转化为可操作的结构条件，而非事后修补的附加规范。具体而言，在算法目标函数设定、数据采集范围界定以及分类标签设计等关键节点，应同步引入公平性、可及性、非歧视性与公共利益等原则，形成与性能指标并行的价值约束。例如，在公共服务分配算法中，不应仅以效率最大化为唯一目标，而应通过多目标优化结构，将弱势群体保障、区域平衡与长期社会影响等因素纳入决策框架^[17]；在数据治理层面，应限制对敏感特征的直接或间接利用，防止技术结构在无意识中再生产社会偏见。

进一步而言，还需在技术运行与治理层面构建“价值显影机制”，使已经进入系统的伦理原则能够在网络中持续被看见、被讨论与被修正。由于行动者网络是动态演化的，技术系统在不同情境中的实际影响可能偏离初始设定，因此应通过持续评估、可解释机制与公众参与渠道，使技术行动的价值后果得以显现。例如，通过算法审计、

影响评估报告与多方监督机制，揭示技术决策在不同群体间产生的差异性影响，从而将“隐性价值后果”转化为可讨论议题；在制度层面，应允许利益相关方对算法规则提出异议与修正建议，使伦理原则在网络运行中保持开放状态，而非被一次性固化。通过这种方式，伦理不再只是嵌入代码中的静态规则，而成为贯穿技术行动链条的动态调节力量，使人类社会的价值反思能力能够持续作用于技术系统。

5 结语

人机关系的演进不仅体现了技术发展的逻辑，更折射出伦理、权力与价值体系的深刻变革。传统的人类中心主义视角强调对技术的单向控制，将机器视为工具和资源，忽略了其在行动网络中的能动性。然而，随着人工智能、算法系统、智能代理的广泛应用，非人行动者在社会网络中日益具备影响力，其行为不仅塑造社会秩序，也参与价值实现与决策生成。在此背景下，单纯的控制范式已难以应对人机互动的复杂性与伦理困境，自主性与依赖性悖论、责任归属错位、可解释性与信任断裂，以及效率与价值冲突等问题日益突出。行动者网络理论为我们提供了理解和分析人机关系的工具：通过承认人类与技术的平等能动性、关注网络互动的动态生成过程，它揭示了人机关系从控制到共生的必然性与可能路径。

推动人机共生不仅是技术优化的需求，更是伦理与社会治理的紧迫课题。在实践层面，保持人类行动在技术网络中的持续在场，确保人机互动的生成轨迹在网络结构中得以持续显现，维持人类经验与技术转译之间的循环关系，积极将伦理原则嵌入技术行动链条内部，这些举措共同构成了人机关系良性发展的框架。总的来看，人机关系的伦理反思提醒我们：技术的能动性不可忽视，人类对技术的管理和共生策略必须同步演进。从行动者网络理论视角出发，“共生”不只是理想模式，而是可操作、可实践的伦理路径。它要求社会在制度设计、价值导向、权力配置与认知

建设上形成协同,从而在高效、智能的技术系统中维持人类主体性与社会价值。未来的人机关系,不应再是单向控制与被动适应的简单叠加,而应成为多维协作、互构赋能的动态网络,在共生中实现技术发展与人类福祉的双重目标。

责任编辑:李琦 校对:李琦 梁思琪

参考文献

- [1] 景星维,王思奇.论智能时代人的价值自觉[J].理论导刊,2023(9):74-80.
- [2] (美)兰登·温纳.自主性技术[M].杨海燕译.北京:北京大学出版社,2014:33-36.
- [3] (法)布鲁诺·拉图尔.我们从未现代过[M].上海:上海文艺出版社,2022:102-104.
- [4] 戴宇辰.“旧相识”和“新重逢”:行动者网络理论与媒介(化)研究的未来——一个理论史视角[J].国际新闻界,2019,41(4):68-88.
- [5] Peter-Paul Verbeek. Moralizing Technology: Understanding and Designing the Morality of Things[M]. Chicago: University of Chicago Press, 2011: 105-108.
- [6] Mark Coeckelbergh. AI Ethics[M]. Massachusetts: The MIT Press, 2020: 77-81.
- [7] 马文武,周文杰.社交机器人情感真实性的伦理争议:“实体论”与“关系论”两种进路及其关系[J].自然辩证法研究,2024(1):73-79.
- [8] 王淑庆.社交机器人的道德承受性地位[J].自然辩证法研究,2023(9):64-70.
- [9] 王亮.基于情境体验的社交机器人伦理:从“欺骗”到“向善”[J].自然辩证法研究,2021(10):55-60.
- [10] 窦笑.社交机器人伦理问题与政策建议研究[J].智库理论与实践,2022(5):103-110.
- [11] 邵雷,石峰.生成式人工智能对社交机器人的影响与治理对策[J].情报杂志,2024(4):154-163.
- [12] Jenna Burrell, Marion Fourcade. The Society of Algorithms[J]. *Annual Review of Sociology*, 2021, 2(47): 213-237.
- [13] 简予繁,黄玉波.人机互动:替代还是增强了人际互动?——角色理论视角下关于社交机器人的控制实验[J].新闻大学,2023(4):75-90+122.
- [14] Umberto Eco. Lector in fabula[M]. Milano: La nave di Teseo, 2020: 85-94.
- [15] Seibt Johanna. Integrative Social Robotics, Value-Driven Design, and Transdisciplinarity[J]. *Interaction Studies*, 2016, 1(21): 111-144.
- [16] Chamayou Grégoire. Drone Theory[M]. New York: Penguin, 2015: 102-105.
- [17] Archie B Carroll. The Pyramid of Corporate Social Responsibility: Toward the Moral Management of Organizational Stakeholders[J]. *Business Horizons*, 1991, 4(11): 39-42.

From "Control" to "Symbiosis": Ethical reflections on human-machine relationships from the perspective of actor-network theory

Guo Jianan

(School of Marxism, Zhengzhou University of Technology and Business, Zhengzhou 450000, China)

Abstract: With the rapid development of artificial intelligence, algorithmic systems, and intelligent agents, human-machine relationships are gradually evolving from traditional one-way control to multidimensional symbiosis. This paper, through an overview of actor network theory and an analysis of its applicability in reshaping human-machine relationships, further points out core ethical issues in this evolution, including the paradox of autonomy versus dependence, misalignment of responsibility attribution, a breakdown in interpretability and trust, and conflicts between efficiency and value. To address these issues, actor network theory offers a new path for developing healthy human-machine relationships: maintaining the continuous presence of human action within the technological network, ensuring the continuous manifestation of the generative trajectory of human-machine interaction within the network structure, maintaining the cyclical relationship between human experience and technological translation, and actively embedding ethical principles into the technological action chain. This aims to achieve the synchronous coordination of technological agency and social value in networked interaction, clarifying that human-machine symbiosis is not only a goal of technological optimization but also an inevitable requirement for ethics and governance, providing theoretical and practical references for the sustainable development of intelligent systems in future society.

Keywords: human-computer relationship ethical dilemma; actor-network theory; control and symbiosis; value consensus; power structure