

人机共融时代道德教育的困境与出路

靳思远

【摘要】 在人机共融时代，道德教育旨在不断丰富和发展人类文明形态，持续推动人类社会道德进步，实现人的精神自洽。整体而言，人机共融时代道德教育的发展面临着社会道德标准畸化使人类深陷道德选择困境、智能机器的道德主体地位争议挑战人类道德权威、智能机器难以实现人类价值对齐等多重挑战。对此，应从正确认识和把握人机共融时代道德教育的发展趋向、推进人工智能人类伦理和人工智能机器伦理同步完善、在生活世界中关注与引导情景化的道德教育实践等方面入手，促进人工智能技术影响下道德教育效能的整体性提升。

【关键词】 人机共融时代；具身智能；道德教育；道德进步

【作者简介】 靳思远，浙江大学马克思主义理论创新与传播研究中心研究员，马克思主义学院讲师（杭州 310058）

在全球新一轮科技革命和产业变革浪潮之下，人工智能技术日新月异、迭代升级，正在重塑人类社会生产和生活方式。从人类历史演进角度来看，人工智能的技术革新与数字文明演化进程相互碰撞、相互交织。随着生成式人工智能、具身智能等新一代人工智能技术成果在人类生产生活中的应用持续向纵深发展，人类和人工智能技术之间的存在边界日益模糊，人类社会生活进入人机共融时代。基于此，加强对人工智能技术进步与道德发展之间关系的理论阐释，论证人机共融时代道德教育的价值追求，揭示人机共融时代道德教育的主要困境，并据此提出道德教育的发展趋向和实践进路，具有重要的理论和现实意义。

一、人机共融时代道德教育的价值追求

当前，人工智能技术已由构想转向实用阶段，成为引领新一轮科技革命和产业变革的战略技术。回顾人工智能技术的发展历程，从计算智能到感知智能，再到认知智能，随着智能体所具备的能力形态和水平逐渐升级，其对人类生产生活的影响也更具深刻性和变革性。当人类尚惊叹于计算智能时代围棋机器人AlphaGo战胜围棋世界冠军、职业九段棋手李世石的出色表现时，基于大语言模型的生成式人工智能（Generative Artificial Intelligence）的出现标志着智能体已在一定程度上具备与人相类似的逻辑思维能力。2025年初，由杭州深度求索人工智能基础技术研究有限公司推出的AI应用DeepSeek一经上线，就立即火爆全球。随后，多地政务部门、学校、医院和企业等纷纷探索在其应用场景中接入DeepSeek大模型，试图借助人工智能的技术优势，进一步提升其业务办理的实际效能。在国家政策层面，面对新一代人工智能技术快速演进的新形势，习近平总书记强调要“牢牢掌握人工智能发展和治理主动权”，“推进人工智能全学段教育和全社会通识教育”^[1]。2025年，我国首次将“具身智能”（

Embodied Intelligence) 写入政府工作报告^[2], 标志着具身智能从技术概念正式上升为国家战略。具身智能让机器拥有了与人相类似的感知和行动能力, 使得机器能够主动感知物理空间、理解人类意图。“真正的智能无法脱离身体存在, 具身化是人工智能走向通用化的必经之路”^[3], 一旦具身智能实现通用化, 人工智能将真正成为辅助人类实现更好生活的好帮手。由此可见, 人工智能对社会生活的影响速度之快、程度之深、范围之广均前所未有。如果说以往人们更多是将人工智能视为一种辅助社会生活的技术工具, 那么, 当前人工智能已然成为继电力、互联网之后的重要社会基础设施, 融入社会生活各领域各行业各环节, 正在对经济社会发展产生颠覆性影响。在这个意义上, 人类社会正在步入一个人类智能和人工智能共生共存、相互影响的人机共融时代。

科学技术是推动社会变革和道德进步的动力。从人类历史来看, 每一次技术上的突飞猛进都会带来社会生产关系和生活方式的巨大变革, 进而改变社会的价值秩序与伦理规范。一方面, 科学技术作为第一生产力, 在推动经济发展的同时, 还通过社会经济基础的变化作用于道德, 影响着人们的思想道德观念; 另一方面, 社会道德观念会随着科技发展和生产力变革而不断进步。在唯物史观视域下, 道德具有历史性、阶级性。社会的道德观念并非永恒不变, 不同的时代、不同的社会制度、不同的阶级有不同的道德观念。作为一种社会意识形态和行为准则, 道德是生产关系的产物。“物质生活的生产方式制约着整个社会生活、政治生活和精神生活的过程”, ^{[4]591} “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]99} 道德观念是社会意识形态的重要组成部分, 由社会经济基础所决定。当发生重大科技变革时, 社会生产力革新也会带动新的道德观念的产生。人工智能作为推动新质生产力发展的关键要素, 在推动旧道德体系变革的同时, 也驱动新道德体系的产生, 显示出推动道德进步的巨大力量。新一轮人工智能技术的迅猛发展及其在社会生活领域的广泛应用, 在优化产业结构、促进产业结构转型升级的同时, 不仅极大地冲击了传统道德体系, 还对人类提出了许多新的伦理道德之问。人工智能技术发展过程中存在的深度伪造、算法偏见、平台操控、隐私泄露等伦理失范现象给国家和社会和谐稳定带来巨大风险, 动摇了公众对技术驱动个体和人类福祉的信任根基。面对“人机共融”的深度数字化生存境遇, 人工智能技术对人类社会和个体生命的影响愈加深远, 一方面, 社会治理迫切需要建立起新的价值秩序, 以适应日新月异的技术发展状况; 另一方面, 人们需要重新审视人类智能和人工智能的关系, 及时调整和更新自身的道德观念和道德行为。

在人机共融时代, 道德教育的价值旨归和终极自标是不断丰富和发展人类文明形态, 持续推动人类社会道德进步。纵观人类历史, 尽管我们承认人类社会文明在整体上呈现出向前演进的进步趋势, 但人类社会道德水平通常滞后于科学技术的发展水平。科技进步在给人类社会带来巨大福祉的同时, 也加剧了人与自然、人与社会、人与人之间的矛盾, 给人类生存带来了诸多风险和不确定性。基于此, 卢梭对科技之于道德进步的作用持悲观态度, 提出了科技发展和道德进步存在二律背反定律的观点。在他看来, 科技的发展和革新非但没有带来

社会道德水平的提升，反而会引发诸多社会道德问题。“随着科学与艺术的光芒在我们的地平线上升起，德行也就消逝了。”^{[6]11}事实上，科技进步并不直接等同于社会文明进步，科技进步所带来的力量可以被用于善恶两个方面。一旦技术进步所产生的力量被用于恶的方面，将会给人类社会带来空前危机。当前，由技术变革造成诸多发展困境也是现代性危机的一个重要根源和特征。在这种情况下，人类使用技术工具不仅没有提升社会道德水平，实际上可能还会降低。“除非伴随着道德的进步，技术的进步所带来的伤害会超过带来的好处。”^{[7]44}由此，技术进步过程中的道德教育命题亟需引起关注。在社会数字化进程向纵深发展和智能大众化时代背景下，不断实现“道德上的更好”仍代表着人类社会向前发展的价值追求。面对人工智能给人类社会生活带来的巨大变革，应该在人工智能技术不断演进过程中对其技术特性、发展趋势和社会影响予以更多价值关切，始终秉持技术向善的伦理原则，最大限度地克服科技发展存在的局限性，充分发挥其在促进人类文明跃升中所起到的赋能作用，不断推动人类社会道德进步螺旋式上升。

人机共融时代的道德教育还应致力于不断满足人的精神文化需求，实现人的精神自治。人的发展应是“人以一种全面的方式，就是说，作为一个完整的人，占有自己的全面的本质”^{[8]189}。在人机共融状态下，道德教育对于人的全面发展的关注重点从：“知识域”转向“价值域”。对于人类社会整体而言，由于新一轮人工智能技术革新所带来的物质财富增长成效显著。但是，人类社会的发展和进步不能仅仅停留在物质层面，还应更多关注精神层面的提升。在我国，中国式现代化的关键特征之一是“物质文明和精神文明相协调”^{[9]19}，强调在物质生产力水平不断提升的同时，还应关注丰富与满足人民群众日益增长的精神文化需求。在人机共融时代，当人工智能作为驱动新质生产力发展的关键要素在改变社会整体面貌和运行逻辑的同时，也在整体上影响着人们的精神文化需求。在精神文化需求领域方面，人工智能及其技术应用在社会生活领域的普及使得人类社会的精神交往实践由线下空间拓展到数字领域。人机情感交互在给人类提供替代性情感补偿方案的同时，也带来了前所未有的道德风险。在精神文化需求层次方面，由于人工智能发展与超级计算、深度学习、传感网、脑科学等高新科技相融合，智能体在信息搜集、分析与处理等方面的能力显著提升，能够高效回应人类的求知需求，推动人类认知层级实现从“事实”到“价值”的质性跃升。道德教育为人类的人格全面发展和精神自由实现创造了条件，帮助人们不断在生活世界中自觉寻求自我超越，完善自我理想人格，进而达到自我道德理想的美好境界。面对人工智能及其技术应用给社会道德秩序带来的冲击，应该通过道德教育帮助人们筑牢价值基石，在正确伦理道德规范的指引下进行价值判断、作出价值选择，从而成为内在精神品质与外在行为表现相一致的自治主体。

二、人机共融时代道德教育面临的主要困境

受到人工智能技术二重性特征影响，新一代人工智能技术在推动社会价值秩序不断调整与向前发展的同时，也给现有社会道德体系带来巨大冲击。从人工智能的技术特性及其社会

影响角度来看,人机共融时代道德教育的发展面临着社会道德标准畸化使人类深陷道德选择困境、智能机器的道德主体地位争议挑战人类道德权威、智能机器难以实现人类价值对齐等多重挑战。

(一) 人工智能冲击社会道德标准,使人类深陷道德选择困境

道德选择是人类存在的基本方式,也是人类社会生活的永恒主题。在社会生活中,个体基于对真善美的不懈追寻进行各种各样的道德决策,体现出人作为道德主体的能动性和高贵性。道德选择的实质是“价值观的表现形式”,“是人在一定的道德意识支配下,根据某种道德标准,在不同的价值准则或善恶之间所做的自觉自愿的抉择。”^{[10]344}道德选择的显著特征是自觉自愿性,即个体的道德行为是出自于其自由意志的理性选择,而不是受到外界力量干预的被迫选择。从行为生成的因果性和内在联动性来看,只有基于一种自觉自愿的道德选择,个体才能主动对其行为负责。随着人工智能的不断升级和机器学习技术的迅猛发展,相关技术及其应用对社会道德标准以及个体自由意志的影响日渐深刻。

一方面,多元价值冲突易使个体陷入道德相对主义困境。在人机共融状态下,作为道德主体的个体在数字技术加持下,依靠数字化身与智能体展开在线交互,需具备更高的道德自制力。然而,当人机交互中道德规范的遵循与否完全取决于个体的自主选择时,极易出现个人主观感受、情感偏好和利益考量凌驾于道德约束之上的情况,形成道德无关紧要的错误认知。在互联网已经成为社会生活基础设施的今天,赛博空间的信息传播组织架构具有去中心化、扁平化、开放性、多节点等显著特征。每个用户都是信息生产与接收的统一体。由于个体成长经历和价值观念都存在差异,其在网络信息生产和传播中所表达出来的观点具有价值多元化特征,使社会舆论场呈现出“众声嘈杂”状态。这在一定程度上解构了社会主流价值的权威性,为道德相对主义的滋生和泛滥提供了“技术温床”。道德相对主义者消解了道德坚守之于个体道德选择的重要性,忽视了真善美的伦理导向性,完全把个体选择视为正当,而无视其是否存在负面倾向或带来消极影响,可能导致极端个人主义价值观对主流价值观的侵蚀。

另一方面,同质信息沉迷易使个体陷入道德绝对主义困境。大数据、机器学习、深度学习等技术是当前人工智能发展的重要技术底座,为智能体赋予信息搜集、整理与分析能力奠定技术支撑。在看到新一代人工智能强大的信息处理能力的同时,也要充分认识到,算法依据用户喜好为其提供信息内容看似“投其所好”,但长期沉浸在同质化信息内容所营造的“信息茧房”之中,还会使用户加剧对某一特定价值观的固化认知,进而陷入道德绝对主义困境。表面看来,用户在与智能体互动的过程中是在自主选择信息,他们的认知和行为似乎是自觉自愿的。实际上,他们所接触的信息是由算法经过刻意筛选后推送的,这种信息选择的过程并非真正出自于用户的自由意志。道德绝对主义把某一价值观作为伦理评判的唯一依据,这是造成社会价值观念固化的重要因素。一旦这一观念出现偏颇,极易在该信息接收圈层内引起涟漪效应,导致某种具有明显倾向性的价值偏见固化,进而造成群体极化现象。同时

，长期接收大量同质化的信息内容，个体会在“信息茧房”效应下形成封闭的道德标尺，拒斥多元价值对话，丧失对复杂情境的共情能力，陷入“非黑即白”的认知图式，造成价值判断僵化和伦理感知萎缩。

（二）智能机器存在道德主体地位争议，挑战人类道德权威

随着人工智能技术的迅猛发展，关于“人工智能体是否能成为道德主体”的问题逐渐成为当前学界关注的一个热点与焦点问题。对此，国内外学者众说纷纭、莫衷一是。持否定论者认为，自前的人工智能体尚不能成为完全意义上的道德主体，原因在于其无法像人一样拥有意识^[11]。持肯定论者主张，尽管人工智能体不具备人类所特有的意识、理性和自由意志，但其在深度学习脑科学等技术的加持下，展现出对外界信息的高度感知和自主决策能力，因而具有交互性、自主性、适应性等特征^[12]，能够表现和理解责任。在这个意义上，“智能机器成为道德主体在技术上是可行的，只不过需要时间慢慢发展。”^[13]上述争议体现出人类关于人工智能道德主体性地位问题的思考以及对智能机器发展前景的担忧。

早在20世纪50年代，享有“计算机科学之父”和“人工智能先驱”盛誉的图灵在《计算机与智能》一文中就开创性地提出“机器是否能思考”的哲学命题。时至今日，生成式人工智能在情感计算、深度学习等技术支持下能够理解人类意图，根据人类需求高效处理信息。一直以来，哲学家关于什么是“人”的思考答案指向人的思想性。马克思在揭示人类实践活动的自觉能动性和创造性时，特地强调了意识在实践活动中的重要作用，“蜜蜂建筑蜂房的本领使人间的许多建筑师感到惭愧。但是，最整脚的建筑师从一开始就比最灵巧的蜜蜂高明的地方，是他在用蜂蜡建筑蜂房以前，已经在自己的头脑中把它建成了。”^[14]²⁰⁸人工智能体的出现及其在社会生活各领域中所展现出的强大理解和思维能力打破了只有人类才具备思维的常识。特别是具身智能的兴起与广泛应用，具备人形而更加接近真实人类认知和行为模式的人工智能的出现正在深刻改变着道德的内涵与道德主体的边界。具身智能拥有像人一样的外在机械身体，能够直接和物理世界中的人或其他智能机器进行交互。当前，尽管具身智能的传感器、执行器等硬件仍存在诸多不足之处，但它也具备一定的机器理解和逻辑能力，展现出本体性特征。在传统的技术伦理中，人工智能被视为由人类所创造和使用的工具而存在。随着人工智能体展现出互动性、自主性和适应性，人类主体和技术人工物之间的界限逐渐模糊。当人工智能体具备越来越多诸如计算、语言、学习等曾被认为独属于人类的特性时，是否应当考虑它们的“生命尊严”问题？是否应将其视为一种具备自我学习、适应和决策能力的“准生命体”？当人工智能体越来越多参与到和人类生命、健康相关的决策时，是否应该赋予其一定的道德责任？

由此，有学者提倡，面对人工智能体成为类人生命的趋势，应该改变现有关于人工智能属性的认识，实现从“人类的技术手段”到“人工的智能生命”的实践转向、从“技术伦理”到“生命伦理”的研究范式转向。^[15]这种从实践到范式的双重转向，要求我们重新关注和思考人工智能的伦理地位。“当机器变得足够复杂的时候，它们既不是人类的仆人，也不

是人类的主人，而是人类的伙伴。”^{[16][208]}人工智能体在自主学习与决策、语言理解、情感交互等方面逐渐展现出类主体的特质，使得人类智能和机器智能之间的界限变得愈发模糊。人工智能的拟人化趋势一再引发关于机器智能是否应当享有一定的道德权利、能否成为道德主体的追问。一旦智能机器成为道德主体，势必会削弱人的道德主体性，危及人类中心地位，导致人类遭遇存在价值危机。此外，智能机器的道德主体地位还会冲击传统伦理关系，给人类道德权威带来挑战。比如，AI教唆用户自杀、人与AI“谈情说爱”甚至“结婚”、用户对智能语音助手产生心理依赖等现象进一步冲击了现有社会道德秩序，引导人们重新审视人工智能技术带来的伦理困境。

（三）智能机器缺乏道德实践智慧，难以实现人类价值对齐

人类中心主义观点认为，智能体作为辅助人类追求美好生活和实现自身目的的工具而存在，其发展和使用应坚持造福人类原则，避免机器损害人类整体利益而更好地服务于人类。从技术变革趋势来看，尽管发展具身智能已上升为国家战略，但人形机器人在复杂情境中快速学习、自主决策仍是当前人工智能领域的公认难题。当人工智能体与人类协同开展工作，必须确保机器的安全性和可控性，以免对人类造成伤害。由于当前传感器的精密度、紧急制动系统的反应速度等均未达到强人工智能发展的技术要求，人工智能体做不到像人类一样为自己的一切行为负责，还算不上是一个完全的责任主体。这需要在技术研发过程中同步推进伦理规范完善，确保人工智能体的动机与行为始终与人类价值观保持一致。

近年来，面对人工智能技术变革给社会生活带来的诸多伦理风险，学界专门提出“人工智能价值对齐”这一命题，主张应确保人工智能体与人类价值观念和伦理原则相一致，实现人工智能与人类价值的对齐。那么，智能机器应该如何具备道德判断能力呢？有人提出“人工智能道德增强”的技术构想，即将义务论、美德论、契约论等人类社会道德原则人为地嵌入人工智能系统，构建一种选择可评估、行为可预测的人工道德主体（Artificial Moral Advisors，简称AMA）。^[17]他们认为如此，人工智能体就能依据人类所嵌入的某种道德原则的计算模型作出道德选择和行为决策，并且其处理问题的结果能够被大多数人所接受。上述关于人工道德主体的技术构想展现出当前全球人工智能伦理发展过程中存在的一种明显的结果主义倾向，^[18]即人们主要是根据人工智能体的行为结果来评估它的道德正确性。这意味着，当前人们对人工道德主体的关注仅停留在结果层面，忽视了对于人工道德主体作出道德选择和行为决策的过程是否符合程序正义和人类社会道德原则的关注。事实上，这也是当前人工智能体发展过程中存在的一个显著缺陷，智能机器尚不具备人类在道德行为中所展现的道德实践智慧，只能依据人类嵌入的道德原则作出符合伦理结果主义要求的决策，无法依据实际情况灵活地作出调整和修正。真正意义上的人工道德主体，“是在实时的社会环境中具备实践智慧的道德主体，就意味着人工道德主体是一个在情感动机的触发下具备信用赋能和实践推理能力的智能主体。”^[19]就当前技术发展水平而言，智能机器还不能称为真正意义上的道德主体，故而难以真正实现人类价值对齐。

还应关注到，人工道德主体得以运行的一个底层逻辑是“道德可以计算”，存在着将道德标准量化的现象。而道德本身是一种规范性要求，对于“何为善”“何为正当”等问题的思考夹杂着人类社会复杂的道德判断和情感考量。目前，智能机器对于道德问题的处理尚处于低级阶段，不能有效回应社会生活中复杂的道德困境。在人类所面临的道德两难情境中，不存在那种能明确告知人们如何做才正确的答案，且每个道德选择都牵涉大量问题。面对社会生活的诸多不确定性，智能机器给出的道德建议过于简单，无法有效回应人们的道德困惑。从长远角度来看，智能机器的形式化规则无法替代社会伦理道德规范，而道德量化也不应成为人类社会道德进步的发展方向。

三、人机共融时代道德教育的实践路向

尽管当前人机共融时代的道德教育面临诸多困境，但是道德教育受到人工智能技术影响、回应人工智能带来的伦理问题、推动人类社会道德进步是大势所趋。正如恩格斯所言，“历史同认识一样，永远不会在人类的一种完美的理想状态中最终结束；完美的社会、完美的‘国家’是只有在幻想中才能存在的东西；相反，一切依次更替的历史状态都只是人类社会由低级到高级的无穷发展进程中的暂时阶段”^{[20][22]}对此，应正确认识和把握人机共融时代道德教育的发展趋向，推进人工智能人类伦理和人工智能机器伦理同步完善，在生活世界中关注情景化的道德教育实践，促进人工智能技术影响下道德教育效能的整体性提升。

（一）坚持守正创新原则，正确认识和把握道德教育的发展趋向

身处人机共融时代，面对新一代人工智能及其技术应用给社会伦理道德规范和价值秩序带来的巨大挑战，应该注重发挥道德教育对于推动社会道德进步、构建良好道德秩序的正向作用。对此，我们应坚持守正创新原则，科学认识和把握人机共融时代道德教育现代化转型的“不变”和“变”，正确把握人工智能技术影响下道德教育的发展趋向。

一方面，道德教育的价值追求没有变化，仍是以“道德上的更好”为价值牵引，持续推动社会道德进步和满足人的精神文化需求。习近平总书记提出，“教育是民族振兴、社会进步的重要基石”，^[21]²“育人的根本在于立德”，^[22]³⁴深刻阐明了道德教育之于文明进步和人的发展的重要意义。首先，人机共融时代的道德教育仍需关注人类优秀文明成果的价值传承，为社会有序运行奠定牢固的价值基石。面对道德相对主义、道德绝对主义等社会道德标准畸化所带来的道德选择困境，应该认识到，传统社会规范中的诚信、公平、正义等道德原则仍然适用，在当下仍是道德教育的重要根基，也是构筑和谐社会大厦的重要基石。与此同时，还应关注到数据在具身智能发展过程中的关键作用。正是在对社会秩序和价值准则进行科学数据化的基础上，具身智能才得以与外部现实世界展开高质量交互，并由此作出相应的行动决策。因此，在具身智能技术由设想转向应用的起步阶段，应加强对数据这一关键要素的前置性关注，坚持以数据向善原则构建具身数据平台，在保证数据多样性、符合人类社会基本道德规范的基础上，注重通过技术手段实现数据的自主纠偏，进而不断优化具身智能对真实环境的理解。其次，人机共融时代的道德教育仍需秉持人本主义立场，捍卫与坚守人的

道德主体性。尽管人工智能技术变革使得人类主体的道德实践出现了技术化转向，甚至出现了诸如“智能机器的道德主体地位争议”等挑战人类既有道德秩序的伦理困境，但这并不意味着人类的價值选择完全被技术所遮蔽。道德教育始终致力于维护人类意志在道德领域的独特价值，确保人类不会失去指引前进方向的道德风向标。最后，人机共融时代的道德教育仍需培养人的正确价值观，有效应对各种价值选择困境。人工智能技术带来的伦理困境前所未有，仍需发挥道德教育的价值引领功能，为人类在复杂的道德情境中作出正确价值决策提供保障。

另一方面，道德教育的发展环境发生了变化，要在顺应人工智能技术变革特点的基础上，推动道德教育数字化转型向纵深发展。道德现代化是中国式现代化的重要维度和题中应有之义。在人机共融时代，道德教育现代化转型的一个重要面向就是数字化转型。为了更好地适应人工智能技术加速演进及其在社会生活领域中的广泛应用趋势，有效回应人工智能技术及其应用给人类社会生活带来的诸多伦理风险和挑战，需要着力推动实现道德教育的数字化转型。首先，要关注道德教育的形态变化，深刻把握人工智能技术使道德教育由离身到具身的变化境遇。其次，要关注道德教育的内容变化，根据人工智能技术变革给社会伦理道德规范带来的冲击和挑战，特别是具身智能发展所带来的技术观、人机关系观、哲学意义上的身体观、具身认知观等内容变化，应及时调整与更新道德教育的内容，使之能够有效回应社会公众的思想道德困惑。最后，要关注道德教育的手段变化，依托人工智能的数据采集、追踪和动态分析功能，为教育对象制定个性化的道德培养方案。

(二) 完善伦理规范，建立人工智能“人类伦理—机器伦理”二元框架

在人工智能技术不断升级且日渐融入人类生活世界的时代背景下，应该重视进一步建立健全人工智能高度发展之后的社会伦理道德体系。现有的人工智能伦理框架主要侧重于单一的机器伦理，缺乏对人类使用机器时应遵循的相关伦理规范的考量。比如，“阿西莫夫机器人三定律”“帕斯奎尔四法则”以及“海斯三定律”等均从机器角度出发，强调人工智能发展应秉持对人类负责任的首要原则，避免机器对人类作恶。进入人机共融时代，随着人类与人工智能技术的存在界限逐渐模糊，应在人工智能伦理框架中进一步引入人类伦理视角，从人类合理认识与使用人工智能技术角度确立起人工智能发展的人类伦理，构建起关于人工智能发展的“人类伦理-机器伦理”二元框架。一方面，在进一步完善人工智能机器伦理的基础上，加快推进人工智能人类伦理的确立与发展。在人工智能机器伦理角度，现有关于智能机器发展的伦理规范并非一成不变，而是随着人工智能技术变革及其给人类社会带来的显著变化而不断更新。要结合新一代人工智能技术发展的新变化和新特点，适时调整与完善人工智能机器伦理的内容要求，使之适应人工智能技术发展趋势和人类文明进步的价值要求。阿西莫夫在提出“机器人不得伤害人类，必须服从人类命令，以及保护自己已但不违背前两条”[23]44的“机器人三定律”之后，又结合新的情况提出了一条新的定律，即“机器人不得伤害人类整体，或因不作为使人类整体受到伤害”[24]378。他特地将这条新增的内容概括为

“第零定律”，将其重要性放在前面三定律之前，强调智能机器的发展除了要考虑人类个体的安全之外，更要优先考虑人类的整体利益。由于社会意识具有相对滞后性特征，人类关于人工智能伦理的价值主张会在一定程度上相对滞后于社会实践的具体情况。面对蓬勃发展的人机共创、人机共教等人与机器协同工作境遇，既要依据实际情况不断完善人工智能机器伦理，文要及时确立与发展人工智能人类伦理，使人工智能的技术研发与使用始终朝着“善”的方向发展。

另一方面，确立“人机联合归责”^[25]的治理思路。进入人机共融时代，应对人工智能体进行道德归责以预防其不当应用可能产生的负面效果。然而，无论是对人类或是对智能机器进行单一主体的道德归责都会产生伦理困境，无法彻底解决人工智能技术发展和使用存在的风险和挑战。对于人工智能体的道德归责，既要发挥人的主观能动性，也要将智能机器视为有限道德主体，尽快确立起一种“人机联合归责”的伦理风险治理思路。美国计算机科学家约瑟夫·利克莱德最早提出“人机共生”概念。他在人机关系研究中引入生物学中的“共生”思想，主张实现人机优势互补，从而实现高效协同工作^[26]。在上述技术构想中，人类和智能机器是共生共存的，二者同向发力，促进人类达到理想的生活和工作状态。基于人工智能“人类伦理—机器伦理”的二元伦理框架，“人机联合归责”主张强调应对人类主体和智能机器的交互行为及其过程进行双重道德归责。人类是完全道德主体，智能机器是有限道德主体，二者应共同承担相应的道德责任。还要认识到，智能机器承担责任并不意味着人类可以逃避责任，其作为有限道德主体不能独立承担道德责任，无法和人类主体一样直接接受道德惩罚。因此，不能脱离人对智能机器的影响来讨论人工智能体的道德责任问题。

（三）回归生活世界，关注与引导情景化的道德教育实践

“道德教育不同于知识教育，道德教育在于生活的意义，为人的生活提供终极关怀和安身立命的精神。道德于人的‘做’成性，决定了道德教育只能在生活中进行。”^[27]人工智能时代，生活依然是人之存在的方式，道德教育应该回归生活世界，关切与回应人工智能技术影响下的人的生存境遇和精神需求，不断提升人们的生活旨趣和道德境界，实现人的理想人格完善和幸福生活追求。

一方面，由于人与人工智能技术的存在界限日渐模糊，要更加关注情景化的道德教育实践。在各行各业普遍开展“人工智能+”行动的时代背景下，人机交互从教育娱乐、医疗照护、家庭服务和紧急救援等多元场景逐步渗透到社会生活各领域。当前，具身智能在传感器技术、控制工程与仿生材料等高新科技的支持下，已逐步实现运动感知、多模态操作、灵巧操作等功能。尽管如此，受到真实物理世界的环境复杂性和多变性因素等的影响，具身智能的发展和应用仍处于起步萌芽阶段。未来一段时期内，具身智能的发展应侧重于突破“莫拉维克悖论”，不断提升机器人适应新环境的能力，使之能够通过机器学习所习得的技能不断迁移到新的复杂情境之中。在此趋势之下，应关注到人工智能技术对人类道德生活的影响以及如何应对的问题。“到处是生活，即到处是教育；整个社会是生活的场所，亦即教育之

场所”^{[28]246}在多元场景应用中，人机交互呈现出显著的日常化和情景化特征。一个个对话框、一段段人机即时对话等都对应着多个复杂而又具体的道德教育情境。这要求教育者应学会理解不同的教育情境，并在具体的教育情境中寻求合宜的道德教育方案。对于广大用户而言，应不断提高自身的人工智能素养，坚持以灵活、适应性伦理原则应对人工智能技术发展过程中存在的复杂性和不确定性因素，根据人机交互中的具体情境及时作出正确的价值判断和道德选择。

另一方面，面对人机共融时代复杂的道德实践情境，应注重培养人的道德思维，使之学会做出正确的道德选择。人类作为理性主体的一个重要条件就是具备道德思维能力，并且能够依据这种道德思维能力进行价值判断和决策。在人机共融时代，面对智能机器生成的海量信息、各式各样复杂的道德实践情境以及人工智能技术发展带来的诸多伦理风险和挑战，道德教育尤其需要注重培养人的道德思维能力，基于对“什么样的生活是值得过的生活”“应该追求什么样的生活”等问题的省思，以符合真善美的道德标准来引导个体追问什么样的道德选择才是正当的，什么样的道德选择才是向着人格完善和幸福生活的，并在此基础上培养个体道德选择的品性和能力，使之能够作出合乎道德的选择。

参考文献：

[1]习近平在中共中央政治局第二十次集体学习时强调坚持自立自强突出应用导向推动人工智能健康有序发展 [N]. 人民日报, 2025-04-27(1).

[2]李强. 政府工作报告——2025 年 3 月 5 日在第十四届全国人民代表大会第三次会议上 [N]. 人民日报, 2025-03-13(1).

[3]孙欣.Hi, 具身智能机器人 [N]. 青岛日报, 2025-05-14(1).

[4]马克思恩格斯文集（第 2 卷）[M]. 中共中央马克思恩格斯编译局, 译. 北京: 人民出版社, 2009.

[5][德]恩格斯. 反杜林论 [M]. 中共中央马克思恩格斯列宁斯大林编译局, 译. 北京: 人民出版社, 2018.

[6][法]卢梭. 论科学与艺术 [M]. 何兆武, 译. 北京: 商务印书馆, 1963.

[7][美]弗里曼·戴森. 反叛的科学家 [M]. 肖明波, 杨光松, 译. 浙江: 浙江大学出版社, 2013.

[8]马克思恩格斯文集（第 1 卷）[M]. 中共中央马克思恩格斯列宁斯大林编译局, 译. 北京: 人民出版社, 2009.

[9]中共中央文献编辑委员会. 习近平著作选读（第一卷）[M] 北京: 人民出版

社,2023.

[10]罗国杰.伦理学[M].北京:人民出版社,1989.

[11]张今杰.人工智能体的道德主体地位问题探讨[J].求索,2022(1).

[12]Floridi, L., Sanders, J. W. 'On the Morality of Artificial Agents' [J]. Minds and Machines, 2004, 14(3).

[13]田训龙.智能道德主体的潜在风险探析[J].道德与文明,2024,(4).

[14]马克思恩格斯文集(第5卷)[M].中共中央马克思恩格斯列宁斯大林编译局,译.北京:人民出版社,2009.

[15]舒红跃,李文龙.从技术伦理到生命伦理:人工智能伦理研究的范式转换[J].学习与实践,2025(3).

[16][美]约翰·马尔科夫.人工智能简史[M].郭雪,译.杭州:浙江人民出版社,2017.

[17]Giubilini, A. , Savulescu, J. 'The artificial moral advisor. the "Ideal Observer" meets artificial intelligence' [J]. Philosophy & Technology, 2018,31(2).

[18]吕梁山、梁书豪.人工智能伦理建构中的结果主义倾向及其隐患[J].福建师范大学学报(哲学社会科学版),2025(3).

[19]刘晓力.哲学与认知科学交叉融合的途径[J].中国社会科学,2020(9).

[20]马克思恩格斯选集(第4卷)[M].中共中央马克思恩格斯列宁斯大林编译局,译.北京:人民出版社,2012.

[21]习近平.论教育[M].北京:中央文献出版社,2024.

[22]习近平.高举中国特色社会主义伟大旗帜为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告[M].北京:人民出版社,2022.

[23][美]艾萨克·阿西莫夫.我,机器人[M].叶李华,译.南京:江苏凤凰文艺出版社,2015.

[24][美]艾萨克·阿西莫夫.机器人与帝国[M].叶李华,译.南京:江苏凤凰文艺出版社,2015.

[25]赵子明.人工智能体道德责任归责问题探析[J].中国矿业大学学报(社会科学版),2024(3).

[26]J. C. R. Licklider. Man-computer symbiosis[J]. IRE Transactions on Human Factors in

Electronics, 1960,1(1).

[27]鲁洁, 冯建军. 让道德教育成为最具有魅力的教育——鲁洁教授专访[J]. 苏州大学学报(教育科学版), 2020,8(2).

[28]陶行知. 陶行知全集（第三卷）[M]. 成都:四川教育出版社, 1991.