• 专题：算法设计的伦理反思 •

编者按：

以机器学习为代表的算法作为新一代人工智能的底层技术逻辑或核心驱动力已经引起了广泛关注。然而算法在实践应用中也被揭示出可能出现侵犯公众隐私、信息控制、算法偏见、歧视甚至危害公共安全等负面问题。有鉴于此，如何确保算法“向善”而不违背人类社会基本的伦理价值或道德原则，成为需要认真思考的重要问题。算法会产生怎样的伦理后果，一方面受到算法应用方式的影响，另一方面则可能源于算法自身的设计特征。正因如此，越来越多研究开始主张内嵌伦理或道德的算法设计，这也成为许多国家和国际机构发布的人工智能治理规范中的重要构成。然而，“伦理算法”或“道德算法”的构想固然吸引人，其能否实现则面临着许多挑战，例如人类伦理道德原则本身的内在张力，由此凸显出对算法设计的伦理问题进行审慎反思的重要性。

本专题的三篇论文从不同角度对这一问题进行回应。第一篇闫瑞峰的论文对算法设计所涉及的功利论、契约论、义务论与德性论四种伦理立场及其引发的治理争论进行了细致梳理，这为我们思考算法设计中的伦理争议提供了一个整体性框架。第二篇李大山的论文聚焦于自动驾驶中的道德两难问题，作者通过区分“一阶道德算法”与“高阶道德算法”，提出并论证了“自动驾驶无需高阶道德算法”这一命题，这对当前许多人呼吁的“道德算法”是一个很有意思的回应。第三篇张海柱的论文转向公共治理场景中的算法应用，作者通过对算法内在政治性的分析提出了“算法公共性”命题，并将算法公共性的实现诉诸于算法的参与式设计，从而在社会政治层面上对算法设计伦理问题进行了初步讨论。三篇论文角度各有侧重，观点立场也不尽相同，但均为我们反思算法设计中的伦理复杂性提供了有益借鉴。

（专题策划：张海柱）

算法设计伦理治理的立场、争论与对策

The Standpoint, Controversy and Countermeasures of Ethical Governance of Algorithm Design

闫瑞峰 /YAN Ruifeng

（北京工业大学马克思主义学院，北京，100124）
(College of Marxism, Beijing University of Technology, Beijing, 100124)

摘　要：算法伦理问题直接产生于算法设计实践，而算法设计的伦理问题源自不同算法设计者的伦理

基金项目：国家社会科学基金青年项目“数字经济时代资本无序扩张的伦理规制研究”（项目编号：22CZX044）；研究阐释党的十九届五中全会精神国家社科基金重大项目“深入推进科技体制改革与完善国家科技治理体系研究”（项目编号：21ZDA017）。
收稿日期：2022 年 7 月 30 日
作者简介：闫瑞峰（1987–）男，河北邯郸人，北京工业大学马克思主义学院讲师，研究方向为科技伦理、资本论理和政治伦理。Email: yanrfccps@163.com

立场，是现实世界与技术界互动交融的结果。根据算法设计主体的差异化价值取向，可将相关伦理立场分为功利论、契约论、义务论、德性论等四大类型。算法设计的四种伦理立场在治理中极易导致四种特性鲜明的伦理倾向，即功利论下技治主义的弱正义性、契约论下规则主义的强形式公正性、义务论下规范主义的弱目的性以及德性论下自律主义的不可控性。算法设计的伦理治理应以构建负责人的多元主体参与的治理网络为指引，在实践中要遵守算法向善的基本价值立场，坚持公共善的核心伦理导向和负责人的可持续发展观，实现算法设计道德化的全流程覆盖并完善相关伦理治理体系。

关键词：算法设计　伦理　立场　治理

Abstract: The ethical problem of algorithm comes directly from the practice of algorithm design, and the ethical problem of algorithm design, in turn, comes from the ethical position of different algorithm designers, which is the result of the interaction between the real world and the technical community. According to the differentiated value orientation of algorithm design subjects, relevant ethical positions can be divided into four types: utilitarianism, contract theory, obligation theory and virtue theory. The four ethical positions of algorithm design can easily lead to four distinctive ethical tendencies in ethical governance, namely, the weak justice of technological governance under utilitarian theory, the strong formal justice of ruleism under contract theory, the weak purpose of normative theory under obligation theory and the uncontrollability of self-discipline under virtue theory. The ethical governance of algorithm design should be guided by the construction of a governance network with the participation of multiple subjects of the person in charge. In practice, we should abide by the basic value position of the algorithm towards good, adhere to the core ethical guidance of public good and the sustainable development view of the person in charge, realize the whole process coverage of the moralization of algorithm design, and improve the ethical governance system of algorithm design.

Key Words: Algorithm design; Ethics; Position; Governance

中图分类号：B82-05; TP301.6　　**文献标识码：**A　　DOI: 10.15994/j.1000–0763.2023.06.001

一、算法设计伦理问题的溯源

当今社会，随着互联网、大数据、人工智能与经济社会的融合式发展，算法技术在生产、分配、交换、消费等领域的广泛应用，为人类社会带来诸多伦理问题。算法伦理问题可以基于使用过程，从“自主性特征、应用性场景和归责性困难三个方面理解”，[1]涉及哲学文化、社会正义、技术安全等多重维度。从溯源学意义上来讲，算法伦理问题的产生与算法实践主体的价值立场和偏好、道德水平和心理、行为习惯和倾向等因素息息相关，这些人类行为特性赋予了算法价值负载性特征。也就是说，“由于算法具有价值负载性，基于不同价值判断的人通常会设计出截然不同的算法，或者说这些人会选择不同的算法方式来解决问题。”[2]截止目前，“即便是人工智能算法展现出隐喻意义上的不同程度学习能力，可能能够通过所谓的图灵测试，但这种学习能力归根结底仍然属于工具能力，人类对算法始终处于控制支配地位。”[3]

在现实社会中，算法实践主体主要指代算法设计者，因此算法伦理问题直接产生于算法设计者的实践。所谓算法设计，主要指人们为实现特定目的或意图，借助一系列信息技术工具和数字手段，对计算方式、计算过程和计算程序进行选择、设计和调整的综合实践。根据学界对算法设计的概念、内涵和作用的研究来看，算法设计具有预见性、程序性、延迟性、价值负载性以及逻辑可塑性等特征，其中价值负载性和逻辑可塑性是构成系列算法设计伦理问题的关键性根源。因此，算法设计因受行为主体的操控而具有最根本的价值负载性，不同利益主体的价值观念和伦理立场会导致差异化的算法设计实践，这就不得不将相关问题放置于现实世界和技术世界进行统筹性考量，并从中进一步溯源算法设计伦理问题的生成肌理。

算法设计伦理问题源于现实世界，同时又反作用于现实世界。一般而言，所有算法设计伦理问题都源自现实世界，同时又经过一系列带有主观目的性的技术实践反作用于现实世界，由此完成“现实世界–技术世界–现实世界”的实践逻辑过程。因此，“算法并不是按照世界本身的样子运行，而是一种现实世界的加强版”，[4]是经由人为改造并根据其需要将现实世界投射至技术世界，进而再次将人为设计好的技术系统地投射至现实世界。进一步讲，算法设计通过算法技术将现实伦理问题投射至技术世界，这一过程不仅伴随伦理问题的“复制”效果，同时也释放出强烈的伦理“创造”功能。伦理问题的技术“复制”更多地是将现实中本身存在的伦理问题映射至技术世界，而伦理问题的“创造”功能主要指在技术投射过程中释放出新的伦理问题，比如人工智能技术带来的人类自主性丧失的新型伦理问题。这些新问题极有可能向极端化方向发展，因此在算法技术加持下所带来的新型伦理问题会更加突出和复杂。

整体而言，算法设计是一种人工技术物，一经创造出来便负载着特定的价值，而设计者、审查者、决策者及其他利益相关者的价值观念、利益立场、道德水平等因素，会深度影响乃至决定这一技术物的价值特性。算法设计的价值特性在现实中会直接形成基于不同价值观念的伦理立场，在此过程中进一步延伸其自身的伦理内涵和伦理实践。

二、算法设计的四重伦理立场

算法设计的伦理立场是研究相关伦理问题的核心议题，不同伦理立场直接决定了算法设计的实践倾向和结果。在不同社会背景中，设计者通常来自承载多元化价值的个体、群体或组织，这些主体所持有的不同价值立场深度影响其自身的道德信念、道德心理和道德实践，由此构成异质性、多元化的算法设计伦理立场。从算法设计的伦理立场的类别来看，大致可将其分为四大类，即功利论立场、义务论立场、契约论立场、德性论立场。这四种伦理立场在社会中是一种共存关联的形态，各自都有其存在的根据，也都有相应的拥护者和支持者。

1. 算法设计的功利论立场

功利主义以实际功效和利益作为衡量行为实践的主要道德标准和道德判断依据，并因以行为结果为基本导向而具有浓烈的目的论色彩，相关思想具有深厚的历史渊源。古希腊哲学家伊壁鸠鲁将快乐作为人类的至善追求，边沁创立了功利主义理论学说并提出趋乐避苦是人类的最高行为准则，穆勒进一步丰富和深化了这一理论。穆勒认为，功利原则或最大幸福原则是任何行为的目的，功利既有量的差别，又有质的差异，同时极力主张总体的功利而非仅仅个人的功利。[5]整体而言，功利主义者认为追求功利就是追求幸福和善，这不仅包括个人层面的功利性追求，而且突出以增加幸福总量为其最高目标。一方面，集体主义的功利论以追求“最大多数人的最大幸福”为宗旨，强调凡是促进最大多数人利益的行为都是道德的；另一方面，个人主义的功利论是一种极端利己主义的表现，强调一切行为都以自己的利益为出发点，并聚焦于个人利益最大化的道德诉求。因此，由于功利论的伦理立场以相关行为是否带来更多的福祉作为善恶判断标准，相关算法设计实践会以实际的功效或利益为基本目的，很少考虑动机和手段是否合道德。

功利主义把人的本质简单化和绝对化了，无法解释复杂多变的社会生活，这种以结果为导向的行为甚至有时隐藏着一种基于强权或“多数人暴政”的逻辑。对于算法设计来说，算法虽是工具性的机器，但它同样不能摆脱其所处的复杂社会场域，人的主观性对算法设计的影响是决定性的，当前以数据为主导的功利主义热潮就恰恰说明了这一点。在现实中，相关主体通常将海量数据和算法设计结合起来，来满足人们的需求和提高社会的运行效率，因此，为了基于大数据的算法所带来的便利和满足，需要让渡自由和隐私来换取。这种现象暗藏着人们对数据自由的崇拜，但是数据的自由并非意味着人的自由，恰恰相反，这种自由会

造成对人的隐私权和自主性的侵犯，从而导致人的齐一性，使人处于隐私安全隐患之中。从另外一个角度讲，“电车难题”伦理困境的根源性之一，很大程度上就是在算法设计中对功利与公平的辩争：算法设计应该以最大化地减少交通事故或风险为依据，还是应以每个人公平地承担事故及相关风险为参考？并且，“根据现有算法的种种局限，无人驾驶算法的未来方向应当是跳出功利主义与利已主义框架，且能够兼容应然与实然需求的第三种方向的算法。”[6]

对于基于功利论立场的算法设计实践而言，以产业发展、人类便利、社会效率等为“善”的标准，容易导致极端集体功利主义和极端利已主义伦理困境，可能隐藏人对数据和算法的过度崇拜，由此对算法治理的正义性形成巨大挑战。

2. 算法设计的契约论立场

契约论是规范伦理学的一种，注重规则在社会中的决定性作用。这种学说以自由平等原则为基点，强调“意志自由、选择自主、责权平等、责任自担”。[7]契约伦理思想在西方源远流长，古希腊时期便已萌芽并在政治领域应用较多，强调国家和法律的出现源自于人们之间的相互约定。近代以来，一批西方思想家将契约论作为反封建斗争的有力思想武器，卢梭认为“社会法则是一种神圣的权利，是其他一切权利的基础。这种权力并非源自于自然，所以它是以协议为基础的。”[8]这种协议（契约）一旦形成，个人就必须服从所约定的内容。罗尔斯以契约论为核心理论基点，提出人们订立契约的目的主要在于“确立一种指导社会基本结构设计的根本道德原则（正义原则）”。[9]契约论在现当代的人类社会发展中具有广泛而深刻的影响，在市场经济社会中更是如此，而算法设计在伦理依据上对其也有极大的依赖性。

一般认为，在市场经济条件下，相关主体应该遵守市场规则，各方应确立一种契约关系，明确规定享有的权利和应尽的义务，由此构成契约论的核心要义。契约本应是一种公开、公平、透明、平等的交易规则。然而，在算法无处不在的数字化时代，契约关系取代了传统的交易或合作形式，很多情况下变成了一种“形式契约”，劳动者或用户在跟企业平台订立协议过程中并不能享受一种具有实质意义的平等地位，因为通常情况下契约规则的制定权完全在平台企业一方，企业则出于自身利益最大化的考虑而将由其制定的不公正契约“强加”给对方，造成一种扭曲的不平等契约关系。这集中表明部分平台企业与用户之间所建立的契约关系的主观性和随意性，这种由新技术催生出来的“形式契约”并不具备人性化、合理化的特征，反而处处昭示着企业推卸责任的心态和不受约束的利益链条。从劳资关系层面来讲，数字劳动者逃离了装配生产线，却又深陷APP、智能算法等数字机器体系所制造的数字漩涡；企业借助无处不在的数字监控和数字化考核不断侵蚀劳动自主性，把数字劳工的劳动强度推向极限。[10]然而，对于普通用户来说，他们在现实中往往难以在这种形式化的契约中得到公正对待：一是用户往往不会认真阅读甚至不阅读协议条款内容，从而在让渡自身权利时并不自知或自明；二是用户处于被动接受协议的一方，其对算法技术及其运作方式一窍不通，特别是算法设计者的目标和意图往往隐藏在算法黑箱中，导致算法的不透明性等问题。

在算法设计的伦理治理中，契约论主张在企业和用户之间建立一种公开透明、平等公正的契约关系，这种理想对政府、企业、个人提出了较大的挑战。政府应从技术监督等角度监测和发现算法设计中潜藏的问题，并要从法律层面保证数据所有者的合法权益；企业要以开放的心态面向社会，通过向大众介绍其算法在内容分析、用户标签、计算规则等方式，加强企业、劳动者、用户之间契约关系的公平性；用户须对算法技术保持科学全面的认识，当遇到不合理的契约关系时需主动剥离，积极保护自身的合法权益。

3. 算法设计的义务论立场

义务论是一种尤为注重行为规范的经典伦理学说，是规范伦理学的突出代表。从理论层面讲，“义务论重视道德原则和行为，认为

道德行为的善恶要以道德原则为基础。”[11]因此，义务论者将行为动机放置于第一位，并将其作为衡量一切行为的终极道德标准，较为关心人们应该做什么以及应有的行为规范，强调道德义务和道德责任的价值理性，要求行为主体必须严格按照特定的道德原则或某种正当性去行动。在现实中，义务论者认为算法设计需要制定某种道德原则或按照某种正当性规则去实践，强调算法设计者的道德义务和责任以及履行义务和责任的重要性，其对于行为结果的道德性质关注不足。义务论与德性论关系密切，但前者更强调公共理性精神和公共伦理规范。因为算法设计者的道德责任是建构算法伦理的首要因素，而道德责任是为他人或社会所承担的道德义务。[12]

对算法设计设置伦理规则，主要存在两种不同的视角。其一，从具体的应用场景出发去制定伦理规则。这种观点是从结果正义的角度出发去纠正和改良算法设计，其所涉及的领域主要包括商业和公共决策两个方面。因此，应针对算法设计的主要应用场景，提出具体的伦理规制。例如，通过确立权责一致原则，明确企业、社会、个人之间的权责划分；通过确立安全原则，保护个体的信息安全和隐私安全；通过确立分配正义原则，避免数据垄断的发生；通过确立消费正义原则，确保消费者的利益不受算法设计的影响。从社会的具体应用场景入手来制定伦理规则，可以及时发现和纠正问题，确保社会整体的公平正义。其二，将道德规则嵌入算法设计中。这种观点则主张从算法伦理的正当性出发，去诉诸一种技术层面的伦理规制，及时规避风险。“道德物化”是当前技术伦理领域中的一个重要研究方向，其基本内涵是把道德善嵌入技术中，目前，这种理念大多应用在物理空间，但随着数字化时代的到来，这种“道德物化”理念需要移植到算法领域。有的学者提出“道德算法”的建构，而所谓道德算法，是“指那些在道德上可接受的算法或合乎伦理的算法，它们使自主系统的决策具有极高的可靠性和安全性。从这种意义上讲，道德算法是实现人工智能功能安全的一项基本原则和技术底线。”[13]

总之，基于义务论的算法设计伦理立场聚焦于如何使算法具有道德属性，主张通过设计出符合公共道德要求的算法，便于其在具体应用时遵循相关伦理准则和规范，最终使相关实践达到符合特定伦理道德的目的。这种观点要求算法设计本身符合伦理的要求，这虽然在很大程度上具有一定的合理性，但也容易让算法技术走向“完全的道德能动性”的误区。

4. 算法设计的德性论立场

德性论强调人的内在品质以及相应的道德实践和道德评价，并极为注重个体道德品质的形成和培养。中国传统儒家所倡导的“君子”人格、内圣外王等理论，就是德性论的典型代表思想。因为在德性论者看来，人的道德品质与其实践活动息息相关，恰如亚里士多德所言：“一个人的实现活动怎样，他的品质也就怎样。”[14]由于德性论强调向善的德性追求和修养，认为算法设计的伦理水平主要由算法设计者和算法企业的主体德性所决定，因此主张通过提升算法设计者的道德水平，为算法设计赋予更多的价值合理性。由此推论，算法设计者应该在算法开发过程中，通过运用技术手段将道德和公共善嵌入算法设计中，以此推动正向社会价值的发展。

由于算法设计的德性论立场更多地从算法设计者的责任视角出发，认为算法设计的伦理水平最终由算法设计者和算法企业主体的德性决定，其观点主要包括两点。其一，要提升算法设计者的道德水平。德性论主张通过提升算法设计者的主体道德水平，为算法设计赋予更多的价值合理性，并将此作为解决算法伦理问题的终极路径。这种观点认为，在价值层面，算法是技术人工物，其背后操作者是人，因此，算法的善恶并非在于技术，而是取决于人，算法设计者需要做的只是在算法设计中增加伦理的维度，让算法在决策中将伦理道德因素充分考虑进去，从而实现算法的道德化。其二，通过技术手段提升算法设计的合道德性。在实践层面，德性论立场认为应将公共善通过技术手段移植进算法设计当中。但是，要实现算法的

绝对善是不现实的，正如“自由”一样，要利用算法实现绝对的自由也是不切实际的；同时，问题的关键在于，将自由等价值物化地融入算法之中，并将这种被道德化了的算法设计与人们的社会生活联系在一起，在技术上还有很长的路要走。

在很多情况下，德性论被认为是解决算法设计伦理问题的“终极方法”，但由于个体的道德水平在现实中参差不齐，加之受到不同主体的道德认知差异、人性的不稳定性以及其他外部因素的影响，现实中极难实现既定的理想效果，其所面临的挑战也具有极端复杂性特征。

三、算法设计的四大伦理治理争论

算法设计产生的系列伦理问题不仅会形成四大伦理立场，而且在此基础上还会进一步从治理层面形成四重伦理争论。这些争论又构成算法设计治理的基本伦理场域——基于整合多维伦理立场和价值诉求的治理倾向。

1. 功利论下技治主义的非正义性

在功利主义价值取向下，技术对个人或集体的最大效用或效益成为关切的核心，这极易导致技治主义的伦理治理极端，尤为注重专业人员、专业机构、专业团体的重要作用。技治主义在很大程度上确实可以达到预期的高效率，但算法设计伦理治理的技治主义无法在程序、结果等层面实现真正的公平——即具有弱正义性特征，因为“纯功利主义算法函数中没有计算的因素：较不利者的优先性、风险与不确定性、应得与无辜等”。[15] 在功利主义主导下的算法设计实践中，存在两大显性伦理问题。一是伦理标准的悖论问题。例如，自动驾驶技术中的“电车难题”揭示出算法设计的伦理困境，极大地提高了在传统伦理框架内解决这类难题的难度。二是责任的豁免问题。功利主义做法使得算法设计游走于自主性和被动性之间，无法或很难在责任分配和豁免等层面对相关损失和伤害进行有效界定。

从技术层面看，算法技术的局限性问题主要体现在技术的不完备方面，包括其所存在的透明性、可解释性、可责性等伦理难题，这将导致算法设计实践难以达到公约化的伦理水平或状态。例如，“受算法模型的不完备、输入数据的结构性偏差以及技术越位等因素的影响，在城市智能治理过程中存在算法失灵风险”，[16] 从而引发一系列伦理治理风险。同时，算法本身的复杂性、专业性和封闭性让一般人难以理解这项技术，从而致使公众对算法伦理规约的参与度较低，无法更好地在这一技术发展过程中表达个人诉求与建议。因此，在治理层面“如何开发技术的工具理性并对其运用加以规制是决定技术适配治理的关键因素”，[17] 进而实现价值理性与工具理性的有机融合，成为化解技治主义治理伦理争论的必然之路。

2. 契约论下规则主义的强形式公正性

对于算法设计的伦理治理而言，契约论下的规则主义因强调行动自由和契约关系而具有强烈的形式公平色彩。在实践中，由于建立契约关系的双方在资源、地位等方面存在较大的位差，强势的一方通常因具有相对的议价优势而将其主观意图设计到契约规则之中，并以柔性的、看似合规合理的方式“逼迫”另一方接受其所提出的不公平契约条件，因此这一主张在现实中往往难以达到真正的实质性公正效果，而更多时候恰恰与公平正义背道而驰。

在现实社会中，这种基于自由契约导向的规则主义是导致公正问题的直接根源，涉及数据权力、隐私与信息安全、风险与责任等问题。一方面，平台或企业的算法系统从个人和用户中收取大量的个人数据，这种行为甚至是在主体不知情的情况下进行的，从而对数据权属问题造成巨大的伦理考验。另一方面，算法系统通过不透明的计算、剖析、解释与预测而不知不觉地介入或侵犯个人隐私，由此带来系列数据隐私、风险与责任等问题。平台所使用的算法通常会优先考虑企业的经济利益，因此企业通常会将自身利益最大化的诉求施加给算法设计者，同时忽略或减少对消费者、用户诉求的公平性考虑。与此同时，平台或企业按照上述设计意图生成一份完整的契约或合同，在签约过程中用户别无选择地被迫自愿放弃一些

权利，如数据权属、数据使用等权利，这样一来便会最终形成一种平台主导下的具有强烈形式公正性的“自由契约”。在这种治理情境下，企业在政府、企业、公众三角关系中明显处于相对的优势地位，这势必会与公平正义的治理理念相违背。

3. 义务论下规范主义的弱目的性

在算法治理理念方面，存在一些持义务论观念的群体。义务论将行为动机的善恶或正当性作为核心道德判断依据，强调行动者的实践应以自身的义务心和责任感为出发点，而较少关注行为结果的道德性质，其中道德规则或准则被当做主体行动的主要道德依据。由此可以看出，以规范主义为核心的义务论与以目的论为核心的功利主义具有本质的差异，前者注重行为动机的道德属性而忽略结果的善恶，后者注重行为结果的道德性质而忽略动机的善恶，因此义务论下的规范主义在道德层面具有明显的弱目的性特征。对于一个社会而言，理想的公共善可以看做一个全面性的道德系统，其中动机之善、程序之善和结果之善都是不可或缺的重要组成部分。

对于算法设计的伦理治理，如果仅仅将行为动机作为道德善的出发点而强调规则的唯一性，那么这将无法保证治理结果的道德善。故而，在算法设计的伦理治理中运用义务论理论时，要特别注意防止步入道德“形式主义”的极端，因为这将会导致忽略行为结果的道德善，而是应该在主体行为的动机、程序和结果等层面实现伦理设计的全覆盖，由此将公共善的理念贯彻于算法设计实践的始末。

4. 德性论下自律主义的不可控性

不论对于算法治理还是其他社会治理，德性论经常被认为是一种理想的伦理治理策略，持这种观点的人通常认为完全可以通过对个体德性的教育、引导、感化等途径，使社会全体成员普遍实现较高的道德水平，以此来保证每个个体都可以在道德上达至理想化的自律状态。这一逻辑预设极易导致一种行业自律主义的治理理念，认为只要通过对算法设计者或相关者的道德水平进行提升，便能够使他们在设计实践中按照公共理性和公共道德标准从事相关业务。德性论被很多人寄予厚望，并将其作为解决算法设计的终极伦理方法。这种观点确有其合理性，但其道路却异常艰辛，因为现有经济社会发展程度还远远没有达到“道德自律主义”对上层建筑所要求的水平，况且每个个体的道德水平在现实中难以全面掌控，这直接为规范伦理和强制性法律创造了存在的必要性。在自律主义的情境下，不但要求算法设计从业者和算法产品要按照行为主体内心的道德律令——道德自律去实践，而且还要强调多元道德主体之间的相互协作性：即按照公共理性的要求将公共善嵌入算法设计实践当中，并在政府、企业、公众等群体之间在相互信任、利益均衡的基础上实现多元化的道德协作。这在目前看来兴许还只是“空中楼阁”，尚不能作为算法设计伦理治理的唯一手段。

四、算法设计的伦理治理对策

算法设计所存在的差异化伦理立场，以及由此带来分化性的多极伦理治理倾向，使其陷入理论和实践的双重困境，如何对这些困境进行有效突破，是当前乃至未来各界面临的重要议题。尤为值得注意的是，即便算法设计的四重伦理立场和相应的四极伦理治理倾向都存在各自的弊端，但其有益思想资源都为解决伦理治理难题提供了重要理论参考。有鉴于此，本研究认为，算法设计的伦理治应该结合现实发展需要，有效融合功利论、契约论、义务论和德性论的有益思想，通过构建负责人的多元主体参与的治理网络，实现政府、专业机构和团体、企业、公众等主体的共同参与和协同治理。

第一，遵守算法向善的基本价值立场。面对算法设计的多元价值立场，必须坚持遵守算法向善的基本价值立场，强调以人类命运共同体为中心的价值理念，凝聚算法设计利益相关者的基本伦理共识，在此基础上共塑基于多元主体的正当性共同利益观，坚决避免算法作恶的倾向。

第二，坚持公共善的核心伦理导向。算法

技术的研发与应用应在公共善的导引下进行，算法技术的设计者、开发者及应用者不得损公肥私、损人利己，同时也要确保利益相关者的合法、正当的权益。例如，平台企业不能因自身不正当利益而违规收集、滥用、泄露公共数据，不得做出大数据杀熟、信息茧房、算法歧视等侵犯公众利益的行为；同时也要通过多重政策保障和治理举措，充分保障利益相关者的正当利益诉求。

第三，秉持负责人的可持续发展观。算法技术作为新兴科技的重要代表，是人类文明的结晶，各界应当对此持一种科学的可持续性发展伦理观。因为可持续性伦理建立在人类活动与自然界的普遍联系、和谐共生和善意合作基础上，强调整体性道德思维、短期利益与长远利益相统一的伦理意识以及高质量发展与高品质生活有机统一的发展价值观。[18]按照发展伦理思维，亟需推进算法伦理的道德创新，破解关键性道德难题或悖论，建设技术与道德协同并进的可持续发展的良性生态。

第四，实现算法设计道德化的全流程覆盖。算法设计是一种系统性技术–社会互动过程，涉及到的利益群体尤为广泛，涵盖技术研发、部署、应用等多个环节，因此必须对算法设计道德化进行全流程覆盖。其一，应在算法设计、部署、应用等阶段，按照负责任的理念实现“前瞻性”道德嵌入，而且在应用过程中还要对道德伦理规范进行“敏捷性”迭代更新与升级，以此来适应不同的道德应用情境。其二，也要对算法进行分类化的道德化处理，根据不同主体、不同情景采取针对性的处理方式，实现“因地制宜”的精准化施策效果。其三，秉持风险最小原则。坚守底线伦理思维，为算法设计实践活动设置“红绿灯”和划定技术边界，尽可能减少算法对社会造成的伤害或降低相关风险。一方面，应当尽量减少不必要的或禁止与相关业务无关的数据收集行为，将算法作恶的风险扼杀在摇篮；另一方面，通过技术化手段，实现数据隐私化的设计处理，如信息加密、匿名化等措施。

第五，完善善算法设计的伦理治理体系。其一，加强算法伦理教育工作，提升相关主体的伦理素养。在全国高校开展科技伦理教育，对确定的或潜在的从业人员普及算法及相关的科技伦理知识；在企业内部通过专业化的教育培训工作，提升算法从业人员的职业伦理素养；同时也要加强算法的科普工作，提升公众的伦理意识和素养。其二，建立和完善算法伦理监管体系。积极发挥国家及各级地方伦理委员会的监管作用，从政府层面加强对算法实践的伦理监督；构建企业算法伦理治理体系，建设企业算法伦理委员会，制定企业算法伦理治理的基本准则和规范；畅通公众监督渠道，提升社会对算法产品及算法设计主体的监督效果。其三，加速算法伦理的法治建设，明晰算法从业人员（包括企业本身及企业法人、管理者、设计者、技术人员等群体）的责任和权利，使相关主体能够在敬法、畏法中依法而行。其四，加强市场文化建设，倡导公平公正、守法诚信的市场契约精神。

结　语

在人类社会发展历程中，一些重要技术发明的问世和应用，都或多或少地会对传统社会伦理秩序和伦理文化产生一定的挑战，人类也几乎都可以在自身的不断努力中实现与技术的“和谐相处”。目前，算法技术的发展和应用带来的伦理问题对人类社会的影响也越来越广泛、深刻，不仅关涉生产生活方式的变革性发展，同时也对人类文化文明的重塑起到至关重要的作用，对这项技术进行科学合理的运用将会为人类带来更多福祉，如若对其处理不当则会造成严重社会问题。算法设计的伦理问题作为一项复杂的系统工程，对其科学治理不仅需要集合各类优秀伦理思想资源并在此基础上进行融合式创新，而且还必须进行跨学科、跨界的全方位协同与合作，以此在超越单极伦理理论立场和单方主体的局限性中实现负责任的可持续性发展。

［参 考 文 献］

[1] 孙保学. 人工智能算法伦理及其风险 [J]. 哲学动态，

2019,(10) : 93–99.

[2] Kraemer, F., Van Overveld, K., Peterson, M. 'Is There an Ethics of Algorithms?'[J]. *Ethics and Information Technology*, 2011, 13: 251–260.

[3] 肖红军. 算法责任: 理论证成、全景画像与治理范式 [J]. 管理世界, 2022, 38 (4) : 200–226.

[4] Ignas, K. *Algorithmic Governance: Politics and Law in the Post-Human Era*[M]. Cham: Springer Nature Switzerland AG, 2019, 29.

[5] 约翰·斯图亚特·穆勒. 功利主义 [M]. 刘富胜 译, 北京: 光明日报出版社, 2007, 20.

[6] 隋婷婷、张学义. 功利主义在无人驾驶设计中的道德算法困境 [J]. 自然辩证法研究, 2021, 37 (10) : 112–117.

[7] 李欣隆. 交易伦理视阈下公民道德责任强化研究——契约论、义务论、美德论的视阈 [J]. 道德与文明, 2022,(2): 170–176.

[8] 让–雅克·卢梭. 社会契约论 [M]. 杨国政 译, 西安: 陕西人民出版社, 2004, 2.

[9] 约翰·罗尔斯. 正义论 [M]. 何怀宏、何包钢、廖申白 译, 北京: 中国社会科学出版社, 1988, 6.

[10] 黄再胜. 人工智能时代的价值危机、资本应对与数字劳动反抗 [J]. 探索与争鸣, 2020, 367 (5) : 124–131.

[11] 王淑芹. 诚信道德正当性的理论辩护——从德性论、义务论、功利论的诚信伦理思想谈起 [J]. 哲学研究, 2015,(12) : 72–77.

[12] 郭林生、李小燕. “算法伦理”的价值基础及其建构进路 [J]. 自然辩证法通讯, 2020, 42 (4) : 9–13.

[13] 李伦、孙保学. 给人工智能一颗“良芯 (良心)”——人工智能伦理研究的四个维度 [J]. 教学与研究, 2018, (9) : 72–79.

[14] 亚里士多德. 尼各马可伦理学 [M]. 廖申白 译注, 北京: 商务印书馆, 2003, 37.

[15] 王珀. 无人驾驶与算法伦理: 一种后果主义的算法设计伦理框架 [J]. 自然辩证法研究, 2018, 34 (10) : 70–75.

[16] 张龙辉、肖克. 城市智能治理中的算法失灵及消解策略 [J]. 电子政务, 2022,(7) : 98–112.

[17] 阙天舒、吕俊延. 智能时代下技术革新与政府治理的范式变革——计算式治理的效度与限度 [J]. 中国行政管理, 2021,(2) : 21–30.

[18] 乔法容. 可持续性发展: 发展伦理观的深刻革命与价值重构 [J]. 伦理学研究, 2021,(3) : 97–104.

[责任编辑 李斌]