



# “硬法—软法”范式下医疗人工智能伦理治理路径探析

徐 辉, 王 译

湘潭大学法学院, 湖南 湘潭 411105

**摘要:**以硬法规范为主干的传统法律科技伦理规则治理具有人机交互“不信任”、决策“不可控”、安全性“不可靠”、医疗资源分配“不公平”等客观局限,无法提供善治之良法所需。为此,强调“开放—控制”一体化理论结构的“硬法—软法”二元法模式能够为医疗人工智能伦理治理提供新的治理框架。理论上,“硬法—软法”范式能够完成医疗人工智能伦理风险规制的合法性论证,确保医疗人工智能伦理风险决策的回应性;实践中,“硬法—软法”范式已然成为国外医疗人工智能伦理治理的通行治理框架。未来,“硬法—软法”范式下医疗人工智能伦理治理路径应当通过不断规范软法的生成机制,健全硬法软法的良性互动机制,充分发挥硬法和软法各自的治理优势并实现制度互补,以构建安全、可信、负责的医疗人工智能伦理治理框架。

**关键词:**医疗人工智能;人工智能伦理;通用人工智能;软法;硬法

中图分类号:R-052

文献标志码:A

文章编号:1671-0479(2024)04-359-010

doi:10.7655/NYDXBSS240169

## 一、问题的提出

随着以ChatGPT为代表的基于大语言模型技术的通用人工智能快速迭代发展,人工智能应用于医疗保健与公共卫生领域的场景也愈加广泛。截至2023年9月底,国内外超过260个医疗健康大模型人工智能(generalist medical artificial intelligence, GMAI)应用于医疗保健、科学研究、公共卫生和药物开发等领域,例如英伟达推出的生物医药基础模型BioNeMo、英特尔研发的生成式模型Aurora genAI、百度基于“文心”大模型底座的“灵药大模型”等。与此同时,随着人工智能广泛应用于医疗领域,其所提供的个性化、精准化、交互性、超时空性的病因推断、精准医疗、疾病预测,给公民生命权、健康权带来价值和优势的同时<sup>[1]</sup>,医疗人工智能技术与生态发展的不成熟衍生出一系列伦理问题与风险亦是绕不过的现实存在<sup>[2]</sup>,造成了社会公众广泛的信任危机、伦理危机和社会危机。不同于医务人员在从医过程中需要遵循的各项医学伦理,人工智能虽然被冠以“人”之名,但其并非人类,不具备人格

或人的情感、意识、道德,而且科技伦理教育课程建设刚刚起步<sup>[3]</sup>,因此其伦理状况也无从评估。尤其是,不同于人工智能在其他领域的应用,在医疗保健与公共卫生领域,医疗人工智能的应用意味着人类对自然对象的控制和改造开始从人类外部转移到人类自身<sup>[4]</sup>,直接关系到公民的健康权和生命权。因此,如若不对医疗人工智能伦理风险进行治理,便将其直接投入规模化的临床实践与社会应用中,不仅可能会直接或间接地引发诸多不可控的伦理风险,甚至可能会给人工智能在医疗领域的可持续发展制造障碍,进一步恶化技术与社会间的信任关系。由此可见,我们既要紧抓人工智能技术发展的契机,推动、容纳和接受这一新兴技术为医疗卫生保健领域提供新的机遇、突破和创新,更要认识到这一新兴技术对医疗领域可能造成的系统性伦理风险。

目前科学界逐渐认识到,单靠“以技治技”的治理逻辑不仅不能解决医疗人工智能实践中出现的技术伦理问题<sup>[5]</sup>,甚至可能会再次造成技术伦理的遮蔽,因此转而强调法律形塑医疗人工智能伦理发

**基金项目:**国家重点研发计划重点专项“智慧司法科学理论与司法改革科技支撑技术研究”(2020YFC0832400)

**收稿日期:**2024-04-20

**作者简介:**徐辉(1998—),男,江苏盐城人,硕士研究生在读,研究方向为数字检察;王译(1988—),男,湖南邵阳人,博士,副教授,研究方向为检察制度,通信作者, hnfjd@126.com。

展方向<sup>[6]</sup>，“以制适技”的法治范式规制理念和路径开始突出<sup>[7]</sup>。但是，风险社会的逻辑背景下现代性导入了新的风险参量<sup>[8]</sup>，因而医疗人工智能伦理治理的法治体系需要探求新的伦理制度补给。一方面，以传统硬法规范为主干的法律系统面对医疗语境下人机交互不信任、不可靠、不可控和不公平等风险特征的医疗人工智能伦理，存在“一立法就落后”的常规误区与人工智能技术特性及硬法规范差异性的多重矛盾。另一方面，以命令控制型规制为主的传统政府治理，遵循的是对抗、威慑和法律服从的制度逻辑，使得以政府为主导的监管模式在应对以信任关系为存续纽带的医患关系时容易陷入两难境地。

综上所述，治理视角下医疗人工智能伦理治理需要兼顾法律控制、开放自愿、协商行政、公众参与。因此，本文尝试通过“硬法—软法”范式的运行机理和作用方式，为医疗人工智能伦理危机治理逻辑提供一种新的框架，不断引导技术向善，发展安全、可信、负责的医疗人工智能伦理。

## 二、风险审视：医疗人工智能伦理风险的诱发机制

技术不仅是推动人类步入风险社会的动因，也是风险社会的主要表征<sup>[9]</sup>。风险社会下，医疗人工智能在医疗和护理教育、科学研究与药物开发、疾病诊断和临床保健等医疗和卫生保健领域实践应用过程中的作用日益凸显，同时医疗人工智能亦存在违背既定医学伦理规范的僭越风险，不断增加医疗人工智能伦理规制的复杂性，形成了多种挑战。

### (一) 算法黑箱的存在导致人机交互的“不信任”

从工具属性上看，随着医疗人工智能技术的迭代发展，机器学习等人工智能算法可以通过模拟临床决策者的认知过程，辅助临床医师进行诊断和决策，及时做出疾病预警与检测、防控救治以及临床诊疗，从而提高临床医师诊断效率和准确度，例如人工智能在放射医学和医学影像、结核病和肿瘤学等领域的辅助诊断。虽然医疗人工智能的应用可以提高疾病诊断和临床决策的客观性，确保临床诊断与决策结果的正确性、疾病治疗与预防效果的有效性<sup>[10]</sup>，对提高患者的健康水平和生命质量起到积极作用，但也不可避免地造成医师对医疗人工智能的依赖程度增加，导致医师的主体性和自由裁量权可能面临被侵蚀的风险<sup>[11]</sup>，进而为算法权力的隐形扩张提供了可操作空间。算法权力的扩张意味着原本辅助医师作出临床决策的人工智能逐渐成长为影响医师临床决策、取代医师临床决策的一种权力。这种医师对自身专属权力的让渡，要求医师保证医疗人工智能所作出的临床诊断与决策结果的

真实性、完整性和准确性，并对其进行合理解释。然而，算法黑箱致使医疗人工智能临床诊断与决策过程缺乏透明性，使得医师受制于专业知识壁垒以及专属所有权<sup>[12]</sup>，无法就医疗人工智能临床诊断和决策过程中的原理和机制向患者履行告知义务，更做不到对可能产生的错误预测履行再判断义务。

可靠性和透明度是医患关系中建立信任的关键，而医疗人工智能的不透明性可能会破坏这种具有天然脆弱性的医患信任关系<sup>[13]</sup>。从本质上看，人机交互语境下的医疗人工智能应用其实质是医患之间的信任。面对人体机理的复杂本质，医疗人工智能技术的复杂性和算法的不透明性使其应用饱受质疑<sup>[14]</sup>。出于认知的局限，当患者难以理解人工智能临床诊断和决策的原理和机制时，并不会将临床诊断和决策的错误归因于人工智能技术，而是转嫁给人因因素，即使用医疗人工智能的医师，这不可避免地会导致患者与医师之间面临信任崩塌的风险。由此可见，医疗人工智能算法的黑箱性质在医患之间制造着鸿沟，这道鸿沟将在彼此之间持续不断地制造警惕、矛盾和不信任。

### (二) 人文精神的缺位导致医疗人工智能决策存在“不可控”

医学正在回归人类，回归社会、回归人文发展<sup>[15]</sup>。正如先贤所言“医乃仁术”，“术”与“仁”相伴相生。自然语言处理技术(natural language processing, NLP)作为人工智能技术的一项最重要的底层技术，其通过整合与集成人类输入提示、人类反馈强化学习以及基于思维链的应用推理等多项技术<sup>[16]</sup>，不断提升对人类情感的识别、理解、表达和生成功能的模拟。该项技术对于医疗人工智能辅助医师作出临床诊断至关重要。不仅如此，NLP还能模仿医师与患者之间的交流方式，从而给人工智能算法注入类似人类“情感”的品质，使其生成和提供的内容看似和人类一样，并据此给予患者更拟人化的情感疏导和支持，实现对患者疾痛及精神负担的“同理共情”。从这个角度看，“既有理智又有情感”的人工智能技术似乎符合人性的需求，实现了对个体生命尊严与价值的尊重。然而，正如世卫组织《卫生领域人工智能的伦理与治理》指南所述，人工智能会带来重大风险，可能会危及个人尊严。暂且不论人工智能情感技术还停留在鹦鹉学舌的功能模仿阶段，相较于人类真实情感的自然生发，这种基于机器学习与训练、情感数据量化与分析的“虚情假意”往往缺少人文关怀<sup>[17]</sup>，可能带来负面影响，存在“情感操纵”的可能性。如2023年3月，一名比利时焦虑症患者同一款名为ELIZA的聊天机器人进行数周的对话后自杀身亡<sup>[18]</sup>。此外，人机关

系的构建可能会从根本上改变医患关系,形成“医务人员—人工智能—患者”的不良医患关系。医患之间直接交流的减少,使得医患关系变得疏远、冷漠,进而影响医患之间感情的表达和传递,即所谓的“医疗公害”<sup>[19]</sup>。由此可见,医疗人工智能人文关怀的缺失不仅造成了“人一技”问题的双向异化,还引发了不可控的伦理风险。

(三)虚假信息的生成导致医疗人工智能安全性“不可靠”

已有研究发现,人工智能在3%~27%的情况下可能会生成虚假信息(又称人工智能的“幻觉”)<sup>[20]</sup>,且具有速度快、体量大和逼真度高等特征<sup>[21]</sup>。人工智能虚假信息生成的缺陷对于被规整之其他生活场景的应用似乎并无大碍,但是在医疗场景下,人工智能“幻觉”的缺陷却十分有害,它会客观上加大临床医生和医护专业人员误诊误治的可能性,给患者生命和身体健康造成实质性危害。尤其是,医疗人工智能的应用势必会加快患者和非医学专业人士将人工智能用于医疗目的的趋势,在缺乏专业医学知识储备以及医务人员先验把关和提示的情况下,患者和非医学专业人士自行贸然使用人工智能生成的治疗建议进行自我诊断,很可能带来诊断误读或误用的风险。例如医疗人工智能沃森在训练时存在“多起治疗建议不安全、不正确的例子”,甚至给一名癌症患者开出了可能导致“严重乃至致命性出血”的药物<sup>[22]</sup>。可见,当前的医疗人工智能无法掌握医学知识和临床决策的全部复杂性,也无法完全复制临床医生的经验和细致的判断,导致医疗人工智能安全性“不可靠”。尽管可以通过进一步提示,诱导医疗人工智能识别出自身的错误,但必须审慎对待人工智能在医学应用时产生的“幻觉”缺陷,否则很有可能导致医护人员出错,给患者带来人身损害。

(四)数据采集偏见与算法偏见导致医疗资源分配“不公平”

只有对技术的历史一无所知的人才会相信技术是完全中立的,每种技术都有自己的议程,都是等待被揭示的一种隐喻<sup>[23]</sup>。作为集成与整合强算法、强算力、大数据三位一体的人工智能技术,其诱发社会公正风险主要源于数据采集偏见与算法偏见两个方面。医疗人工智能的应用本应发挥缓和医疗资源分配不平衡、消除社会公正风险的作用,但实际情况是,由于数据采集偏见与算法偏见的存在,反而再现了“医疗歧视”,加剧了医疗鸿沟。宾夕法尼亚州立大学一项研究表明,在与“残疾”相关的描述或病例中,大模型会在86%的时间将情绪得分从积极转换为消极<sup>[24]</sup>,如将残疾患者识别为“非人类”“没有生命”的对象。此外,医疗人工智能用

以训练的医疗数据以及算法大都来源于高收入国家,因而任何人工智能的回应都可能被已占据经济、技术或者某种资源的优势群体所优先占有<sup>[25]</sup>,最终进一步造成医疗资源分配的不平等、不公平。值得担忧的是,强人工智能时代下,随着技术中心主义在医疗领域的排他场景不断深化<sup>[26]</sup>,医疗人工智能装备对“残疾人的不友好”在以后可能会被认为是“残疾人群体”的自身缺陷,而忽视“技术不合格”是导致相关医疗资源分配不均的真正原因。2020年《新英格兰医学杂志》就曾发文呼吁紧缺医疗资源分配应遵守四项原则,包括最大化健康收益原则、平等对待生命原则、促进工具价值原则以及优待弱势群体原则<sup>[27]</sup>。因此,医疗人工智能数据采集偏见与算法偏见的存在可能会成为更为隐蔽地造成人类社会不平等的新方式,成为医疗人工智能向前发展时极易滑向的伦理深渊。

### 三、医疗人工智能伦理风险硬法治理的现状及其客观局限

(一)医疗人工智能伦理风险硬法治理现状回顾

新兴科技的迭代进步及其应用进程的快速推进使社会治理已经迈入人工智能时代,相关法律、伦理、社会问题的研究氛围也相当活跃<sup>[28]</sup>。目前,我国医疗人工智能相关的法律体系尚未完全建立起来,尚缺乏统一的标准化的质量标准、准入机制、评估体系和保障体系,因而医疗人工智能伦理风险治理的硬法规范散见于各个法律法规之中,主要有《互联网诊疗监管细则(试行)》《互联网诊疗管理办法(试行)》《中华人民共和国基本医疗卫生与健康促进法》《人工智能辅助诊断技术管理规范》《涉及人的生命科学和医学研究伦理审查办法》《人工智能医用软件产品分类界定指导原则》《中华人民共和国数据安全法》《中华人民共和国民法典》等基础性、综合性、全局性的法律、行政法规和部门规章。上述法律法规对数据、算法等内容提出了相关规范性指导,共同规范了医疗人工智能设施的建设、运营和使用,构建了医疗人工智能伦理治理的大体制度框架。但是客观地分析,医疗人工智能所具备的颠覆性技术能力及其带来的不信任、不可控、不可靠、不公平等伦理风险,导致医疗人工智能应用场景下所存在的社会变量远超传统政府独占控制机制与硬法规范框架所构建的调整范围,不仅造成现行硬法规范所构建的医疗人工智能伦理治理领域场意涵不清,边界不明,在保护效果上严重不足,难以适应医疗人工智能的未来发展,而且造成硬法规范的引导、调整作用不强,无法涵盖医疗人工智能伦理治理所需。以医疗人工智能伦理风险所造成的医疗损害责任纠纷为例,依据《中华人

《中华人民共和国民法典》第1218条规定,“患者在诊疗活动中受到损害,医疗机构或者其医务人员有过错的,由医疗机构承担赔偿责任”,但是根据第1221条规定,我国判断医务人员诊疗过失的基本标准是“当时医疗水平”标准,即以实施诊疗活动之时的医疗水准确定的诊疗义务为判断标准<sup>[29]</sup>。一方面,2020年最高法出台的《最高人民法院关于审理医疗损害责任纠纷案件适用法律若干问题的解释》第16条明确规定,“对医疗机构或者其医务人员的过错,应当依据法律、行政法规、规章以及其他有关诊疗规范进行认定”(以下简称“诊疗规范标准”);另一方面,无论是“当时医疗水平标准”抑或“诊疗规范标准”指向的都是安全性和有效性已经得到确立的、较为普及的疗法,并不能涵盖那些刚起步的、有效性和安全性尚未确立的、创新型的甚至是“试验性”的疗法<sup>[30]</sup>,如医疗人工智能的应用。如此一来,医疗场景下与医疗人工智能伦理风险相关的医疗损害责任纠纷问题将会变得极为复杂,不可避免地导致“责任空场”,造成传统硬法规范所制定的责任形式难以套用到医疗人工智能的场景中来。

## (二) 医疗人工智能伦理风险硬法治理的客观局限

### 1. 不确定性伦理风险打破了硬法规制的前提预设

不确定性揭示的是世界的复杂性和多样性,而技术作为人造的复杂系统更会增加不确定性。有学者指出,机器学习技术的运作“似乎杜绝了所有形态的确定性”<sup>[31]</sup>。从风险的视角看,基于多模态大语言模型的医疗人工智能所具备的不信任、不可控、不可靠、不公平等风险特征,使其不仅在“输入—输出”的过程中存在着人类无法窥探的“隐层”,即算法黑箱的存在,而且输入的预训练数据与输出的信息还引发了数据歧视、算法偏见、虚假信息生成等风险,导致其产生的伦理问题也具有极大的不确定性和风险性。在风险社会中,对于医疗人工智能崛起所带来的不确定性伦理风险及其引发的法律挑战的事实变化,硬法规范所构建的治理框架却依旧采取“假定、处理、制裁”模式,致力于通过明确清晰的语言设定一套具体的前提预设,要求相关主体为或不为某种行为,并赋予一种具体后果的法令,试图通过预设的制度框架、强制性规范以及“命令—控制”规制模式以调整医疗人工智能的不确定性伦理风险。然而实际情况是,在风险的不确定性领域,硬法规范有时不可避免地存在“立法僵滞”和“立法空白”的缺陷,导致以硬法规范为主干的部门法科技法律伦理规则面对医疗人工智能不同治理主体以及应用的目标、诉求与场景差异所造成的伦

理治理难题,要么予以规避或否认医疗人工智能的不确定性伦理风险,要么在承认医疗人工智能的不确定伦理风险存在的同时却仍诉诸科学理性以寻求确定性知识,由此产生“不确定性悖论”,导致部分硬法规范束之高阁,甚至“失效”。硬法规范这种单纯“向后看”确认事实变化的前提预设,无法为未来医疗人工智能伦理治理提供规范指引,导致“公地悲剧”现象频发。但需要强调的是,尽管硬法规范缺憾使得其治理医疗人工智能的不确定性伦理风险存在一定限制,但这并不必然意味着硬法规范对医疗人工智能相关伦理风险的规制是不必要的。事实上,随着人工智能技术的发展,相关伦理问题将会逐渐从不确定性变得越来越确定,硬法的规范功能也会随之上升。

### 2. 缺乏回应性无法应对医疗人工智能技术伦理嬗变之现实

在医疗保健领域,新兴技术的发展速度、传播范围和复杂程度都远超以往,已表现出深度数智化的发展趋势。从宏观上看,医疗人工智能的广泛应用引起了诊断、治疗过程时空概念的变化,即人机协作诊疗模式改变了传统医疗的诊治活动场景。医疗场景时空变化的剧烈与迅速对传统社会伦理、道德、观念的挑战是全方位的,使得以硬法规范为主干的传统法律科技伦理治理规则面对技术伦理嬗变之现实有着前所未有的需求性和紧张性。传统规范主义模式下,社会和法律制度的变迁受控于法律实施的结果,而医疗人工智能伦理治理的制度需求显然超越了当下法律规则的制度供给,加剧了医疗人工智能伦理治理的“步调难题”(pacing problem)。硬法规范治理医疗人工智能伦理风险出现“步调难题”主要源于两个方面:一是立法的高成本和低效率;二是硬法监管惯性思维的存在。首先,为维护法律自身的权威、稳定、连续与体系化,法律制定和修改大都是由国家正式立法机关经过提案、审议、表决、公布等程序通过,并且前期会配套如过滤机制、顺序机制、立法调研机制等运作机制来保证立法的科学性和民主性,等到相关法律颁布,人工智能技术可能已经完成了数次迭代改进。以ChatGPT为例,从GPT3.5到GPT4.0再到Sora,才不过仅仅经过了1年时间,因而两者存在无法适配的难题。相较于硬法立法的高成本和低效率,软法的特点则是非正式性<sup>[32]</sup>,因而其能够快速对技术变化进行回应,实现对医疗人工智能伦理风险更加敏捷、灵活的治理。另外,行政监管部门为精准适用法律,对相关医疗人工智能伦理问题的治理较为“死板”,存在惯性思维,这种以“不变”应“万变”的形式法治面对急剧变化的医疗人工智能伦理问题总是相互掣肘、难以两全。

#### 四、“硬法—软法”范式下医疗人工智能 伦理治理的理论之维和实证之维

(一)“硬法—软法”范式下医疗人工智能伦理治理的理论基石

1. “硬法—软法”范式能够通过建立软硬并重的二元法模式完成医疗人工智能伦理风险规制的合法性论证

从宪法和法律价值的角度出发,宪法至上和法律中心主义是“硬法—软法”范式的合法性基础,即行政机关的治理行为必须依据法律作出,且不得对宪法中的基本权利造成减损。这是因为中国法治建设奉行的其实是“一元多样的混合法”模式,亦即以宪法为所有法规范之“元”,各类法规范都应服从宪法;在宪法之下,硬法规范构筑了法律体系的基本骨架,在此基础上,各类软法规范在国家管理、社会治理中各就各位,以不违反硬法规范规定的方式实行政治共同体的自律和社会共同体的自治<sup>[33]</sup>。基于“软硬并重”的二元法医疗人工智能伦理治理模式同样尊重宪法的至高地位,遵从宪法共识。因此,为防止过多地强调国家强制力在立法和执法中的作用而对基本权利造成减损,一方面,基于硬法规范的医疗人工智能伦理治理可以通过权力制约与司法审查的方式侧重解决更为普遍、共性、基础的人的权益保障问题,另一方面,为将宪法的最高法律效力规范作用于所有组织和个人,防止软法规范游离于法治之外,造成“软硬并重”的二元法医疗人工智能伦理治理模式合法性的自我削弱,可以通过比照硬法规范设定软法的价值目标、把软法纳入统一的法律体系、软法“立法”时召开听证会等方式予以监督。

相较于硬法的司法中心主义模式,人机协作诊疗模式对传统医疗模式在“行为—结果”上的时空断裂的冲击,不可避免地造成了因果链条难以追溯、举证责任困难、诉讼主体不明确以及责任分配复杂等法律形式主义的适用局限,导致当事人(尤其是患者)往往难以通过硬法程序规定的争议解决方式(如诉讼、仲裁、行政复议)获得及时、有效、低成本的救济。此时,软法的一个重要优势则可以体现出来,软法程序可以将医疗人工智能伦理治理的难题引出司法程序之外,在遵从合法性基础之上,通过协商、和解、申诉、调解等争议解决方式由共同体内部的监督救济(如医疗人工智能伦理监督委员会)或双方协商化解医疗人工智能伦理风险引起的矛盾纠纷。当然,作为一种法律领域内的权利和义务,软法权利和义务在某些方面仍然要接受硬法的规制,因而,软法权利和义务的争议并非绝对排斥司法审判的最终裁决。

2. “硬法—软法”范式能够通过开放的调整机制与多元的规制场域,确保医疗人工智能伦理风险治理的回应性

在法治现代化语境中,医疗人工智能伦理风险的法律治理需要保障以全社会的创新收益为基本点,保障治理制度在利益冲突解决、社会问题控制和科技公共管理等方面的作用<sup>[34]</sup>。医疗人工智能作为“社会—技术”系统<sup>[35]</sup>,加之医疗语境下人工智能所独有的语境敏感性、非封闭性和独特的信任关系三个特征,创新、应用和扩散的实质是一种技术与社会共生交互、相互联结的持续协同演化过程,伦理风险治理必然需要政府、医疗机构、企业、研发人员、病患与消费者等存在利益相关性的各类行动者和发起者之间广泛协商、协议、理解和支持,以确保达成广泛共识的医疗人工智能伦理风险治理的一致行动性。但是,缺乏法律控制的公众参与极易导致位于强势一方的政府、医疗机构、企业等群体有组织地利益压倒和规则俘获<sup>[36]</sup>。对此,“硬法—软法”范式能够很好地解决该问题。一方面,硬法通过提供具有约束力和可执行的法律规定划定各方的权利与义务,并配以强制力的监督保障参与主体权益,避免产生失衡的权利博弈结构,实现在组织运行层面的多主体自愿平等持续性对话,促进达成医疗人工智能伦理治理的共同理解和认可;另一方面,软法通过国家政策、行业标准、道德准则等提炼医疗人工智能伦理治理的价值原则与诉求,给予各方利益诉求回应,确保不同治理主体不同诉求的畅通表达。“硬法—软法”范式下依靠多元主体参与,共同实现医疗人工智能朝着人类福祉、公平公正、隐私安全、可控可信、责任担当等价值目标轨迹发展的伦理愿景,进而消除社会的担忧和疑虑。

此外,作为源于法律多元主义的软法理论,其在构建中所追求的和谐、灵活、利导、软约束、契约型法、外包法等多元目标价值和多元制度设计能够补充和修正硬法规范在功能和作用上难以应对具有不确定性、复杂性、适应性和突现性等特征的医疗人工智能伦理风险的缺陷<sup>[37]</sup>。虽然软法规范所具有的软约束力不是正式的法律制裁和惩罚,也不算司法裁判强制执行的依据,但是却属于国家认可或默认的外在强制力。因此,行政公共部门在治理医疗人工智能伦理风险时不仅可以遵循硬法规范,还可以在硬法规范失效时,根据价值导向选择匹配适用,由非正式国家立法机关的专业性非营利组织所制定的医疗人工智能伦理准则作为裁量基准,实现制度补给与技术指导。

(二)“硬法—软法”范式下医疗人工智能伦理治理的域外实践

2021年,亚利桑那州立大学的一份人工智能软

法治理报告显示,截至2020年,全球范围内共有634个人工智能软法项目<sup>[38]</sup>。其中,美国是最具代表性的国家(26%),欧洲、英国和国际联盟占62%。从国际层面上看,“硬法—软法”范式已然成为全球医疗人工智能伦理治理的通行监管样态<sup>[39]</sup>。

### 1. 欧盟

欧盟正处于医疗领域的人工智能革命边缘,如何在人工智能创新发展和风险治理之间实现平衡,是当下欧盟所面临的艰巨任务。从欧盟层面来看,受制于立法从属原则,欧盟一直缺少在欧盟层面的医疗领域相关立法,缺乏专门针对医疗人工智能伦理风险治理的具体法律框架,只能通过《医疗器械条例》《通用数据保护条例》部分涵盖医疗人工智能伦理风险的监管。因而,在人工智能伦理风险治理的初期,欧盟通常先采用专家建议、政策文件或伦理指南的方式进行“软”治理,例如《欧洲机器人技术民事法律规则》《欧盟人工智能》《可信赖的人工智能道德准则》以及《走向卓越与信任——欧盟人工智能监管新路径》等。随着《欧洲人工智能培养方法》出台,欧盟开始从具有非强制约束力指导方针的软治理,转向要求制定人工智能法案(AIA),尝试建立新的人工智能产品和服务监管框架,通过对人工智能软治理的同时探索提供硬法规范支持,进而规制人工智能风险。AIA的目标在于改善人工智能内部市场的运作,充分保护《欧盟基本权利宪章》所规定的基本权利和价值观,以建立对人工智能信任的生态系统,使欧盟成为全球人工智能监管的领军者。

总的来说,欧盟的医疗人工智能监管一直采用的是较为包容和弹性的“软”监管模式,但从总体趋势上看,欧盟对于医疗人工智能的监管正由“软”逐渐走向“硬”。

### 2. 美国

美国作为人工智能的发源地,为鼓励人工智能创新和发展,维护自己在人工智能领域的领导地位,2019年以前联邦政府都尽量避免制定具有强制执行效力以及具体义务要求的硬法规范,而选择采用软性的原则性倡议或规范,以避免妨碍人工智能创新和增长的不必要的监管或非监管行动。但随着人工智能医疗器械软件在医疗领域日益广泛地应用,以及其逐渐产生的负面影响,美国的监管态度有所变化,逐步由大力推进医疗人工智能创新发展转向支持安全、可信、负责的医疗人工智能技术发展的监管策略。2019年4月,美国食品与药品监督管理局(FDA)执行了名为《人工智能医疗器械软件变更监管框架提议》的新监管方法,旨在通过制定相关变更控制规范,实现对人工智能医疗器械风险的全生命周期监管。之后根据该提议,FDA在接受了利益相关者的建议后,于2021年1月再次发布

了《人工智能医疗器械软件行动计划》。同年7月美国联邦问责署(GAO)正式制定了具有强制性约束力的《人工智能问责框架》,迈出了美国联邦层面上的人工智能监管第一步。此后,美国联邦层面人工智能监管立法井喷式增长,仅去年一年里就提出了181项人工智能相关的法案,并且通过了25项法案,增长率高达56.3%。2023年7月,美国参议院提交的《人工智能生物安全风险评估法案》指出,要将人工智能在医疗保健和医学方面的应用风险监管纳入美国国家卫生战略;10月,拜登政府签署发布《安全、稳定、可信的人工智能》行政命令,该行政命令要求对国家公共健康和安全构成严重风险的医疗人工智能系统进行模型训练时需通知联邦政府,而且必须分享安全测试结果和相关数据,以确保不会对公民生命和健康造成威胁。

除了联邦政府以及各州政府颁布法令外,美国还向学术协会、研究机构等第三方组织寻求合作,通过组织层面“软治理”与系统层面的“硬监管”的耦合<sup>[40]</sup>,建立医疗人工智能系统伦理方针和伦理标准,进而实现医疗人工智能系统的良法之治。

### 3. 英国

为维持自身在人工智能领域的领先地位,英国采取的是与美国相似的监管策略,即相对宽松、非法定义务的原则性监管模式,倾向于依靠现有法律法规而不是正式的人工智能专项立法。随着人工智能技术潜在风险的不断衍生,英国的监管模式也正在经历着由“软”到“硬”的变化过程,开始逐渐强调监管的重要性。

2019年,为了回应人工智能在医疗领域应用所引发的公众信任危机,英国信息专员办公室与英国国家卫生研究所举办了一系列“公民陪审团”,以获取各个社会阶层的公众对于在医疗领域使用人工智能的看法与意见<sup>[41]</sup>,此次会议证明了多元主体参与协商的医疗人工智能监管体系的合理性;同年5月,英国国家卫生与临床优化研究所(NICE)和英国国家医疗保健服务(NHS)联合发布了《数字健康技术证据标准框架》,对人工智能在医疗领域的应用制定了10项原则性规定<sup>[42]</sup>;2021年9月,英国医疗和健康产品管理局(MHRA)制定了一项名为《作为医疗设备的软件和人工智能变更计划》的监管改革计划,旨在以指南的形式提供一个健全的监管框架,用以加强对医疗设备中的人工智能(AI-MDs)不断增长的算法偏见、算法黑箱风险的监管监督<sup>[43]</sup>;2022年11月,英国科学、创新和技术部(DSIT)下属的监管地平线委员会(RHC)颁布了《医疗人工智能监管规定》,该文件旨在提升患者、社会公众在医疗人工智能全生命周期监管中的参与程度,以改善监管方、制造方与使用方之间的信任关系。

2023年以来,随着ChatGPT为代表的通用人工智能在医疗领域的迅速崛起及其可能带来的伦理风险,英国开始更新监管策略。在此背景下,2023年第二季度,英国《BS30440:人工智能在医疗保健中应用的验证框架》正式颁布,将原散见于各个法律中的医疗人工智能监管整合成一个全面的框架,为了确保标准的多样性和包容性,该标准是由学术界、政府机构、医疗机构、专业组织组成的专家委员会磋商形成的<sup>[44]</sup>。

综上所述,欧盟、美国、英国基本形成一种软硬混治的医疗人工智能伦理治理格局。一方面,均在初期通过提供非法义务的软法在宏观上引导医疗人工智能产业自律,督促企业自发制定标准规范自身行为,以便灵活、敏捷地回应医疗人工智能伦理风险;另一方面,为弥补软法治理强制力不足的缺陷,各国开始将一些软法转化为硬法或直接颁布硬法提供具体规范,逐步转向更为成熟的“软法与硬法双轨并行、刚柔并济”的新型治理模式<sup>[45]</sup>。

## 五、“硬法—软法”范式下医疗人工智能伦理治理之未来路径

(一)“软硬混治”:医疗人工智能伦理法治体系的制度诉求

虽然各国在政治立场、意识形态、利益分配等方面均存在很大差异,但是就医疗人工智能伦理治理的需求是一致的。面对具有伦理风险的医疗人工智能,尚没有任何单一国家主体依靠强制力可单独或共同包揽医疗人工智能伦理风险问题的解决,给医疗人工智能伦理风险治理留下了一定的空白与盲点,形成了不同程度上的“法治真空”<sup>[46]</sup>。而法律的缺位恰恰容易使无节制的裁量、有组织的利益以及混乱的公共意见掌握医疗人工智能伦理风险规制的话语权<sup>[36]</sup>。这种影响的多样性使得医疗人工智能伦理风险治理格局更加复杂化。由此可见,传统“国家间立法”模式所注重的硬法一元主导下的刚性约束方式,已经无法单方面垄断治理医疗人工智能伦理风险这一全球难题,需要其他非国家行为体针对医疗人工智能应用的各个场景中产生的伦理风险,共同制定软法方案,补齐硬法治理的制度短板。因此,“硬法—软法”范式下通过将软法自治与硬法强制引导结合起来形成的软硬交错、刚柔并济的二元法体系进化,能够不断填充医疗人工智能伦理风险治理的法律空白,更好地回应社会关切与需要。

其一,相较于硬法的静态法,软法既可以是静态的法规范,也可以是动态的公共治理方式、治理手段<sup>[47]</sup>。对于医疗人工智能伦理风险治理,神经网络模型的发展极大地增强了机器学习能力,导致硬

法静态式的“硬性治理”无法及时、准确、灵活地回应医疗语境下人工智能算法黑箱的存在、人文精神的缺位、虚假内容的生成、自动化偏见所导致的人机交互“不信任”、决策“不可控”、安全性“不可靠”、医疗资源“不公平”等伦理风险。相较之下,软法制定的自主性、灵活性、开放性、多样性能在前期利益相关方动态合作博弈中,不断试错和自我优化,提供“软性治理”,为安全、可靠、负责的医疗人工智能开发和使用预先提供治理框架,为未来法律法规制定时能够充分了解医疗人工智能技术前沿的实际经验和伦理风险奠定基础。

其二,法律规范的作用在于指引人们的行为。当前,我国医疗人工智能伦理治理正处在初创阶段,普遍存在法律不严、执法不力、审批不严、标准不一、企业合规程度不够等问题,法律法规与医疗人工智能标准脱节,无法为医疗人工智能科技行为提供规范指引作用。在法律规范指引缺位的情况下,行业协会、社会组织、民间团体等非国家行为体制定的行业标准和技术标准,如《国家新一代人工智能标准体系建设指南》《人工智能伦理治理标准化指南》等作为一种软法,是相关硬法在鼓励“发挥行业协会、科学研究机构和学术团体”之下的利益指引与矫正,不仅能够提供指引医疗卫生领域内的各机构、单位或从业人员从事医疗人工智能技术操作时需要遵循的行为规范或准则,还能够为行政管理部门开展行业执法提供实证性标准。

其三,软硬混治通过自律自治和他律强制相结合的规制方式,能够不断调动公与私两个方面规范医疗人工智能伦理风险的积极性。医疗人工智能作为一种“社会—技术”系统,要有一种多元化的治理方式。一方面,行业协会、企业组织、社会组织、民间团体等非国家行为体通过民主协商、达成共识形成自律公约或合规指南规范,促进医疗人工智能以合法合规与合乎道德的方式开发、使用,例如中国网络社会组织于2021年联合发布的《互联网信息服务算法自律公约》和《互联网行业从业人员职业道德准则》。另一方面,硬法通过整合国家强制力保障软法协商博弈的有序性和公平性。软法硬法相互协调、相互配合、相互促进,通过整合国家强制与社会自治两种资源,共同推进医疗人工智能伦理治理。

(二)规范软法的生成机制,补齐软法正当性短板

从立法的权力来源看,各种非国家行为体制定软法中存在不规范的现象,导致软法存在比较严重的“正当性”缺失问题,严重阻碍了“硬法—软法”范式中的软法治理效能。因此,必须规范非国家行为体制定软法的程序以及软法的内容,补齐软法的正当性短板。

从软法制定程序上看,由于软法的制定是一种

利益导向机制,即利益相关方都能在软法形成的过程中表达自己的利益诉求,对利益相关方的利益都能够有所考虑,因而企业之间达成的自律规范、行业标准很可能受竞争策略、营销模式、消费偏好等利益博弈的影响,而带有一定的价值偏向性。在缺乏监督的情况下很可能产生有组织的利益俘获,进而完全由医疗人工智能行业的利益群体所主导,导致患者、社会公众程度的参与程度较低,利益不能得到充分的代表。因此,应当加强对软法协商机制的内外部监督,从源头上保证软法创制的民主性、创制过程的民主参与性与程序正当性,最大限度地协调好“权利—权利”“权利—利益”和“利益—利益”之间的关系,进而保证各方主体的权益都能够得到充分的表达。

从软法制定的内容上看,软法的自主性、灵活性、开放性、多样性是其能够及时、准确、敏捷地回应医疗人工智能伦理风险的关键,但是也造成其存在着创制层级较低、缺乏统一的标准、创制主体专业化程度不高、随意性强和内容模糊等问题,严重制约了软法的实施效力以及权威性。为此,可以设立具有行政和咨询职能的机构、专家组、伦理委员会等伦理治理的专项主体,提供软法背书与法理依据<sup>[48]</sup>,增强软法内容的科学性。例如2019年,国家药品监督管理局联合国家相关部委、科研机构、医疗机构等共同成立人工智能医疗器械创新合作平台,下设立技术法规、数据治理、网络安全等十个工作组,通过合作平台推动各工作组开展多层面、多角度科学研究。此外,可以通过出台规范性文件、建立备案审查机制以及负面清单的方式,对软法规范进行监控,防止软法成为医疗人工智能企业躲避政府监管的“避风港”。

### (三)健全硬法软法的良性互动机制,促成软法硬法配套协调

软法发挥作用的一个重要机制在于其能够“依靠与硬法之间的复杂关联性,利用硬法的国家强制力的某种暗示或影响,促使软法目标的实现”<sup>[49]</sup>。软法代表了现代法治的发展方向,它虽然是非常理想化的模型,但医疗人工智能的不可靠、不可信、不公平和不可控的社会伦理风险影响更为广泛和深刻,并不如其所设想的那般完美,单靠软法肯定会独木难支,甚至适得其反。此时,与软法在创制和实施机制上截然对立,从而在很大程度上能够避免它在现实应用中局限性的硬法机制就有用武之地了<sup>[50]</sup>。可见,“硬法—软法”范式的有效应用依赖于硬法软法的良性互动以及软法硬法的配套协调。如前文所述,医疗人工智能作为一种开放、多层次和多要素的新型社会—技术系统,其伦理风险的治理属于典型的公共治理场域,离不开产业行业、专家团体、民间组织、科研机构等非国家行为体制定的软法提供更为多元、

理性的价值共识,共同影响未来医疗人工智能伦理发展。但是软法的软约束力或无约束力,使得软法无法像硬法那样因具有国家强制执行而使自身顺利施行。为了确保医疗人工智能伦理软法能够发挥其应有的秩序构建作用,可以借助硬法的刚性提升医疗人工智能软法的“硬度”。具体而言,可以在医疗人工智能伦理治理领域构建起软法硬法之间的良性互动机制,促成软法硬法相互影响、相互转化。

一是在条件成熟时,将软法所制定的医疗人工智能伦理行业标准或技术标准直接转化为硬法,进一步明确认定伦理标准或技术标准的法律效力。例如欧盟《可信人工智能伦理指南》《算法责任与透明治理框架》倾向于规定医疗人工智能伦理的混合性条款,即借鉴软法的开放协调程序以定期修订条款,通过模糊硬法与软法的边界,使软法和硬法的双重手段在实践中优势互补、共同推动医疗人工智能伦理建设进程<sup>[51]</sup>。

二是修改硬法,在相关硬法中明确规定违反医疗人工智能伦理所要承担的法律后果。增加诸如“违反本法授权制定之医疗人工智能伦理标准的,依照本法/条例/办法第XX条规定处理”之类的法律责任指引条款,产生违反医疗人工智能伦理硬法的效果,提升软法的实施效力<sup>[52]</sup>。例如德国《社会保障法》第5卷第92条中所规定的决定某人是否生病并因此无法工作的实质性标准,就是一个由医生和医护职业者组成的行业协会所制定的。

三是调动硬法和软法的互补机能。硬法和软法共同构成了医疗人工智能伦理治理的规范体系,并且以各自的方式发挥着作用。为确保医疗人工智能遵从增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险、保持公开透明等科技伦理原则,既需要国家话语权为背书的硬法提供自上而下的“刚性着落”,也需要企业、学术团体、医学伦理学会、民间组织在内的机构发布大量的伦理原则及指南等相关软法,提供自下而上的“软性控制”。因此,在医疗人工智能伦理治理现代化背景下,应当充分调动硬法和软法的互补机能,通过刚性制度约束和柔性控制之合力共同破解医疗人工智能伦理风险治理困境<sup>[53]</sup>,以符合社会对值得的、更好的“公共善”的要求和期待<sup>[54]</sup>。

### 参考文献

- [1] 张兆阳,赵允伍,王晓松,等. 主动健康视角下的“互联网+社区健康管理”[J]. 南京医科大学学报(社会科学版),2023,23(2):138-143
- [2] 赵丹娜,赵蓝蓝,陈任,等. 某三级公立医院医务人员对人工智能医学应用伦理问题的认知研究[J]. 南京医科大学学报(社会科学版),2022,22(5):465-470



- [3] 郑和园,王灿. 课程思政背景下新时代科技伦理教育的整体要求、现实困境及提升路径[J]. 南京医科大学学报(社会科学版),2023,23(2):178-182
- [4] 李正风,刘瑶瑶. 科技伦理治理要准确把握新科技革命及其伦理问题的新特点[J]. 科学通报,2024(13):1677-1680
- [5] 瞿晶晶,王迎春,赵延东. 人工智能社会实验:伦理规范与运行机制[J]. 中国软科学,2022(11):74-82
- [6] MITTELSTADT B. Principles alone cannot guarantee ethical AI[J]. Nat Mach Intell,2019(1):501-507
- [7] 石佳友,徐靖仪. 医疗人工智能应用的法律挑战及其治理[J]. 西北大学学报(哲学社会科学版),2024,54(2):91-103
- [8] 乔治·瑞泽尔. 后现代社会理论[M]. 谢立中,译. 北京:华夏出版社,2003:208
- [9] 乌尔里希·贝克. 风险社会[M]. 张文杰,何博闻,译. 南京:译林出版社,2004:190-224
- [10] LIANG H Y, TSUI B Y, NI H, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence[J]. Nat Med,2019,25(3):433-438
- [11] 毋文文. 法治国家的健康观[J]. 南京医科大学学报(社会科学版),2023,23(5):415-422
- [12] 徐辉,李俊强. 生成式人工智能辅助司法裁判的可能及边界限度[J]. 太原理工大学学报(社会科学版),2023,41(6):24-32
- [13] 刘月树. 我国医患信任问题探析:因由与重建[J]. 南京医科大学学报(社会科学版),2012,12(4):282-285
- [14] SHORTLIFFE E H, SEPÚLVEDA M J. Clinical decision support in the era of artificial intelligence[J]. JAMA,2018,320(21):2199-2200
- [15] 杜治政. 论现代临床医学发展的人文走向[J]. 医学与哲学(人文社会医学版),2008,29(6):2-6,11
- [16] SINGHAL K, AZIZI S, TU T, et al. Large language models encode clinical knowledge [J]. Nature, 2023, 620 (7972): 172-180
- [17] 谢瑜,王潇毅. 人工智能情感的伦理风险及其应对[J]. 伦理学研究,2024(1):132-140
- [18] WALKER L. Belgian man dies by suicide following exchanges with chatbot, The Brussels Times [EB/OL]. [2024-04-25]. <https://www.brusselstimes.com/430098/belgian-man-commits-suicidefollowing-exchanges-with-chatgpt>
- [19] 郭永松. 生命科学技术与社会文化:生命伦理学探究[M]. 杭州:浙江大学出版社,2009:35
- [20] METZ C. Chatbots may ‘hallucinate’ more often than many realize, The New York Times [EB/OL]. [2024-04-16]. <https://www.nytimes.com/2023/11/06/technology/chatbots-hallucinationrates.html>
- [21] 舒洪水,彭鹏. ChatGPT场景下虚假信息法律风险与对策[J]. 新疆师范大学学报(哲学社会科学版),2023,44(5):124-129
- [22] 李明子. 沃森遇冷,国内AI医疗过热[J]. 中国新闻周刊,2018(44):68
- [23] 尼尔·波兹曼. 娱乐至死,童年的消逝[M]. 桂林:广西师范大学出版社,2009:15
- [24] LYERLY E. Study finds AI language models show bias against individuals with disabilities[J]. Disability Compliance For HE,2023,28(12):9
- [25] LI H Z, MOON J T, PURKAYASTHA S, et al. Ethics of large language models in medicine and medical research [J]. Lancet Digit Health,2023,5(6):e333-e335
- [26] 元光. 人工智能的社会公正风险:核心问题、理解视角与呈现类型[J]. 政治学研究,2023(6):64-75,232
- [27] EMANUEL E J, PERSAD G, UPSHUR R, et al. Fair allocation of scarce medical resources in the time of COVID-19[J]. N Engl J Med,2020,382(21):2049-2055
- [28] 邱绍继. 人工智能、法律解析与未来法律实践[J]. 政法论丛,2022(4):83-93
- [29] 周友军. 侵权法学[M]. 北京:中国人民大学出版社,2011:258
- [30] 郑志峰. 诊疗人工智能的医疗损害责任[J]. 中国法学,2023(1):203-221
- [31] 泰娜·布策. 被操弄的真实:演算法中隐藏的政治与权力[M]. 叶妍伶,罗亚琪,译. 新北:台湾商务印书馆,2021:116
- [32] ABBOTT K W, MARCHANT G, CORLEY E A. Soft law oversight mechanisms for nanotechnology [J]. Jurimetrics, 2012: 279-312
- [33] 罗豪才. 软法与治理评论—第一辑[M]. 北京:法律出版社,2013:37
- [34] 罗亚海. 生成式人工智能法律治理现代化研究[J]. 湖北大学学报(哲学社会科学版),2024,51(1):144-153
- [35] 孙启贵,汪琛,王加宇,等. 医疗人工智能发展的社会—技术分析启示[J]. 自然辩证法研究,2021,37(3):48-53
- [36] 董正爱. 环境风险的规制进路与范式重构——基于硬法与软法的二元构造[J]. 现代法学,2023,45(2):112-124
- [37] 梁剑兵,张新华. 软法的一般原理[M]. 北京:法律出版社,2012:28
- [38] GUTIERREZ C I, MARCHANT G, CARDEN A, et al. Preliminary results of a global database on soft law mechanisms for the governance of artificial intelligence [C]//2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G). Geneva, Switzerland.IEEE,2020:33-37

- [39] PALANIAPPAN K, LIN E Y T, VOGEL S. Global regulatory frameworks for the use of artificial intelligence (AI) in the healthcare services sector [J]. *Healthcare (Basel)*, 2024, 12(5): 562
- [40] 张惠彬, 何易平. 平台算法监管的困境与出路——基于美国算法监管模式的研究[J]. *科学学研究*, 2024(7): 1409-1428
- [41] POWELL J, DEETJEN U. Characterizing the digital health citizen: mixed-methods study deriving a new typology [J]. *J Med Internet Res*, 2019, 21(3): e11279
- [42] Evidence Standards Framework (ESF) for Digital Health Technologies. Available online [EB/OL]. [2024-04-21]. <https://www.nice.org.uk/about/whatwe-do/our-programmes/evidence-standards-framework-for-digital-health-technologies>
- [43] DARZI A. The regulation of artificial intelligence as a medical device [R]. London, UK, 2022
- [44] SUJAN M, SMITH-FRAZER C, MALAMATENIOU C, et al. Validation framework for the use of AI in healthcare: overview of the new British standard BS30440 [J]. *BMJ Health Care Inform*, 2023, 30(1): e100749
- [45] 张欣, 宋雨鑫. 全球人工智能治理的格局、特征与趋势洞察 [J]. *数字法治*, 2024(1): 199-212
- [46] 廉睿, 卫跃宁. “硬法之维”到“软硬共治”: 民族地区生态治理的理路重构 [J]. *学习论坛*, 2019, 35(5): 82-88
- [47] 姜明安. 完善软法机制, 推进社会公共治理创新 [J]. *中国法学*, 2010(5): 16-24
- [48] 汪琛. 医疗人工智能伦理治理的问题、困境与求解 [J/OL]. *科学学研究*, 1-18 [2024-04-19]. <https://doi.org/10.16192/j.cnki.1003-2053.20240313.001>
- [49] 姚迈新. 软法之于公共治理的作用机制探析 [J]. *探求*, 2009(5): 39-43
- [50] 程迈. 软法概念的构造与功能 [J]. *金陵法律评论*, 2009, 18(1): 103-119
- [51] 朱明婷, 徐崇利. 人工智能伦理的国际软法之治: 现状、挑战与对策 [J]. *中国科学院院刊*, 2023, 38(7): 1037-1049
- [52] 刘长秋. 作为软法的行业标准研究——以卫生行业标准为视角 [J]. *北京理工大学学报(社会科学版)*, 2013, 15(2): 108-116
- [53] 任学丽. 社会信任模式变迁视阈下的医患信任困境及出路 [J]. *南京医科大学学报(社会科学版)*, 2018, 18(3): 176-181
- [54] 何雷. 法治对国家治理现代化的硬规制、软约束与矫治功能探析 [J]. *领导科学*, 2016(5): 23-25

(本文编辑:姜鑫)

## An analysis of the ethical governance path of medical artificial intelligence under the paradigm of “hard law soft law”

XU Hui, WANG Yi

School of Law, Xiangtan University, Xiangtan 411105, China

**Abstract:** The traditional legal and technological ethical rules which based on hard law norms have inevitable objective limitations when governing medical artificial intelligence ethics with risk characteristics such as lack of trust in human-computer interaction, uncontrollable decision-making, unreliable security, and unfair distribution of medical resources, which make it impossible to provide the necessary for good governance. Therefore, the dualistic legal model of “hard law soft law” that emphasizes the integrated theoretical structure of “openness-control” can provide a new governance framework for the ethical governance of medical artificial intelligence. Theoretically, the “hard law soft law” paradigm can demonstrate the legitimacy of ethical risk regulation in medical artificial intelligence and ensure the responsiveness of medical artificial intelligence ethical risk decision-making. In practice, the “hard law soft law” paradigm has become a common governance framework for ethical risk management in medical artificial intelligence abroad. In the future, under the paradigm of “hard law soft law”, the ethical governance path of medical artificial intelligence should continuously regulate the generation mechanism of soft law, and enhance the benign interaction mechanism of “hard law soft law”. We also recommend giving full play to the governance advantages of hard law and soft law to achieve institutional complementarity, intending to construct a safe, trustworthy, and responsible ethical governance framework for medical artificial intelligence.

**Key words:** medical artificial intelligence; artificial intelligence ethics; artificial general intelligence; soft law; hard law