

论医疗人工智能伦理标准体系的构建

黎 燚 彭飞荣*

(中国计量大学)

摘 要: 随着人工智能在医疗领域的应用日益深入,伴随而来的伦理风险和法律挑战亟需关注。当前医疗人工智能面临失控性、社会性、侵权性、歧视性、责任性等多方面的伦理风险。考虑到专门针对医疗人工智能的法律规制尚处于起步阶段,本文提倡先构建一套伦理标准体系,以填补当前监管空白,确保技术的健康发展。医疗人工智能伦理标准体系的构建应满足全面性、可操作性、适应性以及普适性与地域性等原则,并优先重点建设伦理原则、数据管理、算法透明、责任界定、监督评估等核心领域的伦理标准。通过促进多方共建、加强国际合作和强化教育培训等配套措施,推进我国医疗人工智能伦理标准体系的建立和完善。

关键词: 医疗人工智能, 伦理风险, 伦理标准, 标准构建

DOI编码: 10.3969/j.issn.1674-5698.2024.07.012

On the Construction of the Ethical Standards System for Medical Artificial Intelligence

LI Yi PENG Fei-rong*

(China Jiliang University)

Abstract: As artificial intelligence becomes increasingly integrated into the medical field, the accompanying ethical risks and legal challenges urgently demand attention. Medical artificial intelligence currently faces a multitude of ethical risks including the issues of control, societal impact, infringement, discrimination, and responsibility. Considering that legal regulations specifically addressing medical artificial intelligence are still in their nascent stages, this paper advocates for the establishment of an ethical standards system to fill the current regulatory gaps and ensure the technology's healthy development. The construction of a medical artificial intelligence ethical standards system should adhere to the principles of comprehensiveness, operability, adaptability, as well as universality and regional specificity, with a focus on developing ethical standards in core areas such as ethical principles, data management, algorithm transparency, responsibility delineation, and supervision and evaluation. By promoting collaborative efforts among multiple stakeholders, strengthening international cooperation, and enhancing education and training, it aims to advance the establishment and refinement of China's medical artificial intelligence ethical standards system.

Keywords: medical artificial intelligence, ethical risks, ethical standards, standards development

0 引言

随着第四次工业革命的来临,人工智能

(Artificial Intelligence, AI)技术已经成为推动社会进步和经济发展的重要力量。近年来,人工智能技术正广泛地与各个行业相融合,医疗健康领域

作者简介: 黎燚,硕士研究生,研究方向为知识产权法。

彭飞荣,通信作者,法学博士,教授,国家知识产权培训(浙江)基地副主任,研究方向为知识产权法、标准化法。

正是其部署和应用的重点领域之一。目前人工智能在医疗领域的应用主要包括智能诊疗、智能医疗机器人、智能影像识别、智能药物研发、智能健康管理等方面^[1]。医疗人工智能的出现给传统医疗服务带来了革命性的变化,极大提高了诊疗效率和准确性,为广大患者带来了福音。不幸的是,伴随技术进步而来的还有各种伦理挑战。数据隐私的泄露、算法偏见的产生、决策过程不透明以及责任归属模糊等问题的出现,都在不同程度上引发了社会公众对于医疗人工智能的争议和担忧。如何妥善应对医疗人工智能伦理问题,促使其朝着健康的方向发展,是当下亟需解决的问题。

人工智能伦理是开展人工智能研究、设计、开发、服务和使用等科技活动需要遵循的价值理念和行为规范^[2]。医疗人工智能伦理问题是当前学界研究的热点,研究主题正从原则规范向治理体系演进^[3]。目前我国在人工智能领域的立法工作处于起步阶段,尚未建立起全面的人工智能规范体系。考虑到人工智能技术发展迅速、更新迭代快,依靠立法对其进行规制难免存在滞后。而在法律出台之前,标准能够在一定程度上替代规则,承担起丰富人工智能技术伦理价值内涵的任务^[4]。因此本文主张先行建立医疗人工智能的伦理标准体系对其进行治理,待时机成熟再进行立法。

本文在分析医疗人工智能面临的主要伦理风险的基础上,结合国内外的相关研究和实践经验,提出优先重点建设核心部分的伦理标准,之后再逐步对伦理标准体系进行完善的构想,以期对医疗人工智能伦理问题的解决提供一种可能的方案。

1 医疗人工智能面临的主要伦理风险

全国信息安全标准化技术委员会秘书处组织制定和发布的《网络安全标准实践指南——人工智能伦理安全风险防范指引》指出,开展人工智能相关活动需要关注失控性、社会性、侵权性、歧视性以及责任性等5个方面的伦理安全风险。本文将从上述5个方面展开分析当下医疗人工智能面临的主要伦理风险。

1.1 失控性风险

失控性风险指的是医疗人工智能的行为和结果超出了预设和可控的范围而引发的相关风险。首先,随着算法的不断发展,出现了深度学习算法之类的高级算法,其决策过程极其复杂,即使是开发者也难以完全理解其内部工作机制。这种“黑箱”效应使得医疗人工智能的决策难以被预测和追溯,其风险自然也就不可控了。其次,基于机器学习的医疗人工智能系统可能会进行自我学习并调整其行为模式,这种自主性可能导致其作出开发人员设计预期之外的医疗决策,造成不可预见的后果。此外,医疗人工智能系统本身可能存在安全漏洞,如果被黑客侵入或者恶意篡改,就可能导致医疗机构失去对系统的控制,进而可能对患者的人身造成损害。

1.2 社会性风险

社会性风险是指因不合理使用医疗人工智能,如:滥用、误用,而产生的相关风险。许多医护人员在使用医疗人工智能系统之前并没有经过全面的培训,很可能因不完全理解系统的工作原理和风险而错误地应用这些技术,危及患者健康。同时,对医疗人工智能的过度依赖可能会逐渐削弱医生的专业判断能力,患者对医生的信赖程度也会不断降低。此外,当前医疗人工智能的相关应用程序、智能穿戴设备(智能手表、智能手环等)都非常容易获取,越来越多人将其用于日常的个人健康测试或监控,但其结果的正确性往往无法得到保证,盲目地信任此类人工智能的医疗决策可能会因错误信息而导致健康受到损害。

1.3 侵权性风险

侵权性风险是指医疗人工智能对人的基本权利造成侵害的风险。当下医疗人工智能的应用可能会涉及对患者健康权、隐私权、自主权等基本权利的侵害。首先,医疗人工智能可能会因其训练数据不足、算法缺陷或其他技术问题而导致不正确的诊断,可能对患者健康造成损害。其次,健康数据相比于其他个人数据,具有更高的敏感性和私密性^[5]。如果医疗人工智能系统在处理和分析患者健康信息时,没有做好相应的安全措施,就可能导致

数据泄露,致使患者的隐私权遭受侵犯。最后,诊疗过程中患者自主权的核心在于患者的知情同意权利是否得到尊重。医疗人工智能系统的决策过程往往缺乏透明性,患者无法理解其决策依据,也就无法理解智能诊疗选择的后果,进而削弱患者的自主决策能力。

1.4 歧视性风险

歧视性风险是指医疗人工智能对人类特定群体抱有偏见,缺乏公平公正,而引发的相关风险。医疗人工智能系统的运行主要依靠数据训练和算法决策,数据集质量不佳或算法设计存在问题,均有可能导致算法决策带有偏见。首先,由于数据收集者主观上的有意选择,或者是客观上样本太少、难以获取,均可能致使数据在性别、年龄、种族、地区或其他重要因素上缺乏多样性。医疗人工智能系统采用此种数据集进行训练可能会导致算法带有系统性偏见。其次,医疗人工智能的设计者,出于主观偏好或者利益导向等原因,在设计算法时没有秉承公正的理念,也可能导致算法最终输出的结果带有偏见。此外,由于医疗人工智能属于新兴技术,有些国家或地区并没有足够的资金和技术支持来搭建医疗人工智能基础设施。医疗人工智能资源分配不均,可能在落后地区和低收入群体中形成人工智能医疗资源的应用“洼地”,违背“健康公平”的理念^[6]。

1.5 责任性风险

责任性风险是指医疗人工智能相关方的责任难以界定所引发的风险。在传统医疗模式中,医疗机构和医护人员一般是医疗事故责任承担的主体。但是人工智能技术的介入,使得责任划分变得复杂起来。一方面,医疗人工智能系统的决策过程往往是不透明的。当人工智能出现误诊、漏诊等情况时,难以追踪和解释导致错误的各种原因,因果关系的复杂性导致责任难以归属。另一方面,医疗人工智能系统的应用涉及包括设计者、制造者、使用者等在内的多方主体。一个以人工智能为中介而导致的伦理问题,往往不是“一己之力”,而更多的是“众人合力”^[7]。这种情况下界定各方的责任边界变得十分复杂。对于人工智能本身是否需要

承担责任的问题,法律对此当前并无规定。学界一般认为基于当前的技术发展阶段,医疗人工智能只是一种工具,不具备法律主体资格,在将来医疗人工智能有自由意志属性之后再赋予其法律人格为时未晚^[8]。本文赞同这一观点。

2 医疗人工智能伦理标准体系的核心内容构建

面对医疗人工智能伦理风险引发的一系列问题,当务之急是构建起相应的伦理标准体系,防止问题的进一步扩大,影响到医疗人工智能产业向前发展。在伦理标准体系的构建过程中应遵循以下原则:

(1)全面性原则,医疗人工智能的伦理标准体系应当覆盖其整个生命周期,建立从研发、生产、测试、应用到退役下线的全过程伦理标准体系。

(2)可操作性原则,医疗人工智能伦理标准的设定应当是符合实际并且具体可行的,确保其能够在实践中得到有效落实。

(3)适应性原则,医疗人工智能技术的更新速度极快,其面临的伦理问题可能也会随之发生变化,因而伦理标准体系需具备一定的灵活性和适应性,便于根据技术的发展进行动态调整。

(4)普适性与地域性原则,在确保医疗人工智能伦理标准具有国际通用性的同时,考虑到本国或者地区在文化、法律和社会背景上的不同,允许其存在一定程度的地域性差异。

鉴于建立医疗人工智能的全生命流程伦理标准体系是一个复杂且庞大的工程,本文主张优先建立伦理标准体系中的伦理原则、数据管理与隐私保护、算法透明度和可解释性、责任与合规性以及监督与评估机制等几个核心部分的伦理标准,以应对当下日益频发的伦理问题。

2.1 伦理原则标准化

目前国内外都尚未形成统一的医疗人工智能伦理原则。国内方面,我国传统医学伦理包括有利、尊重、公正和不伤害等四大原则。2022年3月,中共中央办公厅、国务院办公厅发布了《关于加强

科技伦理治理的意见》，提出了以下5项科技伦理原则：增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险和保持公开透明。国际上比较有代表性的是世界卫生组织（WHO）在2021年6月发布的《卫生领域人工智能的伦理与治理》指南，该指南确定了医疗卫生领域人工智能的6项伦理原则：（1）保障人类的自主性；（2）增进人类福祉和保护安全及公共利益；（3）确保透明度、可解释性和理解性；（4）培养责任感和实行问责制；（5）确保包容性和公平性；（6）促进具有响应性和可持续性的人工智能发展。

伦理原则是制定具体伦理标准的基础和方向，因此制定一套通用的标准化伦理原则势在必行。借鉴上述已有的相关伦理原则，本文提出以下5项医疗人工智能基本伦理原则：

（1）以人为本原则。医疗人工智能的开发和应用应当首先考虑和保护人类的利益，将增进人类健康和福祉作为根本目的。注重保障患者的生命安全和人格尊严，确保医护人员和患者的决策自主性。

（2）公正性原则。医疗人工智能应当确保对所有患者公正无私，不因患者在经济状况、性别、年龄、种族等方面的差异而对其进行差别对待。同时，公正性原则也要求对医疗资源进行合理分配，确保所有人都能平等地享受到医疗人工智能带来的便利。

（3）隐私保护原则。医疗人工智能系统应当确保涉及患者隐私的个人健康信息获得妥善管理和保护，对个人健康数据的使用必须依法依规进行，保证患者对自己健康信息的处理拥有完全的知情权和选择权。

（4）责任与问责原则。当医疗人工智能引发医疗事故时，应当有明确的责任归属和问责机制，确保患者权益得到保护。应当明确包括开发者、制造者、医疗机构及医护人员等主体在内承担的不同责任，确保所有责任均可追溯。

（5）透明度与可解释性原则。医疗人工智能系统的设计和应用应当具有透明性，并且能够以人类可以理解的方式进行解释，使得智能系统的相关人员能够理解其运作方式和决策依据。

2.2 数据管理与隐私保护伦理标准

数据管理与隐私保护伦理标准的构建，主要是为了确保医疗人工智能应用过程中的数据收集、存储、处理和共享符合伦理要求，确保不侵犯患者的个人隐私。医疗人工智能的相关数据在收集和选取上应当符合公平性，在构成上需满足代表性、均衡性和规模性等要求。对于不符合要求的数据集，应当对其进行进一步的处理和校准，减少其中的偏见和不平衡后再用于医疗人工智能系统训练。在获取医疗数据时，应告知患者有关数据收集、存储和处理的目的，保障患者知情权，并取得其同意。患者对其个人健康数据应当拥有访问、修改和删除等权利。在医疗数据的存储、使用和传输的过程中，有关主体必须采取包括数据加密、访问控制、身份验证和安全审计等在内的安全措施，防止数据的丢失、泄露、滥用和篡改，防止患者隐私受到侵犯。同时，应当防范医疗数据垄断情况的出现，建立数据流通的激励机制，健全共享共通的数据标准规范，促进不同的医疗人工智能平台协同运作^[9]。

2.3 算法透明度和可解释性伦理标准

算法透明度和可解释性伦理标准，是保证医疗人工智能算法的决策过程和结果能够被理解和解释的相关伦理标准。透明度伦理标准，要求公开医疗人工智能算法的模型结构、训练数据、输入输出、性能评估等信息，使得医护人员、患者、监管机构等相关方能够评估算法的可靠性和适用性。考虑到部分算法可能涉及商业秘密，可以根据医疗人工智能伦理风险高低的不同，设置不同程度的披露范围要求。算法可解释性是算法模型的客观属性^[10]，可解释性伦理标准要求医疗人工智能的设计人员在模型构建时应尽量选用具有高解释性的算法模型（如：决策树模型），使得其决策过程更容易理解和解释。医疗人工智能的研发设计主体应肩负起解释算法的义务，必要时形成书面报告向行政主管部门报备^[11]。算法透明度和可解释性的提升有助于增加医护人员和患者对于医疗人工智能的信任和接受程度。

2.4 责任与合规性伦理标准

责任与合规性伦理标准主要在于明确医疗人工智能应用中各方的责任与义务,确保智能系统的设计、制造和使用符合伦理规定。医疗人工智能系统的开发者和制造者,需对其设计、制造的系统和设备的安全性、可靠性及有效性负责,确保系统能够在预定的临床环境中安全有效地运行。医护人员作为人工智能系统的使用者,需对使用该系统做出的临床决策负责,包括对人工智能提供的建议进行评估和监督,并在必要时进行干预。监管机构应确保医疗人工智能系统在上市前经过严格的评审,满足所有伦理和安全要求,并对市场上的人工智能系统进行持续的监督。患者也应对自身安全和健康负责,在接受智能诊疗服务时,需全面了解可能存在的风险,积极参与决策过程。此外,建立完善的责任追溯机制有助于解决潜在的责任划分问题。医疗人工智能算法的决策过程和结果应是可追溯的,相关数据、模型、参数和操作过程应进行记录和存档,以便在需要进行审计、调查和验证。

2.5 监督与评估机制

建立医疗人工智能伦理标准的监督和评估机制是保障其得到有效落实的重要手段。在外部监督上,政府监管机构应对医疗人工智能建立合伦理性评估程序,审核医疗人工智能系统的开发和使用是否符合伦理标准,定期对医疗人工智能应用进行效果评估和伦理合规审计,并对违规行为进行处罚和纠正。单位内部的伦理审查监督则主要由伦理审查委员会负责。根据科技部、教育部、工业和信息化部等十部门在2023年10月联合印发的《科技伦理审查办法(试行)》的规定,从事生命科学、医学、人工智能等科技活动的单位,研究内容涉及科技伦理敏感领域的,应设立科技伦理(审查)委员会。医疗人工智能相关单位的伦理审查委员会在人员构成上应当多样化,除医护人员之外,还应有研发人员、伦理专家、法律顾问和群众代表等相关方。伦理审查委员会应当在医疗人工智能系统开发早期阶段和上市前,进行伦理审查,评估其设计、数据使用、隐私保护等是否符合伦理标准要求。在智能系统投入使用后,跟踪评估医疗系统的实际使用效果,包括其安全性、有效性、公平性

和对患者隐私的影响。此外,医护人员、患者、社会公众等有关人员也可以积极行使监督权,对医疗人工智能应用过程中违反伦理标准的情况向相关部门进行反映。

3 医疗人工智能伦理标准体系建设的配套建议

3.1 推动多方共建

医疗人工智能的应用涉及多方利益主体,其伦理标准体系的构建需要政府、行业协会、医疗机构、研发机构、高校以及社会公众等多方主体的共同参与。政府部门应当发挥主导作用,制定相关政策、指导原则和行动框架,为医疗人工智能伦理标准体系的建设提供指导和支持。医疗人工智能有关行业协会应当积极制定本行业的伦理标准,推动成员共同遵循伦理标准,提升行业整体的伦理水平。医疗机构和患者作为医疗人工智能系统的使用方,可以反馈在实际使用中遇到的技术和伦理问题,为伦理标准制定提供一手资料。研发机构可以通过研究成果为伦理标准讨论提供科学依据。高校应积极开展医疗人工智能伦理的学术研究,为伦理标准的制定提供理论支持,同时加强对伦理人才的培养。社会公众可以通过参加听证会等方式,积极为医疗人工智能伦理标准的制定建言献策,促使伦理标准更加贴合解决现实问题的需要。

3.2 加强国际合作

当下人工智能治理已成为一项全球性的议题,伦理治理问题是其中的重要组成部分。我国医疗人工智能伦理标准体系的建设不能一味“闭门造车”,而应以开放的心态,积极拥抱国际伦理共治。欧美等国在医疗人工智能技术的开发和应用方面早于我国,同时历来重视人工智能伦理的建设,在医疗人工智能伦理标准体系的建设上我国可以借鉴这些国家的成功经验。我国可以通过举办国际会议、研讨会等形式,邀请全球医疗人工智能领域的专家、学者和有关政府管理人员来华交流,共同探讨医疗人工智能的伦理标准制定问题。支持中国的研究机构、高等院校和企业与国际伙伴

建立合作项目,进行医疗人工智能伦理合作研究。同时,我国也可以利用本国在医疗人工智能伦理领域的研究和应用成果,为国际伦理标准制定提供理论和实践支持。我国作为发展中国家,积极参与有关国际科技伦理的制定工作,有利于打破由发达国家单方面主导国际科技伦理制定的局面,能够更好地推动实现医疗人工智能发展福祉由世界各国共享^[12]。

3.3 强化教育与培训

医疗人工智能伦理标准的贯彻实施有赖于相关人员伦理意识的提高,因此针对不同主体的伦理教育和培训工作十分有必要。对于医疗人工智能设计者和开发者而言,伦理教育能够增进其伦理自觉,使他们在人工智能系统的设计和开发中能够优先考虑到伦理问题^[13]。对于医护人员来说,在培训过程中能够更全面地了解医疗人工智能系统的基本原理、潜在风险和伦理问题,在使用智能系统时能够更加审慎,确保患者的隐私权和自主权得到尊重。医疗人工智能的监管者也有必要加强伦理教育,以便了解该领域的最新技术进展和伦理挑战,制定合适的伦理政策和监管措施。对患者和公众进行伦理教育的目的则在于提高他们对医疗

人工智能的功能和风险的认知,使他们能够在医疗决策中做出知情的选择。伦理教育和培训可以通过开展专题讲座、宣讲会、在线课程等多种方式进行,通过持续的伦理教育,逐步提高有关主体的伦理自觉性。

4 结语

人工智能技术正在重塑医疗行业的未来,给医疗领域带来了前所未有的机遇和挑战。我们在享受医疗人工智能带来的诊疗便利和效率提升的同时,必须警惕潜在的伦理风险。医疗人工智能领域的伦理问题不仅关系到技术的可持续发展,更关乎公众的健康福祉和社会的公平正义。本文提出的伦理标准体系建设方案,旨在为医疗人工智能的发展提供明确的伦理指引和实施框架,从而促进技术的健康发展和伦理风险的有效管控。医疗人工智能伦理标准体系的建设和完善是一个持续且充满挑战的过程,但只要 we 坚持在伦理治理的实践中不断探索与改进,就一定能够推动实现医疗人工智能技术的良善发展,为人类社会带来更多福祉。

参考文献

- [1] 万昆. 人工智能技术带来的就业风险及教育因应[J]. 广西社会科学, 2019(06):185-188.
- [2] 中国信息通信研究院. 人工智能伦理治理研究报告(2023年)[R]. 北京: 中国信息通信研究院, 2023.
- [3] 汪琛, 孙启贵, 徐飞. 医疗人工智能伦理研究的发展趋势——主题分布、知识基础与未来展望[J]. 自然辩证法通讯, 2023, 45(12):18-29.
- [4] 蔡星月. 人工智能的“标准之治”[J]. 中国法律评论, 2021(05):94-103.
- [5] 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(02):91-103.
- [6] 李锟鹏, 杨海秀. 医疗人工智能临床应用的伦理审视[J]. 医学与哲学, 2023, 44(07):27-31.
- [7] 冯永刚, 席宇晴. 人工智能的伦理风险及其规制[J]. 河北学刊, 2023, 43(03):60-68.
- [8] 陈欣, 金权. 我国医疗人工智能侵权规制的演进、特征及难点探析[C]//上海市法学会.《上海法学研究》集刊2023年第6卷——2023年世界人工智能大会青年论坛论文集. 上海: 上海海事大学法学院, 2023:13.
- [9] 周文康, 费艳颖. 医疗人工智能前沿研究: 特征、趋势以及规制[J]. 医学与哲学, 2021, 42(19):38-44.
- [10] 苏宇. 优化算法可解释性及透明度义务之诠释与展开[J]. 法律科学(西北政法大学学报), 2022, 40(01):133-141.
- [11] 徐明, 韦俨芸. 数字时代医疗人工智能的算法逻辑、风险及其应对[J]. 中南民族大学学报(人文社会科学版), 2024, 44(01):146-154+187.
- [12] 吴红, 杜严勇. 人工智能伦理治理: 从原则到行动[J]. 自然辩证法研究, 2021, 37(04):49-54.
- [13] 侯莹, 史励柯, 侯建平, 等. 智能诊疗领域的算法伦理与算法治理研究[J]. 中国医学伦理学, 2021, 34(04):445-449.