

医疗人工智能算法歧视的生成逻辑与法律规制 路径研究

李恒 徐着雨

【摘要】 算法作为医疗人工智能应用的核心，在提升数据价值的同时，可能引发歧视问题，亟需法律规制。本文阐释了医疗人工智能算法歧视的内涵，分析其生成机制，包括健康医疗数据的歧视性、算法逻辑缺陷以及医患互动偏差；针对现有的算法问责模式、数据保护模式、综合治理模式规制理论的不足，提出了优化法律规制的路径，包括制定场景化的法律规制方案、规范健康医疗大数据使用、增强算法可解释性，以及融合健康数据权利保障与技术规制，推动医疗人工智能算法的公正使用。

【关键词】 医疗人工智能；算法歧视；健康医疗数据；法律规制

Doi:10.3969/j.issn.1673-7571.2024.12.002

【中图分类号】 D920；R197

Research on generative logic and legal regulation path of medical AI algorithm discrimination

LI Heng, XU Zhuoyu. China Rule of Law Strategy Research Institute, Health Rule of Law and Policy Research Institute, East China University of Political Science and Law, Shanghai 200042, China

Corresponding author: LI Heng, Email: liheng1985310@sina.com

基金项目：国家社会科学基金委员会一般项目（22FX118）；上海市白玉兰人才计划浦江项目（23PJC042）

作者单位：200042 上海，华东政法大学中国法治战略研究院，华东政法大学卫生健康法治与政策研究院

通信作者：李恒，Email: liheng1985310@sina.com

- in four key legal area[J].European Pharmaceutical Law Review (EPLR),2021,5(1):21-36.
- [21] 欧洲议会. 欧盟人工智能法案 [EB/OL].(2023-06-01)[2024-05-05].<https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>.
- [22] GERKE S,MINNSEN T,COHEN G. Ethical and legal challenges of artificial intelligence-driven healthcare[J].Artificial Intelligence in Healthcare,2020:295-336.
- [23] U.S.Food & Drug.Administration, Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan.[EB/OL].(2021-05-22)[2024-04-09].<https://www.fda.gov/media/122535/download>.
- [24] 腾讯网. 拜登签署的美国首个 AI 安全行政令备受争议 [EB/OL].<https://new.qq.com/rain/a/20231103A034Z000>.
- [25] GERKE S,MINNSEN TIMO, COHEN G.Ethical and legal challenges of artificial intelligence-driven healthcare[M].Artificial Intelligence in Healthcare,2020:295-336.
- [26] DANAI I P,MARIOLINA E. Soft law behind the scenes: transparency,participation and the european union's soft law making process in the field of climate change[J].European Journal of Risk Regulation,2022,14(2):292.
- 【收稿日期：2024-05-09】
【修回日期：2024-06-10】
(责任编辑：刘慧铭)

【Abstract】 As the core of medical artificial intelligence (AI) application, algorithms may cause discrimination, while improving the value of data, which is urgent need of legal regulation. This paper explains the connotation of medical AI algorithm discrimination and analyzes its generation mechanism, including the discrimination of health and medical data, the logic defect of algorithm and the doctor-patient interaction bias. Aiming at the shortcomings of the existing regulation theories, including algorithm accountability model, data protection model and comprehensive governance model, this paper proposes a path to optimize legal regulation, including formulating scenario-based legal regulation schemes, standardizing the use of health and medical big data, enhancing the explainability of algorithms, and integrating health data rights protection and technical regulation, so as to promote the fair use of medical AI algorithms.

【Keywords】 Medical artificial intelligence; Algorithm discrimination; Health care data; Legal regulation

近年来,人工智能(AI)在医疗卫生健康领域得到广泛应用,算法作为其核心,显著提升了健康医疗数据的处理能力,释放了数据的潜在价值。然而,医疗AI算法的自动化处理程序可能蕴含偏见。实践中,智能诊疗系统可根据患者经济状况推荐诊疗方案,利用健康画像、精准推荐等新技术提供个性化服务,可能导致医疗不公平^[1]。2021年9月,国家新一代人工智能治理专业委员会发布《新一代人工智能伦理规范》(以下简称《规范》)。《规范》第十三条明确规定要避免偏见歧视,在数据采集和算法开发中加强伦理审查,在算法设计中增强人工智能系统的可信赖性,充分考虑差异化诉求,避免可能存在的数据与算法偏见,努力实现人工智能系统的普惠性、公平性和非歧视性。为此,有必要深入分析医疗人工智能算法歧视的生成机制和既有规制理论、模式,为治理医疗人工智能算法歧视问题提出法律建议。

1 医疗人工智能算法歧视的缘起

1.1 基本内涵

医疗人工智能算法歧视是随着算法、区块链等新兴技术在医疗领域纵深发展而衍生出的特殊歧视类型。目前,理论界和实务层面均未在算法歧视的内涵上达成统一观点。

算法是数学与计算机科学领域中的运算程序,是被用于执行计算指令、处理数据,并进行自动推理的技术规范^[2]。歧视可以解释为对事物或现象做出的不公正的价值评判。因此,可以将算法歧视界定为:在大数据、人工智能和自动化决策等技术支持下,机器所做出的决策,该决策会对个体或群体产生不合理的区别对待。

医疗人工智能算法在计算机系统运行过程中能够依托其高度算法化的整合分析能力,对健康医疗数据进行整合分析并生成诊疗方案供医患双方选择。但由于受性别、种族、年龄、人类价值观念、算法透明度等因素的影响,医疗人工智能算法运行的技术逻辑不断演化,加

之健康医疗数据代表性不足等,导致医疗人工智能算法的决策引起健康不公。例如,利用人工智能技术诊断黑色素瘤时,皮肤病的图像数据中主要记录的是白人患者的临床特征,这可能导致黑人患者诊疗效果不够精确,损害黑人的平等医疗权,甚至诱发误诊、漏诊等风险^[3]。

因此,可以将医疗人工智能算法歧视理解为算法应用于医疗领域后,以显性或隐性方式产生的不合理的差异化医疗服务行为,具有产生原因的多元性、产生过程的高度隐蔽性、损害行为的反复性、责任难以追溯性等法律特征。

1.2 形成机制

医疗人工智能算法运行的技术逻辑独特,可以根据医疗人工智能算法歧视的形成机制不同,将医疗人工智能算法歧视成因分为3类。

1.2.1 健康医疗数据本身存在歧视 健康医疗大数据是医疗人工智能算法平稳运行的关键,其来源、获取目的和使用方式等因素均可引发歧视问题。

第一,健康医疗数据本身可能存在数据缺失或数据代表性不足,

导致患者歧视或性别歧视。健康医疗数据来源广泛，包括患者、医疗服务提供者、保险公司、制造商、政府，以及智能手机和可穿戴设备等移动设备^[4]。但是大数据中可能存在罕见疾病数据缺失或代表性不足等问题，医疗人工智能系统可能无法识别患者是否患有算法模型中预设疾病类型外的疾病或健康风险。医务人员容易忽略每个患者的特殊性，造成相同诊疗而病情发展却大相径庭的现象^[5]。此外，以往的药物研发常忽视性别差异^[6]，研究人员一般情况下将男性健康数据作为女性健康数据的替代量，较少根据受试者性别分层分别进行试验分析和报告^[7]。这使得医学研究的健康数据以男性为主，忽略了疾病表现和不良反应在不同性别的差异。基于此类健康医疗数据开发的医疗人工智能系统，可能因此存在误诊、漏诊风险，潜在威胁女性患者的健康权益。

第二，健康医疗数据的分类标准存在歧视，使用分类后的数据进一步导致诊疗结果偏差。训练医疗人工智能算法的数据集主要依靠人工分类，可能带有歧视的观念而影响数据分类结果。例如，如果穿白色外套的白人男性可能被归类为医生，而穿同样衣着的女性则更有可能被标识为护士^[8]。所以，即使医疗健康数据集自身不存在准确性和完整性瑕疵，也会因标准不一而影响数据集的整体质量，进而导致算法模型出现歧视风险。

第三，健康医疗数据主体以及数据收集者的认知与行为也会导致

数据收集误差，进而引发歧视风险。例如，若健康医疗数据主体出于特殊目的刻意掩饰某些关键健康信息，那么这种不积极配合治疗的行为可能引起数据集失真，进而影响决策结果。此外，健康医疗数据收集者的忽视或认知偏差也会导致歧视风险。例如，当进行健康医疗数据统计时，可能倾向于以住院患者为研究对象，忽视了患病但未住院群体，导致结果不准确。

第四，健康医疗数据可能因为数据的潜在关联性而引发歧视性后果。医疗人工智能算法系统在训练和运行时均对数据集采用去身份化和匿名化技术以避免歧视。然而，匿名化处理无法实现完全匿名^[9]，去身份化也可以在必要情形下利用识别符再次建立联结^[10]。因此，人工智能算法系统可以基于已收集的与健康无关数据，推算出健康数据主体甚至关联主体的身份、健康状况等信息，例如年龄、作息规律、性别等。人工智能算法技术使用者可能借助此关联性评估数据主体或其关联群体的健康状态，不合理地利用算法技术进行雇员筛选、精准推送等以最大限度谋取利益，进而引发歧视问题。

1.2.2 医疗人工智能算法运行的逻辑缺陷 医疗人工智能算法的运行逻辑不当，可能造成对特定群体的歧视。首先，看似中立的医疗人工智能算法设计标准，实则会因替代性变量的选取不当而蕴含歧视。例如，美国医疗保健系统使用商业算法指导健康决策，这种算法表现出明显的种族偏见，问题的根源在

于其使用健康成本作为健康需求的替代，但患者付出的健康成本与经济实力等因素关联密切，并不能真实反映其健康状况，更无法准确描述其应获得的医疗服务类型，因此该算法使得白人从高风险护理项目中获益，迫切需要额外护理的黑人患者数量反而锐减，最终使得数百万黑人在医疗护理服务上遭受系统性歧视^[11]。其次，医疗人工智能算法系统具有黑箱属性，可能在自动化运行过程中做出难以预料的歧视性决策行为，损害患者健康，并导致患者权益救济困难。最后，由于算法是人类设计的，涉及人类的心智模式，健康数据的运用方式、程序设计目的均可能渗入设计者价值偏好，而可能被无意识地嵌入算法程序设计中。加之医疗人工智能算法不断迭代升级，医疗人工智能系统的自动化决策程度逐渐提升，医患双方的自主性随之被弱化。若不能及时调整算法程序或允许监管力量介入，训练的健康数据可能捕捉到歧视的历史模式，导致健康公平失衡现象出现算法化继承，形成歧视闭环，加剧既有的不平等。此外，技术运营者即使采取了介入式措施，严格把控数据质量，也难以甄别蕴含在算法自动化运行程序中的隐性歧视数据，依然难以避免歧视性结果的发生。

1.2.3 医患双方与医疗人工智能算法的互动型偏差 医务人员偏见和患者偏见是影响医疗人工智能算法公平性的重要因素。

一方面，医疗人工智能算法自动化决策流程弱化了健康数据主体

的数据控制能力，部分患者可能会对诊疗可靠性持怀疑态度。这种不信任可能会导致一些患者不愿意充分暴露所有疾病相关信息，甚至抗拒使用治疗型人工智能设备实施诊疗服务。此外，健康状况受多种因素影响，包括经济条件、地理位置、历史文化等，这些因素与健康成本关联性强，体现群体的健康差距，也会潜在影响患者的健康水平，进而影响患者接受智能医疗服务的可及性及可负担性。

另一方面，医疗人工智能算法的应用虽然释放了医务人员的部分工作压力，但易滋生过度依赖机器智能现象，弱化医务人员主体性。当医务人员过度依赖机器智能时，可能忽视患者个体特征，间接影响诊疗精度。与此相反，一些医务人员对医疗人工智能系统生成的诊疗方案和建议持审慎态度，出于对举证困难和责任归属等问题的担忧，仅选择智能诊疗系统的决策支持作用，以避免医疗人工智能算法的可解释性难题。这两种情境差异会加剧健康公平失衡，增加患者遭受歧视损害的风险。

2 防止医疗人工智能算法歧视的法律理论模式

针对医疗人工智能算法歧视现象，不同国家和地区也在实践中形成了不同保护模式。根据治理手段不同，当前算法歧视治理模式可分为以下3类。

2.1 算法问责模式

采取算法问责模式的国家主要以美国为代表。2017年初，隶属

于美国计算机协会的公共政策委员会发布行业声明《算法透明度和责任声明》（*Statement on Algorithmic Transparency and Accountability*），声明中设定了7项增进算法可解释性和可靠性的原则，要求以元规制为基础治理偏见问题，构建妥当的算法外部控制模式^[12]。2021年，哥伦比亚特区政府提出《停止算法歧视法案》的议案（*Stop Discrimination by Algorithms Act of 2021*）；2022年2月，美国民主党议员提出了《2022年算法问责法案》（*Algorithmic Accountability Act of 2022*）^[13]。上述两部议案虽然暂时未生效成为法案，但是对于算法歧视提出了一系列原则和防止歧视的规则，例如都对鼓励算法的透明度和问责制进行了相应规定。对于算法中有可能出现算法偏见的变量，例如基于个人特征（性取向、残疾、疾病、基因、年龄、收入来源、信用信息等），出现算法决策系统不能解释的偏见，建议进行算法审计，确定算法合格性和信息可用性，尽量纠正或减少算法偏见。

2019年4月，美国食品药品监督管理局（FDA）发布了人工智能在医学应用中的监管框架，采取了广泛全面的监管方案，赋予联邦机构更多自主权以促进创新。2021年美国政府问责署出台了人工智能问责框架，进一步强调以问责推动人工智能的安全可靠应用。2022年1月11日，美国卫生部正式发布了《值得信赖的人工智能（TAI）手册》，目的是促进采用合乎伦理的、成功的人工智能，强调人工智能要遵循

公平、公正、透明、可解释、负责、稳健、可靠、安全和保护隐私等原则，同时注重整个生命周期的监管，从而降低风险^[14]。

与其他仅禁止某些行为的反歧视法不同，《美国残疾人法案》（*The Americans With Disabilities Amendments Act*）直接针对残疾人融入日常生活提供了公平性保障，使残疾人不因其生命活动缺陷丧失就业、获得平等医疗服务的机会^[15]。《平价医疗法案》（*Affordable Care Act*）第六章第1557节敦促将算法公平性作为设计、实施和验证人工智能的关键因素，明令禁止在某些医疗保健项目和活动中以种族、肤色、民族血统、性别、年龄或残疾为由进行歧视^[16]。

综上，以算法问责为核心的治理模式突出强调技术公平性，并且主要依赖于算法影响评估制度，避免或规制算法不公平现象。从医疗人工智能算法运行的技术逻辑和治理目的看，该制度可以对医疗人工智能算法运行的全生命周期进行动态风险评估，改变了以往仅根据损害结果溯源推断责任主体的情形，有利于创建合理的算法透明度规则，增进医患信任。

然而，从上述治理措施的实施主体和法案内容可窥见算法问责治理模式也存在一些不足之处。其一，治理主体分散且多元。在联邦层面暂未形成统一立法，治理以行业的职责为主。其二，算法问责模式倾向于结果公正，容易忽视人工智能算法自动化运行过程的公正。美国的算法治理主要依赖外部审计和影

响评估，对算法运行的内部环节并未深入审查、评估。因此，人工智能算法运行过程中的歧视问题易被忽视，当发生歧视性损害时，由于算法逻辑架构复杂，依然难以进行追责。例如，意大利国家卫生局曾因采取人工智能模型预判新型冠状病毒感染风险和被感染人数时出现系统偏差，导致伦巴地区被误判为高风险区域。此次错误判断致使公民人身自由严格受限，商业活动被迫停止，产生了巨大的经济损失。然而，由于人工智能算法的复杂架构，该案的归责至今难以认定^[17]。

2.2 数据保护模式

欧盟采取的是以个人数据赋权为核心的严格保护模式，该模式将治理规则内置在数据保护框架中，通过强化个体权利来避免算法损害。2018年5月生效的《通用数据保护条例》（*The EU General Data Protection Regulation*）是欧洲第一部承认算法歧视现象并制定打击算法歧视规则和程序的立法，为治理算法歧视问题提供了更为具体的监测和执法工具包^[18]。该条例分别规定两条解决算法歧视的重要方式：一种是数据清洗，即从用于自动决策的数据集中删除特殊类别^[19]。具体可解释为只有当算法基于受法律保护的类别进行歧视时，才会出现法律问题。在欧盟法律中，这些受保护的类别包括性别、种族或民族血统，以及残疾、年龄、性取向和宗教或信仰，这些特定种类数据均禁止被自动化决策系统处理。另一种是赋权，即解释权、知情权、更正权、数据携带权、免于

自动化决策约束权等新型权利。具体而言，就是赋予数据主体在进行自动决策或特征分析时获得“关于所涉逻辑、意义和预期后果的有意义信息”的权利，要求这些信息必须“以简洁、透明、易懂和易获取的方式”提供，并且数据主体有权对决定提出质疑^[20]。实践中，歧视案例也时有发生，Coleman案的当事人（一名残疾儿童的母亲）声称自己受到骚扰和其他歧视，因为她的雇主拒绝为她提供休息时间和灵活的工作安排以照顾她的孩子，欧盟法院在该案件审理中承认，禁止直接歧视不仅限于残疾人本身，还包括与残疾人有关联的人，如他们的照顾者^[21]。

此外，针对医疗人工智能算法技术，欧盟强调的是基于可信赖原则对医疗人工智能实行分级监管。2019年欧盟委员会发布《人工智能道德准则》，明确值得信赖的人工智能应满足的条件，包含非歧视性和公平性、社会和环境福祉以及问责制^[22]。2024年3月13日，欧洲议会通过了《人工智能法案》（*The European Union Artificial Intelligence Act*），该法案作为世界范围内第一部具有约束力的人工智能综合性法规，特别强调基于风险的方法对人工智能进行分类治理^[23]。其中法案第六条第2款特别将生物分类的人工智能系统，即根据个人的种族、性别、政治观点等敏感属性进行分类的人工智能系统列为高风险系统之一。

可见，欧盟采取的是数据保护模式以应对医疗人工智能算法歧视

风险。该模式的优势主要体现在通过个体赋权，如事前知情同意权、事中及事后的算法解释权等方式实现对人工智能算法风险的全生命周期控制，并促进全流程数据保护的可行性。然而，仅依靠数据保护模式并不能充分解决医疗人工智能算法运行中的偏见、歧视和可解释性不足等诸多问题。第一，过于强调数据保护，一定程度上会制约医疗人工智能算法技术发展。例如健康医疗大数据收集的最小化原则和医疗人工智能算法模型要求的数据充分性、健康医疗数据处理的透明度和医疗人工智能算法的黑箱属性之间必然存在一定冲突，导致诊疗准确率、归责问题依然未得到充分解决。第二，对数据保护的规定过于原则化，缺乏可操作性，导致实质公平难以落实。不得收集个人特质相关联数据的要求，导致歧视出现时，算法系统无法证实其与个人健康数据相关的歧视。因为机器学习算法的技术本质具有迭代升级速度快且呈动态变化的特征，所以会在运行过程中衍生出新型具有关联性的健康数据，最终歧视性数据依然会进入到算法运行程序中，导致医疗人工智能算法的公平性仅停留在表面，算法歧视的现象仍然得不到矫正。

2.3 综合治理模式

世界卫生组织和联合国等国际组织为减少医疗人工智能的不平等和歧视，在规定平等和不歧视基本原则基础上，从伦理、技术、法律等多维度切入，形成综合治理模式。

世界卫生组织推出《患者安全权利宪章》，明确规定患者享有不

受歧视的权利，根据人权标准的规定，每位患者在任何地方、任何时间都必须能够获得安全的健康照护，无论其年龄、性别、民族或种族、语言、宗教、残疾、社会经济地位或任何其他状况如何。此外，世界卫生组织在2021年6月发布《医疗卫生中人工智能的伦理治理》指南，为医疗人工智能的临床应用提供了伦理框架，并且将确保包容性和公平性作为伦理治理的重要原则之一^[24]。

2024年3月21日，联合国大会通过关于引导人工智能造福全球、加快实现可持续发展的里程碑式决议“抓住安全、可靠和值得信赖的人工智能系统带来的机遇，促进可持续发展”，该决议被认为是全球首次对人工智能达成共识的决议。决议指出，不当或者恶意地涉及、开发、部署和使用人工智能系统，可能会有导致歧视的风险；特别提出要促进制定和实施有效、具有国际互操作性的框架、做法和标准，用于培训和测试人工智能系统，以帮助保护个人免受一切形式的歧视、偏见、滥用或其他伤害，并避免在人工智能系统的整个生命周期内强化或固化具有歧视或偏见的应用程序和结果。例如，要分析和减轻数据集中存在的偏见，并以其他方式打击算法方面的歧视和偏见，同时又不会无意或过度影响其他用户和受益方的正向发展、获取和使用。同时，决议促请联合国会员国采取具体措施，消除性别数字鸿沟，将残疾、性别和种族平等视角纳入政策决定及其指导框架的主流。

3 我国医疗人工智能算法歧视规制的理论模式分析

我国人工智能技术走在世界前列，法律规制方面主要采取个体赋权和设定平台义务并行的综合治理模式以应对医疗人工智能算法歧视风险。《中华人民共和国个人信息保护法》《中华人民共和国数据安全法》和《互联网信息服务算法推荐管理规定》从数据保护和强调平台义务两个维度为应对算法歧视风险做出了尝试。《中华人民共和国个人信息保护法》第24条规定了拒绝自动化决策约束权，第54条和55条分别规定了个人信息处理审计和个人信息保护影响评估制度以应对潜在的歧视风险。《中华人民共和国数据安全法》则在第8条对数据处理活动做了概括性规定，要求合乎法律规定与伦理规范。《互联网信息服务算法推荐管理规定》第4条明确规定算法服务提供者应遵守的原则，其中包括公平公正原则。

目前已经有部分专家学者就人工智能治理草拟学者建议稿，其中有个别条款对人工智能算法歧视的问题提出了针对性规定：人工智能开发者、提供者有义务采取合理措施保障使用者平等，确保产品和服务的普惠性和公平性，避免出现性别、信仰、地域、年龄、民族、经济能力等方面的偏见和歧视。同时，围绕医疗人工智能算法歧视的治理，也有部分学者提出了反歧视理论，禁止一切形式的歧视。不歧视是实现获得条件的首要准则，要求在医疗资源的分配上不得以任何理

由对任何人加以歧视，即卫生设施、商品以及服务必须面向所有人^[25]。

我国医疗人工智能算法歧视法律规制问题，应秉持审慎和包容的态度。针对未来人工智能技术发展的不确定性，人工智能立法必须预留包容空间以应对未来技术发展，从而实现对人工智能的科学、有效治理^[26]。然而，我国的现有法律规范也显露出我国医疗人工智能算法歧视的治理模式存在问题。第一，对医疗人工智能算法歧视的规制方案缺乏针对性，仍处在一般性算法问题法律规制和个人信息保护法律规范的映射下，忽略了医疗场景下的风险特殊性和歧视发生过程的隐蔽性。第二，当前法律法规尚未形成统一的医疗人工智能算法歧视法律认定标准，更无法对医疗人工智能算法歧视的构成要件做出法律解释。因此，各方主体在监管医疗人工智能算法歧视风险时，难以精准研判某项智能诊疗行为是否存在法律意义上的歧视风险。这一歧视认定难题势必严重影响医疗人工智能算法歧视风险的治理效力。第三，对个人健康数据的保护主要依靠算法影响评估制度，但该制度聚焦于事前，对个人健康数据权益保护的整体审视不足。例如，医疗算法黑箱可能存在的不可解释性，并且健康数据并非都可以量化，例如患者疼痛程度，算法模型的精确性会因此受到影响。因此，事前评估机制难以预测智能诊疗风险，一旦发生漏诊或误诊，患者的举证难度大，责任主体难以追溯，权利救济将严重受阻。

4 医疗人工智能算法歧视法律规制的优化路径

上述医疗人工智能算法歧视治理模式提供了应对算法歧视的原则和基本框架，我国可以借鉴相应规定，在现有立法基础上，针对存在的法律问题构建我国医疗人工智能算法歧视的法律规制体系。

4.1 制定场景化的医疗人工智能算法歧视法律规制

医疗人工智能算法技术应用领域广泛，健康数据的价值发挥体现在多种差异化的应用场景中，可以借鉴欧盟的场景化法律规制模式，并结合应用目的和不同医疗场景中的风险程度予以区分。

一方面，法律应根据医疗人工智能算法的风险程度制定差异化规制方案，促进医疗人工智能算法技术规范发展。针对中低度风险的医疗人工智能算法技术，例如健康管理、智能导诊、临床决策辅助、医疗保险等技术带有健康促进目的，不宜过多进行法律干预，应更多强调医方在安全保障和医疗护理服务平等分配等方面的注意义务，即医师有义务对诊疗系统推出的诊疗结果、诊疗方案进行复核。同时，医疗机构应当引入第三方机构定期测评医疗人工智能算法系统的风险有无异常，医方和第三方机构有义务就测评情况向技术监管部门持续报告。针对高度风险技术，例如全自动化医疗手术机器人等自动化程度极高的医疗人工智能算法系统，因其自主化程度不断加深，健康数据的选取标准和机器学习过程中所吸纳数据的质量难以做到实时审查，

可能对个人生命权、身体健康权造成不可逆损伤，法律必须规定使用时有人类医生的全过程参与。

另一方面，对于依靠健康数据实现不正当目标的场景需严格禁止医疗人工智能算法技术应用。当前，一些用人单位利用人工智能算法技术分析个人健康数据信息或利用其他关联数据推测评估个人及其关联群体的健康状况，以此为根据筛选雇佣工，此类使用场景毫无健康促进目的且具有歧视性和强烈的商业逐利性，背离了医疗人工智能算法公平性目标，应当严格禁止。此外，技术监测部门需适时跟踪研判医疗人工智能算法技术自动化运行过程，并有权严格限制过度推测行为，要求医疗机构及医务人员将医疗人工智能系统的输出结果仅作为医疗活动的参考，降低责任主体认定难度。

4.2 规范健康医疗大数据的使用以降低歧视风险

健康医疗数据收集和使用的规范性，是医疗人工智能算法运行安全可靠的基础。

第一，法律应明确规定健康医疗数据的收集边界和使用目的。医疗人工智能算法设计者应使用多样化和具有代表性的健康医疗数据，将健康数据主体的性别、籍贯、年龄等健康因素纳入考量范围，最大限度保障健康数据集真实、无偏见。此外，健康数据收集者需遵守《中华人民共和国个人信息保护法》中的目的限制原则，即要明晰健康数据收集的边界，对个人健康数据的收集仅限于实现健康目的。研发

者、运营者、使用者等主体均需明确医疗人工智能算法技术的价值发挥不仅限于疾病诊疗，也应体现于医患双方获取的健康数据信息对称程度，从而保障患者健康数据安全和患者知情权。

第二，被用于训练医疗人工智能系统的健康数据集必须经过全方位审核以保证真实性和可靠性。在医疗人工智能系统设计人员的构成上，需要充分考虑平衡度，将年龄差距、性别比例等因素纳入遴选研发人员的必备考要素中，最大限度避免单一设计主体的价值偏好植入系统而导致歧视结果。此外，设计者须坚持平等保护原则，将保护未成年人、老年人、残疾人等弱势群体的健康权益等特殊考量因素纳入医疗人工智能算法应用的设计程序中，以降低误诊、健康公平失衡等风险。

第三，需建立数据的健康公平性和算法歧视评估制度。医务人员、健康数据编码者、产品设计者应结合自身专业能力对健康数据的公平性和技术公平性进行全生命周期评估，全流程减少甚至消除偏见性数据的使用，在运行过程中若发现歧视因素时应立即预警并由医务人员和专业技术人员介入，避免歧视结果发生。

4.3 增强医疗人工智能算法的可解释性

医疗人工智能中的医患关系仍然建立在信任的基础上，医生需要信任可解释的模型，与患者进行沟通^[27]。

第一，医疗人工智能算法体现的是健康数据间的相关性而非因果

性,因此立法应细化算法审查规则,要求审查机构坚持过程和结果审查相结合,定期且频繁地对健康数据的替代性变量进行审查,检查是否存在替代性变量分类错误、相关性不足等问题,进而提升医疗人工智能算法诊疗结果的准确率。

第二,迅猛发展的医疗人工智能算法技术并非实现健康促进目标的灵丹妙药,应将技术的准入和使用纳入全流程治理,进而提升诊疗准确率,降低歧视风险。在技术设计环节,医疗人工智能算法技术研发机构有义务将技术设计原理、理念目标等内容提交技术监管机构审查,增强医疗人工智能算法技术的可解释性。研发者需要将技术中立原则贯穿于整个算法决策流程中,必要时引入第三方审查机构对医疗人工智能算法技术的用途进行跟踪研判,避免在用户使用医疗人工智能算法产品和智能医疗服务时做出带有偏见性和歧视性的决策,从而降低最终追责的阻碍。

第三,在医疗人工智能算法自动化决策过程中,研发人员和使用者等有义务密切关注医疗人工智能算法的技术瓶颈,充分认识医疗人工智能算法的工具属性,并且应当在涉及人类健康的智能程序设计上为人类医生有意义地参与智能诊断活动提供技术支持。

第四,医疗人工智能产品生产应形成技术公平的行业规范,通过技术正当程序加强医疗人工智能算法决策的透明度和可责性。例如,令运营者采取统一的可视化技术手段,对技术透明度进行统一设置,

保证医疗人工智能算法决策全过程的可追溯性,保护健康数据主体信息的完整性和私密性,进而降低归责时的举证阻碍。

4.4 坚持健康数据权利保障与技术规制的融合规制路径

卫生健康法律发展和医学科技规范化是促进健康公平的重要因素,二者可以相互促进,医疗人工智能专业研发人员和政策制定者必须使用技术和法律手段,共同应对医疗人工智能算法歧视风险。

一方面,维护并促进公众健康,减小、消除健康差距,实现健康公平是现代政府的重要价值目标与内在职责^[28]。立法机关需推进医疗人工智能立法,明确禁止歧视并将公民数字素养提升列为医疗人工智能研发者、使用者等主体的义务,促进健康数据主体良好数字素养形成,突破“信息茧房”的桎梏。同时,赋予个体医疗算法权利和健康数据权利,令患者有权自主化选择、拒绝医疗人工智能算法决定和要求人工介入。此外,在规制医疗人工智能算法时,需要明确规定医疗人工智能算法歧视的构成,避免歧视问题泛化。医疗领域算法歧视的实施主体为技术控制者,客体是所有可能的应用对象,包括个体或群体,实施行为表现形式为医疗人工智能算法系统的自动化决策。

另一方面,医疗人工智能算法技术具有极强自主性,单纯依靠立法手段规制健康数据的收集使用可能能力有不逮^[29]。技术监管部门应强化审查义务,以伦理审查推动科技伦理治理。同时,研发机构须对

研发设计人员课以高度注意义务,令其将尊重自主原则、公正原则等生命医学伦理原则内嵌于算法设计程序中,以保障医患和谐关系和助推健康公平。在责任承担方面,医务人员需要主动融入现代医学技术环境中,设计者、使用者等各方主体需根据医疗人工智能算法技术智能化程度灵活配置注意义务,对于智能医疗算法设计者和产品生产者采取严格责任;至于运营者的责任承担,若自动化程度会触及低度敏感个人健康数据,并且造成患者健康信息权益损害的风险性较小时,只要运营者证明已经穷尽技术、法律手段则无需承担责任;反之,则应按严格责任进行客观归责,使患者权利得以救济。

5 结论

医疗人工智能算法的法律规制应关注促进技术发展和避免算法歧视之间的平衡。治理医疗人工智能算法歧视风险的最终目标并非阻止医疗人工智能算法发展,而是促进医疗人工智能算法向善。目前,医疗人工智能算法歧视风险的生成路径多元且隐蔽,亟需更加完善的法律法规予以治理。为此,应结合医疗人工智能算法治理困境和我国实际情况,制定具有我国特色的医疗人工智能算法歧视规制方案,推动现代医学技术与卫生健康法律的互动与融合,为健康中国战略目标实现夯实基础。

参考文献

[1] 张荣,徐飞.智能医学算法决策的

- 伦理困境及风险规制[J].医学与哲学,2022,43(8):10-15.
- [2] 石颖.算法歧视的缘起、挑战与法律应对[J].甘肃政法大学学报,2022(3):58-68.
- [3] ADAMSON AS,SMITH A.Machine learning and health care disparities in dermatology[J].JAMA Dermatology,2018,154(11):1247-1248.
- [4] HOFFMAN S,PODGURSKI A. Artificial intelligence and discrimination in health care[J].Yale Journal of Health Policy, Law and Ethics,2019,19(3):1-49.
- [5] 徐着雨,岳远雷.医疗人工智能算法风险防范的法治化思考[J].医学与哲学,2023,44(11):67-71.
- [6] SEYDEL C.The missing sex[J].Nat Biotechnol,2021,39(3):260-265.
- [7] CIRILLO D,CATUARA-SOLARZ S, MOREY C,et al.Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare[J].NPJ Digital Medicine,2020,3(1):1-11.
- [8] WELCH T. Algorithms of oppression:how search engines reinforce racism and pattern discrimination[J].The Velvet Light Trap University of Texas Press,2020,86(1):72-74.
- [9] 赵精武.个人信息匿名化的理论基础与制度建构[J].中外法学,2024,36(2):326-345.
- [10] 满洪杰.隐私保护与利益分享:临床数据医学研究的规制转向[J].中国政法大学学报,2023(3):94-102.
- [11] OBERMEYER Z,POWERS B, VOGELI C,et al.Dissecting racial bias in an algorithm used to manage the health of populations[J]. Science,2019,366(6464):447-453.
- [12] COUNCIL A.Statement on algorithmic transparency and accountability[R].New York:ACM,2018.
- [13] GURSOY F,KENNEDY R, KAKADIARIS I.A critical assessment of the algorithmic accountability act of 2022[J/OL]. SSRN Electronic Journal,2022: e4193199.
- [14] BAKER-BRUNNBAUER J.TAII framework for trustworthy AI systems[J].ROBONOMICS:The Journal of the Automated Economy,2021,2(17):1-12.
- [15] NEWMAN M,ISENHATH F.The Americans with disabilities amendments act of 2008[J].Federal Lawyer,2008:12-13.
- [16] HILLER JS. Healthy predictions:questions for data analytics in health care[J].American Business Law Journal,2016,53(2):251-314.
- [17] ARAGONA B. Algorithm audit:Why,what,and how?[M]. London:Taylor and Francis,2021:98.
- [18] HOOFNAGLE CJ,VAN DER SLOOT B,BORGESIU FZ.The European Union general data protection regulation:what it is and what it means[J].Information & Communications Technology Law,2019,28(1):65-98.
- [19] GOODMAN B.Discrimination,data sanitisation and auditing in the European Union's General Data Protection Regulation[J]. European Data Protection Law Review,2016,2(4):493-506.
- [20] SELBST A,POWLES J.Meaningful information and the right to explanation[J].International Data Privacy Law,2017,7(4):233-242.
- [21] HACKER P.Teaching fairness to artificial intelligence:existing and novel strategies against algorithmic discrimination under EU law[J].Common Market Law Review,2018,55(4):1143-1186.
- [22] 刘艳红.人工智能的可解释性与AI的法律责任问题研究[J].法制与社会发展,2022,28(1):78-91.
- [23] WOISETSCHLÄGER H,ERBEN A, MARINO B,et al.Federated learning priorities under the European Union Artificial Intelligence Act[J].arxiv preprint arxiv:2402.05968.
- [24] 隗冰芮,薛鹏,江宇,等.世界卫生组织《医疗卫生中人工智能的伦理治理》指南及对中国的启示[J].中华医学杂志,2022,102(12):833-837.
- [25] 满洪杰.健康医疗大数据治理的健康权面向[J].求是学刊,2024,51(2):103-112.
- [26] 张凌寒.中国需要一部怎样的《人工智能法》:中国人工智能立法的基本逻辑与制度架构[J/OL].法律科学(西北政法大学学报),2024(3):3-17.
- [27] ZHANG Y,WENG Y,LUND J. Applications of explainable artificial intelligence in diagnosis and surgery[J].Diagnostics,2022,12(2):237.
- [28] 苏玉菊,徐海波.新时代健康公平观的基本内涵与实践价值[J].贵州社会科学,2022(6):96-102.
- [29] 许可.驯服算法:算法治理的历史展开与当代体系[J].华东政法大学学报,2022,25(1):99-113.

【收稿日期:2024-05-17】

【修回日期:2024-06-11】

(责任编辑:刘慧铭)