

适格诊疗决定中心论与医疗人工智能 医疗损害责任判定

吴梦婷 唐鸿宇

(中国科学技术大学 公共事务学院,安徽 合肥 230026)

[摘要] 医疗人工智能在临床实践中的应用日益深入,但在政策法律颗粒度适配、医疗透明度要求与诊疗决定归责逻辑三个方面仍存在学理争议和系统性矛盾。现有法律体系未明确医疗人工智能介入诊疗活动后的具体定位与责任划分,传统医疗损害归责路径陷入困境。适格诊疗决定应当认定为由医师与患者合意作出,医疗人工智能的当下定位仍限于医疗建议提供者,而诊疗决定的适格性判断关键在于程序合规。在医疗人工智能致损的归责方面,生产者承担医疗人工智能产品安全责任,医师作为医疗人工智能的实际使用者承担采纳其建议后的诊疗损害责任,患者自主采纳人工智能建议致损则需自担风险。

[关键词] 医疗人工智能;适格诊疗;医疗损害归责;“医—机—患”关系;医师负责制

[中图分类号] D926 **[文献标识码]** A **[文章编号]** 1008-7699(2026)01-0049-13

一、问题的提出

医疗卫生行业与自然人身体权、健康权等基本人权紧密关联,在该行业领域中使用人工智能,对患者尊严和医师权威提出了严峻挑战。具言之,用于医疗的人工智能正在侵蚀医疗双方主体价值,^①呈现出以“技治主义”取代“人本主义”的倾向。

学界对于医疗人工智能存在两种观点,一是广义的医疗人工智能,即从技术层面以深度学习算法出现为节点,衍生多种神经网络并在医疗领域广泛应用的人工智能应用,如智能组织自动化机器人等;^②二是狭义的医疗人工智能,即实践中以辅助为本职,利用人工智能技术为医务人员的诊断与治疗活动提供辅助决策支持的系统,亦称为“诊疗人工智能”。^③前一个观点虽然能较好地解释人工智能的运作机制,但同时带来了一个更为棘手的问题——技术黑箱,目前尚无技术专家能够阐明其决策的具体流程。对于循证医学而言,黑箱问题的存在反而更为符合医疗决定的制定过程,但更多反对者认为循证医学需要对结果进行批判性评估,^④而这并非医学所追求的科学决策形式。

[收稿日期] 2025-11-19

[基金项目] 国家重点研发计划(2024YFB3108200)

[作者简介] 吴梦婷(1983—),女,北京人,中国科学技术大学公共事务学院特任副研究员、国家工业信息安全发展研究中心客座研究员,博士。

^① 郭高维:《医疗机器人应用的伦理风险及其治理路径》,《东南大学学报(哲学社会科学版)》2025年第S1期,第31-32页。

^② 孙启贵、汪琛、王加宇、叶斌:《医疗人工智能发展的社会-技术分析 with 启示》,《自然辩证法研究》2021年第3期,第48-53页。

^③ 郑志峰:《诊疗人工智能的医疗损害责任》,《中国法学》2023年第1期,第203-221页。

^④ Justin Blackman & Richard Veerapen, *On the Practical, Ethical, and Legal Necessity of Clinical Artificial Intelligence Explainability: An Examination of Key Arguments*, BMC Medical Informatics and Decision Making, Vol. 25, No. 1, 2025, pp. 1-20.

而后一个观点虽能详细论证起到辅助作用的医疗人工智能的归责体系,却忽视了现实中并非所有医疗人工智能均被设计为辅助医师作出医疗决定的角色。试图代替医师作出医疗决定的人工智能实体或大模型层出不穷,智能临床决策支持系统是一个典型案例^①,更不用说医疗人工智能在肿瘤监测、癌症识别、败血症预测等医用领域中的“出色表现”。^②当下人工智能系统在医疗决定作出中的角色已无法再用“辅助”作为定性概念,而是处于一种难以计算与量化的复杂情形,必须视具体情况作具体分析。相关田野调查表明,临床医师与人工智能的医疗互动呈现持续的知识整合与认知调适,这一过程催生出融合直觉与理性的新型临床决策模式:医师与人工智能组成“决策混合体”。双方都对确保患者最佳治疗结果起着关键作用,而这种协作式调适(最佳医疗决定的作出)只能在自然人(在某种程度上)理解所使用的人工智能系统时产生。^③

可以预见的是,未来人工智能在医疗领域极有可能摆脱“辅助”定位,出现越来越多脱离医师控制、给出“医疗(诊疗)决定”的医疗人工智能或大模型。目前,我国法律体系对于医疗人工智能的规制与归责体系尚不完备,学界对于医疗人工智能的归责体系众说纷纭。在法律领域内,空谈技术提升或推崇在技术层面融入法律或道德标准,既无法真正解决医疗人工智能所引发的责任问题,也难以在实践中有效适用,必须另寻合适的切入点。笔者关注到一个有趣的现象:法学家们普遍倾向于保守适用本体论,即从人工智能的角度讨论医疗人工智能的责任归属,但最终的落脚点往往局限于产品责任,^④或是陷入“理性的思辨”,从而脱离了法学实质性的论证,甚至寄希望于技术的自治,“以科学理解和临床试验确保医疗人工智能的安全和有效”^⑤。而其他的人文社会学者却逐渐意识到,现有方案均以个体行动者为中心,难以应对人工智能辅助决策这一复杂系统行为带来的归责挑战。^⑥“人机耦合”是必然趋势,也或许是一条打破“医—机—患”层面信任挑战和临床应用困境的良好途径,^⑦“人机”法律责任机制自然也随之改变。值此数字化转型时代,关乎医疗人工智能的核心议题在于:1. 医疗人工智能是否可以独立作出诊疗决定? 2. 医疗人工智能的应用如何契合透明度的伦理要求? 3. 医疗人工智能参与诊疗引发损害时,法律责任应如何在医师、开发者、医疗机构等多主体间合理划分? 这些问题是诊疗领域应用人工智能技术所必须厘清的关键问题与重要突破口,以应对人工智能所带来的全新挑战与治理需求。

二、医疗人工智能发展进程中的三大矛盾

在解决复杂问题时要抓住主要矛盾,有主要的矛盾就一定有非主要的矛盾,有矛盾的主要方面就一定有矛盾的非主要方面。^⑧同时,只要解决了主要矛盾,其他矛盾自然就迎刃而解。当下人工智能在诊疗决定方面存在至少三个矛盾:一是政策发展与法律滞后之间的矛盾,二是技术黑箱与医疗透明之间的矛盾,三是诊疗决定与归责不清之间的矛盾。

① 为了与“诊疗人工智能”做区分,本文行文中使用“医疗人工智能”概念,下同。

② Ugo Pagallo, et al, *The Underuse of AI in the Health Sector: Opportunity Costs, Success Stories, Risks and Recommendations*, Health and Technology, Vol. 14, No. 1, 2024, pp. 1-14.

③ Roanne van Voorst, *Redefining Intelligence: Collaborative Tinkering of Healthcare Professionals and Algorithms as Hybrid Entity in Public Healthcare Decision-Making*, AI & Society, Vol. 40, No. 5, 2025, pp. 3237-3248.

④ 王轶晗、王竹:《医疗人工智能侵权责任法律问题研究》,《云南师范大学学报(哲学社会科学版)》2020年第3期,第102-110页。

⑤ 李润生:《论医疗人工智能“黑箱”难题的应对策略和规制进路》,《东南大学学报(哲学社会科学版)》2021年第6期,第83-92页。

⑥ 何丽:《人工智能辅助医疗决策的归责难题新解》,《自然辩证法研究》2023年第6期,第65-71页。

⑦ 王昱洲、廖新媛:《人机耦合:一种解决医疗人工智能信任问题的方案》,《自然辩证法通讯》2025年第9期,第1-9页。

⑧ 唐正东:《〈矛盾论〉与马克思主义中国化的理论创新》,《南京社会科学》2023年第6期,第22-30页。

(一)政策要求与法律滞后之间的矛盾

如前所述,政策正在大力推动“人工智能+”与各行各业深度融合,但法律却没有表现出“紧跟技术发展脚步”的趋势。公共政策呈现出多样化形态,人工智能治理的发展在很大程度上由政策驱动或调控。通常,中央率先出台统摄性政策文件,地方政府依据中央要求,通过更细化的政策落实执行。^①在人工智能治理领域本应平衡的“政策—法律”关系因法律的缺位而逐渐失衡,进而引发司法实务中“有法难依、存案难判”的现实困境。

聚焦医疗领域,以《全国医院信息化建设标准与规范(试行)》为代表的医疗人工智能发展政策,重点推进人工智能在医院运营、诊疗诊断、档案管理等方面的应用落地。具体包括:一是基于患者个人信息、遗传信息及生理指标等数据,构建疾病风险的预测模型(疾病风险预测);二是通过血管摄影、心血管造影、CT、PET、B超、核磁等医学影像技术,实现心脑血管、肺部、乳腺、肝病、眼底、心脏疾病、脑部等疾病病灶识别和分类(医学影像辅助诊断);三是融合语音识别、自然语言处理与临床数据分析模型,生成结构化诊疗建议(临床辅助诊疗);四是整合院内诊疗数据与院外日常健康数据,依托移动终端与可穿戴设备,建立动态个人健康档案,为糖尿病、高血压等慢性病患者提供个性化健康干预方案(智能健康管理);五是利用物联网与数据挖掘技术,自动采集与运营数据并分析运营指标,优化医院服务管理(医院智能管理);六是集成语音识别、自然语言处理与机器学习技术,为医疗人员提供智能化决策支持与信息服务(虚拟助理)。但与此同时,法律上却缺少针对人工智能作出医疗决定的修订或补充,导致对医疗人工智能是否均应适用产品责任的理解出现偏差。^②在《中华人民共和国民法典》(以下简称《民法典》)第四章产品责任的体系下,若探讨医疗人工智能适用产品责任能否有效规制医师使用行为,此路径存在明显局限性。其根源在于未区分医疗人工智能在出厂和使用两个阶段的法律属性差异,将使用阶段的医疗人工智能简单定性为产品并不妥当。对于使用医疗人工智能的行为,应追究行为主体(医师)的责任,但现行《医师法》尚未建立相应责任认定机制。《医师法》中唯一涉及“智能”的条款是第47条关于“国家采取措施,通过信息化、智能化手段帮助乡村医生提高医学技术能力和水平”的规定,但该条款与医疗人工智能责任并无直接关联。

综上,我国尚未建立针对人工智能应用的法律责任制度框架,亦缺乏关于人工智能的基础性法律规制,致使医疗人工智能领域的法律体系存在明显缺漏。当前“政策先行”的治理模式进一步凸显了该领域的立法需求,亟需尽快厘清法律责任、填补法律漏洞、构建系统化的法律实施机制。

(二)技术黑箱与医疗透明之间的矛盾

人工智能系统的内部工作机制(如机器学习模型的决策逻辑)和具体输出的原因难以解释,其中既包括整体运行原理的“全局不透明”,也包括单个结果的“局部不透明”。^③同时,医疗人工智能系统存在其他潜在缺陷,如训练数据偏差导致结果不准确、跨人群适用性问题以及持续学习导致性能下降等。又因医疗人工智能的“思维黑箱”,医师难以了解其思维过程,若两者作出的诊疗决定不一致,将在一定程度上影响医师对病人病情的自主判断。换言之,现在更常用的深度学习模型,本

^① 张宇帆:《论人工智能政策与法律协同治理》,《东方法学》2024年第5期,第187-200页。

^② 候曼曼:《诊疗式人工智能的医疗产品责任认定》,《东方法学》2025年第3期,第64-75页。

^③ Eva Schmidt, Paul Martin Putora & Rianne Fijten, *The Epistemic Cost of Opacity: How the Use of Artificial Intelligence Undermines the Knowledge of Medical Doctors in High-Stakes Contexts*, *Philosophy & Technology*, Vol. 38, No. 1, 2025, pp. 1-22.

质上是一种“黑箱”结构,缺乏可解释性和透明度,可能影响其在高风险医疗场景的可信应用与监管合规。^① 一篇发表在《美国预防医学杂志》的实证研究论文表明,向 ChatGPT-4 和 Bard(谷歌推出的大模型)提出 56 个医疗相关问题,并由两名医生对其回答的准确性进行评估,若出现分歧则由第三名医生加以裁决,前者的准确率与不准确率均为 28.6%、部分准确率为 42.8%,后者表现较好,三项数值分别为 53.6%、17.8%和 28.6%,得出“人工智能大模型尚不能在医疗领域替代专业自然人医师”的结论。^② 在针对识别脓毒症的预测模型中也出现了类似的结果,对于 2552 名经临床确诊的脓毒症患者,该模型仅对其中 843 例(33%)发出了脓毒症警报,漏诊率高达 67%,而在 6971 次脓毒症模型发出的脓毒症警报中,仅有 843 次(12%)是正确的,假阳性率高达 88%。^③

与法学同为人类社会三大最古老学科(另一门学科是神学)的医学也有着类似不可解释的“自由心证”阶段,医学中体现在医师作出诊疗决定之时,而法学则体现在法官裁决案件之时。正如同当下人工智能前沿的技术人员所认为的那般,不可解释性可能是人工智能具备类人“智力”^④的一种体现,过高的可解释性要求反而会损害其决策功能与工作效率。如果不能充分识别可解释性所带来的收益是否能够抵消准确性的下降给医疗事业造成的损失,那么对医疗人工智能的可解释性推崇将延伸成为一种偏见。与此同时,如果人类无法准确地对自身行为作出解释,那么对人工智能的可解释性要求就成了一种“双重标准”。^⑤ 因此,如同司法所推崇“公平公正公开”“公正高效办理每一个案件”一般,医疗透明虽然是当代医疗建设的重点,但其内涵仍需要匹配医学伦理的四大原则:尊重原则、不伤害原则、有利原则和公正原则。首先,从医师角度出发,需保障患者知情同意权,确保患者对身体、治疗方法及治疗手段行使自主决策权。若涉及医疗人工智能的应用,必须明确告知患者并由其决定是否使用,以充分尊重患者的意思自治。其次,从患者角度而言,医疗透明就意味着构建对医疗决定的信任。在以往家长式诊疗决定模式下,医患信任主要是源于医师的特殊职业角色,然而随着诊疗决定模式向协作互动型过渡,患者对医师的信任基础已从职业角色转移到对具体医疗决定的认同与支持。^⑥ 医疗透明并非求解医疗决定的可解释性,而是“医—机—患”之间就医疗决定达成一致与彼此信任的可解释性。

医疗透明的愿景与医疗人工智能所固有的“决策黑箱”形成一种天然的矛盾,在技术浪潮不可逆的情况下,这一对矛盾关系基本决定了医疗人工智能实践的未来走向。

(三)诊疗决定与归责不清之间的矛盾

“医疗”与“诊疗”“决策”与“决定”这两对概念需作区分,在不同情况下加以使用。医疗(Medical)指的是医疗卫生机构和医务人员提供的医疗卫生服务,是群众看病就医、保健康复的核心内容,与群众关系最直接,群众感受最具体。^⑦ 诊疗(Clinical),则更倾向于一种法律表述,在《民法典》尚未出台前,“诊疗规范中心论”的立法导向已经贯穿于《侵权责任法》“医疗损害责任”章,集

① 钱波、李富江、郑常乐、张道强:《医疗大模型发展现状与展望》,《数据采集与处理》2025年第3期,第562-584页。

② 本刊讯:《研究显示人类医疗专家比人工智能工具更可靠》,《数据分析与知识发现》2024年第5期,第28页。

③ Daniele Veritti, et al, *Behind the Mask: A Critical Perspective on the Ethical, Moral, and Legal Implications of AI in Ophthalmology*, Graefe's Archive for Clinical and Experimental Ophthalmology, Vol. 262, No. 3, 2024, pp. 975-982.

④ 潘斌:《增强还是替代:人工智能与人类智能关系的哲学反思》,《理论月刊》2024年第7期,第5-14页。

⑤ 安然:《医疗人工智能可解释性争论的厘清与辨析》,《科学与社会》2024年第3期,第156-168页。

⑥ 汪青松、罗娜:《替代还是支持:AI医疗决策的功能定位与规范回应》,《探索与争鸣》2023年第5期,第100-110页。

⑦ 《学习〈决定〉每日问答|为什么要促进医疗、医保、医药协同发展和治理》,中华人民共和国中央人民政府官网, https://www.gov.cn/zhengce/202410/content_6982363.htm, 2025年8月14日访问。

中体现在第58条第1项、第60条第1项以及第63条,^①后被《民法典》所吸收。“诊疗”概念的范围小于“医疗”概念,一是必须由医师这个主体作出的医疗行为才能视为“诊疗”行为;二是“诊疗”是一种行为,从司法实践上看,在卫生行政部门的日常监管中,“破皮”通常是认定是否从事“诊疗行为”的直接证据;^②三是诊疗仅聚焦医师具体动作,而医疗不仅包含具体治疗行为(Treatment),还可以囊括任何与医疗相关的抽象性内容,在黑格尔哲学中可称此为“具体共相”概念。^③决策(Decision-making)强调多阶段的理性推演过程,包含信息收集、方案评估、价值判断等环节。决策内容的概念,直白地说,就指决策所思考的决策方案的相关因素,而严格地说,决策内容是指决策思维所内含要素的总和。^④决定(Decision),侧重单一环节的结论输出,是对特定问题的即时回应。在诊疗过程中,医生决定患者的治疗方案。根据康德的自律意志理论,决定因“人的目的”所受感性偏好主导,所以在医疗人工智能带来的法律责任场景讨论中,使用“诊疗决定”是对制定病患治疗方案最为恰当的描述。

医疗人工智能的发展为作出诊疗决定的主体提供了另外一种可能,相应的归责体系理应作出调整。一是因为人工智能技术发展迅速,技术更新迭代在速度与效率上均远胜于法律法规,因此学者普遍支持通过软法路径先行治理人工智能。^⑤二是人工智能作为一种全新的技术手段,在一定程度上颠覆了传统法学对于责任机制的理解,人工智能的法律主体问题接踵而至。“否定说”观点如“人的主体性源自人类心智中理性因素与非理性因素完整且有机融合的功能显现,而这种意向性专属于人类的唯一性使得人工智能不可能成为法律意义上的‘人’”^⑥;“支持说”观点如“在迈向通用人工智能的道路上,其涌现出的独特的思维和理解能力、打通虚拟与现实边界的可交互性以及向人脑规模发展的模型参数量等特征使其在法律层面上应被赋予相应的独立法律主体资格”^⑦。三是当下提出的归责方案均存在局限性。方案一以因果关联为核心依据,但人工智能与医师共同作出诊疗决定的因果链条复杂导致争议不断;方案二基于亚里士多德责任理论,因相关主体无法完整认知和控制决策过程,容易出现“责任空场”等不利于司法救济的情形,患者的相关权利难以主张救济,不符合“以人为本”的医疗立场;方案三根据功能角色归责,存在角色责任无法完全区分及责任分配不平衡、没有获得普遍认同的问题。^⑧综上,受医疗人工智能的影响,作出诊疗决定的强制性与诊疗决定的责任模糊之间出现矛盾,即“生命医学与人工智能伦理原则整合后,矛盾在于既希望医疗人工智能技术的安全性 with 责任明确,又希望医疗人工智能能够成为可靠的辅助工具从而降低自身的工作负荷”^⑨的具象化表现。

① 熊静文:《诊疗规范中心论与医疗过失的判定——以〈民法典·侵权责任编〉相关条款的拟定为依归》,《浙江社会科学》2019年第7期,第38-47页。

② 朱晓婕、袁轲:《“诊疗活动”的司法认定标准——布康人体生态健康研究院有限公司诉上海市静安区卫生和计划委员会不服行政处罚决定案》,《上海法学研究》2019年第16卷,第88-93页。

③ 王建军:《论黑格尔“具体概念”的内在机制》,《哲学研究》2014年第2期,第60-64页。

④ 潘建均、颜基义、王玉民、陈泽玮:《决策内容与结构形式》,《中国软科学》2020年第3期,第10-18页。

⑤ 曾雄、梁正、张辉:《人工智能软法治理的优化进阶:由软法先行到软法与硬法协同》,《电子政务》2024年第6期,第96-107页。

⑥ 何邦武:《人工智能法律主体质论》,《东方法学》2024年第6期,第56-69页。

⑦ 孙那:《确立人工智能法律主体地位的再思考》,《法学论坛》2024年第5期,第112-121页。

⑧ 房帅帅、孙岩:《模糊认知地图:医疗决策人工智能归责问题的难点及其应对》,《中国医学伦理学》2025年第12期,第1552-1558页。

⑨ 汪琛:《医疗人工智能伦理治理的问题、困境与求解》,《科学学研究》2025年第2期,第414-422页。

三、适格诊疗决定的构成与责任归属

在医疗人工智能的持续冲击下,破“法律滞后”之局的关键在于对齐政策要求、规范诊疗决定的制定流程,破“医疗透明”之局的关键则在于突破技术黑箱、明确诊疗决定的适用规则。对此,笔者在“诊疗决定”之上进一步提出“适格诊疗决定”这一概念,聚焦诊疗决定的形成过程,通过对诊疗决定主体与流程开展全方位剖析,为构建医疗人工智能的责任认定机制寻求新的突破口。

(一)适格诊疗决定作出主体

医师是适格诊疗决定的主体,而医疗人工智能的出现会扩充决定主体吗?对这一问题应当舍弃技术角度的无谓观察,即避免讨论医疗人工智能是否可以得出与医师近似的准确的诊疗决定,以免陷入“强不知以为知”的无效论述困境。《中华人民共和国医师法》(以下简称《医师法》)第二十五条明确规定:“医师在诊疗活动中应当向患者说明病情、医疗措施和其他需要告知的事项。”也就是说,如果医师在诊疗活动中使用了医疗人工智能,应当向患者尽到告知义务。《法国公共卫生法典》第 L1111-4 条同样就医师告知义务作出规定:“患者应当根据专业人员告知的信息和提出的建议作出决定,医疗机构工作人员有义务尊重当事人的意愿,但如果患者因拒绝或中止治疗而使其生命处于危险之中,相关工作人员应当在合理的时间内告知患者决定的后果及其严重性,并采取措施维护临终病人的尊严。”此时,医师不再是诊疗决定的最终决定者,而是诊疗意见的提供者、解释者,患者有权质疑、要求修改,甚至否定或拒绝诊疗意见中的部分或全部内容。医疗人工智能的数据集交互和算法复杂性虽然使知情同意问题复杂化,传统知情同意出现向泛知情同意、动态同意转变的倾向,^①但决定主体依旧并非提出意见的医师个体,而是医师与患者的合意,甚至依照法律规定,实施手术、特殊检查、特殊治疗的主导权在患者手中(或须获得其近亲属的明确同意)。类似律师提供专业法律知识 with 意见,而非主导此次法律服务,如何采纳、是否采纳、采纳多少建议是在当事人主导下的合意所决定的。

在医师缺位的情况下,作出诊疗决定的场景发生了改变。信息化时代人们寻求医疗帮助的第一反应可能并非前往医院寻求医师的帮助,而是以自力救济为优先选项,通过互联网引擎、在线医疗平台或医疗人工智能等进行自我诊断,自主诊疗无果后再前往医疗机构求诊,此类患者亦被称为“百度病患”。在后续医师接诊后,患者的个性因素(如偏执、多疑或认知开放性)在医患信任调整中起到主导作用,医生普遍对“百度病患”存在两种心理状态,一是提升角色期待,二是对固执患者的防范心态。^② 当今大多数在线医疗平台均以医疗大模型(如灵医大模型、阿里健康 AI 等)或者通用大模型(如 Deepseek、混元通用大模型等)为基础,针对提问者的问题作出回应。此时便形成了没有医师参与的诊疗决定,医疗人工智能自然是作出该“诊疗决定”的行为主体。但此“行为主体”非彼“专业主体”,根据《医师法》第八条之规定,国家实行医师资格考试制度,医师资格考试又可以细分为执业医师资格考试和执业助理医师资格考试。也就是说,作出诊疗决定的主体必须通过上述考试,获取执业资格,而无论何种人工智能均不可能以适格主体的身份参与该考试,进而无法获得医师执业资格,其作出的“诊疗决定”仅满足形式要件,本质上并不能称之为“诊疗决定”,更为恰当的称谓是“医疗建议”,这与家人叮嘱“咳嗽时吃梨润肺”等日常建议并无本质区别。

^① 石佳友、徐靖仪:《医疗人工智能应用的法律挑战及其治理》,《西北大学学报(哲学社会科学版)》2024年第2期,第91-103页。

^② 吕小康、徐恺琳:《网络时代的“好病人”:医生眼中的“百度病患”与网络医学信息》,《社会科学研究》2025年第1期,第100-108页。

针对上述观点,可能有学者会以“当下医疗人工智能已被明确定义为辅助诊断工具及临床决策支持系统”为由予以反驳,其法律依据是《人工智能辅助诊断技术管理规范》,并得出“医疗人工智能具有人工智能的共性,因而不可能取得医疗主体地位,亦无法构成医事犯罪的主体”的结论。^①立法者习惯于套用传统部门法的理论范式,惯性地以各部门法的逻辑来控制风险,^②但这种方式在数字时代,尤其是人工智能技术革新浪潮中已难以适用。因为人工智能技术的发展尚未受到法律的充分制约,所以对具有辅助诊断作用的人工智能加以规范是为了实现技术的标准化应用,而非对未来技术发展进行定性限制,这也会违背法律固有的滞后性特征。易言之,《人工智能辅助诊断技术管理规范》仅针对当下的人工智能辅助诊断技术,而非代表未来人工智能在医疗领域仅限于辅助诊断功能。相关数据显示,人工智能算法在卵巢子宫内膜异位症的MRI诊断中,模糊C均值聚类算法准确率高达94.3%,远高于传统MRI的63.2%,同时在剖宫产和阴道分娩检测中,机器学习的敏感性达94%,^③医疗人工智能有望做到自主进行诊疗决定,这不是对人工智能主体性的承认,而是在医疗这样关乎人类生命、身体健康等根本利益的特殊领域,伦理与正当性不应简单地让步于单纯的速率与准确性。

(二)适格诊疗决定作出过程

可解释性不是适格诊疗决定作出的必要条件,自然人作出决定时同样存在“黑箱”现象。诊疗决定作出过程的是否适格不在于实质而是在于形式,具体可分成两个部分:医师向患者提出诊疗方案与患者和医师间达成诊疗决定。

《医师法》第二十二条规定了医师的权利,该条第一款为“在注册的执业范围内,按照有关规范进行医学诊查、疾病调查、医学处置,出具相应的医学证明文件,选择合理的医疗、预防、保健方案”。医师没有诊疗决定的最终决定权,其享有的是诊疗方案的选择权与制定权。医疗人工智能针对患者诊疗的协助理应同属诊疗的组成部分,医师需在诊疗方案中予以注明,以保障患者的知情权。医疗人员应当及时向患者说明诊疗过程将存在医疗人工智能参与,同时就诊疗方案占比,为诸如病灶识别系统、医学专家系统和医疗机器人系统等设置不同级别的知情同意标准。^④这正是人工智能辅助医疗决策系统得以减轻医师和科研人员工作压力、高效处理繁琐的医疗数据分析工作、提高诊疗效率的核心意义所在。“人类简史三部曲”作者尤瓦尔·赫拉利曾预测,未来关键医疗决定可能由大数据算法主导,这一趋势已在现实生活中逐渐显现。

再观察患者与医师达成诊疗决定阶段,该阶段强调信任的重要性。“谁决策,谁负责”的权责配置模式,实质上是将诊疗决策权更多地向患者倾斜。然而,此种权力配置模式给患者施加了较多的负担。当面临关乎生命存续的决策场景时,患者往往会受到潜意识及非理性因素的干扰。进一步而言,医师作为核心的信息供给主体,其行为模式的差异可能致使患者作出截然相悖的决策选择,而该决策的最终责任却全然由患者承担。^⑤基于不信任,患者可能对医师提供的诊疗方案产生一种无关乎治疗效果的排斥,更不用说在诊疗方案中融入人工智能提供内容的情形了。有学者指出这是诊疗领域中信任与不信任的错置,即从伦理角度看,存在以不正义方式信任值得信赖的对象和

① 皮勇:《论医疗人工智能的刑法问题》,《法律科学(西北政法大学学报)》2021年第1期,第134-148页。

② 宋亚辉:《风险控制的部门法思路及其超越》,《中国社会科学》2017年第10期,第136-158页。

③ Hassan M. Elbiss & Fikri M. Abu-Zidan, *Artificial Intelligence in Gynecologic and Obstetric Emergencies*, *International Journal of Emergency Medicine*, Vol. 18, No. 1, 2025, pp. 1-8.

④ 游俊哲、丁岱睿:《数智时代医疗人工智能的潜在法律风险及规制措施》,《中国科技论坛》2025年第3期,第129-140页。

⑤ 孙保学:《人工智能辅助医疗决策并未挑战尊重自主原则》,《伦理学研究》2019年第6期,第81-86页。

以不正义方式怀疑不可信赖的对象的情况,需要区分人类对医疗人工智能的信任是基于其外在或内在可信赖性,还是基于错误或不合理的信念。^① 需要指出的是,在引入医疗人工智能后,患者对于诊疗决定的信任并不必然延伸至人工智能,因为“在许多情况下,患者并不直接与人工智能互动;他们必须信任提供诊疗方案的医师能够明智地采用人工智能,并审慎地遵循人工智能给出的建议”^②,信任依旧未超出自然人间的范畴。

当医患双方就诊疗方案达成合意,该合意便可视为完成适格诊疗决定的法定程序。此后,医师按照该诊疗决定执行,并在告知患者并获得患者同意的情况下可再作适当修正。再次强调,是否可以被认定为适格诊疗决定的核心在于程序,而非结果,不是无效的诊疗结果就意味着诊疗决定的不适格,亦非有效的诊疗结果就意味着作出了一个适格的诊疗决定。结果论只会造成后续医疗损害事件中各方责任判断的混乱。信任与合意虽为抽象概念,但正如机器学习模型的不透明性与多种常见医疗干预措施的不透明性一样,虽无法作具体解释,却效果显著且被广泛接受。^③

(三)基于适格诊疗决定的医疗损害责任

现阶段无论是适格诊疗决定作出的主体还是作出的过程均与选择了医疗人工智能参与诊疗决定的医师直接相关,若出现医疗损害或医疗纠纷,应构建“谁选择,谁担责”的医疗损害责任制度。我国《民法典》对医疗损害责任采取广义界定,至少包括违反诊疗义务的损害赔偿责任(诊疗损害责任)、违反告知义务的损害赔偿责任(告知损害责任)、医疗产品责任以及侵犯患者隐私权或个人信息的损害赔偿责任四种类型。^④ 鉴于诊疗决定的讨论不涉及最后一种损害赔偿责任,故后续不作论述。

美国自上世纪五十年代开始探索主诊医师负责制,其初衷是推动医院卫生资源的合理配置与医疗人才的培养,实行后不仅大幅提高了医师的工作积极性,更使医疗服务的数量及质量得到明显提升,但该制度尚未上升至法律层面。^⑤ 后续医师团体逐渐意识到,主诊医师负责制不仅是对诊疗方案的负责,更需对偶发性的医疗损害后果承担相应责任。^⑥ 按照这个逻辑,主诊医师不仅要对医疗人工智能引发的医疗损害后果承担责任,其责任实质在于将医疗人工智能决策引入诊疗方案并采信其输出结果。同时,主诊医师负责制也保护了医师的相关权益,若不是因为主诊医师诊疗行为而产生的医疗损害,责任自然不会落于医师之身,法律亦没有就类似情形规定医师的无过错责任。此时医疗损害责任若源自产品自身,需从医疗损害责任中剥离医疗产品责任,针对不同情形适用不同责任规则,还需结合两者共同致人损害的情况处理,故有学者建议出台相关标准和建立鉴定委员会。^⑦

主诊医师负责制明确了医师就使用医疗人工智能行为仅承担诊疗损害责任或告知损害责任,

^① Georg Starke & Marcello Ienca, *Misplaced Trust and Distrust: How Not to Engage with Medical Artificial Intelligence*, Cambridge Quarterly of Healthcare Ethics, Vol. 33, No. 3, 2024, pp. 360-369.

^② Madeline Sagona, et al, *Trust in AI-assisted Health Systems and AI's Trust in Humans*, npj Health System, Vol. 2, 2025, pp. 1-5.

^③ Joshua Hatherley, Robert Sparrow & Mark Howard, *The Virtues of Interpretable Medical Artificial Intelligence*, Cambridge Quarterly of Healthcare Ethics, Vol. 33, No. 3, 2024, pp. 323-332.

^④ 程啸:《侵权责任法》(第3版),法律出版社2021年版,第632页。

^⑤ 徐僖等:《主诊医师负责制的实践与探讨》,《医院管理论坛》2020年第7期,第36-37页。

^⑥ 詹剑、聂业、欧册华:《论院内急救气管插管术麻醉医师负责制的法律问题及对策》,《医学与法学》2017年第2期,第78-82页。

^⑦ 钟晓雯、高洁:《“人工智能+医疗”的风险研判及治理路径》,《广西师范大学学报(哲学社会科学版)》2024年第4期,第29-41页。

医疗产品责任应当由医疗人工智能的开发者或研发人员等主体承担。在医疗场所内开展的人工智能辅助诊疗活动中,医疗机构作为诊疗服务的组织者与管理者,应承担超出医师责任的补充义务与监督责任。例如,医疗机构需对所采购医疗人工智能设备的合规性、适配性进行审慎审查,建立定期维护与性能校验机制,并对医师开展人工智能使用规范的专项培训。若因医疗机构未履行上述义务导致医疗损害,即便医师无主观过错,医疗机构也应承担相应的补充责任。这一责任分配既符合“医—机—患”三元关系中医疗机构的中间枢纽定位,也通过强化机构监管义务,进一步筑牢患者接受医疗的安全底线。有学者就使用医疗人工智能触犯医疗事故罪的角度指出,犯罪过失本质是违反注意义务,在医疗人工智能介入下,医务人员需对其负有注意义务,并不能依据信赖原则豁免,并就“当时的人工智能医疗水平”作出四级分类:诊疗惯例、临床指南、专家共识和医学文献。^① 笔者认同该观点的前半部分,即医疗损害的发生是医师注意义务的缺失,亦不能通过信赖原则豁免,但过度严苛地规定“医疗水平”可能会导致医师们不敢适用甚至滥用医疗人工智能,出现诊疗速率慢或诊疗总体水平下降等不利于患者的结果。应当适当降低注意义务标准,以鼓励医师合理利用医疗人工智能。在主诊医师负责制与“医师—患者”合意决定诊疗的双重保障下,无需通过刑法手段强化医疗人工智能使用的事故责任。这不仅符合“互益原则”^②,而且有利于促进医疗人工智能在实践中的创新与发展。

针对使用医疗人工智能而造成的医疗损害责任承担问题,全球多数法域仍处于立法空白阶段。欧盟和西班牙在医学领域人工智能应用的监管虽已有雏形,但仍缺乏专门且明确的规范。欧盟的人工智能规范与基本权利宪章的核心内涵保持相对一致,而西班牙在产品责任、数据保护等方面虽有相关规范,但无针对医疗领域人工智能的具体内容。^③ 美国《二十一世纪治愈法案》明确指出,如果一个人工智能产品不提供诊断或治疗服务,也没有给出明确的诊断结论或具体的治疗方案,而仅提供健康监测、咨询等服务,这类产品就不属于美国食品药品监督管理局的监管范围,^④这表现出美国大力推动医疗人工智能发展、占领人工智能技术高地的欲望。

四、医疗人工智能造成损害的归责模式

医疗人工智能产品与一般产品不同,适用过错责任有利于技术创新和责任认定。^⑤ 《民法典》第一千二百二十二条规定了诊疗活动中的过错推定的情形。需要注意的是,医疗人工智能在不同应用场景下的差异将直接影响医疗损害责任主体的认定。通过从医疗人工智能技术发展进程中区分内部研发与临床应用两个维度,可系统性地厘清不同情形下的归责模式。

(一)时间轴:医疗人工智能在诊疗决定中的角色转变

首先是产品阶段。当下的医疗人工智能可分为医疗器械用途的医用人工智能和提供医疗服务的医用人工智能,前者将人工智能技术嵌入医疗设备或软件,自动化提供数据或依医护人员指令完成特定医疗行为;后者则强调服务属性,可独立与患者或医护人员交互提供服务,^⑥如域外开发了

^① 霍俊阁:《人工智能视域下医疗事故罪的适用嬗变与调适》,《中国人民公安大学学报(社会科学版)》2023年第2期,第61-70页。

^② Jonathan Adams, *Defending Explicability as a Principle for the Ethics of Artificial Intelligence in Medicine*, *Medicine, Health Care and Philosophy*, Vol. 26, No. 4, 2023, pp. 615-623.

^③ Óscar Andrés Molina, et al, *What Is Spanish Regulation on the Application of Artificial Intelligence to Medicine Like?* *Humanities and Social Sciences Communications*, Vol. 11, 2024, pp. 1-7.

^④ 李润生:《论医疗人工智能的法律规制——从近期方案到远期设想》,《行政法学研究》2020年第4期,第46-57页。

^⑤ 王利明:《生成式人工智能侵权的法律应对》,《中国应用法学》2023年第5期,第27-38页。

^⑥ 罗仟合:《伦理法视域下医用人工智能的治理研究》,《云南大学学报(社会科学版)》2025年第3期,第133-144页。

一款针对家长在孩子牙齿创伤后寻求信息和安慰的医疗聊天机器人。^①但无论是已有的医疗人工智能还是未来可能出现的“人工智能医师”,均无法脱离“产品”的范畴,所以产品责任应当贯穿其从生产至销毁全生命周期。同时,产品责任中的无过错原则能够督促生产者尽最大努力保证人工智能产品的安全。产品责任中的三类产品缺陷(设计缺陷、制造缺陷和警示缺陷)亦可在人工智能产品生产中加以适用,我国现行产品责任的救济制度较完善,可以为医疗人工智能产品的规定提供参考。^②

其次是医疗决策提出阶段。医疗人工智能通过大量数据训练后可就患者实际情况提出医疗决策,需就医师是否参与再作区分。当医师未参与医疗决策时,相当于“机一患”双方直接对话,如“百度病患”情形,主导医疗人工智能提出医疗决定乃至诊疗决定的是患者,而医疗人工智能仅是对患者方给出的信息作出推断,诊疗决定的决定权在患者手中。当医师参与医疗决策时,医师提出建议形成诊疗方案后再与患者沟通,医师作为专业知识者参与避免了人机直接对话,形成“医一机一患”的三元关系。在此过程中,虽然最终诊疗决定权仍由患者行使,但医师的专业意见对决策产生重要影响。由此可见,医疗人工智能始终保持着医疗决策建议者的角色定位。

医疗人工智能能否成为医疗决策的主导者?在医疗人工智能未失控的情况下,其并不会成为真正意义上的主导者,但可能会出现患者“唯范是从”的情形。所谓“唯范是从”,指的是患者忽视寻求医师建议的主观能动性,一味根据医疗人工智能给出的医疗决策自行治疗,最终导致损害结果的发生。这种情形本质上是自然人对自主权利的放弃,应自担风险,不能将此时医疗人工智能视为最终造成损害结果的医疗决策的主导者。若未来医疗人工智能可自主执医,便意味着在执业医师考核时将其与自然人同等看待,应当一视同仁地适用主诊医师负责制,而无需单独设置“主导者”身份用以归责。

(二)生产者:因医疗人工智能自身原因造成医疗损害的归责主体

在适用产品责任的前提下,因产品造成的医疗损害赔偿应当由生产者承担。考虑到人工智能固有的“技术黑箱”属性,需要对生产者就何种原因承担责任予以进一步说明。

一是医疗人工智能的安全性。安全,作为当今技术发展的核心要义,在归责判断中占据举足轻重的地位。参考我国《数据安全法》对于数据安全的定义——“通过采取必要措施,确保数据处于有效保护和合法利用的状态,以及具备保障持续安全状态的能力”,医疗人工智能安全可定义为:“通过采取必要措施,确保医疗人工智能运行处于有效和可控状态,以及具备保障持续安全状态的能力。”这种安全状态具有双向性:若医疗人工智能使用者的健康安全得到持续保障,亦可反证该系统的安全合规性。

二是医疗人工智能的准确性。医疗领域与生命健康权紧密相关,简单的医疗行为就可能对个体造成严重损害,不能抱有侥幸心理。医疗人工智能除了要经过监管者的严格审核,还要求生产者实时更新医疗数据,以作出更准确的诊疗决定。以医疗机器人“达芬奇”为例,其投入市场后引发多

^① Mihriban Gökçek Taraç & Tuğba Nale, *Artificial Intelligence in Pediatric Dental Trauma: Do Artificial Intelligence Chatbots Address Parental Concerns Effectively?* BMC Oral Health, Vol. 25, 2025, pp. 1-9.

^② 张安毅:《人工智能侵权:产品责任制度介入的权宜性及立法改造》,《深圳大学学报(人文社会科学版)》2020年第4期,第112-119页。

起产品责任诉讼,部分法院认为制造商应负有对机器人的“持续训练义务”。^①虽然生成绝对准确的诊疗决定并不存在,但生产者至少应承担基本的注意义务,以确保医疗人工智能能够作出相对准确且最优的诊疗决定。

三是医疗人工智能的告知义务。2025年国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局联合发布《人工智能生成合成内容标识办法》,明确规定人工智能生成合成服务(参照《互联网信息服务深度合成管理规定》第十七条第一款,包括智能对话、智能写作等模拟自然人进行文本的生成或者编辑服务)提供者需履行告知义务。若生产者未明确告知相关内容系医疗人工智能生成,则应承担产品服务瑕疵责任。

(三) 医师:因诊疗决定中采纳医疗人工智能造成医疗损害的归责主体

《医师法》第二十四条规定,医师实施医疗、预防、保健措施,签署有关医学证明文件,必须亲自诊查、调查,并按照规定及时、完整地填写病历等医学文书;若诊疗过程中使用医疗人工智能,需将具体应用情况如实记录入医学文书,并依法承担相应责任。

相较于后端运营的生产者,医师(医疗机构)是医疗人工智能的前端使用者,有权决定医疗人工智能输出结果的应用方式,应当就采纳医疗人工智能输出结果的行为承担医疗损害赔偿责任。鉴于人工智能应用导致过错认定面临阻碍,欧盟尝试借助产品责任这一无过错责任形式,减轻被侵权人的举证负担。无过错责任虽然在人工智能侵权场景中较过失责任可能更具效率优势,但与私法自治理念相悖,适用时需审慎权衡。人工智能侵权仍应采过错归责原则,且不应贸然规定为过错推定责任。针对人工智能特性引发的过错证明难题,法秩序应通过专门的信息披露规则予以解决。“重大医疗瑕疵原则”的主要目的不在于惩罚医师,而在于解决患者举证困难问题。我国适用该原则时,应重点评估医疗机构隐匿、篡改、伪造或者销毁病历的行为是否影响因果关系等基本事实的查明。若该证明妨碍行为导致基本事实无法认定,则可适用该原则推定医疗行为与损害结果存在因果关系,此时应由医疗机构或医师证明其重大医疗瑕疵与患者的损害无关联。若医师依照法律、行政法规、规章以及其他有关诊疗规范中关于医疗人工智能参与诊疗决定的信息披露规定并按照标准流程作出适格诊疗决定,则不应就采纳医疗人工智能行为承担医疗损害责任。

当医疗损害源于人工智能的设计缺陷、制造缺陷或警示缺陷,且医师已履行合理注意义务时,责任应由生产者承担。若医疗损害系医师与医疗人工智能生产者的共同过错导致,则应根据双方过错程度承担按份责任。此外,医师在诊疗过程中发现人工智能存在异常输出或潜在风险时,负有向医疗机构及生产者及时反馈的义务,未履行该反馈义务导致损害或损害扩大的,需承担相应的法律责任。这一要求既强化了“医—机—患”关系中的医师把关责任,也为生产者优化医疗人工智能产品提供了实践反馈渠道。为进一步明确医师使用医疗人工智能行为的相应责任,《医师法》后续修订中可在第二十二条款中增设专项条款,规定“医师有权在执业活动中使用人工智能”,基于权责一致原则,基法律责任亦应同步明确。

(四) 患者:对自主采纳医疗人工智能建议造成医疗损害承担责任

根据汉斯·伦克(Hans Lenk)的责任理论,责任是待确定的。即使是因为自身行为所引发的损害结果,该个体同样是一个适格的责任承担者。有许多学者认同患者在一定程度上也要为医疗

^① 钱思雯:《弱人工智能时代的法律回应——构建以产品责任为核心的责任分配体系》,《中国科技论坛》2019年第9期,第76-84页。

人工智能参与其诊疗决定而承担相应责任:从技术发展的角度,若在医疗人工智能使用中过度苛责医疗机构或诊疗类人工智能生产者,将妨碍人工智能技术的发展。^①从自由意志的角度,即使医疗人工智能不是自由意志的自然延伸,患者也参与了自由意志(诊疗决定)的形成。^②从因果关系认定的角度而言,由于患者自身参与了诊疗决定过程,甚至拥有否定权,最终的医疗损害结果与患者的自主选择行为存在因果联系。但不必苛求任何一方的“预见义务”,并非所有的损害结果均有对应的责任主体,也不是所有的损害结果必然与损害赔偿相挂钩。这一点在医疗等不确定因素多于确定因素的领域表现得尤为突出,^③这间接导致制定归责体系的困难。根据《民法典》第一千一百七十四条规定,若损害是因受害人故意造成的,行为人不承担责任。在上述生产者履行自身职责后,若患者自主采纳医疗人工智能建议造成损害结果,应当自行承担,生产者(含医疗人工智能)则不承担责任。

诸如监管者等其他角色应虽然其与诊疗决定所引发的医疗损害无直接利害关联,但仍可能因为未尽到监督、管理、审批医疗人工智能的职责而承担间接或其他责任。

五、结语

医疗人工智能在临床中的应用既是技术革命对医疗领域的结构性重塑,又向法学界提出了人机权责分配的时代命题。当诊疗决定模式从“医师主导”演进为“人机协同”,法律的回应既需恪守“以人为本”的伦理底线,亦需为技术创新预留制度空间,责任归属认定应慎之又慎。展望未来,破解医疗人工智能的法治困境,需从三方面着力:其一,加快人工智能专项专条立法以及与《民法典》《医师法》的衔接修订,明确医疗人工智能的法律定位与诊疗决定的效力边界,填补“政策先行、法律滞后”的空白;其二,司法实践应坚持“程序优先于结果”的裁判导向,以“适格诊疗决定”要件作为责任认定的基础,避免陷入对技术黑箱的无谓讨论与追责;其三,医疗行业需推动内部人工智能的透明化治理,建立备案制度、医师使用医疗人工智能记录等配套机制,确保技术进步始终运行在法治轨道内。让医疗人工智能既赋能诊疗效率,又坚守患者本位,通过法治与创新的协同,护航健康中国建设。

① 金贤婷:《诊疗人工智能致害的相关理论争议、归责模式与责任承担路径》,《医学与法学》2025年第5期,第39-47页。

② 朱振:《归责何以可能:人工智能时代的自由意志与法律责任》,《比较法研究》2022年第1期,第39-54页。

③ 李坤海、徐来:《人工智能对侵权责任构成要件的挑战及应对》,《重庆社会科学》2019年第2期,第55-65页。

The Patient-Centered Model of Qualified Medical Decision-Making and the Determination of Medical Malpractice Liability from Medical Artificial Intelligence

WU Mengting, TANG Hongyu

(School of Public Affairs, University of Science and Technology of China, Hefei 230026, China)

Abstract: The application of medical artificial intelligence in medical treatment is becoming increasingly profound, yet it raises theoretical questions from three perspectives: the alignment with policies and laws, transparency requirements in medical treatment and the accountability logic for diagnosis and treatment decisions. However, the current legal framework does not clearly define the specific role and liability allocation of Medical Artificial Intelligence (MAI) in clinical decision-making, leaving traditional medical malpractice liability pathways at a crossroads. In fact, qualified diagnostic decisions should be made through mutual agreement between physicians and patients, and the current role of MAI remains limited to providing medical recommendations. The validity of these decisions hinges on procedural compliance. Regarding liability in cases of harm caused by MAI, manufacturers should bear liability for the safety of AI products, while physicians, as actual users of MAI, should be accountable for damages resulting from their recommendations. If patients independently adopt AI suggestions and suffer harm, they should bear the risk themselves.

Key words: medical artificial intelligence; qualified diagnostic decision; liability allocation of medical malpractice; “doctor-machine-patient” relationship; physician responsibility system

(责任编辑:董兴佩)