

• 专家论坛 •



于广军 国家“百千万”人才工程入选者，国家有突出贡献的中青年专家，享受国务院政府特殊津贴。现任香港中文大学（深圳）医学院副院长兼附属第二医院院长、博士生导师、国家健康医疗大数据研究院（深圳）院长，兼任中国医院协会信息专业委员会副主任委员。曾先后获得上海市领军人才和上海市优秀学术带头人称号。主持国家重点研发计划重点专项、国家高技术研究发展计划（“863”计划）、国家自然科学基金重大研究计划、国家自然科学基金面上项目等多项科研项目。作为第一作者或通信作者发表学术论文 80 余篇，主编专著 7 部。获国家科技进步奖二等奖、上海市科技进步奖一等奖、中国医院协会医院科技创新奖及科技部“十一五”国家科技支撑计划优秀执行团队奖等奖项。

DOI: 10.16781/j.CN31-2187/R.20250443

基于人工智能的临床辅助决策支持系统问责制度探索

彭颖¹, 杨超然², 于广军^{1,2*}

1. 香港中文大学（深圳）医学院, 深圳 518172

2. 香港中文大学（深圳）附属第二医院, 深圳 518172

[摘要] 随着人工智能（AI）在医疗领域的快速发展，临床辅助决策支持系统（CDSS）逐步融入临床诊疗流程。然而，CDSS 在提升诊疗效率与精准性的同时也引发了复杂的责任归属界定问题。本文围绕 CDSS 问责困境的国内外现状展开系统讨论，从责任主体模糊、算法“黑箱”特性、数据偏倚、法律与监管制度落后与部署环境异质性 5 个方面深入剖析了 CDSS 问责困境的形成机制，并提出以全生命周期治理为核心的协同治理框架：主张通过技术路径增强 AI 的可信度，通过法律与监管路径构建风险分级监管与无过错赔偿基金等创新机制，通过组织路径强化机构内控，通过伦理与教育路径重塑“人机协同”的职业新范式。破解 CDSS 问责困境既要保障患者安全与医疗公平，也要为身处技术变革中的医务人员提供清晰的责任边界与坚实的职业保障，从而推动医疗 AI 技术健康发展。

[关键词] 人工智能；临床辅助决策支持系统；医疗责任；问责；医疗伦理

[引用本文] 彭颖, 杨超然, 于广军. 基于人工智能的临床辅助决策支持系统问责制度探索[J]. 海军军医大学学报, 2025, 46(8): 963-969. DOI: 10.16781/j.CN31-2187/R.20250443.

An exploration of accountability frameworks for artificial intelligence-driven clinical decision support systems

PENG Ying¹, YANG Chaoran², YU Guangjun^{1,2*}

1. School of Medicine, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, Guangdong, China

2. The Second Affiliated Hospital of The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, Guangdong, China

[Abstract] With the rapid advancement of artificial intelligence (AI) in healthcare, clinical decision support system (CDSS) is gradually integrated into clinical workflows. While CDSS enhances diagnostic and therapeutic efficiency and accuracy, it has also sparked complex debates regarding accountability. This article systematically examines the current global landscape of accountability challenges associated with CDSS. It delves into the mechanisms underlying these challenges, focusing on 5 key dimensions: ambiguity in liability attribution, the “black-box” nature of algorithms, data bias, backward

[收稿日期] 2025-07-02

[接受日期] 2025-07-25

[基金项目] 癌症、心脑血管、呼吸和代谢性疾病防治研究国家科技重大专项(2024ZD0524100, 2024ZD0524102), 深圳市医学研究专项资金项目(C10120250085), 深圳市基础研究专项(JCYJ20240813160307011). Supported by National Major Science and Technology Project for Prevention and Treatment of Cancer, Cardiovascular and Cerebrovascular, Respiratory, and Metabolic Diseases (2024ZD0524100, 2024ZD0524102), Shenzhen Medical Research Fund Project (C10120250085), and Shenzhen Basic Research Project (JCYJ20240813160307011).

[作者简介] 彭颖, 博士生. E-mail: 224050082@link.cuhk.edu.cn

*通信作者 (Corresponding author). E-mail: guangjunyu@cuhk.edu.cn

legal and regulatory systems, and heterogeneity in deployment environments. To address these issues, a collaborative governance framework centered on whole-lifecycle governance is proposed. Specifically, the framework advocates: enhancing AI trustworthiness through technical pathways; constructing innovative mechanisms such as risk-based tiered regulation and no-fault compensation funds through legal and regulatory pathways; strengthening institutional internal controls through organizational pathways; and reshaping a new professional paradigm of “human-AI collaboration” through ethical and educational pathways. The study argues that resolving CDSS accountability dilemmas is not only critical for fostering trustworthy AI in medicine but also essential for safeguarding patient safety, ensuring healthcare equity, and providing clinicians with clear liability boundaries and robust professional protections amid technological transformation.

[Key words] artificial intelligence; clinical decision support system; medical liability; accountability; medical ethics

[Citation] PENG Y, YANG C, YU G. An exploration of accountability frameworks for artificial intelligence-driven clinical decision support systems[J]. Acad J Naval Med Univ, 2025, 46(8): 963-969. DOI: 10.16781/j.CN31-2187/R.20250443.

随着人工智能 (artificial intelligence, AI) 技术在医疗领域的快速发展, 临床辅助决策支持系统 (clinical decision support system, CDSS) 日益成为智慧医疗体系构建的重要支撑。当前 CDSS 已在中国部分区域实现规模化部署, 但其整体发展与先进国家相比仍存在差距^[1]。伴随着 CDSS 应用的深入和推广, 一系列结构性困境随之暴露, 尤其是在责任归属与问责机制构建方面引发了广泛争议。当诊疗出现偏差时, 传统的归责方式在新的诊疗模式下难以为继, 特别是由于 AI 的“黑箱”特性, 临床医生常常对其内部逻辑缺乏充分理解与掌握, 却又被传统法律框架定位为“最终责任保障者”^[2]。面对这一争议, 国际社会开始积极探索解决方案。2023年 WHO 发布《卫生领域人工智能的监管考虑》^[3], 强调在 AI 系统应用过程中必须明确各参与方 (开发者、供应商、医疗机构、使用者等) 的法律边界, 防止在出现损害事件时产生责任“真空”。在这一全球性共识的指引下, 欧盟在其 2024年3月正式批准的《人工智能法案》^[4]中明确将 CDSS 等医疗 AI 系统划分为“高风险系统”, 这意味着相关产品的开发者必须在上市前进行严格的合格评定, 并建立覆盖整个产品生命周期的风险管理和上市后监督体系。欧盟委员会提出的《人工智能责任指令》草案^[5]试图通过设立因果关系推定、要求开发者和使用者披露证据等创新机制, 直接回应 AI “黑箱”给使用者维权带来的举证困难。在中国, 顶层设计与监管框架正逐步设立。在国家科技部发布的《新一代人工智能伦理规范》^[6]等宏观伦理原则的指导下, 国家药品监督管理局 (National Medical Products Administration, NMPA) 基于《人工智能医疗器械注册审查指导原

则》^[7]将核心 CDSS 产品纳入医疗器械监管体系, 要求其遵循全生命周期质量控制; 国家互联网信息办公室等部门联合发布的《生成式人工智能服务管理暂行办法》^[8]更是强调服务提供方 (如 AI 平台、医院部署方) 为归责重点, 而非用户。然而, 这种从产品准入到临床实践的治理理念传导仍存在巨大鸿沟: 在具体的法律责任认定上, 尚缺乏与《民法典》《医师法》相配套的司法解释; 在医疗机构的实践中, 也远未形成能够有效吸纳、评估和监督 AI 应用的伦理适应与风险共担机制。

1 CDSS 问责困境的国内外现状

CDSS 在责任归属与问责机制上的困境是一个全球性挑战, 但在不同国家和地区的法律文化与应用阶段下其表现形式与问题焦点存在显著差异。

在以美国为代表的西方国家中, CDSS 的应用起步早、程度深, 其问责难题已通过具体的争议性事件和深入的司法探讨充分暴露。以曾在美国数百家医院使用的 Epic 脓毒症预测模型为例, 该模型宣称其能够有效预测脓毒症的发病情况。2021年, 密歇根大学研究团队对该预测模型进行了严格评估, 结果令人震惊: 该模型对脓毒症的识别率远低于其宣称的性能, 甚至显著低于简单临床筛查标准^[9]。这种显著的效能不足使临床医生面临一个典型的“责任陷阱”: 如果医生遵循了模型报告的可能错误的高风险警报而投入资源进行排查时, 可能造成对模型漏报的真正危重患者延误诊治, 或对真正的低风险患者进行不必要干预, 从而导致医生可能因医疗过失面临诉讼; 如果医生基于专业判断而选择忽略一个他们认为不准确或无关紧要的风险警报但该患者后续确实发生脓毒症时, 医生则可能被指

控忽视系统警告,同样面临法律风险。与此同时,Epic公司作为系统开发者,为规避责任可以辩称其模型仅为辅助工具,最终决策由医生做出。IBM沃森肿瘤专家项目则暴露了问责链条的另一端——算法设计与训练数据的责任。该项目因训练数据存在严重局限性和偏见(如过度依赖特定医疗中心的非规范化病历、模拟病例数据,缺乏真实世界多样性和严谨循证数据),导致系统有时会推荐出不符合临床指南甚至具有潜在安全风险(如药物剂量错误、方案不适用)的治疗方案^[10]。这引发了关于开发者算法设计责任与医疗机构审查责任如何划分的复杂争议。这些事件深刻反映了在成熟的法律环境下,CDSS存在责任主体多元化且权责边界模糊的特征,系统性风险往往被不当转嫁或分散,最终在临床一线或算法源头形成难以填补的“责任鸿沟”^[11]。

相比之下,我国的CDSS应用尚处于规模化推广的初期,问责困境更多体现在宏观的制度性缺失与微观的实践性规范不足^[12]。首先,CDSS的法律属性正从模糊走向明确,但新的合规挑战随之而来。过去其法律地位确实存在争议,但NMPA于2022年发布的《人工智能医疗器械注册审查指导原则》^[7]明确将具备辅助诊断或决策功能的软件界定为医疗器械(通常为第二类或第三类)进行严格监管。这意味着CDSS的开发、验证、审批和上市后监管都必须遵循医疗器械的法规路径,其法律属性已不再是主要障碍。然而,这一转变也给开发者和医疗机构带来了巨大的合规压力,司法实践中如何依据其医疗器械的属性来裁定各方责任仍缺乏成熟的判例指引。其次,数据质量与区域发展不平衡所引发的“数据原罪”在新的法律框架下责任更加聚焦。我国优质医疗资源集中的大型三甲医院是AI模型训练数据的主要来源。当这些模型部署到数据特征与质量迥异的基层医院时其泛化能力不足的问题尤为突出,这一问题不仅是技术风险,更直接关系到法律责任。这意味着,如果模型因训练数据存在偏见或采集不合规而导致误诊,其责任不仅涉及数据提供方(医疗机构)的法律合规性,还可能追溯至算法设计者,责任链条更为复杂。我国2025年发布的《人工智能生成合成内容标识办法》^[13]强调了提高信息透明度与可信度,减少误导、虚假和生成内容滥用风险的重要性。最后,法律地位的明

确化与技术不确定性共同加剧了医疗实践中的“防御性应用”心态。一方面,部分医生担心CDSS作为“持证上岗”的医疗器械,其建议更具“权威性”,若不遵循可能面临更大的法律风险,从而产生过度依赖。另一方面,开发者虽然受到更严格的监管,但仍倾向于通过用户协议中的免责条款来转嫁临床应用的最终风险。这种双向的责任规避最终损害的是患者的根本利益,如何依据我国《民法典》《医师法》等现有法律,构建一个既能鼓励创新又能保障患者安全且明确划分开发者、医疗机构和医生之间权责的协同治理框架,并将其落实到机构的准入、培训和监督实践中,存在巨大的探索空间。

2 CDSS 问责困境的形成机制：一个多维度的结构性难题

2.1 责任主体模糊

CDSS的决策是一个典型的“多元参与者”协作过程,由开发者、集成商、医疗机构、医生及患者多方协作完成,这种多方协作的模式在提升决策质量的同时也模糊了责任边界^[14]。正如Tizhoosh和Pantanowitz^[15]在数字病理学等具体实践中所揭示的,责任事实上分布在一个包含数据、算法、软硬件、用户和机构的复杂生态系统,一旦出现不良医疗后果,追责便会陷入“谁是过错方”的困境。

这种模糊性集中体现在对医生角色的界定上,法律和伦理普遍要求医生对最终决策负责,将CDSS仅视为辅助工具。这种将医生定位于“最终责任保障者”的本质是算法带来的责任缺口的具象化^[16]。当系统建议与医生判断相悖时医生将面临两难:若采纳系统建议后出错,医生是否会因“盲从”而承担全部责任;若未采纳系统建议而出错,医生又是否会因“忽视”先进技术提供的潜在正确选项而被追责^[17]。医生往往对AI的内部运作无法做到充分理解与掌控,却被要求承担超出其控制能力的责任,使得其“守门人”角色难以获得清晰的免责或担责界定。

2.2 算法“黑箱”与可解释性缺失

现代CDSS,特别是基于深度学习模型的系统,其内部运作机制常被视为算法“黑箱”。尽管这类系统能够在特定条件下输出高精度的预测或建议,但其依赖多层神经网络与复杂权重参数的计算过程,导致其推理

逻辑难以被人类直观理解与追溯。这种内在的可解释性缺失 (lack of explainability/interpretability) 被普遍认为是阻碍 CDSS 有效问责的关键技术瓶颈^[18]。当系统输出错误结论时,其“黑箱”特性使得开发者、监管者和临床医生难以回溯并解析导致错误的具体逻辑路径,因此无法有效判定错误的根源究竟是模型设计缺陷、特定数据输入触发的偶然错误,还是训练过程中的偏差。这种由可解释性缺失导致的不可审计性 (unauditability) 在技术层面上严重阻碍了问责路径的清晰界定^[19]。其核心困境在于:由于缺乏透明的决策依据,无法可靠地定位责任主体 (开发者、数据提供方、使用者) 或精确识别失误发生的具体环节 (数据预处理、模型训练、部署应用或临床交互),使得“是谁”以及“在哪个环节犯了错”这两个问责的关键问题变得难以解答。

2.3 数据偏倚与训练数据集代表性不足 算法性能高度依赖训练数据的质量与代表性。当前医疗数据的采集、标注与清洗过程缺乏标准化规范,易引入结构性偏倚 (structural bias)。若训练数据集集中于特定人群 (如特定种族、性别、地域或医疗机构),模型在泛化至其他群体或临床环境时其准确性可能显著下降,并加剧健康不平等。更隐蔽的风险在于代理指标偏倚,即当算法使用与真实目标 (如健康需求) 相关但受到社会不平等因素污染的替代指标 (如医疗费用) 影响时,可能出现系统性歧视弱势群体现象。一个典型案例是美国 UnitedHealth Group 开发的护理需求预测算法,因将医疗费用作为健康需求的代理指标,导致黑人患者需比白人患者病情严重近 2 倍才能获得同等优先级护理,直接放大了种族间的健康不平等^[20]。不具代表性的训练数据集是 AI 产生系统性错误和加剧健康不平等的主要原因之一。特别是在妇产科等涉及弱势群体的领域^[21],数据偏倚可能造成尤为严重的伦理后果。这种数据源头的“原罪”进一步加剧了问责困境。当一个基于偏倚数据集训练的 CDSS 对某位代表性不足的患者做出错误诊断时,责任归属变得异常复杂:这究竟是医生未能识别出模型的局限性,还是数据本身的偏倚导致了必然的错误,抑或是模型泛化能力不足的技术问题。

2.4 法律与监管制度落后 CDSS 应被视为受严格监管的医疗器械还是功能更灵活的软件,不同的

定性意味着截然不同的审批流程与责任划分。传统的产品责任法和医疗过失法在面对机器学习系统时均存在适用上的困难。尽管技术创新的步伐远超法律法规的更新速度,但全球立法者与监管机构积极应对以填补制度空白。例如,欧盟的《人工智能法案》^[4]通过创设基于风险的分层体系将大多数 CDSS 明确归为“高风险 AI 系统”,并为其设定了严格的合规义务,同时其配套的《人工智能责任指令》草案^[5]更试图通过因果关系推定规则来破解算法“黑箱”带来的举证难题。美国 FDA 则通过其“人工智能/机器学习医疗软件行动计划”将 CDSS 作为医疗器械软件进行监管,并提出了预定变更控制计划 (Predetermined Change Control Plan, PCCP) 等创新框架来管理能够持续学习的 AI 系统^[22]。2022 年我国 NMPA 发布《人工智能医疗器械注册审查指导原则》明确将诊断类 CDSS (如影像辅助诊断系统) 纳入 III 类医疗器械管理 (需进行临床试验审批),非诊断类 CDSS (如用药提醒) 按 II 类医疗器械或软件备案管理^[7]。2023 年 7 月国家卫生健康委员会发布的专门针对 CDSS 的顶层管理制度《医疗机构临床决策支持系统应用管理规范 (试行)》^[23]明确 CDSS 为辅助工具,输出结果仅供医务人员参考,最终决策权在于医生。随着生成式 AI 与自适应学习技术的快速渗透,新一代 CDSS 正逐步突破传统辅助工具的被动边界向具备临床推理能力的协同决策主体演进。这种智能化跃迁将挑战现行以“人类决策者为中心”的责任框架,亟须通过动态监管分类与弹性归责机制提前布局制度响应。

2.5 数据异质性引发的泛化失败 CDSS 在实际部署中常面临跨机构性能衰减问题,其主要原因在于训练数据与应用场景间的分布偏移 (distribution shift)。这种偏移可能源于不同医院的患者群体差异、数据采集标准不一或系统接口兼容性障碍等。例如,基于浅肤色人群训练的皮肤病检测模型应用于深肤色患者时灵敏度下降 18.3%~36.7%^[24],漏诊风险倍增。这意味着同一个算法即便在 A 医院表现良好,在 B 医院也可能因上述异质性而失效。而一旦发生医疗事故,复现当时的具体运行环境和决策场景极为困难,导致问责链条因技术环境的不可复现而断裂,使得将责任精准定位到某个具体环节成为几乎不可能完成的任务。

3 CDSS 问责困境应对路径探索的总体思路与具体实施策略

前文系统性剖析了 CDSS 在技术、法律、伦理及实践层面所面临的问责困境,其复杂性与交互性决定了单一维度的治理策略难以奏效。近年来, AI 治理理念日益转向全生命周期管理,如美国国家标准与技术研究院(National Institute of Standards and Technology, NIST)在其《AI 风险管理框架》中强调,可信赖 AI 系统的治理应贯穿其完整生命周期^[25]。

基于此,笔者认为,破解 CDSS 问责困境的关键在于推动问责范式向一种以系统性风险前置为核心的全生命周期治理模式转型。这一治理模式的实现,必须通过一个涵盖技术、法律与监管、组织管理及伦理与教育四大维度的协同框架来系统性地推进。这 4 个维度相互支撑、缺一不可,共同构筑了一个从“事前预防”到“事后救济”的闭环体系:技术路径是基础,旨在通过提升 AI 本身的可解释性、公平性与稳健性为信任和问责提供技术基石;法律与监管路径是保障,旨在通过构建适应性的法规和创新的责任分配机制为各方行为划定清晰的权责边界;组织管理路径是实践,旨在通过强化医疗机构的主体责任确保 AI 能够安全、有效地融入复杂的临床 workflow,实现风险的内部控制;伦理与教育路径则是灵魂,致力于通过重塑职业规范和提升“AI 素养”培养能够负责任地驾驭技术的文化与人才生态。

3.1 技术路径:迈向“可信赖 AI” 技术的内在属性是问责困境的起点,因此,提升技术本身的可信赖度是所有解决方案的基石。具体实施方案应包括以下递进层次。

3.1.1 在数据层建立标准化的“数据身份证” 算法公平性根植于数据源头,因此必须建立标准化的医疗数据治理框架。AI 伦理领域先驱 Gebru 提出的“数据集信息表”(datasheets for datasets)为此提供了理论基础与实践路径^[26]。该框架强制要求数据集创建者系统披露构成分布、采集规范、预处理方法、潜在偏倚及适用场景等核心元数据,本质上构建了一套可审计的数据“血缘”追溯体系。我国在具体操作上,可建议 NMPA 要求所有申请注册的 CDSS 产品提交“数据集信息表”作为技术评审的

必备文件。该表需详细披露数据来源、清洗规则、人群分布(性别、年龄、地域等)、已知偏见及其消解措施,方便医疗机构采购时能与本地患者特征进行量化对比,评估其泛化能力。

3.1.2 在模型层增强可解释性 为破解算法“黑箱”,必须推广可解释人工智能(explainable artificial intelligence, XAI)技术。通过模型可视化、特征归因等方法使 AI 从“黑箱”向“灰箱”(gray box)转变,尽可能使 AI 系统的运作方式和决策逻辑被使用者或受影响者理解^[18]。监管方应要求高风险 CDSS(如治疗方案推荐类)提供“伴随解释”功能。例如,当模型提供某一用药方案时,必须同时输出该决策所依据的关键参数(如检验检查指标、影像特征等)及其权重,方便医生进行批判性评估。

3.1.3 在应用层内嵌不可篡改的“决策日志” 强制要求 CDSS 内置“决策日志”功能,自动、不可篡改地记录每一次决策调用的关键信息。该日志需自动记录的信息至少包括时间戳、调用的患者数据摘要、系统输出的建议、医生最终采纳/修改/拒绝的决策及其简要理由(若可输入)。此日志可为事后进行精准溯源提供完整的技术证据链,也可为医疗纠纷责任溯源提供技术证据。

3.2 法律与监管路径:构建风险分级、责任适配的法规体系 法律的确信是社会信任的关键,必须构建与技术发展相适应的法律监管框架。为避免扼杀创新,我国 CDSS 的法律与监管路径必须摒弃“一刀切”模式,转向精细化的适应性治理。

3.2.1 在准入端实施风险分级的差异化监管 CDSS 的法律监管需在 NMPA 现有医疗器械分类的基础上进一步细化。例如,仅提供信息参考、不直接参与决策的 CDSS 作为较低风险类别,可简化审批流程;而对于能够自动生成治疗或用药指令的闭环系统,则必须经过严格的多中心、前瞻性临床试验验证,并对其算法的持续更新建立严格的变更控制流程。

3.2.2 在问责端构建“控制力-责任”相匹配的动态责任分配模型 开发者需对因算法设计缺陷、数据源偏见导致的系统性错误承担严格责任;医疗机构作为采购和部署方,对未能进行充分本地化验证、未对医护人员进行充分培训、未建立有效监控机制而导致的损害,承担过错推定责任;医生则需

对重大且明显的违规操作负责,例如在 CDSS 已明确提示高风险或与临床指南显著冲突时仍不加判断地采纳。

3.2.3 在救济端探索社会化补偿的创新机制 针对算法“黑箱”下难以精准归责的损害,建议由国家卫生健康委员会和银行保险监督管理委员会牵头,试点设立“医疗 AI 无过错赔偿基金”。其资金来源可设计为“开发者按销售额提取一定比例+高风险 AI 系统应用医疗机构缴纳部分保费+政府财政初期引导”的混合模式。当患者能够初步证明其损害与 CDSS 存在合理关联时,可由基金会先行赔付,再由基金会的专业委员会向查证有过错的一方追偿,从而降低患者的维权成本。

3.2.4 在创新端全面推行“监管沙盒”(regulatory sandbox) 允许自学习、自适应的新型 CDSS 在少数几家资质齐全的教学医院内,于“沙盒”环境中进行有限制的临床应用。在此期间,企业需向监管机构实时、透明地共享模型的性能数据和决策日志,以换取宝贵的真实世界数据来迭代算法,实现安全与创新的动态平衡。

3.3 组织管理路径:强化医疗机构的主体责任 设立内部 AI 治理委员会,统筹管理与实践规范。要将 AI 安全、有效地融入复杂的临床环境,医疗机构必须承担起首要的治理主体责任。为此,相关大型三甲医院与教学中心应成立一个由临床、信息技术、法律、伦理等多学科专家组成的“AI 治理与风险管理委员会”。该委员会的核心职责是构建一个从源头到应用的全流程风险闭环,包括通过制定标准化流程对 CDSS 进行严格的准入审查,通过定期审查“决策日志”和不良事件报告进行持续监督,以及通过制定明确的预案为“人机决策冲突”提供冲突裁定指引。考虑到不同机构的现实能力,该治理模式必须具备适应性。对于中小型医院和社区卫生服务中心,可不设置专门委员会,转为指定一名“AI 医疗安全官”作为核心责任人,并依托上级医联体或区域卫生信息平台获得专业技术评估和伦理审查的支持。

3.4 伦理与教育路径:重塑“人机协同”的职业新范式 技术效能的实现最终依赖于使用者实践,因此,重塑医疗从业者的认知体系与行为模式,是应对 CDSS 问责挑战的根本支柱。这需要从教育、伦理和文化 3 个层面系统性地构建一种新型的“人机

协同”职业范式。在人才培养层面,必须将 AI 素养前瞻性地纳入医学教育与职业再培训体系。其目的不仅在于技术操作的赋能,更在于主动防范因过度信赖 AI 而可能导致的“技能退化”风险,确保医生始终具备独立的核心诊疗能力^[27]。在职业规范层面,中华医学会等权威行业团体应牵头制定与 AI 时代相适应的伦理指南。该指南应超越“医生即最终责任人”的传统定式,将医生的伦理责任边界进行重新界定:医生的核心责任从“全知全能的决策者”转变为“整合多源信息(包括 AI 建议)对患者最终负责的专业判断者”。在文化建设层面,医疗行业内部应着力培育一种“批判性信任”的职业文化,以对抗“自动化偏见”等认知惰性。这种文化鼓励医生在充分利用 AI 优势的同时,始终保持专业的审慎、独立的判断力与不可或缺的人文关怀。这确保了技术始终作为赋能工具服务于医生的专业智慧,而非取代之。

4 结 语

CDSS 的问责困境并非技术发展中的暂时性注脚,而是对现有法律、伦理与社会治理体系的系统性拷问。本文深入分析了这一困境的形成机制,并提出了应对路径的探索方向。通过构建从“被动追责”转向“主动治理”的协同框架,将技术、法律、组织管理与伦理教育 4 个维度的策略融合成一个相互支撑、不可分割的有机整体。未来,AI 在医疗领域的应用必然会继续深化,我们有责任在技术进步的同时,追求更加公平、安全、可靠的社会治理体系。通过持续的政策创新、技术改进和伦理审查,最终实现 AI 与医学实践的和谐共生,构建一个更加智能化、信任化的医疗生态系统。CDSS 问责困境的有效破解不仅将为 AI 技术在医疗领域的深度应用扫清障碍,使患者获得更优质、更安全的服务,同时也将为身处“人机协同”新范式中的医务人员确立明确的责任边界与坚实的职业保障。

[参考文献]

- [1] 郑锐,于广军.临床辅助决策支持系统研究综述[J].中国数字医学,2023,18(11):70-77. DOI: 10.3969/j.issn.1673-7571.2023.11.014.
- [2] PRICE W N 2nd, GERKE S, COHEN I G. Potential liability for physicians using artificial intelligence[J]. JAMA, 2019, 322(18): 1765-1766. DOI: 10.1001/

- jama.2019.15064.
- [3] World Health Organization. Regulatory considerations on artificial intelligence for health[EB/OL]. (2023-10-19) [2025-07-01]. <https://www.who.int/publications/item/9789240078871>.
- [4] European Parliament and Council of the European Union. Artificial Intelligence Act[EB/OL]. (2024-07-12) [2025-07-01]. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>.
- [5] European Commission. Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence[EB/OL]. (2022-09-28) [2025-07-01]. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496>.
- [6] 国家新一代人工智能治理专业委员会. 新一代人工智能伦理规范[EB/OL]. (2021-09-25) [2025-07-01]. https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html.
- [7] 国家药品监督管理局医疗器械技术审评中心. 人工智能医疗器械注册审查指导原则[EB/OL]. (2022-03-09) [2025-07-01]. <https://www.cmde.org.cn/flfg/zdzy/zdzywbk/20220309091014461.html>.
- [8] 国家互联网信息办公室, 中华人民共和国国家发展和改革委员会, 中华人民共和国教育部, 等. 生成式人工智能服务管理暂行办法[EB/OL]. (2023-07-10) [2025-07-01]. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.
- [9] WONG A, OTLES E, DONNELLY J P, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients[J]. JAMA Intern Med, 2021, 181(8): 1065-1070. DOI: 10.1001/jamainternmed.2021.2626.
- [10] STRICKLAND E. How IBM Watson overpromised and underdelivered on AI health care[J]. IEEE Spectrum, 2019, 56(4): 24-31. DOI: 10.1109/MSPEC.2019.8701311.
- [11] MALIHA G, GERKE S, COHEN I G, et al. Artificial intelligence and liability in medicine: balancing safety and innovation[J]. Milbank Q, 2021, 99(3): 629-647. DOI: 10.1111/1468-0009.12504.
- [12] 周杰, 宋扬. 临床决策支持系统应用中的伦理思考[J]. 医学与哲学, 2022, 43(8): 16-19. DOI: 10.12014/j.issn.1002-0772.2022.08.04.
- [13] 国家互联网信息办公室, 工业和信息化部, 公安部, 等. 关于印发《人工智能生成合成内容标识办法》的通知[EB/OL]. (2025-03-07) [2025-07-01]. https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm.
- [14] COHEN I G, AMARASINGHAM R, SHAH A, et al. The legal and ethical concerns that arise from using complex predictive analytics in health care[J]. Health Affairs (Millwood), 2014, 33(7): 1139-1147. DOI: 10.1377/hlthaff.2014.0048.
- [15] TIZHOOSH H R, PANTANOWITZ L. Artificial intelligence and digital pathology: challenges and opportunities[J]. J Pathol Inform, 2018, 9: 38. DOI: 10.4103/jpi.jpi_53_18.
- [16] SANTONI DE SIO F, VAN DEN HOVEN J. Meaningful human control over autonomous systems: a philosophical account[J]. Front Robot AI, 2018, 5: 15. DOI: 10.3389/frobt.2018.00015.
- [17] SCHUITMAKER L, DROGT J, BENDERS M, et al. Physicians' required competencies in AI-assisted clinical settings: a systematic review[J]. Br Med Bull, 2025, 153(1): ldae025. DOI: 10.1093/bmb/ldae025.
- [18] ARRIETA A B, DÍAZ-RODRÍGUEZ N, DEL SER J, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI[J]. Inf Fusion, 2020, 58: 82-115. DOI: 10.1016/j.inffus.2019.12.012.
- [19] AMANN J, BLASIMME A, VAYENA E, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective[J]. BMC Med Inform Decis Mak, 2020, 20(1): 310. DOI: 10.1186/s12911-020-01332-6.
- [20] 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. DOI: 10.1126/science.aax2342.
- [21] RAJKOMAR A, HARDT M, HOWELL M D, et al. Ensuring fairness in machine learning to advance health equity[J]. Ann Intern Med, 2018, 169(12): 866-872. DOI: 10.7326/M18-1990.
- [22] U.S. Food and Drug Administration (FDA). Artificial intelligence/machine learning (AI/ML)-based Software as a Medical Device (SaMD) Action Plan[EB/OL]. (2021-01-28) [2025-07-01]. <https://www.fda.gov/media/145022/download?attachment>.
- [23] 国家卫生健康委办公厅. 国家卫生健康委办公厅关于印发医疗机构临床决策支持系统应用管理规范(试行)的通知[EB/OL]. (2023-07-17) [2025-07-01]. https://www.gov.cn/zhengce/zhengceku/202308/content_6895966.htm.
- [24] ADAMSON A S, SMITH A. Machine learning and health care disparities in dermatology[J]. JAMA Dermatol, 2018, 154(11): 1247-1248. DOI: 10.1001/jamadermatol.2018.2348.
- [25] National Institute of Standards and Technology (NIST). AI risk management framework 1.0[EB/OL]. (2023-01-26) [2025-07-01]. <https://www.nist.gov/itl/ai-risk-management-framework>.
- [26] GEBRU T, MORGENSTERN J, VECCHIONE B, et al. Datasheets for datasets[J]. Commun ACM, 2021, 64(12): 86-92. DOI: 10.1145/3458723.
- [27] TOPOL E J. High-performance medicine: the convergence of human and artificial intelligence[J]. Nat Med, 2019, 25(1): 44-56. DOI: 10.1038/s41591-018-0300-7.