

• 临床科研管理 •

人工智能辅助决策类医疗器械临床研究的伦理审查要点探究

李卫¹ 费科锋² 章培³ 陈少莹⁴ 毛威¹¹浙江医院伦理审查委员会,浙江杭州,310013;²浙江医院人工智能和信息化部,浙江杭州,310013;³宁波市医疗中心李惠利医院伦理委员会办公室,宁波 315000;⁴宁波大学附属第一医院伦理委员会办公室,宁波 315012

通信作者:毛威,Email:maoweilw@163.com,电话:0571-81595231

【摘要】 目的 构建人工智能辅助决策类医疗器械(Artificial Intelligence-Assisted Clinical Decision Support Devices, AI-ACDSD,相当于 AI-CDSS/SaMD)临床研究伦理审查规范,明确伦理审查资料、审查原则、审查要点,更好地指导该类研究的伦理审查工作,切实保护研究参与者的合法权益。**方法** 通过综合运用政策解读、文献回顾法和专家咨询法,深入分析 AI-ACDSD 临床研究存在的伦理风险及其原因,并从数据采集共享、算法风险防控、隐私保护、知情同意和研究方案设计等维度,提出 AI-ACDSD 临床研究伦理审查规范。**结果** AI-ACDSD 临床研究存在知情同意、隐私泄露及算法歧视、算法黑箱等伦理风险,伦理审查委员会应重点审查数据来源的合规性、多样性和充分性,数据质控的可靠性,算法风险防控的有效性、隐私保护的充分性和方案设计的合理性等。**结论** 为提高 AI-ACDSD 临床研究伦理审查质量,应尽快出台 AI-ACDSD 临床研究伦理审查指南,完善机构伦理审查委员会组成或成立区域伦理审查委员会,加强伦理委员系统化培训。

【关键词】 人工智能; 辅助决策类医疗器械; 临床研究; 伦理审查**基金项目:**浙江省科技厅软科学研究计划项目(2024C35037);浙江省卫生信息学会科研计划项目(2025XHAQ-Z01)**【中图分类号】** R19;R-05 **【文献标识码】** A DOI:10.3760/cma.j.cn113565-20260510-00160

Key elements for ethical review of clinical research regarding Artificial Intelligence-Assisted Clinical Decision Support Devices

Li Wei¹, Fei Kefeng², Zhang Pei³, Chen Shaoying⁴, Mao Wei¹¹Ethics Review Committee of Zhejiang Hospital, Hangzhou 310013, China;²Department of AI and Information System of Zhejiang Hospital, Hangzhou 310013, China;³Office of the Ethics Committee, Ningbo Medical Center Li Huili Hospital, Ningbo 315000, China;⁴Office of the Ethics Committee, The First Affiliated Hospital of Ningbo University, Ningbo 315012, China

Corresponding author: Mao Wei, Email: maoweilw@163.com, Tel: 0086-571-81595231

【Abstract】 Objective To establish ethical review standards for clinical research on Artificial Intelligence-Assisted Clinical Decision Support Devices (AI-ACDSD/AI-CDSS/SaMD), clarifying the required review materials, review principles, and key elements of ethical review, so as to better guide the ethical review of such research and effectively protect the legitimate rights and interests of research participants. **Methods** Through policy interpretation, literature review, and expert consultation, this study deeply analyzed the ethical risks and their causes in AI-ACDSD clinical research, and proposes ethical review standards for such research from the perspectives of data collection and sharing, algorithm risk control, privacy protection, informed consent, and research protocol design. **Results** AI-ACDSD clinical research involved ethical risks such as informed consent issues, privacy breaches, algorithm discrimination, and algorithm black boxes. Ethics committees should focused on reviewing the compliance, diversity, and adequacy of data sources, the reliability of data quality control, the effectiveness of algorithm risk prevention and control, the sufficiency of privacy protection, and the rationality of protocol design. **Conclusions** To improve the quality of ethical review for AI-ACDSD clinical research, guidelines for ethical review of AI-ACDSD clinical research should be issued as soon as possible, the composition of institutional ethics committees should be improved or regional ethics committees should be established, and systematic training for ethics committee members should be strengthened.

【Key words】 Artificial Intelligence(AI); Assisted Clinical Decision Support Devices; Clinical research; Ethical review**Fund program:** Soft Science Research Program of Zhejiang Provincial Department of Science and Technology(2024C35037); Zhejiang Health Information Association Research Plan Project(2025XHAQ-Z01)

DOI:10.3760/cma.j.cn113565-20260510-00160

人工智能辅助决策医疗器械(Artificial Intelligence-Assisted Clinical Decision Support Devices, AI-ACDSD)指的是利用人工智能技术处理医学影像、临床数据、基因组数据或生理信号等医学信息,并基于分析结果提供诊疗活动建议,辅助用户(医务人员、患者)进行医疗决策的医疗器械^[1]。这类产品的功能与基于人工智能的临床决策支持系统(AI-CDSS)或基于 AI/ML 的医疗器械软件(SaMD)相对应,但“AI-ACDSD”这一表述更侧重于强调其受中国医疗器械法规监管的医疗器械属性。据统计,该类产品占上市后 AI 医疗器械的 70% 及以上^[2-3]。AI-ACDSD 已广泛应用于临床的医学影像诊断、临床分诊预检及手术规划导航等领域。国家医保局已在放射、超声、康复等项目指南中设立“人工智能辅助”扩展项^[4-5]。有专家预测,成熟的 AI 辅助诊断服务未来有望通过该路径纳入国家医保乙类目录^[4-5]。

随着应用规模持续扩大、商业化闭环逐步打通, AI-ACDSD 临床研究(包括产品研发阶段和临床性能评价阶段)呈现快速增长趋势,自 2016 年起年复合增长率达 24.22%^[6]。然而,这类研究具有数据依赖性强、数据隐私风险高、算法黑箱及算法歧视等显著区别于传统医疗器械临床研究的特点,导致现行《涉及人的生命科学和医学研究伦理审查办法》难以提供有效指导,而《新一代人工智能伦理规范》也仅停留在原则层面,缺乏具体的审查细则。为进一步规范 AI-ACDSD 临床研究的伦理审查工作,切实保障研究参与者的合法权益,本文系统梳理国内外人工智能伦理治理相关法律法规、技术指导文件以及相关文献,聚焦 AI-ACDSD 研究全流程突出的伦理风险与关键问题,明确该类研究伦理送审资料、伦理审查原则,并从数据采集与共享、数据质量管控、算法风险防控、隐私保护与临床研究方案设计等维度,提出具有可操作性的伦理审查要点,以期将伦理规范与治理要求嵌入 AI-ACDSD 临床研究全生命周期。本文着眼于 AI-ACDSD 临床研究伦理审查规范,暂不探讨 AI-ACDSD 应用于临床存在的人类主体的自主性弱化、医疗责权划分、利益冲突和失业危机等问题。

1 研究方法

1.1 政策解读和文献回顾法

2025 年 8 月 5 日至 2026 年 5 月 6 日,持续以“人工智能辅助决策类医疗器械”“人工智能医疗器械”“人工智能”“AI”“大数据研究”分别与“伦理问

题”“伦理风险”“伦理或审查”“知情同意”“隐私保护”“数据安全”“算法黑箱”和“算法歧视”等作为关键词进行布尔逻辑组合检索相关文献、政策法规以及标准,覆盖中国知网(CNKI)、万方数据知识服务平台、维普中文期刊服务平台、百度学术、PubMed 等中外文数据库,以及中国政府网、国家法律法规数据库、国家药品监督管理局官网、国家卫生健康委官网等政策法规数据库。经阅读标题与摘要初筛、全文精读复筛,删除重复以及不相关文献后,最终纳入核心学术文献 17 篇、政策法规及标准文件 16 份。围绕“人工智能辅助决策类医疗器械临床研究伦理问题”“伦理审查资料”“伦理审查要点”和“伦理审查质量提升要点”,对纳入的文献进行综合分析。

1.2 专家咨询

1.2.1 专家遴选标准 (1)具有开展或审查人工智能辅助决策类医疗器械临床研究的经验。(2)高等医学院校生命伦理学专家,医疗机构内从事伦理审查工作专家,科技伦理监管机构人员或人工智能、大数据领域专家。(3)在相关领域具有 5 年及以上工作经验,具有丰富的理论认知与实践积累。(4)具有高级职称或博士学位,具备较高的学术水平与专业判断能力。(5)热心支持本次研究工作的专家。

1.2.2 专家人数与学科构成 根据专家遴选标准,结合同行推荐方式,最终邀请 11 位专家进行咨询,涵盖高等医学院校生命伦理学专家(3 名)、科技伦理监管机构人员(1 名)、医疗机构内从事伦理审查工作专家(6 名)、人工智能及大数据专家(1 名)。专家来自 11 家医疗机构、科研院所和监管机构,地域分布覆盖浙江、上海、北京和福建等地区。

1.2.3 咨询过程 课题组在广泛文献检索和前期研究基础上设计《人工智能辅助决策医疗器械临床研究伦理审查要素专家调查问卷》,问卷内容包括 4 个一级指标、22 个二级指标、59 个三级指标。请专家从重要性程度、熟悉程度、判断依据等对各级、各项指标分别评价。问卷通过电子邮件方式送达专家,专家在 5 个工作日内独立填写后返回。

专家对调查问卷中各指标重要性作出判断的依据,即判断系数 C_a ;专家对指标的熟悉程度,即熟悉程度系数 C_s 。权威程度为判断系数和熟悉程度系数的算术平均值,即 $C_r = (C_a + C_s) / 2$ 。一般来说专家权威程度越高,指标的预测精度就越高。 C_r 值大于 0.7,认为是可接受信度。本次咨询,专家积极系数为 100%,专家的权威系数集中在 0.73~1.0 之间。

1.2.4 意见整合方式 计算各条目的重要性赋值均数($\bar{x}$)、标准差(SD)、变异系数(CV),以重要性赋值均数 > 3.5 且 变异系数 < 0.25 作为条目保留的筛选标准。4 个一级指标、22 个二级指标、59 个三级指标,仅有伦理审查原则中“免费和补偿、赔偿”的均值为 4,变异系数为 0.295,予以删除。除此以外,其他各项指标均保留。

2 AI-ACDSD 临床研究存在的伦理问题

AI-ACDSD 临床研究的核心路径在于利用医疗大数据与算法模型,开发和验证辅助临床诊疗决策的智能系统,其风险主要来源于数据与算法。医疗大数据的采集、传输、存储、处理、共享等,涉及数据来源合规性、知情同意、数据安全与隐私保护^[7]。数据质量、数据代表性、算法性能,又将引发算法偏见、算法黑箱等伦理问题^[7-8]。

2.1 知情同意问题

AI-ACDSD 临床研究需要利用大量影像、病理、生理信号等数据,这些医疗大数据属于个人敏感信息^[9]。

根据我国相关法规规定,除了“利用可识别身份信息的人体材料或者数据进行研究,已无法找到受试者,且研究项目不涉及个人隐私和商业利益”以及“利用匿名化数据”可免除知情同意外,其他情况均应获得研究参与者或其监护人的知情同意^[9-11]。但是,传统“一对一、单项目”的特定同意模式,已难以适应“海量、多元、回溯性分析”的研究新范式。一是传统知情同意工作量巨大,实践中不可行。AI-ACDSD 依赖于万例级数据,若要求对已有医疗数据逐一追溯获取知情同意,将面临操作不可行、研究效率迟滞等现实困境。比如一项“基于人工智能影像分析的颅内动脉狭窄精准诊治决策平台研发与应用”研究,需利用 1 万例患者的既往临床诊疗数据,若采取传统知情同意模式,研究者需逐一联系这 1 万例患者,详细告知研究目的、内容、风险、受益、拒绝参加研究权力等,获其知情同意后方可运用其健康数据。按每例患者知情同意耗时 10 分钟估算,完成 1 万例患者的知情同意共需约 1 666 小时,约 208 个工作日(按每日 8 小时计算)。二是“同意偏差”影响样本代表性与研究结果的科学性。AI-ACDSD 临床研究依赖于大量的回顾性数据,涉及万例级的患者,时间跨度长达数年,部分患者可能失联、离世、拒绝,这就引起“同意偏差”,导致样本缺乏代表性,进而影响研究结果的科学性与公正性。

2.2 隐私泄漏问题

AI-ACDSD 临床研究涉及医疗大数据的采集、传输、存储、处理、共享等多个环节,数据在医疗机构、申办方、数据分析方等多个主体间流转,任何一个环节的疏漏都可能导致严重的隐私泄漏问题。医疗大数据涵盖诊疗记录、影像、基因、生化指标乃至生活方式与环境信息,属于个人敏感性信息,一旦泄露,被恶意利用,不仅对个人造成严重的心理和社会影响,还会侵害群体隐私利益和国家安全。例如,精神疾病患者的信息泄露可能对其社会关系和职业生涯造成毁灭性打击;遗传性疾病信息泄露可能影响家族成员的婚姻和生育选择;基因数据泄露,可能造成恐怖分子和敌对国利用制造种族生物武器,实施定向打击。2015 年 9 月,英国伦敦皇家自由医院信托与谷歌旗下 DeepMind 公司签署数据共享协议,DeepMind 于同年 11 月开始处理 160 万患者的敏感医疗数据用于开发急性肾损伤检测系统。2017 年 7 月 3 日,英国信息专员办公室经过长达 13 个月的调查后,正式裁定英国伦敦皇家自由医院未做好隐私保护,被认定违反了数据保护法^[12]。

2.3 算法风险

2.3.1 算法偏见/歧视 算法偏见,又称算法歧视,是指 AI 系统在处理数据和做决策时,系统性地产生对特定人群(如种族、性别、年龄、地域等)不公平、不公正的结果。AI-ACDSD 临床研究过程中,算法偏见主要源于训练数据、算法设计以及研发者主观偏见^[7-8]。

首先是训练数据的多样性和充分性。AI-ACDSD 基于大数据训练,其算法通过统计学习提取模式,当训练数据存在样本偏差或多样性缺失时,将导致系统在作出临床决策时出现偏见性方案。例如,某全球知名的皮肤科 AI 公司推出的皮肤癌筛查系统在欧美市场表现优异,但诊断深色皮肤(如非洲裔美国人)的准确率显著低于白人患者。根本原因是训练数据集中的深色皮肤样本严重不足,导致模型未能学习到深色皮肤的特征表达^[13]。

其次是算法设计与特征选择。除了数据本身,算法在特征提取和模型优化阶段也可能引入不公平。例如,某医疗 AI 公司开发用于预测患者再入院风险的模型,曾错误地使用了“医疗费用”作为健康需求的代理指标。由于少数族裔群体(如黑人患者)的经济原因,其花费的医疗费用往往低于同等健康状况的白人患者,模型因此误判黑人患者健康状

况较好^[14]。

最后是研究者的主观情感偏向、人类社会所固有的偏见等。例如,罪犯管理分析系统广泛用于美国司法系统,帮助法官判断被告人在释放后是否有可能再次犯罪。调研发现该系统存在严重的种族偏见问题,黑人被告被错误标记为“高风险”的概率是白人被告的近两倍,即使他们实际上并未再犯^[8]。

2.3.2 算法黑箱/不可解释性 算法黑箱,又称算法的不透明性与不可解释性,即输入(如患者的影像或体征)和输出(如诊断结论)是已知的,但中间的逻辑路径却无法被人类清晰解释或追溯。例如,一个 AI 模型可能准确地识别出黑色素瘤,但无法解释它是通过“皮肤纹理”还是“色素分布”做出的判断。算法黑箱在 AI-ACDSD 的实践中,将导致质量不可靠、结果不可信、责权不清的困境^[15]。

算法黑箱的形成主要源于技术自身的复杂性与 AI 自主学习机制的叠加影响^[16]。首先,算法与大数据技术的高度复杂性构成了黑箱的技术基础。现代 AI-ACDSD 通常需要在数十亿乃至万亿级别的数据维度上进行运算,其代码规模庞大、结构复杂,对于用户甚至开发者而言,已无法通过逐行阅读代码来完整理解系统的运作逻辑。其次,算法的自主学习与动态演化特性进一步加深不可预知性。这类系统并非机械执行固定指令,而是在训练过程中通过对海量数据的分析不断调整内部参数和决策逻辑。当代码与大规模数据长时间相互作用,模型的内部表征会随学习过程发生难以预测的变化。最终,这种输入与输出之间自动演化的中间过程,使得算法的行为难以被精确预判。即便我们清楚输入的数据和最终输出的结果,中间经过的复杂运算与自调整机制也往往成为一个难以解析的“黑箱”,从而导致算法缺乏可解释性与透明度。

3 AI-ACDSD 临床研究伦理审查规范

3.1 伦理审查委员会的组成与要求

为保障该类临床研究的伦理审查质量,机构伦理审查委员会应增补人工智能技术、应用、伦理和法律等相应专业背景的专家,或聘请相应独立顾问提供专业意见。对于该类临床研究较多的机构,可考虑设立专门的人工智能临床研究伦理审查委员会^[17-18]。

3.2 伦理审查材料

除《涉及人的生命科学和医学研究伦理审查办法》和《医疗器械临床试验质量管理规范》等法规要

求的常规临床研究伦理审查资料以外,伦理审查委员会还应重点审查如下材料:

3.2.1 数据来源合规性证明文件 自行收集数据,需明确数据采集的机构、时间段、人群、数据类型和采集人等;非自行收集来源数据应注明数据来源,以及数据所在机构伦理批件,签署的数据共享协议;公开数据应提供合法途径获取的说明文件,包括公开数据库的名称、创建者、数据总量和使用情况等^[1,19]。

3.2.2 数据质控资料,包括数据构成与分布说明,数据整理、数据采集、数据标注的操作规范及其质量评估说明^[1]。

3.2.3 算法风险管理资料 明确算法名称、类型、选用理由、基本原则和机制机理,以及过拟合与欠拟合、假阴性与假阳性、数据污染与数据偏倚(如数据扩增)等风险的控制措施。若是公认成熟算法,应提供相关文献资料;若是全新算法,应提供安全性和有效性验证资料^[1]。

3.2.4 数据安全与隐私保护方案 明确数据采集、处理、存储、传输和销毁等全生命周期的数据安全和隐私保护的具体措施^[1]。

3.2.5 科技伦理风险评估情况、防控及应急处理预案,包括人工智能技术预期应用带来的潜在科技伦理风险评估,对科技伦理风险的监测预警措施,对可能产生科技伦理风险的防控计划等^[18]。

3.3 伦理审查原则

3.3.1 增进人类福祉 坚持以增进人类健康福祉为根本宗旨,服务于人与社会及生态的可持续、和谐发展,致力于促进人机协同、提升医疗可及性与健康公平。始终以研究参与者权益与公共利益为核心,研究具有科学价值、社会价值^[18,20-21]。

3.3.2 促进公平公正 坚持普惠性和包容性,公平合理地选择研究参与者,并采取有效措施在数据获取、算法设计、技术开发、产品研发过程中消除偏见和歧视,避免因样本偏差导致模型对特定群体产生系统性歧视。同时保障资源分配、机会获取、决策过程的客观性和包容性,避免部分人群承受着隐私泄露等风险,却无法享受到同等质量的精准医疗服务^[18,20,22]。

3.3.3 确保可控可信 研究的科学和社会利益不得超越对研究参与者人身安全与健康权益的考虑,研究风险受益比应当合理,建立持续监测机制与突发状况应急处置预案,使研究参与者可能受到的风

险最小化。临床决策中,医生始终处于主导地位,最终决策必须由有资质的临床医生确定。不断提升 AI 系统的透明性、可解释性、可靠性、可控性,逐步实现可审核、可监督、可追溯、可信赖^[18,20,22]。

3.3.4 知情同意 个人信息的收集、存储、加工、使用等处理活动应遵照个人信息保护法规定,获取研究参与者或其监护人的知情同意,并建立合理的数据撤销和退出研究机制^[10,23]。

3.3.5 保护隐私安全 采用匿名化、去标识等手段切实保护研究参与者的隐私权,建立覆盖数据采集、存储、处理、共享、访问、传输、使用、销毁等全生命周期的隐私保护措施,确保符合国家数据安全和个人信息保护等有关规定。未经研究参与者授权,不得与其他数据进行匹配,不得向第三方透露^[18,20,22]。

3.4 伦理审查要素材料与要素要点

伦理审查委员会对 AI-ACDSD 临床研究进行审查时,除了遵守《涉及人的生命科学和医学研究伦理审查办法》《科技伦理审查办法》《医疗器械临床试验质量管理规范》的相关规定外,还应重点审查以下内容:

3.4.1 机构资质 医疗机构应具有与所开展的 AI 辅助决策技术相适应的功能、资质、技术能力、设施设备,以及相应风险处置能力。具体按研究风险等级分层要求:低风险研究(辅助分诊、辅助检测和辅助诊断):机构应具开展相关诊疗工作的基本资质、数据采集与处理所需的软硬件设施(如影像设备、医学影像图像管理系统及其计算机硬件平台)以及相应的数据安全能力。高风险研究(辅助治疗、手术导航),除满足上述条件外,临床研究承担科室还应具有开展临床诊疗工作 10 年以上,床位不少于 50 张,年手术量不少于 1 000 例,其技术水平在本省医疗机构中处于领先地位^[24]。

3.4.2 研究者资质 一是研究团队由临床医学、人工智能、医疗器械等多学科背景人员组成^[25]。二是有与开展 AI-ACDSD 临床研究相适应的专业技术人员。其中,涉及辅助诊断技术的,应具有开展相关诊断工作 5 年以上经验的医师或技师;涉及辅助治疗技术的,应具有 10 年以上三级甲等医院相关专业临床诊疗工作经验、具有副主任医师专业技术职务任职资格、熟练掌握本专业开放手术或微创手术技术的医师。三是研究者须经研究方案规定的标准化培训并考核合格后方可参与研究;涉及高风险操作的,操作者还须具备独立诊疗资质及常见并发症处

理能力^[24]。

3.4.3 数据来源合规性审查 若研究数据来源于机构,应获得数据个体的知情同意、数据所在机构的伦理审查委员会的同意,并与机构签署数据共享协议。若数据来源于公开数据,获取途径应合法合规。

3.4.4 数据来源的多样性与充分性 AI-ACDSD 基于大数据训练,当训练数据存在样本偏差或多样性缺失时,将导致系统做出临床决策时出现偏见。因此,应保障训练数据/验证数据集的多样性、充足性以及代表性,以提高算法泛化能力,避免数据偏倚对模型性能和安全性带来不当的影响以及歧视的风险^[1]。

数据的多样性和代表性。一是应结合目标疾病人群分布(如性别、年龄、职业、地域、生活方式)、流行病学特征(如分级、分型、分期)、统计指标(如患病率、发病率、死亡率、治愈率、生存率)以及目标疾病并发症与类似疾病的影响情况等选择数据,确保数据来源的多样性、样本符合临床实际分布和诊疗规范要求^[1,26]。必要时在临床研究中富集相关具有代表性的亚组^[27]。二是连续收集数据,防止“挑选数据”导致的入组偏倚。基于明确且严格的入排标准和临床试验计划,连续收集特定医疗机构中过往某段时间内的患者数据,避免主观挑选病例^[27]。三是避免同一患者同一目标部位的多组样本。为防止“数据伪独立”导致的假重复,在临床研究中,需避免入组同一患者同一目标部位的多组样本数据^[27]。

数据的充分性。研究者应基于统计学原理、试验设计类型及预期的临床效应进行精确估算样本量,应当明确样本规模及确定依据,需要考虑样本量不足对软件的影响及其风险^[1]。

3.4.5 数据集的分布 一是训练集原则上需保证样本分布的均衡性,测试集、调优集原则上需保证样本分布符合真实情况^[1]。而且,为公正评价系统性能,避免过拟合,训练集、调优集、测试集的样本应两两无交集并通过查重予以验证^[1,28]。如果测试集中的样本出现在了训练集中,模型在训练时就已经“见过”这些样本,将导致准确率虚高,但面对全新数据时表现又会大幅下降。就如同考试试题曾出现在模拟试卷中,从而引发“作弊式”分数虚高。三是测评数据库应满足权威性、科学性、多样性、规范性、动态性、封闭性。值得注意的是,公开数据库因其不具备封闭性而不能作为测评数据库^[1]。为提高系统泛化能力,最好在不同地域、不同层级的 3 家机构进行多

中心验证^[1,26,28]。

3.4.6 数据质量控制 AI-ACDSD 的研发和性能评价都高度依赖数据,数据质量是决定系统模型性能、安全性与有效性的核心基石。高质量的数据能够训练出精准、可靠、泛化能力强的模型,从而在临床实践中辅助医生进行精确诊断、制订个性化治疗方案;反之,低质量、有偏见或不完整的数据则会直接导致模型性能下降,产生错误的医疗决策,对患者安全构成潜在威胁。因此,伦理审查委员会应对数据质控的充分性、有效性和准确性进行审核,具体审查要点如下:

一是数据质控是否覆盖了数据全生命周期,如设置数据采集、整理、标注等关键环节的操作规程和质控标准^[1]。二是明确人员资质与职责:明确采集人员、标注人员、审核人员和仲裁人员的资质、选拔依据、培训内容、考核内容等。人员配置时,应遵循仲裁人员的资质要求>审核人员>初级标注人员的原则,并且优先选择数据相应的科室专家或熟悉该领域的一线临床专家,以及具有数据采集、标注和仲裁相关经验的人员^[1,29]。三是数据标记:数据标记规则应基于临床指南、专家共识或权威文献,而且每一例数据样本应由至少两名标注人员在互不知晓对方结果的前提下独立完成标注。可使用自动标注软件进行数据,但自动标注结果不得直接使用,应由标注人员审核后方可使用^[1,29]。四是数据标记的一致性和准确性:当初级标注人员的结论一致时,应对标注结论进行抽样审核;当初级标注人员结论不一致时,宜提交仲裁;采用埋题验证的方式评价标记的重复性(即同一标注员前后多次对同一数据进行标注的一致性程度),如每完成 20 张糖网图像的分类标注后,随机抽选其中一张重新标注。对比标注人员与仲裁结论,评价标记结果的准确性,如每完成 20 张糖网图像的标注,随机抽选一张由仲裁人员仲裁和对比,以统计准确性^[28-29]。

3.4.7 算法风险管理 一是审查算法选择的合理性及其风险可控性。算法包括公认成熟算法和全新算法,其中公认成熟算法是指源自公开文献资料、原理简单明确、上市多年且无不良事件的算法,而全新算法是指源自临床研究、科学研究的新算法^[30]。对于公认成熟算法,伦理审查委员会应审查研究者提交的公开文献资料,确保算法成熟;对于全新算法,伦理审查委员会应重点审查算法安全性和有效性验证资料,确保风险可控,内容包括但不限于压力测试、对抗

测试、安全阈值设置说明、使用多个独立的测试集验证其泛化能力等^[1,30]。二是审查算法风险管控措施。伦理审查委员会应对研究者提交的过拟合与欠拟合、假阴性与假阳性、数据污染与数据偏倚(如数据扩增)等风险的控制措施进行审查,主要审查风控措施的科学性、可操作性、与风险等级的匹配度。

3.4.8 信息安全与隐私保护 伦理审查委员会应重点审查隐私保护措施是否覆盖数据收集、访问、处理、传输、销毁等全数据生命周期,是否建立数据审计日志、数据安全监测机制、制订紧急预案等。首先,应优先采用后台挖掘方式,直接输出结论性结果,从而构建“数据不出库、成果可输出”的安全研究模式^[19]。其次,对于不具有后台挖掘条件的,需落实数据全生命周期各阶段的隐私保护措施,确保数据安全可控,具体如下:数据采集应遵循“最小必要原则”,在确保系统研发与验证所需的情况下,尽可能采集最少数据,避免过度采集^[25]。数据处理,应采用高级的脱敏技术、隐私增强技术或匿名化技术,防止隐私泄露。根据工作需求严格设置访问权限,访问数据时进行身份验证,并定期更换用户登录密码,防止数据的未授权的访问、另存、复制、下载和打印等。数据传输时,应保证数据以安全的方式传输给指定的对象,如使用加密技术、身份验证技术、数据完整性校验技术等。临床研究结束后,应及时销毁数据,不得用于其他研究,未经授权,不得与其他数据进行匹配,并且执行数据标注、计算和存储的设备在停用、退役和退出标注任务前应将其中所有数据彻底删除,无法恢复。再者,应建立数据审计日志,确保数据的访问、标注、修改等可追溯。建立数据安全监测机制,实时检测异常访问或数据泄漏风险。最后,应确保网络安全,如采用防火墙、边界防护、入侵防护等安全措施。定期更新安全系统,对技术人员进行网络安全和隐私保护培训,提高整体防御能力,并制订紧急预案,包括数据泄露应急处理流程、数据备份与恢复方案等^[29,31]。

3.4.9 临床研究方案 临床研究医疗器械的预期用途、使用场景和核心功能进行方案设计,伦理审查重点内容如下:首先,为了避免假阴性和假阳性,研究方案设计应明确最终诊疗决策由医生做出,AI 系统仅提供辅助建议。其次,建议优先选择同品种产品或临床参考标准(即临床金标准)进行非劣效对照设计,次之可选择用户结合软件联合决策与用户单独决策进行优效对照设计;非劣效或优效界值的确定应有充

分临床依据。此外考虑到用户的差异性,可选择多阅片者多病例试验设计。再者,应以特异性、敏感性、接受者操作特征曲线(Receiver Operating Characteristic, ROC)或曲线下面积(Area Under Curve, AUC)为主要指标,亦可选择时间效率等指标作为评价指标^[1,32]。最后,为了保障研究参与者的安全与权益,研究方案中设置安全阈值,当系统模型“超出信任范围”或“异常”时,系统会自动触发人工复核。

3.4.10 知情同意 首先,应获得研究参与者知情同意。根据个人信息保护法规定,除使用匿名化数据可以免除知情同意外,前瞻性收集研究参与者数据以及使用既往临床诊疗数据,均应获得研究参与者的知情同意。前瞻性收集数据,因已明确研究目的、有与研究参与者接触的机会等,应签署该研究特定的知情同意书。对于利用既往临床诊疗已有的医疗大数据,建议参考“生物样本捐献者已经签署了知情同意书,同意所捐献样本及相关信息可用于所有医学研究”,采取“泛知情同意”的形式^[10]。其次,对于泛知情同意,伦理审查委员会还应审查研究目的、数据存储期限等是否覆盖当前临床研究。对于特定知情同意,应充分告知隐私泄露的风险以及 AI-ACDSD 特有的风险,包括但不限于漏诊、误诊、算法偏见、算法不可解释性、系统不稳定、系统故障等。

4 提升伦理审查质量的策略与建议

4.1 出台 AI-ACDSD 临床研究伦理审查指南

由于 AI-ACDSD 临床研究具有数据依赖、隐私泄漏、算法偏见、算法黑箱特点,《涉及人的生命科学和医学研究伦理审查办法》《涉及人的临床研究伦理审查委员会建设指南》已经不能适用 AI-ACDSD 临床研究伦理审查,而《新一代人工智能伦理规范》仍停留在原则层面,缺乏具体的审查细则。建议针对 AI-ACDSD 临床研究的特点出台相应伦理审查指南,明确伦理审查委员会应增补的委员或独立顾问、伦理审查原则、新增的伦理审查材料,以及研究者资质、数据合规性、算法公正性、隐私保护等审查要素要点。如将数据来源合规性声明、算法风险管理资料、数据安全与隐私保护方案等列入伦理审查清单,并明确样本来源的多样性与充分性、数据质量控制、算法风险管理、隐私保护等的审查要点。

4.2 成立 AI 辅助决策类医疗器械临床研究区域伦理审查委员会

机构伦理委员会成员多为临床医学、药学背景的兼职人员,普遍缺乏大数据分析、算法可解释性评

估等专业技术能力,面对 AI-ACDSD 临床研究,难以对算法偏见风险、数据代表性、数据质量等作出实质判断。为保障 AI-ACDSD 临床研究伦理审查质量,建议集区域内 AI、大数据、信息安全等优势力量,成立 AI 辅助决策类医疗器械临床研究区域伦理审查委员会,接受其他医疗机构的委托审查。

4.3 构建 AI-ACDSD 临床研究伦理教育与培训体系

加强教育与培训是提升 AI-ACDSD 伦理治理能力的基础性工程。首先,要加强对伦理委员的系统化培训,重点提升其对 AI 算法偏见、数据隐私、可解释性等风险的识别与评估能力,确保伦理审查的专业性与规范性。其次,高等院校应面向医学、工学、法学及交叉学科学生开设 AI-ACDSD 伦理相关课程,从源头培养具备伦理意识与合规素养的复合型人才。最后,相关学会与协会应充分发挥行业引领作用,通过举办专题论坛、发布伦理指南、开展线上线下宣传等方式,持续提升从业者对 AI 医疗器械伦理风险的认知水平与规范意识,加强行业自律。通过“院校教育奠基、行业宣教普及、委员培训提质”三位一体的联动机制,为 AI 医疗健康应用的伦理合规与可持续发展提供坚实的人才与制度保障。

5 展望

希望国家科技部门、卫生健康行政部门、网信部门加强协同,尽快完善人工智能伦理审查制度和体系建设,明确伦理审查资料、伦理审查原则、伦理审查要素要点等,更好地指导 AI-ACDSD 临床研究伦理审查工作。另外,希望参考美国的做法,引入数据安全与隐私保护专项评估,在项目申报时,要求申请人提交《数据安全与隐私保护方案》,详细说明数据来源与性质、数据全生命周期管理措施、风险防控与应急预案。由专家进行评估,未通过评估的项目不得立项^[33]。借助行政部门力量,以科技项目立项为契机,从源头上保护好研究参与者的数据安全和隐私。

利益冲突 所有作者均声明不存在利益冲突

作者贡献声明 李卫:研究设计、论文撰写;费科锋:研究指导,论文修改;章培和陈少莹查阅资料;毛威:选题设计,研究指导

参 考 文 献

[1] 国家药品监督管理局 国家医疗器械审评中心. 人工智能医疗

- 器械注册审查指导原则[EB/OL]. (2022-03-07)[2025-12-18]. <https://www.cmde.org.cn/xwdt/shpgzg/gztg/20220309090800158.html>.
- [2] Sivakumar R, Lue B, Kundu S. FDA approval of artificial intelligence and machine learning devices in radiology: a systematic review[J]. JAMA Netw Open, 2025, 8(11): e2542338. DOI:10.1001/jamanetworkopen.2025.42338.
- [3] 滚动资讯. 中国发布《医疗健康产业人工智能发展研究报告(2026)》, AI 正重塑医疗全链条, 产业进入规模化发展新阶段[EB/OL]. (2026-04-11)[2026-05-04]. <https://news.10jqka.com.cn/20260411/c675918403.shtml>.
- [4] 国家医疗保障局. 放射检查类立项指南出台 CT、磁共振、X 光收费将执行新规! [EB/OL]. (2024-11-20)[2026-05-04]. https://www.nhsa.gov.cn/art/2024/11/20/art_52_14773.html.
- [5] 国家医保保障局. 国家医保局印发《康复类医疗服务价格项目立项指南(试行)》[EB/OL]. (2024-11-14)[2026-05-04]. https://www.nhsa.gov.cn/art/2024/11/14/art_14_14689.html.
- [6] 卢岩, 陈娟, 张婷, 等. 全球人工智能医疗器械临床试验注册现状分析[J]. 生物医学工程杂志, 2025, 42(3): 512-519.
- [7] 国家人工智能标准化总体组. 人工智能伦理风险分析报告[EB/OL]. (2019-04-26)[2026-05-06]. <https://www.cesi.cn/201904/5036.html>.
- [8] 张卫涛. 算法黑箱的成因、危机及治理路径选择[C]//2025 数字时代的社会结构变迁与治理创新学术交流会议论文集(下). 北京: 中国知网, 2025. DOI: 10.26914/c.cnkihy.2025.020534.
- [9] 全国人民代表大会常务委员会. 中华人民共和国个人信息保护法: 中华人民共和国主席令 第九十一号[Z]. 2021.
- [10] 国家卫生和计划生育委员会. 涉及人的生物医学研究伦理审查办法: 国家卫生和计划生育委员会令 第 11 号[Z]. 2016.
- [11] 国家卫生健康委 教育部 科技部 国家中医药局. 涉及人的生命科学和医学研究伦理审查办法: 国卫科教发[2023]4 号[Z]. 2023.
- [12] 罗芳. 医疗人工智能伦理问题及对策研究[D]. 武汉: 华中科技大学, 2020.
- [13] Guo LN, Lee MS. Bias in, bias out: Underreporting and underrepresentation of diverse skin types in machine learning research for skin cancer detection-A scoping review[J]. J Am Acad Dermatol, 2022, 87(1): 157-159. DOI:10.1016/j.jaad.2021.06.884.
- [14] Obermeyer Z, Powers B, Vogeli C. 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.
- [15] Yao J, Liu Y, Li B, et al. Visualization of Deep Models on Nursing Notes and Physiological Data for Predicting Health Outcomes Through Temporal Sliding Windows[M]. //Shaban-Nejad A, Michalowski M, Buckridge D L, eds. Explainable AI in Healthcare and Medicine: Building a Culture of Transparency and Accountability. Cham: Springer, 2020:187-207.
- [16] 陈子瑜, 程国斌. 医疗人工智能中的算法黑箱及其核心伦理问题[J]医学与哲学, 2024, 45(12): 6-10.
- [17] 耿雯倩, 朱万, 马磊. 人工智能医疗器械临床研究伦理审查要素研究[J]中国医学伦理学, 2022, 35(12): 1352-1356.
- [18] 工业和信息化部, 国家发展改革委, 教育部, 等. 人工智能科技伦理审查与服务办法(试行): 工信部联科[2026]75 号[S]. 2026-03-20.
- [19] 数智驱动科技活动伦理审查指南专家组. 数智赋能医学科技的治理管理(三)数智驱动科技活动的伦理审查指南[J]. 中华医学科研管理杂志, 2025, 38(3): 161-168.
- [20] 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[Z]. 2021-09-25.
- [21] 中共中央办公厅, 国务院办公厅. 关于加强科技伦理治理的意见: 中办发[2022]19 号[Z]. 2022-03-20.
- [22] 国家新一代人工智能治理专业委员会. 新一代人工智能治理原则——发展负责任的人工智能[Z]. 2019-06-17.
- [23] 李佳霖. 《人工智能北京共识》发布[N]. 经济日报, 2019-05-30(015).
- [24] 国家卫生计生委办公厅. 人工智能辅助诊断技术管理规范(2017 年版): 国卫办医发[2017]7 号[Z]. 2017-02-14.
- [25] 王国豫, 主编. 人工智能医学影像伦理手册[M]. 上海: 上海三联书店, 2023:102.
- [26] 国家药品监督管理局医疗器械技术审评中心. 深度学习辅助决策医疗器械软件审评要点: 2019 年第 7 号[Z]. 2019-07-03.
- [27] 国家药品监督管理局医疗器械技术审评中心. 人工智能辅助检测医疗器械(软件)临床评价注册审查指导原则: 2023 年第 38 号[Z]. 2023-11-07.
- [28] 国家药品监督管理局医疗器械技术审评中心. 病理图像人工智能分析软件性能评价审评要点: 2023 年第 23 号[Z]. 2023-07-10.
- [29] 国家药品监督管理局. 人工智能医疗器械 质量要求和评价 第 3 部分: 数据标注通用要求: YY/T 1833.3-2022[S]. 2022-08-17. 实施日期: 2023-09-01.
- [30] 国家药品监督管理局 医疗器械技术审评中心. 医疗器械软件注册审查指导原则(2022 年修订版): 2022 年第 9 号[Z]. 2022-03-09.
- [31] 陆婷, 张燕. 人工智能科研项目中的隐私问题研究[J]. 中国现代医生, 2022, 60(36): 109-111.
- [32] 周吉银, 李红英, 杨阳. 人工智能医疗器械的伦理审查要点[J]. 医学与哲学, 2020, 41(6): 35-39.
- [33] National Institutes of Health. Data Management and Sharing Policy[S]. 2023.

(收稿日期: 2026-05-10)