医学人文

诊断辅助人工智能应用情境下医疗损害责任认定

张晓梅
浙江农林大学文法学院，浙江 杭州 311300

摘 要：诊断辅助人工智能实现了从“规则驱动”到“数据驱动”的技术跃迁，其固有的自主决策性、黑箱性与动态进化性，对以“当时医疗水平”标准为核心的传统医疗损害责任认定构成挑战。对此，应将“诊断辅助人工智能的使用”定位为“新疗法”，以确认其应用的合法性；进而明确该人工智能的输出结果并不直接等同于“当时医疗水平”，且这一标准的内涵亦将动态演进。在此基础上，构建以“审慎再判断义务”为核心的动态过错认定框架，将评价重心转向对医务人员履行审慎使用与结果核验义务过程的审查。这一调整旨在回应当前“人机协作”模式下医疗过错的认定需求，实现技术创新与患者安全保护的规范平衡。

关键词：诊断辅助人工智能；当时医疗水平；医疗过错；审慎再判断义务
中图分类号：D922.16 **文献标志码**：A **文章编号**：1671-0479（2026）03-280-008
doi：10.7655/NYDXBSSS260023

一、问题的提出

以数据驱动为核心的诊断辅助人工智能系统（以下简称诊断辅助人工智能[①]），正在重塑医疗临床实践。此类系统固有的自主决策性、算法黑箱性与动态进化性，在提升诊疗效率与精度的同时，也对以《中华人民共和国民法典》第1221条“当时医疗水平”标准为核心的医疗损害责任认定规则构成挑战[②]。《中华人民共和国民法典》第1218条确立了医疗损害责任的过错责任原则，第1221条进一步将“过错”的判断客观化为“医务人员是否履行了‘与当时医疗水平’相应的注意义务”。对此，传统医疗过错的判断，依赖一个相对稳定、透明且经由专业共识形成的“医疗水平”知识体系；而诊断辅助人工智能的运行逻辑，却是一个持续自我迭代、过程不透明且可瞬时吸纳前沿知识的“黑箱系统”。由此，当客观化的“当时医疗水平”标准适用于诊断辅助人工智能应用情境时，产生了如下困境：①如何确

基金项目：中共浙江省委政法委员会浙江省法学会法学研究课题重点项目“医疗人工智能法律规制与侵权责任归属研究”（2025NA32）
收稿日期：2026-01-15
作者简介：张晓梅（1994— ），女，浙江平湖人，博士，讲师，硕士生导师，研究方向为民商法学，通信作者，zhangxiaomei@zafu.edu.cn。

① 本文所探讨的诊断辅助人工智能，是指通过为诊疗活动提供分析建议以辅助医务人员进行临床决策的智能系统，如基于影像识别的病灶判定、基于数据分析的诊断建议输出等。仅提供医疗参考信息而不参与医疗决策的技术应用，如仅提供医疗流程简化、成像效率提升等参考性信息的，不属于诊断辅助人工智能。此外，直接介入治疗操作或执行治疗方案的智能系统，因其技术原理、风险模式及责任逻辑与诊断辅助存在显著差异，也不在本文讨论之列。

② 我国法上的医疗损害责任是一个类型化的责任体系，依据所违反义务性质的不同，可大体分为医疗技术损害责任、医疗伦理损害责任和医疗产品损害责任。本文所探讨的诊断辅助人工智能应用下的责任问题，主要聚焦于医务人员在使用该技术进行辅助诊断时，是否尽到合理注意义务，这本质上归属于医疗技术损害责任的范畴。其核心判断标准即《中华人民共和国民法典》第1221条所规定的“当时医疗水平相应的诊疗义务”。对于因人工智能系统本身存在缺陷而引发的产品责任，或因数据隐私、算法歧视等引发的伦理责任，虽与本文主题密切关联，但囿于归责原则与构成要件的显著差异，不在本文核心讨论之列。

定“诊断辅助人工智能”在临床应用中的合法性基础与法律定位？这一问题的核心在于，使用一个自主且不透明的“黑箱系统”辅助诊断，其行为本身面临是否具有正当性的质疑。相应地，它也将影响对医务人员一开始就违反注意义务的判断。若其应用本身的正当性存疑，则后续关于责任分配与过错认定的讨论将失去根基。②在责任认定层面，当人工智能的诊断建议与基于传统诊疗方案发生冲突时，医务人员采纳或者拒绝该建议的行为，应当如何运用“当时医疗水平”标准进行评价？算法的“黑箱性”使得诊断辅助人工智能的决策过程难以被医务人员理解和验证，这是否意味着其注意义务的履行基础已然丧失？③在义务层面，当“人机协作”成为新常态的背景下，医务人员的核心角色与注意义务是否已发生改变？如何才能更好地平衡技术赋能与患者安全？

上述困境不仅关涉《中华人民共和国民法典》相关条款在智能时代的规范生命力，更可能因责任边界模糊而同时抑制技术创新与损害患者权益。因此，本文旨在系统回应这一时代性诘问。首先，梳理诊断辅助人工智能的技术特性，并据此论证诊断辅助人工智能应用的合法性及界定其法律属性；其次，探讨其对现有医疗损害责任规则的冲击与挑战；最后，提出针对性的规则调适方案，以期为司法实践与行业规范提供智识参考。

二、诊断辅助人工智能的合法性基础与法律定位

诊断辅助人工智能与传统医疗决策系统存在根本区别。以斯坦福大学开发的MYCIN抗生素推荐系统作为传统医疗决策系统的典型，其本质是“预设规则的集合体”，通过专家编码的200多条规则，进行逻辑推演形成临床结论及诊疗建议[1]。虽然其在预设的狭窄领域内可实现较高决策水平，但无法处理规则之外的复杂情形[2]。而当前基于机器学习的诊断辅助人工智能已实现从“规则驱动”到“数据驱动”的技术跃迁。以通用诊断模型MedFound为例，该系统通过整合870万份真实电子健康记录进行训练，能够从数据中自主习得诊断模式，在多项专科诊断中表现优异[3]。其模型权重的开放共享，使开发者能够针对区域化需求进行微调，显著提升技术的可扩展性[3]。

这一从“规则驱动”向“数据驱动”的转向，赋予诊断辅助人工智能三大核心特性：①自主决策性，即能够基于机器学习与大数据分析，独立处理患者信息并形成诊断建议；②算法黑箱性，其输入与输出间表现为统计“相关性”，而非临床可解释的生物学“因果关系”[4]，且其运作方式难以通过逆向工程还原，导致决策缺乏可追溯性与可解释性[5]；③动态进化性，即能够通过持续接受新数据优化参数、更新方法乃至完成算法迭代[6]。

上述技术特征引发了对临床应用的根本性质疑：使用一个具有自主性、不透明性且持续演进的“黑箱”系统辅助诊断，是否具备正当性？是否自始构成对医务人员注意义务的违反？对此，有必要做出回应。

（一）使用行为本身不具有违法性

“诊断辅助人工智能的使用是否构成注意义务违反”这一质疑，其隐含前提在于将“工具的技术特征”直接等同于“使用行为的可归责性”。然而，这一推定既缺乏充分依据，也与损害责任的基本法理相悖。

一方面，法律评价的焦点在于医务人员行为的合理性，而非所使用工具的静态属性。医疗损害责任遵循过错原则，医务人员的“诊疗义务”是一种手段债务，核心在于其是否提供符合当时医疗水平的专业服务，而非保证治愈的结果债务[7]。若仅因人工智能具备技术先进性或者“黑箱”特性，便径直推定使用行为违法，实则混淆了工具评价与行为评价，不当地滑向以“工具的技术特征”进行归责的“技术决定论”，实质上将架空过错责任原则，使医生责任异化为严格责任。

另一方面，医疗行为固有的不确定性与认知局限，并非人工智能所独有。常规医疗干预和临床决策即使不涉及人工智能，对医疗专业人员自身而言也往往是不透明的[8]。这并不必然否定干预手段的合法应用。例如阿司匹林等药物，即使其作用机制尚未被完全理解，仍被广泛应用并证实有效且可靠[8]；电休克疗法虽对重度抑郁症有显著疗效，但其作用机制至今未明[9]。医学对因果系统的理解仍处于初级阶段，有效干预的能力往往源于经验积累，且先于我们理解干预措施为何有效的能力[10]。因此，在许多情况下，对诊断辅助人工智能系统未必需要强求完全的可解释性，其经验层面的准确性已经足够支持临床应用。关键在于，使用者是否在认知该工具有效性及局限性的前提下，以符合专业标准的方式审慎运用。人工智能已在特定诊断任务中展现出显著的临床价值，其速度与准确性可媲美甚至超越人类医生。虽然此种统计学意义上的准确性，解决的仅是“技术上的可用性”与“临床上的有用性”问题，并未自动回答“法律应如何负责地使用”这一核心议题，但其临床价值本身已为合法性提供了重要支撑，无需以完全的可解释性作为绝对前提。

（二）监管合规视角下的应用正当性

当前，诊断辅助人工智能通常被纳入医疗器械监管框架进行监管，具体职责由药品和医疗器械部门承担，例如我国的药品监督管理局（简称药监

局）、美国的食品和药品监督管理局（Food and Drug Administration，FDA）。一般认为，各国关于药品、器械监管的目标基本一致，都是要保证药品、器械的“安全和有效”[11]。例如我国《医疗器械监督管理条例》第13条规定，医疗器械的注册条件为“安全、有效”。美国《联邦食品、药品和化妆品法案》第505节也将“安全、有效”作为药品、器械监管的基本目标。“有效”是“安全”目标的自然延展，药品、器械用于诊疗疾病，无效药品、器械不但无法达到预期效果，还将诱发误诊或贻误治疗时机，终将危害患者，因而医疗人工智能的监管目标就是要确保其安全、有效，确保有益于患者的生命健康[12]。有学者认为，经过审批（如获得FDA批准）、具有可靠性的医疗人工智能工具，可与FDA批准的药品（如阿司匹林）相类比，二者均属于信用品（credence good）。医务人员无需深究其底层机制或试验细节，仅需依赖审批机构的可靠性即可[13]。依此逻辑，对于这类成熟度高、经过权威审批的人工智能系统，医务人员可在相应程度上信赖其输出[14]，其技术特性本身不构成临床应用的根本障碍。

但仍需审慎对待的是，人工智能的准确性依赖群体数据验证，本质是统计层面的概率有效。但临床场景复杂多变，个体之间存在差异，对“可解释性”的要求需超越单一的统计学准确性，需融入因果机制的探索与场景适配，以确保契合个体患者情况。同时，需通过临床试验验证因果关系，借助监管限定其应用范围，从而在满足临床需求与追求医学长远发展之间取得平衡[15]。

（三）作为“新疗法”的合法性定位

“当时医疗水平”是判断医疗过错的标准。“当时”是判定的时间节点，不能以“当前医疗水平”评估“损害发生时的诊疗活动”[16]。由于医学和医疗活动的发展性，医疗水平处在不断变化之中。因而，在医疗过错的判断上应考虑时间因素，以医疗行为发生时应当具备的医疗水平作为判断标准[7]。从这个意义上说，“当时医疗水平”判断标准是一个动态且灵活的概念，可将医疗过错的法律评估与医学领域的持续发展相融合。而随着诊断辅助人工智能作为“辅助诊疗”手段被纳入医疗器械监管体系，其注册与应用数量快速增长，并在提升癌症早期检出率、缩短脑卒中抢救时间等方面展现出显著优势，于国内多家医院及基层场景中得到验证。当人工智能在诊断任务上的统计表现优于人类时，医生可以向相关裁判机构解释，其使用了比自己诊断能力更优的诊断工具，是在维护患者的最佳利益[17]。其结果是，诊断辅助人工智能作为一种新兴事物，可作为“新疗法”获得合法性容纳。法律应尊重医务人员的专业判断，允许其选择合适的疗法。只要医生在应用时履行了合理的注意义务，法律就应给予其较大的生存空间[18]。

但值得注意的是，医生作为诊断辅助人工智能的使用者，可能无法知晓该系统是否已被正确制造、编程或训练，但这本身并不构成预先排除其应用的理由。当然，如果医务人员明知或应知所用人工智能存在重大缺陷或故障风险而继续使用，则该行为本身构成注意义务的违反。除此之外，对于经过严格审批与临床验证的人工智能系统，其作为现代医疗技术的组成部分，应用的合法性基础应予以确认。这一基本定位，为后续分析其对具体责任规则的冲击设定了前提，即本文探讨的是“合法使用工具时的责任划分问题”，而非“使用非法工具的责任问题”。同时，将人工智能定位为“新疗法”，也意味着其应用过程应遵循新疗法引入所伴生的更高注意义务要求，这为后文“审慎再判断义务”的展开埋下了法理伏笔。

三、诊断辅助人工智能对医疗损害责任的冲击

人工智能的自主决策性、算法黑箱性以及动态进化性，改变了医疗决策的生成与知识基础。这与传统医疗损害责任规则所预设的“人类主导”“过程透明”与“知识相对稳定”等前提之间，产生了深刻冲突。这种冲突直接动摇了医疗过错认定的两大支柱：其一是作为客观判断基准的“当时医疗水平”标准，其二是需要在具体情境中履行和判断的医务人员注意义务。

（一）对“当时医疗水平”确定性基础的动摇

诊断辅助人工智能颠覆了传统“当时医疗水平”所依赖的稳定、共识性的知识生成与传播周期，造成了技术能力与法律标准之间的“时差”与“错位”。详细来说，《中华人民共和国民法典》第1221条将“医疗过错”客观化为“是否履行与当时医疗水平相应的诊疗义务”[16]。此处的“诊疗义务”是一种手段债务[7]，其判断通常具体化为“合理医生标准”[19]，即以同行医务人员普遍具备的“平均”或“一般”医疗水平作为基准[20]。传统医疗水平的演进，依赖临床研究、同行评议、指南制定与专业教育这一透明且漫长的共识形成过程。因此，“当时医疗水平”本质上是一个建立在专业共同体共享、相对稳定且可通过既定途径传播的知识体系之上的规范性概念。然而，诊断辅助人工智能的应用从根本上动摇了这一概念的确定性前提。

以数据驱动为核心的人工智能，凭借其动态进化性，引入了一种全新的知识生成逻辑。它能够瞬时处理海量数据，通过统计关联生成新的诊断模式。其“知识”内嵌于算法参数中，并通过系统更新迭代实现“传播”。这种基于相关性而非因果性的

知识生产机制,与传统医学基于理解与解释的知识生产机制存在根本分野。

更关键的是,人工智能的知识迭代速度已远远超越传统共识的形成周期,导致技术系统所展现的"诊断能力水平"在事实上领先于法律所认定与依赖的"当时医疗水平"。此种"诊断能力水平"与"当时医疗水平"之间的"错位",可能动摇后者作为稳定、共识性判断基准的规范功能。人工智能不仅改变了知识的生产方式,更在知识应用层面制造了法律评价标准与技术现实之间的断裂。

(二)对医务人员注意义务的冲击

我国现行法律框架明确将诊断辅助人工智能定位为"辅助角色",强调每一次机器判断都需经过医务人员的再判断,才能最终转化为直接作用于患者的诊疗措施[14]。这表明,即使人工智能参与或者辅助诊断决策,但仍然是"人类主导责任"。医务人员需通过履行注意义务对患者负责[14]。但值得注意的是,人工智能的"临床潜能"与"算法黑箱性"相结合,在实践中催生了新的风险。

其一是"警觉衰退"风险。人工智能已被证明具有显著的临床应用价值,其被认为在诊断速度、准确率等方面拥有匹敌甚至超越人类的能力。典型如谷歌开发的乳腺癌人工智能检测系统,其在识别难以检测的转移性乳腺癌方面远超人类病理学家,其诊断准确率高达99%,而人类病理学家仅为38%。因而,在人工智能介入诊疗行为后,可能导致医务人员产生过度信赖,使本应进行的独立专业判断流于形式,产生"警觉衰退"(atrophy of vigilance)[21]。

其二是"决策困境"风险。前述"诊断能力水平"与"当时医疗水平"之间的"错位",使医务人员在具体诊断行为中面临决策困境。医务人员在决定是否采纳一个技术上更优但共识度不足的人工智能诊断建议时,缺乏清晰、稳定的法律指引。若采纳人工智能推荐的、超出传统标准但事后证明正确的方案,是否因"超前"而构成过错?若拒绝该建议并沿用效果较差的传统疗法,导致患者丧失最佳治疗机会,是否又因未尽到为患者争取最大利益的义务而构成过错?而算法的"黑箱性"使其内部决策过程无法被医务人员直接理解或验证,医务人员仅能通过观察数据(observational data)进行间接推断[22]。这已非简单的"采纳与否"的两难选择,而是在规范依据模糊、决策过程不透明的条件下,医务人员注意义务的履行本身受到根本质疑的困境。

四、人工智能时代诊疗过错标准的重塑与适用路径

诊断辅助人工智能不仅冲击了以稳定共识为基础的"当时医疗水平"标准,也对医务人员在具体情境下的注意义务内容提出了重构需求。为应对上述挑战,本文将构建一个能容纳技术动态性,并聚焦于医务人员行为合理性的过错认定框架。在展开具体框架之前,有必要先澄清一个根本性的法理困惑:既然人工智能在特定诊断任务上可能展现出"先进性",为何法律仍需强调医务人员的独立判断,并将"过度依赖"视为过错?

(一)过错认定核心法理的澄清

《中华人民共和国民法典》第1221条所确立的"当时医疗水平"标准,其规范内涵指向的是"医疗水平",而非"医学水平"。"医学水平"主要指除具备基本诊疗能力外,还需具备的医学研究水平[16],其必须经过临床实践的充分验证与专业共同体的广泛认可,方能转化为普遍遵循的医疗水平[23]。诊断辅助人工智能凭借其数据处理与学习能力,能够瞬时触及医学前沿。然而,这并不意味着"人工智能的输出水平"可直接等同于法律意义上的"当时医疗水平"。法律允许将诊断辅助人工智能作为"新疗法"应用,与将其输出结果直接认定为判定过错的"当时医疗水平"标准,是两个不同层面的法律问题。

当前,存在一个看似矛盾的法理诘问:既然人工智能在特定诊断任务上展现出"先进性",为何法律仍需强调医务人员的独立判断并批评其"过度依赖"?这触及了智能医疗时代责任分配的根本问题,必须在法理上作出区分。人工智能所展现的"先进性",主要是其算法在特定任务上表现出的高准确率,这是一种基于数据和统计的技术能力;而法律判断所依据的"当时医疗水平",则是一个融合了专业知识、可解释性、可验证性等价值的法律标准。正因为技术上的高准确率难以被法律上的通行标准即时、完全地吸纳,所以在绝大多数临床场景中,人工智能仍被定位为决策验证工具,而非医务人员必须严格遵从的强制性要求[24]。

因此,技术的"先进性"非但不能豁免或消解医生的责任,反而对其专业角色提出了更高的要求。医生需在智能工具与患者个体之间进行审慎权衡,并承担最终的法律责任。"过度依赖"之所以构成过错,并非因为使用了先进工具,而在于医生放弃或严重懈怠了"独立判断"这一核心注意义务。从而,其自身从负责任的决策者降格为技术的被动传递者。而要求医务人员进行"结果核验",并非源于对技术的不信任,而是将技术输出纳入人类专业认知体系与法律评价框架的必经步骤,确保技术效能以负责任且合法的方式转化为患者的切实福祉。下文所构建的以"审慎再判断义务"为核心的动态责任框架,其法理正当性正根植于此。

（二）动态医疗过错认定框架的重塑与适用路径

基于前述法理澄清，并针对人工智能的特性对确定性基础的冲击，本文试图构建一个动态发展的医疗过错认定框架，旨在通过“采纳建议”与“拒绝建议”两类行为分别设立明晰且具有前瞻性的评判标准，实现风险控制与技术激励的平衡。

其一，确立以“审慎核查”为核心的采纳建议归责标准。医务人员采纳人工智能推荐的诊断方案，只要该方案未明显超出法律容许的“当时医疗水平”边界，且医生已结合患者个体情况与自身专业能力进行了独立、审慎评估，则不因其建议来源而直接认定过错。反之，若因疏忽未履行审慎核查义务而盲目采纳建议导致损害，则应承担相应责任。判断的关键在于，审查医务人员是否尽到了与同行相应的“审慎再判断义务”（下文将进一步交代何为审慎再判断义务）。需特别明确的是，鉴于算法的“黑箱性”，医务人员基于合理怀疑（例如人工智能结果与患者整体临床表现存在显著冲突）而不予采纳相关建议，通常不应被认定为存在过错[25]。此标准旨在鼓励对人工智能工具的合理利用，而非助长盲目依赖。

其二，适用动态演进的拒绝建议归责标准。“当时医疗水平”是随医学实践发展而动态演进的规范性概念。在现阶段，医务人员出于合理审慎的考量，拒绝采纳尚未形成广泛共识的人工智能建议，转而选择符合现行诊疗规范的方案，通常不构成过错。然而，人工智能的知识迭代速度远超人类共识与传统诊疗规范的更新周期，这导致法律认定的“当时医疗水平”在事实上滞后于技术已实现的“前沿水平”。随着特定人工智能应用的临床证据持续积累，其效能获得广泛验证并逐渐成为普遍接受的医疗实践，“当时医疗水平”的内涵将随之更新。届时，医务人员若无正当理由拒绝使用已成熟的人工智能系统，或拒绝采纳其经过验证的正确建议，则可能因未达更新后的同行标准而构成过错；相应地，若其遵循了错误的人工智能建议，或可基于该技术已成为通行实践而免除损害责任[26]。这要求医务人员必须持续关注相关领域的发展，因为其专业标准可能会迅速变化[24]。

（三）行业协同治理机制的构建与医疗机构责任

为确保前述动态认定框架的实践效果，需构建由专业组织与医疗机构共同参与的协同治理体系，明确各方在人工智能临床应用中的责任边界。该体系体现了“硬法与软法协同”的治理思路。其核心在于，专业标准（软法）为前沿技术应用提供灵活、专业的指引，而医疗机构的内部审查机制则将软法要求与硬法责任相衔接，实现自律与他律的结合[27]。

一方面，应充分发挥专业组织的规范指引作用。医务人员应积极推动其所属行业协会，针对人工智能的临床应用制定具体的操作指南与评估规范。在美国，以FDA为代表的监管机构侧重于产品准入时的安全性与有效性检查，而专业协会则可进一步提供实施阶段的操作指引，并针对个体患者的人工智能建议给出专业评价标准，从而在行业层面确立“合理使用”人工智能的行为规范[24]。在我国，《中华人民共和国民法典》第1222条规定，如果违反法律、行政法规、规章以及其他有关诊疗规范的规定，则被认为有过错[20]。该条是判断过错的法定事项或标准[20]。虽然医务人员的注意义务与“符合法律、行政法规、规范以及诊疗规范的有关要求”并不相同[28]，但该条确立了诊疗规范在过错认定中的法定参照地位。而专业组织制定的人工智能临床应用评估指南与实践规范，正是对“诊疗规范”的细化与延伸，为司法实践中认定过错提供了更具操作性的行业依据。同时，法律对诊疗规范的强制性援引，也反过来促进了行业标准的普遍遵循。

另一方面，须压实医疗机构的主体管理责任。《中华人民共和国民法典》第1221条过错行为的主体表述为“医务人员”，但结合第1218条“医疗机构或者其医务人员有过错”的表述可见，法律已明确承认医疗机构可作为独立主体，就其整体医疗活动中的过错承担责任，而非仅追究具体医务人员的个体责任。依据医疗组织过错责任理论，判断过错时应关注医疗机构整体提供的医疗服务是否达到合理的行业医疗水平，而非孤立评判个别医务人员的具体行为[7]。基于此，医疗机构必须建立并执行严格的人工智能内部审查与全流程评估机制。在采购和部署外部人工智能时，应确保其符合真实的临床需求与安全保障标准，审慎程度不应低于对其他新型医疗设备的要求，从源头上管控应用风险[24]。

五、“人机协作”模式下医务人员审慎再判断义务的厘清与展开

前述动态过错认定框架，最终需通过对医务人员具体行为的评价来落实。然而，研究表明，医务人员对人工智能应用中的伦理与法律问题认知普遍不足[29]。因此，在“人机协作”成为常态的背景下，明确并强化医务人员的“审慎再判断义务”具有现实紧迫性。

从法理渊源上看，“审慎再判断义务”并非凭空创设的新义务，而是医务人员注意义务在人工智能医疗场景下的具体展开与形态更新。在传统医疗领域中，医务人员应具备与其职业要求相当的诊疗技能与谨慎态度；当人工智能介入诊疗过程后，虽然决策方式发生了变化，但医务人员作为最终责任主体的法律地位并未改变。因此，其注意义务的内

容必然随之更新，从“独立做出正确诊断”扩展为“审慎地使用智能工具并对算法输出进行实质性核验”。正是在此意义上，“审慎再判断义务”构成了医务人员注意义务在智能时代的具体化形态。

在此模式下，医务人员的角色正从独立的决策者转变为人工智能的使用者，并对算法输出负有不可转移的终审责任，医疗人工智能的结论必须经医师审核确认后方可实施[30]。“审慎再判断义务”诞生的逻辑必然性在于，在承认技术价值的同时，必须通过制度性安排确保医生作为“最终责任主体”的法律地位不被技术所架空。其核心并非简单重复人工智能的计算过程，而是要求医生以专业判断，对人工智能的诊断建议进行法律意义上的审查与确认，从而使其转化为一个可归责于医生的“专业诊断”。本章将具体阐释这一义务的双重内涵。

（一）使用审慎义务

“使用审慎义务”要求医生及医疗机构须像对待任何高风险医疗设备一样，对所用人工智能系统的安全、有效运行承担责任[31]。

第一，运行保障义务。医务人员需熟悉所使用人工智能系统的基本功能、工作方式及其数据库的构成与潜在偏差[31]。数据库的真实性与代表性尤为关键，不具代表性的数据将导致输出结果脱离临床实际。以IBM沃森系统为例，其早期因未使用真实患者数据，而仅依赖纪念斯隆凯特琳（MSK）癌症中心医生设计的少量“合成病例”，导致算法给出的建议脱离实际诊疗需求。这凸显了数据源真实性与代表性的关键影响。因此，确保数据库合理有效是基本要求。此外，医务人员需对系统运行状态进行持续监督，包括及时安装更新、补丁及修复程序等，以确保其处于正常工作状态。若自身能力不足，应将维护工作委托专业人士完成[31]。

第二，基于风险分级的管理义务。医务人员的注意义务程度应与所用人工智能工具的风险等级相匹配。根据《医疗器械监督管理条例》，医疗器械按照风险程度被分为低风险的第一类医疗器械、中度风险的第二类医疗器械、较高风险的第三类医疗器械。而根据《人工智能医用软件产品分类界定指导原则》，人工智能医用软件和含人工智能软件组件的医疗器械的管理类别应结合产品的预期用途、算法成熟度等因素综合判定。若用于辅助决策，按照第三类医疗器械管理；若用于非辅助决策，按照第二类医疗器械管理。对于算法在医疗应用中成熟度高的人工智能医用软件，其管理类别按照现行的《医疗器械分类目录》和分类界定文件等执行。该原则体现了人工智能医疗监管以风险水平决定管理严格程度的核心逻辑。算法成熟度高的人工智能医用软件，将根据其实际的功能和成熟度找到对应的传统产品类别进行管理，无需被“特殊关照”或者拔高管理级别。反之，以“诊断辅助人工智能”为代表的第三类医疗器械，医务人员在采纳其判断时应履行更高的注意义务。

第三，数据输入的质量控制义务。人工智能诊断的可靠性高度依赖输入数据的质量。医务人员必须确保输入系统的患者信息（如病历、影像等）完整、准确，这是获得有效输出结果的前提。此时，诊断辅助人工智能仅作为被动工具，接受并执行人类输入的指令，无法拒绝或修正。因而，医务人员负有高度的注意义务，需对诊断辅助人工智能进行规范操控[30]。若输入的医疗数据本身存在不准确、不清晰、不全面等瑕疵，则难以生成高质量的机器判断，此时医务人员应谨慎采纳[14]。

（二）结果核验义务

“结果核验义务”是“审慎再判断义务”的核心环节，它要求医务人员对人工智能生成的诊断建议进行主动、独立的审查，并做出最终决策。该义务的法理基础在于，法律追究的是医务人员提供专业诊断服务的过错，而非单纯搬运人工智能结论的行为。人工智能的数据分析结果或判断，被认为与传统医疗中的验血、B超检查结果具有同质性，仅为疾病诊断提供参考，不改变医务人员作为诊疗核心与责任主体的地位[25]。

该义务可具体化为两方面要求：其一是审查诊断建议的医学依据是否充分，如是否遵循最新临床指南、是否符合“当时医疗水平”；其二是结合患者具体情况进行个性化评估，如考虑其过敏史、基础疾病、并发症等，避免机械套用通用方案[30]。医生的诊疗行为需遵循合理医生标准，专业知识和谨慎态度是不可放弃的法定要求，即使医务人员真诚依赖“黑箱”算法，也不能免除自身责任[26]。这一规则在监管层面亦有呼应。美国《联邦食品、药品和化妆品法案》明确规定，获监管豁免的临床决策支持软件，必须确保医疗专业人员能独立审查建议依据，不得诱导其主要依赖软件决策。该条款正是为了防范医生因盲目信任人工智能而放弃独立判断。

但值得注意的是，人工智能系统的“黑箱性”使得医生无法深入审查其内部逻辑，但这并非免除核验义务的理由，而是改变了义务的履行方式。医务人员应当尽可能地核验结果，而不能以表面数据进行直接判断，否则若因直接采纳人工智能建议而给出错误诊断，造成患者损害，将根据情节承担民事或刑事责任[25]。从实际操作层面看，应结合“结果复验可能性”进行具体分析。例如对于明确标注特征的影像辅助诊断，医生可基于标注信息进行针对性复核；但对于系统可能存在的漏诊，要求医生完全重复检视所有原始数据并不实际，合理的标准应

是在依赖可靠系统的基础上,对显而易见的疑点保持警觉并审慎核查[14]。

六、结 论

诊断辅助人工智能的临床应用,对医疗损害责任规则构成挑战。传统的"当时医疗水平"标准与医务人员注意义务体系,在面临人工智能的自主性、黑箱性与动态进化性时,均显露出适用困境。为此,本文系统回应了三个层次的问题:其一,确立了诊断辅助人工智能应用的合法性基础,明确其使用行为本身不具违法性,且作为"新疗法"可为现行法律框架所容纳;其二,揭示了人工智能对既有责任认定体系的双重冲击——既动摇了"当时医疗水平"标准的确定性,也使医务人员面临规范依据模糊的决策困境;其三,针对性地提出了以"动态过错认定框架"与"审慎再判断义务"为核心的规则调适方案。

本文通过构建一个容纳技术动态性的责任认定框架,将对过错行为的评价重心,从对静态诊疗方案合理性的判断,转向对医务人员履行"审慎再判断义务"之过程合规性的审查。该义务要求医务人员在"人机协作"中承担终局责任,具体体现为对人工智能工具的审慎使用与对算法输出的实质性核验。这一重构不仅是对医疗损害责任认定规则在智能时代的具体调整,更是对智能时代医疗专业精神的再确认。它旨在寻求技术创新与患者安全保护之间的规范平衡,为司法实践、行业治理与医务人员合规操作提供一个清晰的指引框架。唯有在规范中明晰责任,方能在协作中重拾信任,从而推动人工智能在医疗领域负责任、可持续地应用与发展。

参考文献

[1] SHORTLIFFE E H, BUCHANAN B G. A model of inexact reasoning in medicine[J]. Math Biosci, 1975, 23(3/4): 351-379

[2] LAGIOIA F, CONTISSA G. The strange case of Dr. Watson: liability implications of AI evidence-based decision support systems in health care[J]. European Journal of Law and Technology (EJLS), 2020, 12 (2): 245-289

[3] LIU X H, LIU H, YANG G X, et al. A generalist medical language model for disease diagnosis assistance[J]. Nat Med, 2025, 31(3): 932-942

[4] PRICE W N II. Black-box medicine[J]. Harvard Journal of Law & Technology, 2015, 28: 419-467

[5] GIUFFRIDA I. Liability for AI decision-making: some legal and ethical considerations[J]. Fordham Law Review, 2019, 88 (2): 439-456

[6] LIU F, TONG X, YUAN M, et al. Evolution of heuristics: towards efficient automatic algorithm design using large language model [C]. Proceedings of the 41st International Conference on Machine Learning , Vienna, Austria: PMLR, 2024: 1-22

[7] 邹海林,朱广新. 民法典评注—侵权责任编[M]. 北京:中国法制出版社,2020:541-542

[8] VÉLIZ C, PRUNKL C, PHILLIPS-BROWN M, et al. We might be afraid of black-box algorithms[J]. J Med Ethics, 2021, 47(5): 339-340

[9] 埃里克·托普. 深度医疗[M]. 郑杰,朱烨琳,曾莉娟,译. 郑州:河南科学技术出版社,2020: 89

[10] LONDON A J. Artificial intelligence and black-box medical decisions: accuracy versus explainability[J]. Hast Cent Rep, 2019, 49(1): 15-21

[11] 弗雷德里克·M. 阿尔伯特,格雷厄姆·杜克斯. 全球医药政策:药品的可持续发展 [M]. 翟宏丽,张立新,译. 北京:中国政法大学出版社,2016: 6-9

[12] 李润生. 论医疗人工智能"黑箱"难题的应对策略和规制进路[J]. 东南大学学报(哲学社会科学版),2021,23(6):83-92,146

[13] COHEN I G. Informed consent and medical artificial intelligence: what to tell the patient?[J]. Georgetown Law Journal, 2020, 108: 1425-1469

[14] 郑志峰. 诊疗人工智能的医疗损害责任[J]. 中国法学,2023(1):203-221

[15] PIERCE R L, VAN BIESEN W, VAN CAUWENBERGE D, et al. Explainability in medicine in an era of AI-based clinical decision support systems[J]. Front Genet, 2022, 13: 903600

[16] 徐涤宇,张家勇.《中华人民共和国民法典》评注[M]. 北京:中国人民大学出版社,2022: 1288

[17] GIUBILINI A. It is not about AI, it's about humans. Responsibility gaps and medical AI[J]. J Bioethical Inq, 2025, 22(3): 527-537

[18] 赵西巨. 医生对新疗法的使用和告知[J]. 东方法学,2009(6):119-133

[19] 程啸. 侵权责任法[M]. 3版. 北京:法律出版社,2021:644-645

[20] 张新宝. 侵权责任法[M]. 5版. 北京:中国人民大学出版社,2020:200

[21] SMITH H, FOTHERINGHAM K. Artificial intelligence in clinical decision-making: rethinking liability[J]. Medical Law International, 2020, 20(2): 131-154

[22] PRICE W N Ⅱ. Describing black-box medicine [J]. Boston University Journal of Science & Technology, 2015, 21: 347-357

[23] 窦海阳. 法院对医务人员过失判断依据之辨析——以

《侵权责任法》施行以来相关判决为主要考察对象[J]. 现代法学,2015,37(2):167-181
[24] PRICE W N II,GERKE S,COHEN I G. Potential liability for physicians using artificial intelligence[J]. JAMA,2019,322(18):1765-1766
[25] 刘建利. 医疗人工智能临床应用的法律挑战及应对[J]. 东方法学,2019(5):133-139
[26] GERKE S, MINSSEN T, COHEN G. Ethical and legal challenges of artificial intelligence-driven healthcare[A]// Artificial Intelligence in Healthcare. Amsterdam:Elsevier,2020:295-336
[27] 徐辉,王译. "硬法—软法"范式下医疗人工智能伦理治理路径探析[J]. 南京医科大学学报(社会科学版),2024,24(4):359-368
[28] 黄薇. 中华人民共和国民法典侵权责任编解读[M]. 北京:中国法制出版社,2020:210
[29] 赵丹娜,赵蓝蓝,陈任,等. 某三级公立医院医务人员对人工智能医学应用伦理问题的认知研究[J]. 南京医科大学学报(社会科学版),2022,22(5):465-470
[30] 孙运梁,王璨. 医疗人工智能介入过失诊疗行为的结果归属研究[J]. 江西师范大学学报(哲学社会科学版),2025,58(5):137-148
[31] MAROUDAS V P. Fault-based liability for medical malpractice in the age of artificial intelligence: A comparative analysis of German and Greek medical liability law in view of the challenges posed by AI systems[J]. Rev Eur Comp Law,2024,57(2): 135-169

(本文编辑:姜 鑫)

The dilemmas and solutions of applying medical malpractice liability to diagnostic-support AI

ZHANG Xiaomei
College of Humanities and Law, Zhejiang A&F University, Hangzhou 311300, China

Abstract: Diagnostic-support AI has undergone a technological leap from being rule-driven to data-driven. Its inherent autonomous decision-making capacity, an algorithmic black-box nature, and dynamic evolution pose structural challenges to the traditional medical malpractice liability system, which is centered on the prevailing medical standard. To address these challenges, the use of diagnostic-support AI should be characterized as the adoption of a new medical treatment to establish its legal basis. It is necessary to clarify that its outputs are not directly equivalent to the prevailing medical standard, and that the implications of this standard will also evolve. On this basis, this paper constructs a dynamic fault-determination framework centered on the duty of prudent reassessment. This framework shifts the focus of assessment to evaluate whether healthcare professionals have fulfilled their duties of prudent use and substantive verification of algorithmic outputs, thereby responding to the needs of liability determination in the current human-computer collaboration model and achieving a normative balance between technological innovation and patient safety protection.

Key words: diagnostic-support AI; prevailing medical standard; medical fault; duty of prudent reassessment