DOI: 10.3724/SP.J.1224.2020.00241

“新兴医学工程伦理与治理”专刊

医疗 AI 安全风险的伦理与法律保障机制研究

王 玥✉，戴海洋

（西安交通大学 法学院，西安 710049）

摘 要: 人工智能（AI）正在迅速发展并应用到医学领域当中，给广大医务工作者和患者带来了新的机遇与希望。然而，由于对象是人类的疾病与健康，医疗 AI 的应用本身会给患者和受试者健康带来直接威胁，且因为数据和算法缺陷、系统漏洞或缺陷可能引发的安全风险也尤为突出。因此，在已经被广为接受的 AI 伦理原则之外，医疗 AI 还应当强调以患者和受试者为中心创设合理的人机交互，以及保持持续评价的伦理保障机制。此外，应当在我国现行的医疗 AI 法律监管框架下，明确划定医疗 AI 范围，并针对医疗 AI 的特殊性完善其注册和后续监管制度，在保证安全的基础上构建患者和临床医生对医疗 AI 的信任，从而使医疗 AI 能够更好地改善社会医疗保健，保障患者和受试者的安全和最大福祉。

关键词: 医疗人工智能；医疗伦理；医疗损害；医疗数据

中图分类号: D90；R-02 **文献标识码**: A **文章编号**: 1674-4969(2020)03-0241-11

引言

近年来，人工智能（AI）技术已经逐步从实验室中走出并转向落地应用阶段，而医疗健康领域是人工智能应用大量集中的场景之一。据国务院发布的《新一代人工智能发展规划》，“推广应用人工智能治疗新模式新手段，建立快速精准的智能医疗体系”[1]是我国今后人工智能发展的重点任务之一。随着人工智能在成本、质量和接入方面释放出的巨大力量，它的受欢迎程度正在呈爆炸式增长。据预测，到 2021 年，人工智能健康市场规模预计将以 40%的复合年增长率增至 66 亿美元[2]。

2017~2018 年，美国食品药品监督管理局（FDA）已经批准了包括 QuantX 乳腺癌诊断系统在内的 14 个基于人工智能技术的医疗应用[3]。在国内，2020 年 1 月 15 日，国家药品监督管理局宣布，北京昆仑医云科技有限公司开发的“冠脉血流储备分数计算软件”获批上市[4]，这是国内获批的首个应用人工智能技术的三类医疗器械，标志着人工智能技术在我国正式进入临床应用阶段。我国最新颁布的《基本医疗卫生与健康促进法》中也提出，“国家要推进人工智能等的应用发展，运用信息技术促进优质医疗卫生资源的普及与共享。”[5]然而，当前关于医疗人工智能研究成果以及大规模的商业报道均集中在相关应用的专业性和可靠度，却很少关注到 AI 在医疗领域应用的潜在风险及对策。

已经有一些学者开始注意到人工智能技术在医疗卫生领域的应用将会产生的问题和存在的隐患[6]：在医学伦理领域，包桉冰、徐佩提出了医疗 AI 对医生主体地位的挑战[7]，周吉银、刘丹、曾圣雅关注到了公平受益、失业、患者隐私、医疗安全、责任划分和监管等一系列问题[8]。在法学研

收稿日期: 2020–03–20; 修回日期: 2020–05–06

作者简介:✉ 王 玥（1983–），女，博士，副教授，研究方向为网络与信息安全法、医疗法律与伦理。E-mail：wangyue2011@xjtu.edu.cn（通讯作者）

戴海洋（1997–），男，硕士研究生（在读），研究方向为网络与信息安全法。

究中，学者从不同角度表达了对医疗 AI 应用数据隐私或医疗损害责任分配的担忧。曹艳林、王将军梳理了人工智能在医疗领域的应用状况和法律问题[9]，刘建利对医疗 AI 的法律地位、损害责任和医疗数据利用等医疗 AI 临床应用法律问题进行了研究[10]。然而，与患者隐私或者医疗责任相比，关于医疗 AI 系统本身的安全性问题没有得到普遍关注。已有学者建议，应当尽快制定与技术进展相匹配的医疗信息与人工智能系统的行业标准[11]。

有鉴于此，本文将从 AI 在医疗领域应用存在的主要安全风险出发，探究如何通过构建合理的伦理和法律机制应对这些安全风险，增进患者和医务工作者对人工智能的信任。

1　医疗人工智能及其主要安全风险

尽管人工智能的概念早在 1956 年就已经诞生，但时至今日人们还是时常混淆这一概念的内涵①。欧盟委员会在《可信人工智能伦理指南》中给出了关于人工智能系统可供参考的定义，“人工智能（AI）系统是由人类设计的软件（也可能是硬件）系统，它们在给定复杂目标的情况下，通过数据采集感知其环境，解释收集的结构化或非结构化数据，对从这些数据得出的知识进行推理，或处理信息，并决定为实现给定目标而采取的最佳行动，从而在物理或数字维度上行动。人工智能系统既可以使用符号规则，也可以学习数字模型，它们还可以通过分析环境如何受到之前的操作影响来调整自己的行为”[12]。最近发展起来的深层神经网络是一种类脑智能软件系统，它的出现使得人工智能的研究进入了一个新阶段[13]，算法的改进、算力上的突破和数据量的上涨使得人工智能在安防、金融、自动驾驶等各个行业中逐步落地（图 1）。

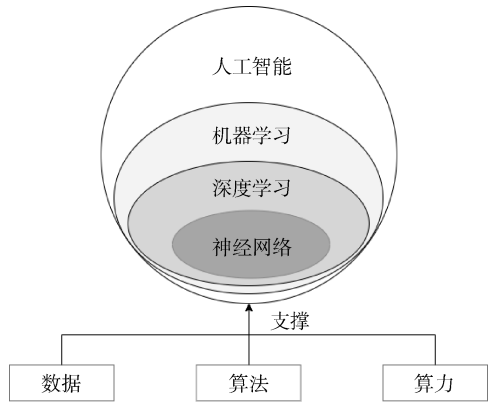

图 1　人工智能相关概念辨析

具体到医疗领域，医疗人工智能（“医疗 AI”）泛指应用在医疗健康领域的人工智能技术，涉及医疗行业各个环节，其目标是利用人工智能为患者进行诊断或治疗、提供健康咨询意见，或为医疗保险和药物研发等环节提供辅助，以提高医疗服务的效率及专业水平，减少医疗成本[14]。它区别于一般的医疗设备软件（SaMD），其独特性在于医疗 AI 应用涉及从数据中进行学习，并且可以通过不断学习改进和优化输出结果。据研究，目前医疗 AI 在 AI 医学影像、AI 辅助诊断、AI 药物研发、AI 健康管理和 AI 疾病预测领域均有较为明显的进步[15]。需要说明的是，机器学习不是人工智能的全部，本文的研究对象仅为基于机器学习的人工智能系统，并且仅限于目前已存在的作为辅助工具的医疗 AI 应用，不包含替代人类做出最终决策的情形。此外，“手术机器人”这类单纯传递操作者指令以进行微创手术的设备也不属于本文所研究的医疗 AI 的范围。

然而，所有的医疗 AI 应用在为医疗机构带来效率的同时，也蕴藏着对患者健康的潜在风险（图 2），这些风险包括应用本身对患者和受试者健康具有直接威胁、因数据或算法的缺陷引起的决策失误风险，以及因系统设计缺陷而难以避免的信息安全风险。

① 或者说，目前人们没有在同一的语境下讨论这一概念，将“机械姬”与“Siri 语音助手”放在一个语境下谈论，脱离人工智能的发展现状，过早地憧憬了能够达到或超出人类水平、完全替代人类决策的强人工智能。

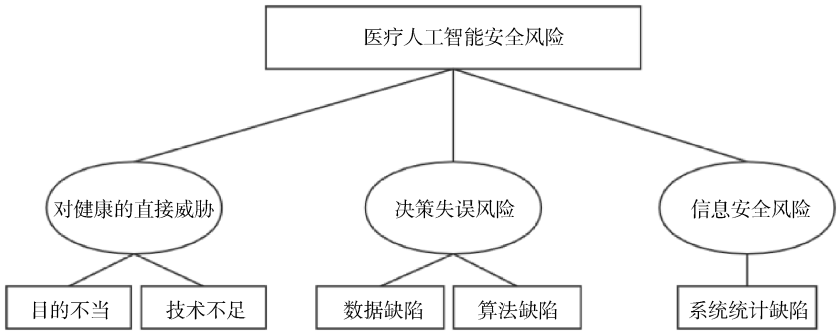

图 2　医疗人工智能安全风险及成因

1.1　应用本身对患者和受试者健康具有直接威胁

斯坦福大学生物医学伦理学中心主任戴维·马格努斯（David Magnus）指出：“医疗系统中内置的算法可能反映了不同的，相互冲突的利益。”医疗 AI 应用的开发，通常会同时考虑多个因素，如果多个因素的排序出现问题，则会对患者或受试者的健康与安全造成直接威胁。例如，如果医院的一套诊疗系统是以节省医保资金开支为目的设计的，进而根据患者的医保情况给出不同的诊疗建议，便会带来影响公正等伦理问题，甚至在特定的情况下直接威胁患者的生命。研究表明，许多开发医疗 AI 应用的独角兽商业公司很少在有同行评审的期刊上发表研究成果[16]，应用缺乏足够的透明度，让人们难以对这些应用的安全性产生充分信任。

刚诞生不久的深度伪造（Deepfake）技术①则已经成为人工智能技术被滥用的最典型场景，令 Facebook 等科技公司和政治家们头疼不已，而此类技术滥用的风险并不会因为应用于“治病救人”的医疗领域而消失。联合国人工智能和机器人中心高级战略顾问贝里泽指出：“人工智能是一个工具，任何一个工具都有可能被用作好的或坏的目的，因此这完全取决于我们如何使用这一工具。”[17]尽管目前公众视野中的各种医疗 AI 应用，不论是辅助诊断还是疾病预测，都是以患者健康和福祉为出发点的，但我们仍然应当对潜在的技术被滥用的情形保持警惕，2019 年发生的“贺建奎事件”再次为我们敲响了警钟，保证新技术被应用于正确目的是监管的永久命题。

1.2　因数据或算法的缺陷引起的风险

基于机器学习的医疗 AI 应用产生的结果与用于研发医疗 AI 应用的训练数据和算法高度相关，产品的安全性和有效性将在很大程度上取决于数据和算法的质量，这将使医疗 AI 应用的输出结果难以避免地包含“偏见”。研究指出，人的偏见、设计引入的偏见以及医疗系统使用数据的方式的偏见都有可能影响健康数据[18]。尽管医疗 AI 应用不会进行价值选择而仅仅是根据数据和算法输出结果，但越来越多的证据表明，用于开发医疗 AI 应用的数据集可能不能代表广泛的人群，算法的设计不合理，AI 系统的“偏见”可能会导致不准确甚至有害的临床建议[19]。还有研究表明，即使是简单的问题也可以将低风险软件转换为严重的安全威胁，从而导致高风险的产品召回[20]。因此，美国医学会（AMA）敦促美国 FDA，在医疗 AI 应用的审批过程中，应该强调偏见是机器学习的一个重大风险[21]。

此外，医疗 AI 系统中的偏见不但对个体有影响，还可能削减弱势群体的整体福祉，进而影响

① deep fake 是英文“deep learning”（深度学习）和“fake”（伪造）的混成词，指基于人工智能的人体图像合成技术的应用，该技术可将已有的图像或影片叠加至目标图像或影片上，目前被大量用于制作色情影片和虚假新闻视频。

社会公平。美国的一项研究表明，在一个用于确定谁可以正常访问高风险医疗管理程序的软件程序中，由于处于特定健康水平的黑人最终产生的成本低于白人，故会让健康的白人先于健康状况较差的黑人进入程序[22]。

值得注意的是，医疗 AI 应用还面临着与其他行业不同的挑战。在临床中，许多事故通常不会被报告，而且其影响难以衡量[23]，这意味着一旦纳入临床实践，智能决策支持系统的正面和潜在负面影响可能很难确定。并且，在其他行业的 AI 应用系统中，由于软件可以方便地更新升级，即使系统出现错误，也可以迅速地进行自我纠正。然而，针对患者健康和福祉开发的医疗 AI 系统，在安全性方面几乎没有试错的余地，这使得在最开始进行正确设计和投入使用前的审慎评估显得尤为重要。

1.3 因系统漏洞或缺陷引起的信息安全风险

由于医疗 AI 产品需要通过软件以及网络来实现其功能，其不可避免地会产生信息安全风险。研究表明，胰岛素泵、心脏起搏器以及达芬奇外科手术机器人均被证明已经存在可能危及人生命健康的信息安全漏洞[24]。对于 AI 系统，常见网络威胁包括但不限于框架漏洞攻击和数据污染，前者是指利用算法所用现成框架本身漏洞进行网络攻击，后者则是指通过污染输入数据进行网络攻击，如对抗样本攻击①。

医疗器械网络安全出现问题不仅会侵犯患者隐私，而且可能会产生医疗器械非预期运行的风险，导致患者或使用者受到伤害或死亡。多个安全研究机构的报告表明，医疗机构是勒索软件等黑客攻击针对的首要目标[25]，破坏性关闭关键设备或篡改患者数据可能会危及生命[26]，这些情况要求我们必须关注医疗 AI 应用在网络安全方面的可靠性。

2 患者和受试者安全的伦理问题及保障机制

伴随着医疗 AI 市场规模的不断扩大，社会资本大量涌入医疗 AI 领域，因追求利益而导致患者或受试者伤害的风险值得警惕。为了开发人类基因编辑等违反伦理的医疗应用、为了获取训练数据进行伤害性的实验、为了通过上市审批而在申报材料中弄虚作假等情形均有出现的可能。因此，必须将医疗 AI 产品研发的全过程纳入到医疗器械管理的制度中，以此构建一个符合伦理的医疗 AI 应用开发流程。

生命医学伦理学的基本原则包括不伤害原则、尊重与自主原则、有利原则和公正原则[27]。而人工智能应当遵循遵守的伦理尽管原则上无一致的共识，但学术界和相关企业已经有了一些初步实践。被提及较多的有阿西洛马人工智能原则②、Google 的人工智能七原则③，以及近期欧盟委员会发布的《可信人工智能伦理指南》[12]，这些准则可以为我们构建符合伦理的医疗 AI 提供指引，前提是正确认识医疗领域与 AI 其他应用领域的区别。

已有研究指出，高新技术的发展可能带来各种不同的医患伦理冲突，如设备依赖导致的医患关系固化、技术主导导致人员关怀弱化，“医生-机器-病人”的生冷关系，妨碍了医患之间的思想交流和情感沟通[28]。为了使医疗 AI 向符合伦理的方向发展，AMA 向 FDA 对促进医疗 AI 提出了五点建议，分别是：1）按照以用户为中心的设计最佳实践进行设计和评估；2）保证透明；3）符合

① 例如，对于一张熊猫的图片，增加人为设计的微小噪声之后，人眼对扰动前后两张图片基本看不出区别，而人工智能模型却会以 99.3%的概率将其错判为长臂猿。参见 GOODFELLOW I, SHLENS J, CHRISTIAN S.Explaining and harnessing adversarial examples[EB/OL]. (2015-03-20) [2018-06-23]. https://arxiv.org/abs/1412.6572.

② 参见 https://futureoflife.org/ai-principles/.

③ 参见 https://blog.google/topics/ai/ai-principles/.

重复性的领先标准；4）识别并采取措施解决偏见，避免引入或加剧医疗保健差异，尤其是在弱势人群上测试或部署新的人工智能工具时；5）保护患者的隐私[21]。

根据上述已经被广为接受的与医疗 AI 相关的伦理原则和建议，结合目前医疗 AI 发展和应用的现状，本文认为，以下三方面伦理保障机制应当被着重强调。

2.1　以患者和受试者为中心

美国“以患者为中心医疗研究所”（IPFCC）将以患者为中心的要求概括为四个核心理念，即尊严和尊重、信息共享、患者参与、合作[29]。医学伦理学的第一原则是不伤害，无论在涉及人类受试者的医学研究中还是在医疗 AI 研发和应用的过程中，对患者和受试者健康的考虑都应优先于科学和社会的需要。临床医生必须完全相信算法的准确性、可靠性和客观性，才能使 AI 成为常规临床应用的一部分。但是，有时用于训练算法的数据有偏差，或者算法有潜在的缺陷，这些都将加剧决策失误的风险。同时，如果临床医生不知道算法是如何得出结果的，他们将无法知道算法是否有偏差，也使得医生在做出临床决策时很难完全信任该技术。在英特尔进行的一份调查中，超过三分之一的受访者表示，患者不信任 AI 在医疗保健中发挥积极作用，而 30%的人认为临床医生也不信任 AI[30]。因此，医疗 AI 的临床应用应当积极征得患者的同意。

以患者为中心的另一要求是医护人员与患者完整地共享信息，这也是知情同意原则的要求。在医疗过程中，用于为治疗提供信息的数据质量对于建立患者对其提供者的信任至关重要。患者应当确保用于最终决策的数据是准确的，并且相信这些数据只会以合理的方式使用。然而，临床医生可能本身也未必明确 AI 应用的工作原理，这可能构成对传统患者知情权的挑战。临床医生应当尽可能充分理解算法的创建方式，严格评估用于创建旨在预测结果的统计模型的数据来源，了解模型如何发挥作用并防止过度依赖这些模型。

2.2　创设合理的人机交互

AI 参与到对患者治疗的“决策”过程是医疗 AI 最大的伦理风险来源。当人们与执行自动化任务的机器进行交互时，即使人们知道或应该知道自动化是错误的，他们仍然依赖自动化的结果，这种现象称为“自动化偏差”。临床决策支持（CDS）可以通过警告潜在错误来提高安全性，但有时也会成为新的风险来源。当用户（医护人员）过度依赖 CDS 时就会出现自动化偏差，从而降低对信息搜索和处理的警惕[31]。将系统应用于混乱的现实世界临床实践时，其性能无疑将受到影响。因此，操作人员必须知道何时信任该系统、何时不信任该系统。如何设计人机交互，对于防止引入新的偏见和错误至关重要。

风险稳态（Risk homeostasis）理论指出，人类冒险行为与所感知的危险程度紧密相关。减少活动的感觉风险，人们会更加大胆[32]。临床医疗实践的过度自动化可能会导致医护人员过于自信，从而增加错误和事故的发生。例如，对英国一家重症监护病房的护士进行的一项研究发现了临床中存在风险波动的证据。该病房在配药过程中实施的安全措施涉及不同同事在给病人配药前进行多次交叉检查，虽然护士接受了双重检查的培训，但安全措施降低了感知的风险水平，故护士认为出错的可能性较小[33]。AMA 认为，FDA 应当考虑与机器学习相关的风险，例如是否使用了完全自动化的系统而不需要人类干预。在这一点上，欧盟委员会伦理小组也有类似的看法，并要求在 AI 决策过程当中必须有“有意义的人类控制”（Meaningful Human Control）[34]。

因此，为了实现这种良好的人机交互，医疗 AI 的生产者应当在项目设计之初就考虑该应用面向的使用者情况，做出不同的设计。目前对 AI 错误决策的所有纠正都依赖于人的介入，但如果

AI 产品计划应用在医疗水平较低的地区，需要考虑到基层医院等机构可能没有经验丰富的临床医生来发现并纠正这种错误。

2.3　保证持续评价

“ 黑箱 ”问题是人工智能应用被关注最多的焦点之一，许多反对者以 AI 决策缺乏可解释性为由拒绝 AI 在医疗领域的应用。然而，有许多普通药品尚无已知的作用机理，但医生依然开处方指示患者使用，各国的监管机构通过严格的实验和审批流程来保证这些药品的安全性和有效性。因此，缺乏可解释性不是阻止医疗 AI 进入临床应用的理由，问题在于能否保证充足和持续的监管。

尽管医疗 AI 的开发者往往用各种努力来创造一个理想的应用，但是仍然必须监控现实生活中发生的事情，以防止“ 偏见 ”危害患者的健康和其他权益。因此，运营者需要监控数据本身是否表现出不平等或其他异常，还应当与医疗保健提供者、患者和管理人员进行交谈，以确定他们是否发现任何公平问题。“ 收集的有关患者健康，诊断和结果的数据已成为医疗系统收集的公开文献和信息的‘ 集体知识 ’的一部分，并且可能会在不考虑临床经验和患者护理的人为因素的情况下使用。”[35]另外一个可考虑的做法是，在设计最终将影响他们生活的算法时，让患者参与其中[36]。

3　医疗人工智能安全的法律保障机制

伦理保障机制依靠个体的自觉遵守来发挥作用，针对患者与受试者的安全这一重要利益的保障，还需要依靠刚性的法律制度来划定边界和管控相关风险，确保医疗 AI 应用安全、有序且可控地进行。

3.1　我国医疗 AI 监管的法律框架

目前我国针对医疗 AI 监管的基本思路，是将医疗 AI 作为医疗器械的一种类型进行监管，相关规范可分为三个层级（表 1）：(1) 关于医疗器械的管理相关的规范，即《医疗器械监督管理条例》及相关配套制度；(2) 针对软件类医疗器械的专门规范，包括《医疗器械软件注册技术审查指导原则》、《医疗器械网络安全注册技术审查指导原则》、《移动医疗器械注册技术审查指导原则》等；(3) 专门针对以人工智能（深度学习）辅助技术的规范性文件，如国家药品监督管理局医疗器械技术审评中心（CMDE）在 2019 年发布的《深度学习辅助决策医疗器械软件审评要点》。需要说明的是，我国还有一些与医疗 AI 相关的规范性文件，包括《人工智能辅助诊断技术管理规范》、《人工智能辅助诊断技术临床应用质量控制指标》，这两份文件尽管也使用了“ 人工智能 ”这一表述，但二者针对的对象实际是机器人手术系统，不属于本文讨论的医疗 AI 的范畴。

表 1　我国医疗人工智能监管规范

类别	名称	发布机构	效力级别
医疗器械管理法规	《医疗器械监督管理条例》	国务院	法规
	《医疗器械分类规则》	国家食品药品监督管理总局	部门规章
软件类医疗器械管理规范	《医疗器械软件注册技术审查指导原则》	CMDE	规范性文件
	《医疗器械网络安全注册技术审查指导原则》	CMDE	规范性文件
	《移动医疗器械注册技术审查指导原则》	CMDE	规范性文件
人工智能（深度学习）辅助技术的规范	《深度学习辅助决策医疗器械软件审评要点》	CMDE	规范性文件

我国的医疗器械管理制度，制定了医疗器械的分类规则和分类目录，将医疗器械按照风险程度分成一二三类实行分类管理，根据医疗器械生产、经营和使用情况，及时对医疗器械的风险变化进行分析和评价，并对分类目录进行调整。对第一类医疗器械实行产品备案管理，对第二类、第三类医疗器械实行产品注册管理，并设置了较为严格的管控机制与程序。根据现行的 2017 年版

《医疗器械分类目录》子目录 21，医疗 AI 归属于医用软件类别的医疗器械，这类医疗器械包括治疗计划软件、影像处理软件、数据处理软件、决策支持软件、体外诊断类软件和其他类共六类。该目录同时明确，医疗信息管理软件如果仅仅是医院管理工具，管理内容是患者信息等非医疗诊断和/或治疗内容，不按照医疗器械管理[37]。如果医疗信息管理软件包含患者诊断、治疗数据和影像，则按照软件处理对象（影像、数据）的不同，分别归为影像处理软件或数据处理软件。此外，《医疗器械分类目录》将仅提供诊断建议的辅助诊断软件归为第二类医疗器械，而对病变部位自动识别的并提供明确诊断提示的软件按照第三类医疗器械管理。

根据《医疗器械监督管理条例》对第二类和第三类医疗器械注册的规定，注册第二类和第三类医疗器械，申请材料中应当包含临床试验报告。这些临床实验要遵守《涉及人的生物医学研究伦理审查办法》中的六个伦理审查标准，即知情同意、控制风险、免费和补偿、保护隐私、依法赔偿和特殊保护，并通过伦理委员会的审查。而在医用软件类别的医疗器械方面，我国还制定了《医疗器械软件注册技术审查指导原则》，对其审查进行了较为细致的规范。此外，我国制定了专门针对现阶段医疗 AI 审批的规范性文件《深度学习辅助决策医疗器械软件审评要点》，基于深度学习技术特点，结合软件的预期用途、使用场景和核心功能，重点关注软件的数据质量控制、算法泛化能力和临床使用风险。

在医疗器械上市之后，目前的监管体系还建立了不良事件监测与召回制度。根据《医疗器械注册管理条例》的规定，要求医疗器械的生产、经营企业开展不良事件监测，就不良事件或可疑不良事件，向医疗器械不良事件监测技术机构报告。此外，如医疗器械生产企业发现其生产的医疗器械不符合强制性标准、经注册或备案的产品技术要求或者存在其他缺陷的，应当停止生产并召回上市产品。

总体来说，我国对于医疗 AI 监管的法律框架已经基本具备。

3.2　我国医疗 AI 安全保障法律机制的待改进之处

3.2.1　明确医疗 AI 的范围

尽管我国已经在《医疗器械分类目录》中对医疗 AI 的范围和分类进行了基本界定，然而“辅助诊断”和“明确诊断提示”的边界并不清晰，在实际注册和评审过程当中难以准确判断。相比较而言，美国 FDA 在《21 世纪治愈法》中，采取反面排除的规定，将与疾病诊断、治愈、缓解、预防和治疗无关的软件排除出了“医疗器械”的范围[38]。之后，美国 FDA 又发布了《临床决策软件指南草案》[39]，其中规定，需要首先判断软件是否：(a) 获取、处理分析来自体外的诊断信号；(b) 旨在显示、分析或打印患者医疗信息或其他医疗信息；(c) 旨在支持医护人员或向医护人员提供有关预防、诊断或治疗疾病或病症的建议。在满足上述 (a) (b) (c) 三项条件的前提下，FDA 再根据 (d) 项条件判断使用者是否依赖软件。在医疗器械的范围内，FDA 表示对危险系数较低的临床决策支持软件并不进行监管，例如执行常规临床实践计算的临床决策支持软件。而对危险系数较大的则会重点监管，例如通过分析睡眠呼吸暂停监测仪监测到的呼吸模式来诊断睡眠呼吸暂停的软件，以及用于诊断脑血肿的和分析近红外照相机信号的软件等[14]。通过综合运用正反面的标准，FDA 将对患者健康影响最大的部分纳入监管范围，这种灵活定义既减轻了监管压力，又给予了医疗器械充足的创新空间，值得借鉴。

此外，我国相关制度中未明确必须作为医疗器械注册的医疗 AI 的范围。由于并非所有的医疗 AI 应用都属于医疗器械，对于属于医疗器械的产品，自然可以适用现行的医疗器械管理制度，并为其进一步构建详细注册标准。然而，由于《医

疗器械注册管理办法》和《医疗器械分类规则》中未明确医疗 AI 应用作为医疗器械的范围，部分性质不明的医疗 AI 应用，如医院管理软件或移动医疗健康软件可能落入监管盲区。其中，对于以移动设备 app 形式提供的医疗 AI 应用，美国 FDA 的做法可供参考。在 FDA 发布的《移动医疗应用指南》中，FDA 认为，如果某个移动医疗应用被视为一种“已受监督的医疗设备”的附件，或者将移动平台转变为“已受监督的医疗设备”，那么这个应用就会受到 FDA 的监管[40]。

另外，还有一部分医疗 AI 的应用，如 AI 制药应用，AI 主要在药物设计环节发挥作用，即通过 AI 缩短设计、合成和实验的周期，药品的临床试验和上市仍然需要按照《药品管理法》审评注册，这种 AI 几乎不会对患者的权利产生任何影响，因此暂时没有必要纳入监管的范围。

3.2.2 针对医疗 **AI** 的特殊性完善其注册和后续监管的制度

1）伦理审查

医疗 AI 与传统的医疗器械不同，很多医疗 AI 的研发企业是信息技术企业，并不具备与医疗器械伦理审查相关的资质与条件。2017 年 10 月，中共中央办公厅和国务院办公厅印发了《关于深化审评审批制度改革鼓励药品医疗器械创新的意见》，明确要完善伦理委员会机制，提高伦理审查效率，其中一项重点内容是，各地可根据需要设立区域伦理委员会，指导临床试验机构的伦理审查工作，可接受不具备伦理审查条件的机构或注册申请人委托对临床试验方案进行伦理审查。目前，上海、山东、广东和深圳等省市已经成立了区域伦理委员会。

区域伦理委员会的设立为不具备伦理审查条件的人工智能企业提供了进行伦理审查的渠道，但还存在一些问题。一方面，目前的伦理审查制度是从受试者角度出发的，关注的主要是受试者的权益和研究人员的资格等，在医疗 AI 产品的注册过程中，并未对医疗 AI 产品本身所实现的效果进行审查。另一方面，区域伦理委员会中缺乏人工智能领域的专业学者，例如上海市临床研究伦理委员会的 46 位委员和专家中，并无人工智能行业的相关从业者，这使得区域伦理委员会可能对医疗 AI 的审查缺乏专业性。

2）数据算法评价制度

对数据和算法的审查评价是保证医疗 AI 应用输出结果准确和有效的最关键环节。《深度学习辅助决策医疗器械软件审评要点》要求着重考虑相关算法的特点，将软件的数据质量控制、算法泛化能力和临床使用风险全面纳入审评范围，并保持持续监管。

需要注意的是，审评要点采用的是针对软件的评审思路与流程，基于风险的全生命周期管理方法考虑软件技术审评要求，包括需求分析、数据收集、算法设计、验证与确认以及软件更新等内容，涵盖算法性能评估、临床评价以及网络与数据安全等要求。然而，对于医疗 AI 与医疗本身的结合程度，以及相关伦理原则和价值观在算法中的体现等问题，还有待改进的空间。

3）网络安全评价制度

医疗器械网络安全是医疗器械安全性和有效性的重要组成部分之一[41]。2019 年网信办安全中心在测评中发现，我国目前医疗器械存在较多安全隐患且无法得到及时修复；厂商在设计时没有考虑相应的安全性问题，而安全性提高需要得到医院等医疗机构的配合。我国国家药品监督管理局医疗器械技术评审中心在 2017 年发布了《医疗器械网络安全注册技术审查指导原则》。国家药品监督管理局于 2018 年 10 月 18 日发布了新版《医疗器械网络安全管理上市前申报指南》草案征求公众意见。相较于 2014 年版的指南，修订后的指南草案纳入了新建议：将医疗器械网络安全风险层级分为“高网络安全风险”和“标准网络安全风险”，并引入了网络安全物料清单（CBOM）的概

念[42]。另外，在 2019 年的测评中，网信办安全中心捕获了大量的明文健康数据，这表明大部分医疗器械制造商均未考虑健康数据传输过程中的保密性问题，也反映出国内医疗器械厂商的信息安全意识不足。

因此，医疗 AI 产品的信息安全性受两方面因素影响，一是应用厂商设计时是否充分考虑到系统信息安全性，二是医疗机构使用时是否充分维护网络安全。为此，我国需要从应用上市前的审批和投入运行后的监管两方面强化来确保医疗 AI 应用的信息安全。

3.3　医疗 AI 应用的上市后监管

医疗 AI 应用上市后的监管涉及三个问题：一是不良事件监测与召回的问题；二是软件更新的问题；三是设备培训问题。

在不良事件监测与召回方面，鉴于医疗 AI 目前的定位是辅助工具，医疗 AI 导致的不良事件是不容易从实际病例中被分离和发现的，监管机构需要进一步研究和确定其不良事件监测的方法和判定标准。

在软件更新方面，《深度学习辅助决策医疗器械软件审评要点》明确软件版本命名规则应涵盖算法驱动型和数据驱动型软件更新，并应列举重大软件更新的全部典型情况。轻微数据驱动型软件更新可通过质量管理体系控制，无需申请注册变更。但是这一界定是否会对患者的安全有实质性的影响，还需要再观察一段时间才能确定。

在设备培训方面，医疗AI目前还是辅助工具，如何正确地使用医疗AI工具是事关患者安全的重要因素，需要在立法规范的层面予以关注，并加大其强制性监管力度。

4　结语

医疗 AI 与软件医疗器械的最大区别在于，AI 涉及从现有数据中学习并优化算法。对医疗 AI 监管的目标，是在保证安全的基础上构建患者和临床医生对医疗 AI 的信任，从而使医疗 AI 能够更好地为改善社会医疗保健提供支撑。伦理的规则在其中非常重要，因为法律不能解决所有问题。正如风险社会理论的创立者贝克所说，“ 法律制度的价值和意义就在于规范和追寻技术上的可以管理的，哪怕是可能性很小或影响范围很小的风险和灾难的每一个细节 ”[43]。在医疗 AI 的监管过程中，法律制度的主要目的是帮助伦理的规范并保障伦理审查程序的实施，确保每一个患者与受试者的安全、健康和最大利益。

参考文献

[1] 国务院关于印发新一代人工智能发展规划的通知（国发〔2017〕35 号）[EB/OL]. (2017-07-08) [2020-03-20] http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.

[2] Accnture. Artificial Intelligence: Healthcare’s New Nervous System [EB/OL]. ([2020-03-20] https://www.accenture.com/fi-en/insight-artificial-intelligence-healthcare.

[3] Topol E J. High-performance medicine: the convergence of human and artificial intelligence[J]. Nat Med, 2019, 25(1): 44-56.

[4] 国家药品监督管理局. 冠脉血流储备分数计算软件产品获批上市[EB/OL]. (2020-01-15) [2020-03-20] http://www.nmpa.gov.cn/WS04/CL2056/373503.html.

[5] 中华人民共和国基本医疗卫生与健康促进法[EB/OL]. (2019-12-19) [2020-03-20] http://www.gov.cn/xinwen/2019-12/29/content_5464861.htm.

[6] 王海星, 田雪晴, 游　茂. 人工智能在医疗领域应用现状、问题及建议[J]. 卫生软科学, 2018, 32(5): 3-5, 9.

[7] 包桉冰, 徐　佩. 医疗人工智能的伦理风险及应对策略[J]. 医学与哲学(A), 2018, 39(6): 37-40.

[8] 周吉银, 刘　丹, 曾圣雅. 人工智能在医疗领域中应用的挑战与对策[J]. 中国医学伦理学, 2019, 32(3): 281-286.

[9] 曹艳林, 王将军. 人工智能医疗应用及面临的主要法律问题[J]. 网络信息法学研究, 2019, (1): 85-103, 334-335.

[10] 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.

[11] 孔　鸣, 何前锋, 李兰娟. 人工智能辅助诊疗发展现状与战略研究[J]. 中国工程科学, 2018, 20(2): 86-91.

[12] Commission H-LEGOAISUBTE. Ethics guidelines for trustworthy AI[EB/OL]. ([2020-03-20] https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419.

[13] 焦李成, 杨淑媛, 刘 芳. 神经网络七十年: 回顾与展望[J]. 计算机学报, 2016, 39(8): 1697-1716.

[14] 孟春婷, 李玲旭, 刑 冲. 医疗人工智能产业监管及法律研究[R]. 北京: 腾讯法务, 2019.

[15] 上海交通大学人工智能研究院, 上海市卫生和健康发展中心, 上海交通大学医学院. 中国人工智能医疗白皮书[EB/OL]. (2019-01-09) [2020-03-20] http://www.cbdio. com/image/site2/20190306/f42853157e261de9c8ff57.pdf.

[16] Cristea I A, Cahan E M, Ioannidis J P A. Stealth research: Lack of peer-reviewed evidence from healthcare unicorns[J]. Eur J Clin Invest, 2019, 49(4): e13072.

[17] 张 立. 人工智能是福是祸？——专访联合国人工智能和机器人中心高级战略顾问贝里泽[EB/OL]. ([2020-03-20] https://news.un.org/zh/story/2018/07/1013 231.

[18] Char D S, Shah N H, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges[J]. N Engl J Med, 2018, 378(11): 981-983.

[19] Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine[J]. N Engl J Med, 2019, 380(14): 1347-1358.

[20] Ronquillo J G, Zuckerman D M. Software-Related Recalls of Health Information Technology and Other Medical Devices: Implications for FDA Regulation of Digital Health[J]. Milbank Q, 2017, 95(3): 535-553.

[21] AMA. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD), Discussion Paper and Request for Feedback[EB/OL]. (2019-06-03) [2020-03-20] https://searchlf.ama-assn.org/undefined/documentDownload?uri=%2Funstructured%2Fbinary%2Fletter%2FLETTERS%2F2019-6-3-Letter-to-Sharpless-re-FDA-AI-MIL.pdf.

[22] 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.

[23] Pear R. Report Finds Most Errors at Hospitals Go Unreported[EB/OL]. ([2020-03-20] https://www.nytimes.com/2012/01/06/health/study-of-medicare-patients-finds-most-hospital-errors-unreported.html.

[24] Arxiv E T F T. Security Experts Hack Teleoperated Surgical Robot[EB/OL]. (2015-04-25) [2020-03-20] https: // www.technologyreview.com/s/537001/security-experts-hack-teleoperated-surgical-robot/.

[25] Fuentes M R. Cybercrime and other threats faced by the healthcare industry[J]. Trend Micro, 2017.

[26] Team MRFF-LTRF. Cybercrime and Other Threats Faced by the Healthcare Industry[EB/OL]. ([2020-03-20] https://documents.trendmicro.com/assets/wp/wp-cybercrime-and-other-threats-faced-by-the-healthcare-industry.pdf.

[27] Beauchamp T L, CHILDRESS J F. Principles of biomedical ethics[M]. USA: Oxford University Press, 2001.

[28] 杨国斌. 现代医学伦理学面临的新挑战[J]. 医学研究生学报, 2012, 25(2): 113-118.

[29] Care IFPAFC. Patient- and family-centered care is working “with” patients and families, rather than just doing “to” or “for” them[EB/OL]. ([2020-03-20] https://www.ipfcc. org/about/pfcc.html.

[30] Esposito J. Commentary: Artificial Intelligence in Health Care Is Just What the Doctor Ordered[EB/OL]. ([2020-03-20] https://www.usnews.com/news/healthcare-of-tomorrow/articles/2018-07-06/commentary-artificial-intelligence-in-health-care-is-just-what-the-doctor-ordered.

[31] Lyell D, Magrabi F, Raban M Z. Automation bias in electronic prescribing[J]. BMC Med Inform Decis Mak, 2017, 17(1): 28.

[32] Wilde G J. Risk homeostasis theory: an overview[J]. Injury prevention, 1998, 4(2): 89-91.

[33] Sanghera I S, Franklin B D, Dhillon S. The attitudes and beliefs of healthcare professionals on the causes and reporting of medication errors in a UK Intensive care unit[J]. Anaesthesia, 2007, 62(1): 53-61.

[34] Technologies EGOEISAN. Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems[EB/OL]. (9 March 2018) [2020-03-20] http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf.

[35] Terry N P. Appification, AI, and healthcare’s new iron triangle [J]. J. Health Care L. & Pol’y, 2017, 20: 117.

[36] Rajkomar A, Hardt M, Howell M D. Ensuring fairness in machine learning to advance health equity[J]. Annals of internal medicine, 2018, 169(12): 866-872.

[37] 食品药品监管总局. 总局关于发布医疗器械分类目录的公告（2017 年第 104 号） [M].

[38] 21st Century Cures Act. 114 - 255 Pub. L. 114-255[R], Section 3060.

[39] FDA. Clinical Decision Support Software Draft Guidance for Industry and Food and Drug Administration Staff[EB/OL]. ([2020-03-20] Clinical Decision Support Software Draft Guidance for Industry and Food and Drug Administration Staff.

[40] FDA. Policy for Device Software Functions and Mobile Medical Applications Guidance for Industry and Food and Drug Administration Staff[EB/OL]. (September 27, 2019)

[2020-03-20] https://www.fda.gov/media/80958/ download.

[41] 国家药品监督管理局医疗器械技术审评中心. 医疗器械网络安全注册技术审查指导原则[EB/OL]. (2017-03-28) [2020-03-20] https://www.cmde.org.cn/ directory/web/WS01/images/0r3Bxsb30LXN+MLnsLLIq9eisuG8vMr1yfOy6da4tbzUrdTyo6gyMDE3xOq12jEzusWjqS5kb2N4. docx.

[42] FDA. Content of Premarket Submissions for Management of Cybersecurity in Medical Devices Draft Guidance for Industry and Food and Drug Administration Staff[EB/OL]. (October 18, 2018.) [2020-03-20] https://www.fda.gov/media/119933/download.

[43] 乌尔里希·贝克, 王武龙. 从工业社会到风险社会(上篇) ——关于人类生存、社会结构和生态启蒙等问题的思考[J]. 马克思主义与现实, 2003, (3): 39.

Ethical and Legal Protection Mechanisms for Security Risks of Medical Artificial Intelligence

Wang Yue✉, Dai Haiyang

(*School of Law*, *Xi'an Jiaotong University*, *Xi'an* 710049, *China*)

Abstract: Artificial intelligence (AI) is rapidly developing and being applied in the medical field, bringing new opportunities and hopes to medical professionals and patients. However, as the objects of medical AI are humans, an AI application can pose a direct threat to the health of patients and human subjects; moreover, the potential security risks from data and algorithm defects, system flaws, or defects are particularly prominent. Therefore, in addition to the widely accepted ethical principles for AI, medical AI should also emphasize an ethical guarantee mechanism centered on patients and subjects, create reasonable human-computer interactions, and maintain continuous evaluation. In addition, the current medical AI should be regulated under a legal regulatory framework to further clearly define the scope of medical AI, aiming at the particularity of medical AI for improving registration and follow-up supervision systems. This will help ensure safety, based on building the trust of patients and clinicians in regard to medical AI. In such cases, medical AI can improve social health care conditions, and protect the security and welfare of the largest number of patients and human subjects.

Key Words: medical artificial intelligence; medical ethics; medical damage; medical data