# 人工智能医疗器械中的伦理问题

唐桥虹，王浩，任海萍*（中国食品药品检定研究院，北京 100050）

**摘要　目的：**梳理人工智能医疗器械的伦理现状与面临的问题，探讨医疗器械领域可遵循的人工智能伦理准则和发展原则。**方法：**查阅国内外已发布的人工智能伦理准则，国内外监管、医学伦理相关规范要求文件。**结果与结论：**人工智能医疗器械在发展与应用过程中仍面临着社会影响、个人数据保护、人工智能算法和医学伦理等问题。人工智能医疗器械在沿着推动医学人工智能伦理标准化、保持数据完整性方向发展的同时，还要保障数据隐私。

**关键词：** 人工智能；伦理；医疗器械

**中图分类号：**TP181；TH77　**文献标识码：**A　**文章编号：**1002-7777(2019)09-1004-05
doi:10.16153/j.1002-7777.2019.09.007

## Ethical Issues in Artificial Intelligence Medical Devices

Tang Qiaohong, Wang Hao, Ren Haiping* (National Institutes for Food and Drug Control, Beijing 100050, China)

**Abstract　Objective:** To summarize the current ethical status and problems of artificial intelligence (AI) medical devices, and discuss the ethical guidelines and development principles to be followed in the field of artificial intelligence medical devices. **Methods:** Published ethical guidelines of AI at home and abroad, relevant regulations and requirements of medical ethics were retrieved and reviewed. **Results and Conclusion:** There are some problems in aspects such as social influence, personal data protection, artificial intelligence algorithm and medical ethics during the development and application of AI medical devices. AI medical devices should protect data privacy during the development in the direction of promoting standardization of medical AI ethics and maintaining data integrity.

**Keywords:** artificial intelligence; ethical issues; medical devices

人工智能（Artificial Intelligence，缩写为AI）一词，在1956年的达特茅斯会议上被首次提出。人工智能作为新一轮产业变革的核心驱动力，已发展上升为国家战略，在自动驾驶、医疗、工业机器人以及教育、金融、互联网服务等领域得到越来越多的应用。在医疗领域的应用主要包括辅助诊断、辅助手术、临床辅助决策、患者信息管理等，对应的人工智能医疗器械产品主要包括独立的医疗软件、AI赋能医疗设备、医疗信息化系统（云医疗）几大类。随着以深度学习为代表的人工智能技术不断发展，我们在积极拥抱人工智能的同时，需要思考人工智能医疗器械在发展与使用过程中面临的伦理问题，充分认识人工智能医疗器械在数据获取、隐私保护等方面带来的影响。

基金项目：中国食品药品检定研究院中青年发展研究基金课题“人工智能医疗器械软件性能评价方法研究”（编号 2018C5）
作者简介：唐桥虹；研究方向：生物医学工程、人工智能、有源医疗器械检定
通信作者：任海萍；研究方向：生物医学工程、医疗器械检定；E-mail：renhaiping@nifdc.org.cn

人工智能本身是技术而不是产品，医疗工作者使用的是被医疗人工智能赋能后的设备或者信息化系统，而不是使用人工智能技术本身。由人工智能发展所带来的伦理问题，在一定程度上可由设计开发者通过遵循一定的原则而规避，因此，在我国推动人工智能发展的关键时期，探讨人工智能发展应遵循的伦理准则，对人工智能的发展有着极为重要的意义。

## 1 人工智能伦理准则及现状

### 1.1 现阶段人工智能伦理共识

随着人工智能伦理的发展，目前国内外主要达成了两个影响较为广泛的人工智能伦理共识，一个是“阿西洛马人工智能原则”（Asilomar AI Principles），一个是国际电气电子工程师学会（IEEE）组织倡议的人工智能设计的伦理准则[1]。

“阿西洛马人工智能原则”于2017年1月初在美国加利福尼亚州阿西洛马举行的Beneficial AI会议上被提出，阿西洛马人工智能原则是著名的阿西莫夫机器人学三定律的扩展版本。阿西洛马人工智能原则目前共23项，分为三大类，分别为科研问题（Research Issues）、伦理和价值（Ethics and Values）、更长期的问题（Longer-term Issues）[2]。其中涉及伦理方面的共13项，主要为1）安全性；2）故障透明性；3）司法透明性；4）责任；5）价值归属；6）人类价值观；7）个人隐私；8）自由和隐私；9）分享利益；10）共同繁荣；11）人类控制；12）非颠覆；13）人工智能军备竞赛。

“阿西洛马人工智能原则”可以理解为人工智能不能单纯地为了利益而创造，而应该为了在确保人类不被替代的情况下通过自动化实现人类繁荣。保持一个尊重隐私但开放、合作的人工智能研究文化也是一个优先考虑的问题，以确保研究人员和政策制定者在彼此交换信息的同时，不会用危害人类的手段与对手竞争。

国际电气电子工程师学会（IEEE）最早于2016年提出了“关于自主/智能系统伦理的全球倡议”，并于2016年12月和2017年12月在全球范围内先后发布了第一版和第二版的“人工智能设计的伦理准则”白皮书（"Ethically Aligned Design"）。该白皮书来自于IEEE自主与智能系统伦理全球倡议项目，在当前版本的《人工智能设计的伦理准则》（第2版）中，白皮书提出了一些相关的议题和建议，希望能够促进符合这些原则的国家政策和全球政策的制定。该伦理准则提出了5个核心应遵循原则[3]：1）人权：确保它们不侵犯国际公认的人权；2）福祉：在它们的设计和使用中优先考虑人类福祉的指标；3）问责：确保它们的设计者和操作者负责任且可问责；4）透明：确保它们以透明的方式运行；5）慎用：将滥用的风险降到最低。该人工智能伦理准则的发布旨在为IEEE正在推动的11个与人工智能伦理相关的标准制定提供建议。

上述两项接受较为广泛的伦理共识，由来自人工智能/机器人研究领域的专家学者以及专业技术学会的研究人员讨论制定而成。同时，由于人工智能在产业发展中的战略性地位和应对人工智能伦理风险的迫切需要，许多国家政府机构、社会团体、产业界等也在制定适用于自身国情的人工智能伦理准则或指南，为人工智能相关企业提供风险把控、评估和应对的系统性指引。

### 1.2 国外人工智能伦理发展现状

为了有效应对AI带来的新机遇，欧盟委员会于2019年4月8日以“建立对以人为本AI的信任”为题，发布了欧洲版的AI伦理准则。该伦理准则提出了“可信任AI”应当满足的7项关键要点[4]，具体包括1）人的自主和监督；2）可靠性和安全性；3）隐私和数据治理；4）透明度；5）多样性、非歧视性和公平性；6）社会和环境福祉；7）可追责性。欧盟委员会在人工智能方面布局已久，早在2018年12月欧盟委员会人工智能高级专家组就发布了《关于可信赖人工智能的伦理准则（草案）》（Ethics Guidelines for Trustworthy AI），该草案为人工智能伦理提出了一个框架，给后续发布的欧洲版AI伦理准则奠定了基础[5]。

国际计算机协会（ACM）下属美国公共政策委员会于2017 年发布《算法透明性和可问责性声明》，提出了7项基本原则[6]：1）意识；2）获取和救济；3）责任制；4）可解释；5）数据来源保护；6）可审查性；7）验证和测试。该声明的重要部分要求开发人工智能的机构能够对算法的过程和特定的决策结果给予一定的解释，即人工智能算法的哪些输入特性会引起某个特定输出结果变化的可解释性。

### 1.3 我国人工智能伦理发展现状

我国也已将人工智能上升为国家战略，在法

律法规和政策体系进行了深入布局。国务院于2017年7月20日印发了《新一代人工智能发展规划》，在战略目标中对法律政策体系建设提出了三步走要求：到2020年，部分领域的人工智能伦理规范和政策法规初步建立；到2025年，初步建立人工智能法律法规、伦理规范和政策体系，形成人工智能安全评估和管控能力；到2030年，建成更加完善的人工智能法律法规、伦理规范和政策体系。规划要求围绕自动驾驶、服务机器人等应用基础较好的细分领域加快立法研究，重点解决人工智能下的民事与刑事责任确认、隐私和产权保护、信息安全利用、问责和追溯机制、潜在危险与评估等方面的法律问题[7]。2018年1月，国家人工智能标准化总体组成立，会上发布了《人工智能标准化白皮书（2018版）》，提出要依托于社会团体和公众对人工智能伦理进行深入思考和研究，并遵循一些共识原则：一是人类利益原则，即人工智能应以实现人类利益为终极目标；二是责任原则，在责任原则下，在技术开发方面应遵循透明度原则，在技术应用方面则应当遵循权责一致原则[1]。2019年4月，国家人工智能标准化总体组发布《人工智能伦理风险分析报告》，进一步明确了人类根本利益原则要从对社会的影响、人工智能算法、数据使用三个方面来考虑[8]。

## 2 人工智能在医疗器械领域的伦理风险

从整体来看，现阶段人工智能在医疗器械领域的应用主要是医学图像AI技术和AI技术赋能硬件，例如目前大量涌现的“智能读片”类AI医疗软件，利用深度学习在具有代表性的医学影像数据库中进行模型训练（多层神经网络）[9]，利用这些模型来解析图像、文本、声音，从而实现对医学图像病症的早期筛查；目前的AI技术赋能硬件，通常内嵌于各类医学影像设备，在设备前期拍摄及后期图像处理过程中实现图像分割、目标检测、图像分类、图像映射、图像配准等功能[10-12]。

上述人工智能还是实现特定功能的专用智能，并不能像人类智能那样拥有真正实现推理、思考和解决问题的能力，因此被叫做弱人工智能。与此对应的强人工智能则是达到类人水平的、能够自适应地应对外界环境挑战的、具有自我意识的人工智能[13]。强人工智能在哲学上存在巨大的争议（主要涉及思维与意识等根本问题的讨论），在技术上存在着极大的挑战，当前鲜有进展与应用，因此，考虑医疗领域人工智能的伦理风险，主要是以数据驱动类AI医疗器械产品为主，从社会影响、个人数据保护、人工智能算法和医学伦理四个方面来考虑其风险影响。

### 2.1 对社会的影响

从现有伦理准则与共识来看，医疗人工智能伦理的标准化工作仍处于起步阶段，行业内对医疗人工智能的内涵、应用模式还未达成准确共识，由此带来的行业内竞争可能会造成人工智能技术在医疗领域的滥用，例如不考虑医学实际情况，在医学成像、病灶识别、手术规划等临床领域盲目使用人工智能技术。造成的结果：一是资源的浪费，目前绝大多数人工智能辅助诊疗结果仍需医生确认操作，这种AI产品是否具有临床意义仍有待商榷；二是增加了医生对先进人工智能产品的依赖性，随着科技的发展，传统诊疗方法将逐步向高科技辅助诊疗转变，而这类辅助诊疗产品往往集中在科技发达、财富集中的国家或地区，这将造成各地区的医疗资源、医疗水平不平衡，如何让每个地区的人都从人工智能创造的福祉中收益，这是需要思考的问题。

### 2.2 个人数据保护风险

以深度学习+大数据为框架的医学人工智能系统需要大量的数据来训练学习算法。深度学习分为两个阶段：训练阶段和应用阶段。训练阶段的最大特点是数据驱动，一是需要大量的训练样本；二是所有样本需要明确的标注（金标准）；这就需要提高样本数据采集的效率和质量。目前，大多数人工智能产品的训练样本主要来自医院患者的各类医学影像数据，少部分来自于人类行为信息的数字化记录等。医学影像及患者行为信息涉及患者数据隐私的伦理问题，使用者在获取患者数据时必须要抹去个人敏感信息，只保留相关医学信息。

关于敏感信息的定义、识别与处理，国内现在还没有明确的标准，主要依赖于企业的自觉，可供企业参考的数据隐私保护准则主要是欧盟在2018年5月生效的《欧洲通用数据保护条例》（GDPR）和美国国会于1996年发布的《健康保险可携带性与责任法案》（HIPAA）[14]。GDPR条例对个人数据、医学健康相关数据给出了明确定义，并对如何在保护数据主体权益的情况下开展工作做

了详细说明[15]。HIPAA法案则既保护个人受保护的健康信息，又确保研究人员可以持续获得必要的医疗信息来进行研究，为达到这两点的平衡，HIPAA有3点重要规定[16]：1）对于不具备身份识别功能的健康信息，可以使用和披露给研究机构；2）在获得病人书面授权的情况下，可以因为研究目的而使用或披露病人受保护的健康信息；3）在某些特殊情况下，HIPAA也允许无须同意授权的信息共享，包括审查委员会或保密委员会批准的情况。

### 2.3 算法方面的风险

在深度学习算法的应用阶段，也面临着诸多风险。一是安全性，算法存在泄漏、参数被非预期修改的情况，且现阶段的深度学习算法是一个典型的“黑箱”算法，当算法被修改时，算法性能的降低或错误的发生将很难被察觉到，医疗领域与人身安全息息相关，这样的风险造成的后果将直接侵害人身权益。二是算法偏见风险，算法的复杂性和专业性，在现阶段很难解释清楚人工智能算法输入的某些特性是如何引起某个特定输出结果发生的[17]，算法的偏见可能是程序员主观认知的偏差，也有可能是输入数据的分布本身不具有整体代表性，同时如果算法在临床应用中通过本地数据进行无监督学习，也有加重这些偏见的风险。

### 2.4 医学伦理风险

对于医学人工智能产品的伦理思考，还应该纳入医学伦理范畴，考虑在医学伦理上如何进行患者隐私、数据保护等，这也是医疗类AI产品与一般AI产品在伦理方面的最大区别。在我国，获取医学临床数据、进行临床试验还必须获得医学伦理审批，医学方面的伦理监管主要依赖于伦理审查制度，而负责伦理审查的组织——伦理委员会则肩负着医学研究伦理审查、受试者权益保护、人的尊严维护等方面的重要职责[18]。原卫生部在1988年发布了《药品临床试验管理规范》[19]，于1995年出台了《卫生部临床药理基地管理指导原则》。随着生物医学科学技术与研究的飞速发展，在临床实践中遇到的伦理难题更为多样化，上述规范性文件已不能满足伦理审查的需求。原国家卫生和计划生育委员会在2016年发布了《涉及人的生物医学研究伦理审查办法》，该办法以涉及人的生物医学研究项目的伦理审查为重点，明确了医疗卫生伦理委员会的职责和任务，强化了对伦理委员会的监管[20]。2017年10月，中共中央办公厅、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》，提出要完善伦理委员会机制，提高伦理审查效率，其中重要一条是提出了区域伦理委员会职能，指导临床试验机构伦理审查工作。区域伦理委员会将肩负起解决多中心临床研究的伦理审查标准不一致、重复审查的问题，解决不具备伦理审查条件的机构直接发起的项目审查问题。

当前医疗人工智能产品在样本数据采集与临床试验阶段都有可能面临医学伦理审查，就需要相关企业熟悉医学伦理规范文件，在产品开发与应用阶段遵守基本医学伦理原则[21]：1）知情同意原则；2）控制风险原则；3）免费补偿原则；4）保护隐私原则；5）依法赔偿原则；6）特殊保护原则。

## 3 人工智能医疗器械应遵循的伦理发展方向

### 3.1 推动医学人工智能伦理标准化

为了促进行业的发展，开放、共享的高质量医学数据集是未来的发展趋势，但是数据获取的来源（前瞻式采集与回顾式采集交叉混合）[22]、数据清洗个人信息的准则/尺度[23]、数据标注的规范格式[24]等现在并无统一的定论，这会使得高质量的数据集难以在各个AI产品开发者间互通。只有相应的规则确定后，人工智能医疗器械产业才能高速发展。

### 3.2 数据完整性与医学伦理的平衡

在满足医学伦理要求保护隐私和个人数据的同时，医疗类AI产品为了获得高质量的数据集，还应尽可能地保留更多的信息（例如既往病史）用于分析处理，避免数据收集纳入偏见的、非典型的，甚至是错误的信息。与此同时，医疗类AI产品在数据收集、数据标注的过程中必须保证其数据的完整性，使得AI产品是可解释的、可信任的。如何最大程度地保留信息，同时避免通过信息追溯到个人，这是医疗类AI产品应遵循的伦理方向。

## 4 结语

在人工智能与伦理方面做深入的探索，围绕人工智能建立一系列标准，形成医疗健康人工智能标准生态，这关系到人工智能的发展。人工智能产业能否得到公众的信任并获得持续性发展，关键在于建立有共识的人工智能伦理准则，在较高程度保持数据完整性的同时，保障数据隐私，在符合

GDPR等要求的前提下，使得来自不同渠道的数据可用于机器学习，从而促进产业的发展。

**参考文献：**


[1] 中国电子技术标准化研究院. 人工智能标准化白皮书（2018版）[S]. 2018.

[2] Future of Life Institute（FLI）. Asilomar AI Principles[S]. 加利福尼亚：Beneficial AI 2017，2017.

[3] The IEEE Global Initiative on Ethics of Autonomousand Intelligent Systems（IEEE）. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems[S]. 2017.

[4] The European Commission's High-level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI[S]. 2019.

[5] The European Commission's High-level Expert Group on Artificial Intelligence. Draft Ethics Guidelines for Trustworthy AI [S]. 2018.

[6] ACM US Public Policy Council and ACM Europe Policy Committee. Statement on Algorithmic Transparency and Accountability[S]. 2017.

[7] 国务院. 国发[2017]35号 新一代人工智能发展规划[S]. 2017.

[8] 国家人工智能标准化总体组. 人工智能伦理风险分析报告[R]. 2019.

[9] Hinton GE，Salakhutdinov R. Reducing the Dimensionality of Data With Neural Networks[J]. Science，2006，313：504-507.

[10] Setio AAA，Traverso A，deBel T，et al. Validation，Comparison，and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in Computed Tomography Images：The LUNA16 Challenge[J]. Med Image Anal，2017，42：1-13.

[11] Gulshan V，Peng L，Coram M. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundusphotographs[J]. JAMA，2016，316：2402.

[12] Becker AS，Marcon M，Ghafoor S，et al. Deep Learning in Mammography：Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer[J]. Invest Radiol，2017，52：434-440.

[13] 莫宏伟. 强人工智能与弱人工智能的伦理问题思考[J]. 科学与社会，2018，8（1）：14-24.

[14] 刘抒悦，高上知，商瑾，等. 美国《健康保险携带和责任法案》中关于生物医学研究的规定及其影响[J]. 中国医学伦理学，2016，29（6）：1011-1014.

[15] 王灏晨. 欧盟《通用数据保护条例》对人工智能发展的影响及启示[J]. 中国经贸导刊：理论版，2018，900（17）：22-24.

[16] 阎娜，王伊龙，李子孝，等. 美国健康保险流通与责任法案对临床研究的影响[J]. 中国卒中杂志，2011，6（12）：971-974.

[17] Tim Miller. Explanation in Artificial Intelligence：Insights from the Social Sciences[J]. Artificial Intelligence，2019，267：1-38.

[18] 安丽娜. 我国伦理委员会的变迁、现状与监管研究[J]. 山东科技大学学报：社会科学版，2019，21（3）：26-32，40.

[19] 张妞，张涛，徐菊华. 中国医院伦理委员会发展的回顾与思考[J]. 医学与哲学：A，2017，38（11）：14-17.

[20] 涉及人的生物医学研究应遵循的伦理原则[J]. 中国心血管杂志，2019，24（2）：194.

[21] 中华人民共和国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[S]. 2016.

[22] 王权，王浩，孟祥峰，等. 人员管理对人工智能医疗器械用数据集质量的影响分析[J]. 中国医疗设备，2018，33（12）：6-9.

[23] 郝烨，唐桥红，李佳戈，等. 数据清洗技术在DICOM格式医学图像质控中的应用[J]. 中国医疗设备，2018，33（12）：10-13.

[24] 王浩，孟祥峰，王权，等. 人工智能医疗器械用数据集管理与评价方法研究[J]. 中国医疗设备，2018，33（12）：1-5.


（收稿日期 2019年6月24日 编辑 王雅雯）