医疗人工智能伦理研究的发展趋势
——主题分布、知识基础与未来展望

Development Trends of Ethical Research on Artificial Intelligence in Medicine: Topic Distribution, Knowledge Base and Future Prospects

汪琛 /WANG Chen[1]　孙启贵 /SUN Qigui[2]　徐飞 /XU Fei[1]

（1. 中国科学技术大学科技哲学系，安徽合肥，230026；2. 中国科学技术大学公共事务学院，安徽合肥，230026）
(1. Department of Philosophy of Science and Technology, University of Science and Technology of China, Hefei, Anhui, 230026; 2. School of Public Affairs, University of Science and Technology of China, Hefei, Anhui, 230026)

摘　要：医疗人工智能伦理已成为学界关注热点，发文数量近年来增长迅速。运用科学计量学方法对医疗人工智能伦理研究文献的主题分布、演进脉络与知识基础进行系统分析，研究发现：（1）医疗人工智能伦理的研究主题正从原则规范向治理体系演进；（2）医疗人工智能伦理的研究视野不再局限于伦理学与技术自身，社会学、管理学、科学技术学等交叉性、系统性研究开始成为主流；（3）模块化的引文网络为相关研究的展开提供了可参照的知识基础；（4）医疗人工智能子域技术伦理的发掘、传统伦理学理论的深度融合、伦理准则的治理转向以及心理学方法的介入等或将成为下一阶段医疗人工智能伦理研究的关注要点。

关键词：医疗人工智能　技术伦理　科学计量学

Abstract: Ethics of artificial intelligence in medicine has become a hot topic in the academic circles, and the publication has increased rapidly in recent years. In this paper, scientometric methods was used to systematically analyze the topic distribution, evolution context and knowledge base in the literature on artificial intelligence in medicine, and the research found that: (1) the research topics on the ethics of artificial intelligence in medicine are evolving from principles and norms to governance systems; (2) the research vision of ethics of artificial intelligence in medicine is no longer limited to ethics and technology itself, and sociology, management, science and technology and other intersecting and systematic research have begun to become the mainstream; (3) the modular citation network provides a reference knowledge base for related research; and (4) the discovery of technical ethics in the sub-domains of artificial intelligence in medicine, the in-depth integration of traditional ethics theories, the governance turning of ethical norms, and the intervention of psychological methods may become the key points focused on the research on the ethics of artificial intelligence in medicine at the next stage.

Key Words: Artificial Intelligence in Medicine; Technical Ethics; Scientometrics

中图分类号：N031　　**文献标识码：**A　　DOI: 10.15994/j.1000–0763.2023.12.003

基金项目：中央高校基本科研业务费资助“合成生物学伦理政策法规框架问题研究”（项目编号：YD2110002004）；中国科学技术大学新文科基金项目“人工智能的基础哲学理论研究”（项目编号：YD2110002012）。

收稿日期：2021年11月25日；**返修日期：**2023年10月7日

作者简介：汪　琛（1996–）男，安徽合肥人，中国科学技术大学科技哲学系博士研究生，研究方向为医疗人工智能伦理。Email: chen0522@mail.ustc.edu.cn

孙启贵（1963–）男，安徽无为人，中国科学技术大学公共事务学院副教授，研究方向为科技与社会。Email: sunqg@ustc.edu.cn

徐　飞（1961–）男，江苏扬州人，中国科学技术大学科技哲学系教授，研究方向为科技哲学与科技史。Email: xufei@ustc.edu.cn

医疗人工智能的伦理问题正逐渐成为学界、产业与政府卫生部门关注的热点领域。[1]医疗人工智能伦理相较于技术伦理的研究起步较晚，但随着医疗人工智能技术成熟度的提升以及商业投资规模的不断扩大，医疗智能化转型的整体趋势已日臻显著，相关伦理与道德问题的关注热度也随之迅速升温。业界开始认识到，医疗人工智能的颠覆性不仅是技术范式层面的演化迭代，同时还隐含着对医疗相关社会体制的革新可能。医疗人工智能伦理向度研究的展开，不仅是为应对医学伦理审查，也是确保智能医疗技术实现正向社会效益的前提保障。

为厘清医疗人工智能伦理研究的发展脉络，需要对既有研究成果的主题分布、知识基础及演进趋势进行系统性分析，洞悉医疗人工智能伦理研究关注的重心变化与演进方向，获取研究前沿的发展动态，[2]为国内医疗人工智能伦理与治理研究建设提供情报支撑。为此，本文采用科学计量学方法，系统梳理最新的医疗人工智能伦理研究，在此基础上尝试提出下阶段医疗人工智能伦理研究可能的发展方向。

一、数据获取与工具选择

1. 数据来源

数据采集自Web of Science数据库，结合已有人工智能计量研究，[3]，[4]采用专业检索式①；语言设置为English，文件类型为ARTICLE 和REVIEW，索引范围包括SCI–EXPANDED, SSCI, A&HCI，得到相关主题论文638篇。

（1）方法工具

本文基于科学计量学研究方法，使用Citespace（5.8.R1）、[5]VOSviewer（1.6.15）、[6]Histcite[7]软件对采集到的文本信息进行系统梳理，采用共现、聚类、突现词检测等计量学算法，挖掘医疗人工智能伦理研究的主题演变历程，并以参考文献作为整体表征，分析医疗人工智能伦理研究的知识基础组成。

二、医疗人工智能伦理研究的整体趋势

发文数量与被引次数能够较为直观的反映该研究主题在学界的受关注度（如图1）。医疗人工智能伦理研究最早始于90年代，第一篇提及医疗人工智能伦理的是南希·麦考利（Nancy McCauley）和穆罕默德·阿里（Mohammad Ala）1992年发表的“医疗保健行业专家系统的使用”，文中以专家系统在医疗保健领域的实践历程为展开，认为医疗专家系统的伦理与法律存在空白，强调不同主体间的责任分配与管理会成为医疗保健领域的关注要点。[8]虽然文章只提及医疗专家系统可能存在伦理问题，但也明确提出技术伦理问题应当得到学界关注。研究主体为医疗人工智能伦理的论文发表则要到1999年的“医疗叙述和患者类比:电子病患记录的伦理意义”，艾克–亨纳·克鲁格（Eike–Henner Kluge）在文中探讨了电子病历对病患隐私的侵权可能以及数据可访问性的伦理影响。[9]往后多年，医疗人工智能伦理的研究论文发表数量一直处于较低水平。

2014年起，医疗人工智能伦理的论文数量开始迅速增加。近年来发文篇数持续提速，2019年医疗人工智能伦理相关主题论文发表首次破百，较2018年（34篇）翻了2.35倍，2020年发文量再度翻倍。截至本文数据统计，医疗人工智能伦理累积发表论文638篇，累积被引次数6496次，H指数为36，相关主题研究规模拓展速度之快，且从发文量拟合曲线可以看出，未来发文量仍会保持较大幅度增加。

三、医疗人工智能伦理研究的关键词分析

① TS=((“artificial intelligence” or AI or “machine learning” or “learning” or “natural language processing” or “Speech Recognition” or “Computer vision” or “Gesture control” or “smart robot” or “Video recognition” or “Voice translation” or “Image Recognition” or “Basic algorithm” or “Smart search” or “Smart recommendation”) and (healthcare or Clinical or medical or health or telemedicine or Diagnosis or health or hospital or Doctor–patient or medicine) and (ethic* OR moral*))；数据采集时间：2021–09–22.

1. 关键词共现分析

为了解不同阶段的热点主题，按照发文量的变化特征，将医疗人工智能伦理研究划分为初阶探索时期（1992–2010）、进阶涌动时期（2011–2016）与加速发展时期（2017–2021），挖掘不同发展阶段的高频关键词，分析医疗人工智能伦理研究的主题演化：

不同阶段的主要关键词分布如表1。细分来看，初阶探索时期的13年间，发文量仅有20篇，由于体量较小，与伦理相关的主要关键词有：ethics（3次）、vitalism（1次）、stigma（1次）、protected health information（1次）、information society（1次）、ethical issue（1次）；进入进阶涌动时期（2011–2016），发文量开始

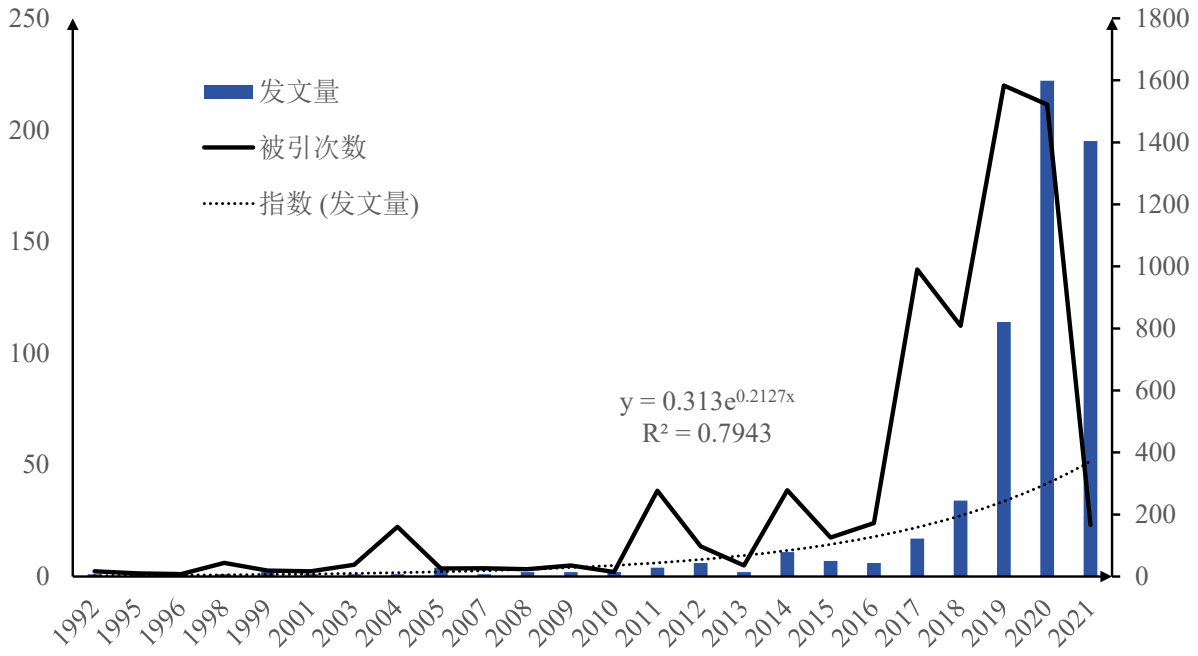

图1 医疗人工智能伦理发文量与被引次数

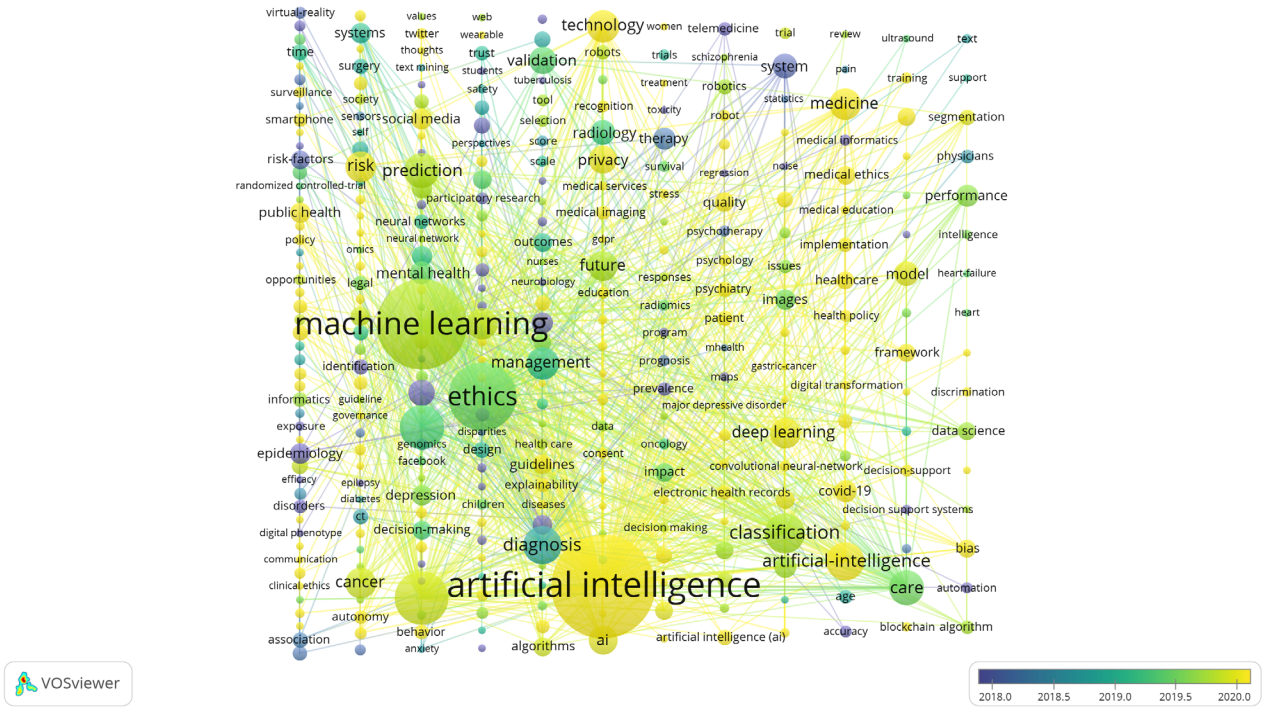

图2 医疗人工智能伦理相关关键词的共现分布

增加，该阶段有16个频次大于2的关键词，对应中心度均小于0.04，可见关键词的数量规模虽然扩大了，其中与伦理相关的关键词主要有ethics training、biomedical ethics、antisocial、ethical code、ethics，均出现1次，但对伦理研究的聚焦度仍相对较低；加速发展时期的5年，伦理相关的研究主题词也更加多样化，单ethics就出现92次，此外还有如bioethics（11次）、medical ethics（5次）、AI ethics（5次）、ethical issue（3次）、research ethics（3次）、clinical ethics（3次），同时regulation（11次）、social media（11次）、health policy（3次）等主题词的出现，说明医疗人工智能伦理已逐渐从理论探讨向实践研究转变，这些主题词的变化更是印证了医疗人工智能伦理研究开始向独立的研究领域迈进。

结合图2、表1可见，医疗人工智能伦理的研究主题分布呈中心扩散趋势，起步阶段多存在于计算智能、辅助腹腔镜手术、生物信息学、专家系统等主题的附带范畴。进入高速发展阶段，医疗人工智能伦理的议题范畴迅速扩大，涉及有可解释性、自主性、数据挖掘、电子健康记录、乳腺癌影像、药物等更多伦理研究的直接相关主题。近年来，机器学习的算法偏见、医疗大数据隐私、技术风险问题成为医疗人工智能研究重点关注方向。

2. 关键词聚类分析

对医疗人工智能伦理相关的638篇论文涵盖的3090个关键词聚类，得到聚类网络的模块值Q=0.5953>0.3，说明主题词聚类结果的社团结构显著；聚类网络的平均轮廓值S=0.8082>0.7，说明研究主题整体较为集中；得到的13个子集聚类模块S值除聚类#0以外均大于0.7，说明聚类结果令人信服；聚类#0的S值为0.535>0.5，说明聚类合理，但关键词分布相对较为分散。

通过本文的梳理整合，结合共现与聚类分析，可以归纳和发现医疗人工智能伦理研究在三个不同的发展阶段主要关注的内容如下：

（1）初阶探索时期主要可划分为三类研究方向：一是强调伦理应当成为智能医疗方法理论的构成要素：通过对技术应用现状的分析，如医疗专家系统、电子病历，[8]，[9]认为技术伦理问题应当引起相关学界的重视。这类技术向度对伦理的关注，为医疗人工智能的伦理研究提供了技术层面的必要性背书；二是探索医疗人工智能伦理可能的应对措施：在探讨技术特性与应用案例的过程中，指出计算智能与生

表1 医疗人工智能伦理不同阶段的高频关键词

序号	初阶探索时期（1992–2010）			进阶涌动时期（2011–2016）			加速发展时期（2017–2021）		
	关键词	频次	中心度	关键词	频次	中心度	关键词	频次	中心度
1	ethics	3	0.09	diffusion	2	0.04	artificial intelligence	209	0.05
2	information	2	0.05	expression	2	0.01	machine learning	137	0.11
3	diagnosis	2	0.01	participatory research	2	0	ethics	92	0.05
4	automation	1	0	drug development	2	0.02	big data	65	0.03
5	vitalism	1	0	epilepsy	2	0.03	health	43	0.01
6	student	1	0	algorithm	2	0.04	classification	41	0.09
7	medicalrecord system	1	0	health	2	0.04	diagnosis	35	0.01
8	chicken	1	0	clinical trial	2	0.03	care	33	0.15
9	decision-support system	1	0	substance abuse	2	0.02	prediction	33	0.04
10	medicaldecision theory	1	0	hiv	2	0.02	technology	32	0.02

命科学的融合过程需要关注可能存在的伦理问题，[10]认为医学中的人工智能系统隐含着特定的理论假设与社会文化态度，[11]并希望通过专业性的课程建设在教育层面实现医学人文的整合发展；[12]三是寻求支撑伦理决策的知识管理框架：认为医疗人工智能的评价需要与临床以及社会环境相整合，作为决策辅助工具，向参与者提供个性化和定制的结构化知识传递，为医师和监护人提供道德决策必要的信息支撑。[13]

（2）进阶涌动时期，该阶段医疗人工智能技术市场的蓬勃前景，使得伦理研究更加注重实践导向的伦理风险分析：一是关注不同医用情境下的技术伦理问题：在探讨人工智能在护理人工智能（AICP）实践过程中，尝试构建伦理与道德评估框架；[14]二是深化技术型塑过程中的伦理认知：在讨论应用性情景的过程中，涉及医疗人工智能和大数据伦理的相关探讨或倡议，如亚历山大·马科维茨（Alexander Markowetz）等探讨了心理信息学背后的大数据伦理与隐私保护问题，但也认为不能过分极化伦理关注。[15]

（3）加速发展时期，智能医疗伦理研究主要可以分为以下三个主要类别：一是智能医疗深层次的伦理问题：如医疗保健中的偏见非公正与算法决策、人机交互可能产生的情感失控、深度学习在医疗实践中的信任关系、家长式行为逻辑产生的道德困境以及心理与精神健康中使用智能算法的伦理挑战；[16]-[21]二是医疗人工智能伦理治理体制的建构：引入利益相关者分析、负责任创新等理论框架，强调医疗人工智能伦理实践必须依托合理的评估指南，技术创新应当与伦理规范平行共进；[22]-[24]三是医疗人工智能伦理的价值论断与认识论基础：关注医疗人工智能计算过程可靠性、可解释性以及透明性，认为医师应当具备必要的人工智能技能与专业知识，与医学信息学、数据科学家合作，是智能黑箱伦理悖论消解的重要前提。[25]-[27]

综上可见，经过近30年的发展积累，医疗人工智能伦理已从技术研究的附带议题，逐渐形成为一类特殊的专业研究领域。研究形式也不再局限于具体伦理问题的对策研究，开始深入挖掘伦理困境产生的原因背景，并结合社会体制，向整合式、系统化与包容性的医疗人工智能治理体系加速转进。随着医疗人工智能迭代速度的不断加快，贯穿算法研发、数据获取、临床实践、商业落地与健康监管与护理的全过程性医疗人工智能伦理体系，是确保智能医疗伦理与技术协同创新重要的研究议题。

表2 国外医疗人工智能伦理关键词的聚类分析

聚类序号	聚类规模	S值	平均年	标签（TF*IDF）	标签（LLR）	标签（MI）
0	84	0.535	2018	artificial intelligence	big data	adolescence
1	59	0.844	2011	artificial intelligence	virtual reality	score
2	58	0.724	2017	artificial intelligence	radiology	automat i c classifiers
3	50	0.797	2012	artificial intelligence	ethics	ethical codes
4	38	0.875	2014	machine learning	machine learning	professionalisation
5	33	0.944	2012	artificial intelligence	algorithm	metabolomics
6	32	0.923	2010	artificial intelligence	critical care	severity scores
7	25	0.961	2013	polymorphism	hiv	reductase gene
8	21	0.967	2012	natural language processing	natural language processing	neurosecurity
9	15	0.983	2008	member states	genomics	artificial intelligence
10	13	1	2013	scavenger receptors	single nucleotide polymorphisms	artificial intelligence
11	10	0.975	2012	domoic acid toxicity	california sea lion	artificial intelligence
14	9	0.99	2004	machine learning	information	mnemic organization

四、医疗人工智能伦理研究的知识基础分析

1. 引文编年图

为了解医疗人工智能伦理研究知识基础的变化脉络，在Histcite中将COUNT设置为50，选取被引文献中被引频次最高的50篇文献按照年代顺序绘制演进编年图（如图3）。从引文编年的演化历程可以看出，医疗人工智能伦理研究的参考引文从2014年开始涌现形成连续性的引文关系脉络，2017–2021年，医疗人工智能伦理开始进入快速发展期，不同引用关系组成的引用链接开始涌现。

分析发现，现行医疗人工智能伦理研究主要的参考文献集中起步于2017和2018年。2017年出现关键节点93和101，93是“放射学实践和研究中的机器学习”，马克·克里（Mark Kohli）等在描述放射科医师在应用机器学习过程中技术要点的基础上，简要提及了可能存在的道德困境与法律风险。[28]101是发表于2016年第38届放射学与社会学交叉会议的论文“大数据与机器学习–驾驭这条总线的战略”，乔纳森克·鲁斯卡（Jonathan Kruskal）等呼吁要确保技术合乎临床伦理性。[29]两篇论文的聚焦主题都与放射学中的机器学习有关，都强调伦理在临床实践中的关切性地位，认同智能医疗伦理是机器学习、大数据与放射性交叉创新中的关键前提。

2018年，节点数量开始迅速增加，其中被引次数最高的前三位：一是派瑟帕·菲利波（Pesapane Filippo）等的“人工智能作为放射学中的医疗设备：欧洲与美国的伦理与监管问题”；[30]二是“2018年人工智能和医学成像：法国放射学白皮书”，提出了十项基于患者利益的人工智能工具开发管理原则；[31]三是帕特里夏·巴尔萨扎（Patricia Balthazar）等的“在大数据、人工智能和预测分析时代保护患者利益”。[32]该年度的主要参考文献多关注医疗人工智能在临床护理、放射诊断等医疗场景中可能出现的各类相关伦理问题。

2019年参考文献节点数量继续增加，150、174和227成为引文关系网络中关系承接和被引次数的代表性文献，分别为尼科尔森·普莱斯（Nicholson Price）等的“医师使用人工智能的潜在责任”，[33]加拿大放射学协会的《放射学中与人工智能相关的伦理和法律问题白皮书》[34]以及盖斯·雷蒙德（Geis Raymond）等的“放射学中的人工智能伦理：欧洲和北美多协会联合声明”。[35]跨国别医疗行业协会的伦

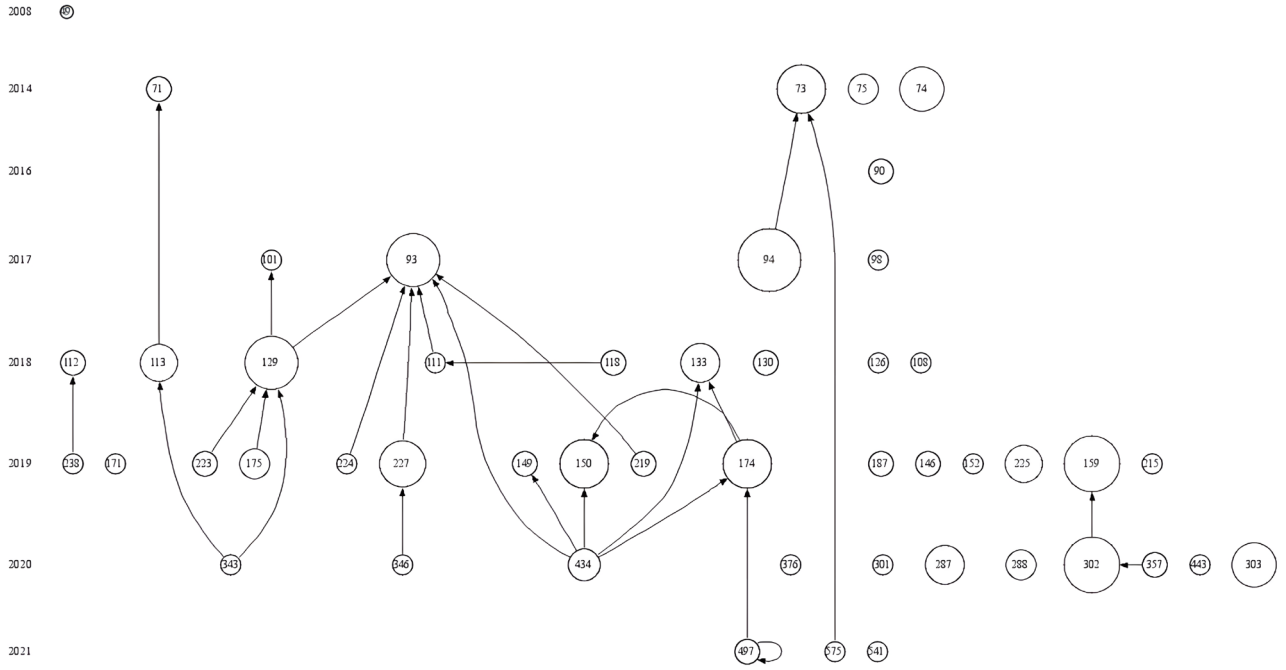

图3 医疗人工智能伦理研究引文编年图

理白皮书与倡议声明陆续发布，说明伦理已经成为医疗人工智能在技术创新、产业落地之外另一要点领域，在智能医疗技术的社会建构过程中发挥着独特的型塑作用。

整体上，参考文献间演化承接的关系脉络说明，伦理研究已逐步从技术研讨中的要素点缀，逐步形成独立的研究主题领域。随着研究的持续推进，医疗人工智能伦理也不再局限于技术或伦理学家学术层面的理论研讨，行业共同体开始成为医疗人工智能临床与商业化伦理准则与设计指南的制定者，研究视角也不再囿于伦理或道德领域，社会、管理、法律等更为综合性、系统性的视角开始成为研究主流。

2. 被引文献聚类分析

为进一步整合医疗人工智能伦理研究的知识基础，对全部有效参考文献作聚类分析，得到聚类模块Q值为0.8298>0.3，表示聚类结果显著，聚类平均轮廓S值为0.853>0.7，每个子聚类S值也都大于0.7，聚类是显著且令人信服的。总的来看，医疗人工智能伦理参考文献的聚类是显著且合理的。结合基础性文献分布、聚类子集中的代表性被引文献，详细探讨不同聚类所代表的知识基础与聚焦主题。

（1）智能算法引发的偏见、隐私与安全问题；算法偏见问题是人工智能伦理研究的重要内容。深度学习与大数据组成的算法黑箱，使得技术偏见的新闻报道屡见不鲜。此外，数据挖掘也常陷入隐私泄露与数据安全的质疑。由于医疗的社会敏感性，人作为医疗人工智能的切身受众，医疗人工智能的隐私保护与安全问题显得更为棘手。该类知识基础的文献主要分布于#0、#1和#10三个聚类，代表性文献有济欧·博迈亚（Ziad Obermeyer）等的“分析人口健康管理算法中的种族偏见”，[36]艾琳·卡利斯卡（Aylin Caliska）等的“从语言语料库自动派生的语义包含有类人偏见”[37]以及科比特–戴维斯·山姆（Corbett–Davies Sam）等的“算法决策和公平成本”，[38]分别从算法原理、语料文本、决策过程三个方面，论述人工智能偏见的诱发机制。

（2）大数据与人工智能伦理指南、数据管理与性能评价；按照人工智能伦理的发展逻辑，第一阶段人工智能伦理必要性的讨论已经达成共识，第二阶段人工智能伦理准则在主要内容上已保持基本一致，现阶段人工智能伦理开始进入到伦理体系的研究。[39]医疗保健作为人工智能交叉创新的代表领域，医疗人工智能伦理研究需要承接既有人工智能伦理指南与管理原则的研究基础，结合医疗独特的场景语境，逐步推进医疗人工智能伦理体系建设。该主题中的知识基础文献散见于#0、#1、#2与#10，代表文献有安娜·乔布（Anna Jobin）等的“人工智能伦理指南的全球图景”[40]、唐安（Tang An）等的“加拿大放射学家协会–放射学人工智能白皮书”[41]以及马克·威尔金森（Mark Wilkinson）等的“科学数据管理与职责的公平

表3 国外医疗人工智能伦理参考文献的聚类分析

聚类序号	聚类规模	S值	平均年	标签（TF*IDF）	标签（LLR）	标签（MI）
0	122	0.808	2018	artificial intelligence	health	smartphone
1	46	0.812	2016	artificial intelligence	psychotherapy	design ethics
2	46	0.894	2016	artificial intelligence	ethics	noncommunicable diseases
3	33	0.91	2017	artificial intelligence	data	automatic classifiers
10	22	0.918	2018	artificial intelligence	digital psychiatry	digital psychiatry
11	21	0.958	2015	digital phenotype	digital psychiatry	tept

指导原则”。[42]代表性引文也说明，智能医疗伦理是人工智能伦理的分支子集，医疗人工智能伦理应当立足一般性、共识性人工智能伦理准则，逐级明晰不同医疗技术应用场景下的伦理价值形态。

（3）智能医疗预测（决策）引发的伦理与法律问题；大数据是人工智能驱动的基础要素，人工智能伦理的主要内容中也与大数据伦理紧密相关。数据规模的扩大以及运作逻辑的转变，使得大数据技术存在可信度存疑、信息误判、数据内容偏见等一系列伦理问题。该主题是聚类#1的延展互补，在聚类#3和聚类#10中也有体现。代表性引文包括格伦·科恩（Glenn Cohen）等的“在医疗保健中使用复杂的预测分析所引起的法律和伦理问题”[43]、格伦·考普斯密斯（Glen Coppersmith）“社交媒体的自然语言处理作为自杀风险的筛选”[44]以及布伦特·米特斯塔德（Brent Mittelstadt）等的“大数据伦理：生物医学领域的当前和可预见问题”。[45]医疗保健大数据是医疗人工智能发展的根基支撑，大数据伦理也随着“数据＋算法”的运作模式，延续成为医疗人工智能伦理研究的知识基础。

（4）医学心理、精神与情感适应性（#11）；机器学习能够重新定义心理健康的诊断模式，提早识别可能存在的心理或精神疾病加以干预，在心理健康领域有着极高的应用前景和创新潜力，[46]但同时也可能存在着伦理风险。[47]从聚类规模看，人工智能心理学已然成为医疗人工智能伦理重要的研究内容，相关的实践成果也成为伦理研究的实证参照来源，代表文献有约翰·托罗斯（John Torous）等的“利用个人智能手机定制应用程序评估重度抑郁症患者的患者健康问卷”，[48]尤卡·昂内拉（Jukka Onnela）等的“利用基于智能手机的数字表型来增强行为和心理健康”[49]以及罗伯特·吉本斯（Robert Gibbons）等的“开发计算机化的抑郁症适应性测试”。[50]

3. 知识基础小结

结合引文编年与参考文献聚类，本文认为医疗人工智能伦理的知识基础可分为以下三类：

（1）人工智能技术类：医疗人工智能伦理的兴起缘起于技术专家对于非技术要素的关注。技术类引文一方面能够拆解智能医疗的技术黑箱，透视技术伦理的诱发机制，从原理层面发掘医疗人工智能伦理的源始塑因；另一方面，医疗人工智能的一般性伦理准则已逐步得到各界认同，向后需要深度扎根技术条件，结合具体智能医疗技术的运作机理，向技术原理与临床情景结合的专业向伦理细则深化研讨。因此，技术类参考文献不仅是医疗人工智能伦理的起步索引，同时还是智能医疗伦理重要的基础材料，是医疗人工智能伦理重要的知识基础组成。

（2）医疗学科类：医疗人工智能是大数据、机器学习、知识表征、专家系统等技术方法，与医疗保健、公共卫生领域的交叉性创新成果。医疗人工智能的知识基础中，存在大量医学与大数据、人工智能等新兴信息技术结合应用的参考文献，一类是运用大数据或人工智能研究医学问题，通过实践案例检验，为医疗人工智能伦理提供论证的科学性基础，印证了医疗人工智能伦理议题的关注必要性与潜藏的技术伦理风险；另一类是技术综述中将伦理引出作为附带议题，可以理解为人工智能与医学研究者的专业背书，是推进医疗人工智能伦理研究专业性、系统化与规模化的原始奠基。

（3）医疗大数据与人工智能伦理：医疗大数据与人工智能伦理自身相关的文献已经成为当前研究重要的知识基础，一方面说明，医疗人工智能伦理研究在引文脉络上有牢靠的沿承，现有研究也是在既有提出问题的基础上，进一步深化拓展，逐步形成了紧密联结的引文链接网络；另一方面经典文献在领域内多次被引，也印证了医疗人工智能伦理研究的专业化与体系化。此外，代表性参考文献的研究主题也说明，医疗人工智能伦理是医疗大数据伦理在智能化时代新的内核延展。医疗大数据是医疗人工智能的基本技术基础与驱动条件，医疗大数据伦理也自然在医疗人工智能伦理研究中延续下来，共同组成医疗人工智能伦理研究的

知识参考。

五、研究总结与未来展望

1. 医疗人工智能子域技术伦理的发掘研究

医疗人工智能每个子域技术开发与应用的实践域境相差甚异。因此，医疗人工智能伦理应当划分为一般性伦理与细分性伦理两大模块。一般性伦理问题讨论的重点应当是普适性的伦理疑难，如数据预测的风险可控性、数据平台权限配置、机器学习的算法偏见、智能医疗资源分配公平性，人工智能伦理的相关研究基本起步于此，下一步需要理清一般性伦理的诱发机制，究竟是技术原理层面存在的条件缺失，还是技术在社会环境中被动性的外源驱动异化？技术内生与外部环境对医疗人工智能伦理形态塑造的权重几何？都需要进一步的研判。

细分性伦理则要求依照不同医疗现实语境下可能出现的伦理问题，如面对人体健康的复杂机制，智能可穿戴设备提供的健康预测能否保障精确性？可能产生的情绪恐慌与过分焦虑如何克服？将智能技术引入患者诊疗与医院管理，医师的主体地位存在被算法消解的风险，传统医学伦理的人文关怀如何保障？弱化家长式诊疗模式，增大病患、消费者的自主性，是否会引发更加棘手的责任归属问题？这一系列医疗人工智能细分技术可能产生的道德困境都是亟待关注和研究的关键议题。

2. 伦理学理论的创新与发展

人工智能在医学领域的深度介入引发的伦理问题，其最重要的特殊性在于智能人工物对医师主体地位的替代。目前，医疗人工智能伦理开始进入加速发展阶段，专业性的研究体系正在同步推进建设，提出了一系列医疗人工智能伦理准则，其中有很大一部分源承于传统医学伦理原则，但由于医疗主体的转换、交互模式的改变、医疗信息形式的转化，需要重新审视传统医学伦理规范的适用性、可靠性与完备性。医疗智能化转型的“去人化”，使得医疗人工智能伦理问题在商业化过程加速推进的背景下，与传统伦理观念的“撕裂”风险显得尤为尖锐。因而医疗人工智能伦理下一阶段的研究，需要基于医学伦理与技术伦理理论，深化对医疗人工智能伦理与社会问题的创新阐释，一方面需要根据不同医疗场景，细分技术实施的伦理适用准则，另一方面还应针对整体性的伦理困境，如智能医疗的准道德主体地位的确立与否，医疗人文精神如何在智能化时代得以续存，做出伦理向度的应对准备。这些研究的开展，不仅是人工智能伦理研究持续推进的体现，还能够反哺传统伦理学的理论体系，加速伦理学自身的创新发展。

3. 伦理准则向治理体制的实践转化

从技术的社会型塑看，伦理与道德是建构医疗人工智能社会–技术系统必要的非人行动者。应当明确，医疗人工智能伦理与道德是智能医疗技术必要的支撑模块，需要伦理学家、人工智能学者、技术企业、政府机构等利益攸关方的共同协商参与。医疗人工智能伦理已从单纯的学术议题，逐步转向智能医疗实践的政策规范与管理体制的实践研究。医疗人工智能伦理不仅是技术的规范确立与研究人员的素养教育，同时也是医疗人工智能法律、管理与政策重要的理论基础。为实现医疗人工智能伦理与治理体制的对应配套，不仅要补足医疗人工智能临床与产品的伦理审查机制，还存在诸多过渡性伦理议题亟待解决，如人工智能战略布局中，医疗人工智能伦理有怎样的战略定位？如何通过法律与行业准则，在医疗人工智能技术设计中嵌入伦理价值？伦理与体制层面如何与医疗人工智能创新协调并进？实现医疗人工智能伦理准则与社会治理体系的双向加持，是确保医疗人工智能技术可靠性与社会可行性重要的外部协同要素，也是防治医疗人工智能技术目标与实践脱轨异化的制度保障，这一系列交叉性议题还需要学界进一步的关注研究。

4. 拓展医疗人工智能伦理研究的心理学视角

为保障医疗人工智能伦理的审查与规范有效性，需要封闭性的社会实验作为中介环节，在一系列的试点反馈中，不断调整和优化医疗

人工智能伦理的指导原则与规约制度。界定医疗人工智能伦理准则有效性的实践成果，一方面来自于有限边界内的社会后果评价，另一方面更需要关注技术受众的心理与情感反馈。心理学研究能够为医疗人工智能伦理的实践评估提供病患或健康消费者观点的实证参考，为技术伦理提供实践性的证据支持。医疗人工智能伦理规范与治理体系的建设完善，不仅需要“自上而下”的伦理原则作为框架性规约，还需考量各利益有关群体“自下而上”的伦理诉求。医疗人工智能伦理不能被简单地视为强制性的约束限制，更需要关注技术受众的情绪感受，检验伦理规范的实践有效性。心理学方法通过个体心理可接受程度的变化，研判技术调整的可能方向，能够为医疗人工智能伦理提供基于特定医疗情景伦理实践的心理评价，应当成为医疗人工智能伦理下一阶段的重要辅助。

［参 考 文 献］

[1] Ellahham, S., Ellahham, N., Simsekler, M. 'Application of Artificial Intelligence in the Health Care Safety Context: Opportunities and Challenges'[J]. *American Journal of Medical Quality*, 2019, 35(4): 341–348.

[2] Murphy, K., Ruggiero, E. D., Upshur, R., et al. 'Artificial Intelligence for Good Health: A Scoping Review of the Ethics Literature'[J]. *BMC Medical Ethics*, 2021, 22(1): 1–17.

[3] 刘龑龙、史冬梅、刘进长、唐莉、王金鹏. 基于文献计量学的人工智能领域研究现状及热点分析[J]. 科技管理研究, 2021, 41（10）: 38–48.

[4] 刘艳秋、韩俊敏、王建国、华连连. 人工智能专利技术分布、演化及合作创新网络分析[J]. 中国科技论坛, 2021,（3）: 64–74.

[5] 陈悦、陈超美、刘则渊、胡志刚、王贤文. CiteSpace知识图谱的方法论功能[J]. 科学学研究, 2015, 33（2）: 242–253.

[6] Eck, N. J. V., Waltman, L. 'Software Survey: VOSviewer, A Computer Program for Bibliometric Mapping'[J]. *Scientometrics*, 2010, 84: 523–538.

[7] 罗昭峰. 引文分析软件histcite简介［EB/OL］, http://blog.sciencenet.cn/blog–304685–383399. html. 2021–09–27.

[8] McCauley, N., Ala, M. 'The Use of Expert Systems in the Healthcare Industry'[J]. *Information & Management*, 1992, 22(4): 227–235.

[9] Kluge, E. H. 'Medical Narratives and Patient Analogs: The Ethical Implications of Electronic Patient Records'[J]. *Methods of Information in Medicine*, 1999, 38(4–5): 253–259.

[10] Vasilakos, A. V., Spyrou, G. 'Computational Intelligence in Medicine and Biology: A Survey'[J]. *Journal of Computational and Theoretical Nanoscience*, 2008, 5(12): 2365–2376.

[11] Spyropoulos, B., Papagounos, G. 'A Theoretical Approach to Artificial Intelligence Systems in Medicine'[J]. *Artificial Intelligence in Medicine*, 1995, 7: 455–465.

[12] Kanter, S. L., Wimmers, P. F., Levine, A. S. 'In-depth Learning: One School’s Initiatives to Foster Integration of Ethics, Values, and the Human Dimensions of Medicine'[J]. *Academic Medicine*, 2007, 82(4): 405–409.

[13] Frize, M., Yang, L., Walker, R. C., et al. 'Conceptual Framework of Knowledge Management for Ethical Decision-Making Support in Neonatal Intensive Care'[J]. *IEEE Transactions on Information Technology in Biomedicine*, 2005, 9(2): 205–215.

[14] Luxton, D. D. 'Recommendations for the Ethical Use and Design of Artificial Intelligent Care Providers'[J]. *Artificial Intelligence in Medicine*, 2014, 62(1): 1–10.

[15] Markowetz, A., Błaszkiewicz, K., Montag, C., et al. 'Psycho-Informatics: Big Data Shaping Modern Psychometrics'[J]. *Medical Hypotheses*, 2013, 82(4): 450–411.

[16] Starke, G., Clercq, E. D., Elger, B. S. 'Towards a Pragmatist Dealing with Algorithmic Bias in Medical Machine Learning'[J]. *Medicine Health Care and Philosophy*, 2021, 24(3): 341–349.

[17] Grote, T., Berens, P. 'On the Ethics of Algorithmic Decision-making in Healthcare'[J]. *Journal of Medical Ethics*, 2019, 46(3): 205–211.

[18] Gilbert, F., Cook, M., O’Brien, T., et al. 'Embodiment and Estrangement: Results from a First-in-Human “Intelligent BCI” Trial'[J]. *Science and Engineering Ethics*, 2019, 25(1): 83–96.

[19] Hatherley, J. 'Limits of Trust in Medical AI'[J]. *Journal of Medical Ethics*, 2020, 46(7): 478–481.

[20] Kühler, M. 'Exploring the Phenomenon and Ethical Issues of AI Paternalism in Health Apps'[J]. *Bioethics*, 2021, 36(2): 194–200.

[21]Gooding, P., Kariotis, T. 'Ethics and Law in Research on Algorithmic and Data-Driven Technology in Mental Health Care: Scoping Review'[J]. *JMIR Mental Health*, 2021, 8(6): 1–12.

[22]Wangmo, T., Lipps, M., Kressig, R. W., et al. 'Ethical Concerns with the Use of Intelligent Assistive Technology: Findings from a Qualitative Study with Professional Stakeholders'[J]. *BMC Medical Ethics*, 2019, 20(1): 1–11.

[23]Buruk, B., Ekmekci, P. E., Arda, B. 'A Critical Perspective on Guidelines for Responsible and Trustworthy Artificial Intelligence[J]. *Medicine, Health Care and Philosophy*, 2020, 23(3): 387–399.

[24]Jongsma, K. R., Bredenoord, A. L. 'Ethics Parallel Research: An Approach for (early) Ethical Guidance of Biomedical Innovation'[J]. *BMC Medical Ethics*, 2020, 21(1): 1–9.

[25]Ursin, F., Timmermann, C., Stege, F. 'Explicability of Artificial Intelligence in Radiology: Is a Fifth Bioethical Principle Conceptually Necessary?'[J]. *Bioethics*, 2022, 36(2): 143–153.

[26]Felder, R. M. 'Coming to Terms with the Black Box Problem: How to Justify AI Systems in Health Care'[J]. *Hastings Center Report*, 2021, 51(4): 38–45.

[27]Durán, J. M., Jongsma, K. R. 'Who is Afraid of Black Box Algorithms? On the Epistemological and Ethical Basis of Trust in Medical AI'[J]. *Journal of Medical Ethics*, 2021, 47: 329–335.

[28]Kohli, M., Prevedello, L., Filice, R., et al. 'Implementing Machine Learning in Radiology Practice and Research'[J]. *American Journal of Roentgenology*, 2017, 208(4): 1–7.

[29]Kruskal, J., Berkowitz, S., Geis, J., et al. 'Big Data and Machine Learning——Strategies for Driving This Bus: A Summary of the 2016 Intersociety Summer Conference'[J]. *Journal of the American College of Radiology*, 2017, 14(6): 811–817.

[30]Pesappane, F., Volonté, C., Codari, M., et al. 'Artificial Intelligence as a Medical Device in Radiology: Ethical and Regulatory Issues in Europe and the United States'[J]. *Insights Imaging*, 2018, 9: 745–753.

[31]Group, S. F. R. I. A., Community, F. R. 'Artificial Intelligence and Medical Imaging 2018: French Radiology Community White Paper'[J]. *Diagnostic and Interventional Imaging*, 2018, 99(11): 727–742.

[32]Balthazar, P., Harri, P., Prater, A., et al. 'Protecting Your Patients' Interests in the Era of Big Data, Artificial Intelligence, and Predictive Analytics'[J]. *Journal of the American College of Radiology*, 2018, 15(3): 580–586.

[33]Price, W. N., Gerke, S., Cohen, I. G. 'Potential Liability for Physicians Using Artificial Intelligence'[J]. *JAMA*, 2019, 322(18): 1765–1766.

[34]Jaremko, J., Azar, M., Bromwich, R., et al. 'Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology'[J]. *Canadian Association of Radiologists Journal*, 2019, 70(2): 107–118.

[35]Geis, J. R., Brady, A. P., Wu, C. C., et al. 'Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement'[J]. *Radiology*, 2019, 293(2): 436–440.

[36]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.

[37]Caliskan, A., Bryson, J. J., Narayanan, A. 'Semantics Derived Automatically from Language Corpora Contain Human-like Biases'[J]. *Science*, 2017, 356(6334): 183–186.

[38]Corbett-Davies, S., Pierson, E., Feller, A, et al. 'Algorithmic Decision Making and the Cost of Fairness'[A], Matwin, S., Yu, S., Farooq, F. (Eds.) *KDD 17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*[C], New York: Association for Computing Machinery, 2017, 797–806.

[39]陈小平. 人工智能:技术条件、风险分析和创新模式升级[J]. 科学与社会, 2021, 11（2）: 1–14.

[40]Jobin, A., Ienca, M., Vayena, E., et al. 'The Global Landscape of AI Ethics Guidelines'[J]. *Nature Machine Intelligence*, 2019, 1: 389–399.

[41]Tang, A., Tam, R., Cadrin-Chenevert, A, et al. 'Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology'[J]. *Canadian Association of Radiologists Journal*, 2018, 69(2): 120–135.

[42]Wilkinson, M., Dumontier, M., Aalbersberg, I. J., et al. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship[J]. *Scientific Data*, 2016, 3: 167–172.

[43]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*, 2014, 33(7): 1139–1147.

[44]Glen, C., Ryan, L., Patrick, C., et al. 'Natural Language Processing of Social Media as Screening for Suicide Risk'[J]. *Biomedical Informatics Insights*, 2018, 10: 1–11

[45]Mittelstadt, B. D., Floridi, L. 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts'[J]. *Science & Engineering Ethics*, 2016, 22(2): 303–341.

[46]Graham, S., Depp, C., Lee, E. E., et al. 'Artificial Intelligence for Mental Health and Mental Illnesses: An Overview'[J]. *Current Psychiatry Reports*, 2019, 21(11): 1–18.

[47]Martinez-Martin, N., Kreitmair, K. 'Ethical Issues for Direct-to-Consumer Digital Psychotherapy Apps: Addressing Accountability, Data Protection, and Consent'[J]. *JMIR Mental Health*, 2018, 5(2): e9423.

[48]Torous, J., Staples, P., Shanahan, M., et al. 'Utilizing a Personal Smartphone Custom App to Assess the Patient Health Questionnaire-9 (PHQ-9) Depressive Symptoms in Patients With Major Depressive Disorder'[J]. *JMIR Mental Health*, 2015, 2(1): e3889.

[49]Onnela, J. P., Rauch, S. L. 'Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health'[J]. *Neuropsychopharmacology*, 2016, 41: 1691–1696.

[50]Gibbons, R. D., Weiss, D. J., Pilkonis, P. A., et al. 'Development of a Computerized Adaptive Test for Depression'[J]. *Archives of General Psychiatry*, 2012, 69(11): 1104–1112.

[责任编辑　李斌]