

《医学人工智能社会治理综合评价指南》 (送审稿)编制说明

一、项目背景

(一) 国内外相关情况简介

2017年7月8日，国务院印发了《新一代人工智能发展规划》，要求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。2018年10月，习近平总书记在十九届中央政治局第九次集体学习时明确指出：“要加强人工智能发展的潜在风险研判和防范，维护人民利益和国家安全，确保人工智能安全、可靠、可控。”2021年9月25日，《新一代人工智能伦理规范》发布，提出了增进人类福祉、促进公平公正、保护隐私安全、确保可控可信、强化责任担当、提升伦理素养等6项基本伦理要求，并提出了人工智能管理、研发、供应、使用等特定活动的18项具体伦理要求。2022年3月20日，国务院办公厅印发《关于加强科技伦理治理的意见》，要求“进一步完善科技伦理体系，提升科技伦理治理能力，有效防控科技伦理风险，不断推动科技向善、造福人类”，将人工智能伦理纳入科技伦理的综合治理中。

随着人工智能技术的应用普及，人工智能技术滥用等问题不断出现，同时也带来了信息安全问题。2023年4月，美国斯坦福大学以人为本人工智能研究所(Stanford HAI)发布的《2023年人工智能指数报告》(Artificial Intelligence

IndexReport 2023）分析了人工智能的影响和年度趋势。报告发现，关于人工智能滥用的事件数量正在迅速上升。通过追踪人工智能道德滥用相关事件的人工智能算法和自动化事件和争议（AIAAIC）数据库，2012年至2022年间人工智能相关争议事件增加近26倍。针对各类人工智能问题，世界主要国家和地区陆续从安全性审查、技术禁用等角度出台相关监管措施以实现人工智能的有效治理。

探索合理、有效的治理方案，实现人工智能健康有序发展已经成为全球重要课题。在第三届“一带一路”国际合作高峰论坛上，中国发布的《全球人工智能治理倡议》（下称《倡议》）赢得国际社会广泛赞誉。《倡议》站在构建人类命运共同体、维护全人类福祉的历史高度，就人工智能发展、安全、治理三个方面公平公正地提出建设性解决措施，为全球人工智能治理提供中国方案。2023年11月首届人工智能安全峰会在英国布莱奇利园举行并发布了《布莱奇利宣言》，美国、中国、日本、德国、印度等20多个国家的政府代表以及联合国、经合组织、国际电信联盟等多个国际组织的代表参会，28个国家和欧盟共同签署了《布莱奇利宣言》，同意通过国际合作，建立人工智能（AI）监管方法。

目前人工智能技术主要造成三类问题。一是信息安全问题。人工智能需要实时更新海量数据用于模型训练。当前，在个人隐私领域中，相关企业、平台尚未达成遵循知情同意原则的共识，存在越权使用现象。因此，用户如果缺乏相关意识，在使用人工智能技术时，很可能导致个人隐私数据泄

露。二是虚假信息问题。尽管技术中立，但认为投放不实消息或片面事实可能会导致人工智能模型产出虚假信息，也就是“深度伪造”。三是数字鸿沟问题。人工智能具备高效的信息检索和整合能力，虽然为人们的工作和生活提供了极大方便，但其产出的真实性和准确性并不牢靠。因此，信息甄别、自主决策能力强的使用者将更善于使用这项工具，而盲从的使用者则可能为其所害。随着两者使用技术的差距不断沉积，数字鸿沟也就此加深。

人工智能技术在医学领域使用同样面临以上的治理问题外，还会面临人工智能技术带来卫生系统资源配置的深刻影响。然而，关于医学人工智能治理评价指标体系，业内一直高度关注，但尚未形成广泛一致的意见。截至2024年6月，深圳市地区尚未发布医学人工智能治理评价相关的指南文件。深圳市缺乏本地化的医学人工智能治理评价指标体系，各医疗卫生机构开展医学人工智能治理试验评价遵循的范式参差不齐，导致各级医疗机构间或区域内医学人工智能治理评价结果管理困难的实际境况，通过管控和治理发现医学人工智能可能带来的潜在风险尚未形成有效的监测与预警机制。

为了满足当前医学人工智能治理评价的标准需求，科学引导深圳市医学人工智能的研发与应用的同时，促进各类医疗机构及其他利益相关方组织共同参与医学人工智能治理活动的有序有效地开展，亟需形成医学人工智能治理评价指标体系，从安全、风险、效用、效率、效益的多元视角加强

医学人工智能的治理影响加以监测与管控。

（二）必要性和意义

深圳市响应国家《倡议》号召，为了更好推动医学人工智能应用的发展，在加强医疗人工智能治理能力的同时，亟需构建一套分类管理、综合评价、贯穿医学人工智能技术开发全生命周期的人工智能治理评价指标体系，指导医学人工智能治理活动的开展，动态评估不同类别医学人工智能在安全、风险、效用、效率和效益层面上的表现，规范建设深圳市医学人工智能治理，以达到对医学人工智能发展潜在风险研判和防范，确保其安全、可靠、可控的发展。推动深圳市智慧城市医疗人工智能治理工作先行示范，助力深圳市粤港澳大湾区、中国特色社会主义先行示范区的“双区建设”。

为了促进医学人工智能健康发展和规范应用，维护国家安全和公共利益，保护公民、法人和其他组织的合法权益，根据《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》《中华人民共和国科学技术进步法》《生成式人工智能服务管理暂行办法》等政策法规，结合人工智能治理的相关行业标准，制定本评价指标体系。通过医疗人工智能治理综合评价，从安全、风险、效用、效率以及效益层面的动态评估结果为规范建设人工智能治理体系提供决策性参考。

《医学人工智能社会治理综合评价指南》的制定、发布和实施，明确了深圳市医学人工智能治理评价的行业标准体系架构，确定了深圳市医学人工智能治理综合评价的建设方

向，为后续深圳市医学人工智能治理评价相关标准的编制提供了依据和规范性指引。

二、工作简况

（一）任务来源

根据深圳市市场监督管理局 2022 年 5 月 11 日发布的《深圳市市场监督管理局关于下达 2022 年深圳市地方标准计划项目任务的通知》，《医学人工智能社会治理综合评价指南》予以立项，序号 134。由深圳市卫生健康委员会提出并归口，深圳市卫生健康委员会作为牵头单位，参与起草单位包括深圳市卫生健康委员会、南方医科大学、中国科学院自动化研究所、深圳市卫生健康发展研究和数据管理中心、南方医科大学深圳医院。

（二）主要编制过程

1. 规划准备阶段

2022 年 5 月，《医学人工智能社会治理综合评价指南》（下称《评价指南》）作为 2022 年深圳市地方标准计划项目正式批准立项。随后，在行业主管部门的指导下，市卫生健康委召集参与起草单位的卫生行政管理研究领域、卫生法学领域、医学伦理领域、卫生经济学评价领域、人工智能领域、数据安全领域的学者专家、行政管理者、卫生技术人员、工程师等，组织成立了《评价指南》编制组，并制定了编制计划方案，形成了明确的分工机制。

编制组开展前期文献研究，收集和整理国内外相关法律法规、政策文本、标准规范和研究论文，整理现有各行业关

于人工智能治理政策要求与标准规范，分析适宜深圳本土医学人工智能治理的规范性要素、技术要点和框架结构。

2. 实地调研阶段

2022 年 11 月-12 月，编制组通过问卷调查、实地走访、关键人物访谈、小组访谈等方式对科研院校、医疗卫生机构、人工智能相关企业及相关政府主管部门的医学人工智能治理的安全、风险、效用、效率、效益的治理评价内涵与内容构成展开实地调研和讨论，并关于深圳医疗人工智能治理评价指标体系的可操作性进行探讨。

3. 标准起草阶段

2023 年 5 月，完成了《评价指南》框架起草。编制组根据前期文献研究成果和专家访谈所获资料的文本分析，2023 年 8 月在前述基本框架的基础上，撰写《评价指南》，调整标准基础格式，形成了《评价指南》初稿。此后，经过多次内部讨论，对初稿文本进行了多轮修订和补充，形成面向人工智能治理相关行业专家咨询的《评价指南（征求意见稿）》

4. 征求意见阶段

2023 年 8 月下旬至 2023 年 12 月上旬，编制组采用线上函询线下面询的方式，向中国科学院自动化所、中国医学科学院医学信息研究所、国家卫生健康委员会、国家卫生健康委员会卫生发展研究中心、深圳市卫生健康发展研究和数据管理中心、深圳大学、香港中文大学（深圳）、深圳市人民医院、深圳市中医院、华中科技大学协和深圳医院、深圳市龙华区人民医院、北京大学深圳医院、华为技术有限公司、

深圳市腾讯计算机系统有限公司、深圳平安医疗健康管理有限公司、深圳海云安网络安全技术有限公司、深圳蓝网科技有限公司、浙江省数理医学学会、河南省标准化与质量研究院、深圳市标准技术研究院等机构从事人工智能研发、卫生法、卫生管理、卫生经济评价、数据安全、算法开发、医疗信息管理、行业标准制定等工作的相关专家学者进行意见征询，并基于征求意见进行反馈和开展深入的咨询研讨。截至 2023 年 12 月上旬，开展线下专家研讨会议和意见征询四轮（线下每轮征询专家人数 4-5 人），开展线上专家研讨会议和意见征询 13 轮（线上每轮征询专家人数 1-2 人）编制组根据专家对《评价指南（征求意见稿）》提出的反馈意见以及修改建议，经过逐条整理与回应，确定采纳与否和修改方式，形成征求意见汇总处理表。在征求意见汇总处理表的基础上，完成《评价指南（征求意见稿）》的完善与修订，并基于以上的修改情况反馈于相应的专家以达成最终的修改共识。

2024 年 4 月 26 日-2024 年 5 月 26 日，通过深圳市卫生健康委员会的门户网站公开征求意见，未收到反馈意见。同时，书面征求市科技创新局、市工业和信息化局、市市场监管局、市医疗保障局、市政务和数据局、各卫生健康行政部门和委属医疗机构等单位意见，共收到反馈意见 5 条，无意见 20 条。编制组对收到的反馈意见进行汇总处理，其中采纳 3 条，部分采纳 1 条，不采纳 1 条，并整理征求意见汇总处理表。

截至 2024 年 10 月，编制组从立项至《评价指南（征求意见稿）》初步确定共形成了 17 个版本，展开多轮内部改稿会，多次审阅标准文本，进一步完善《评价指南》的框架、厘清条目内容逻辑、打磨标准文本格式。经过多轮修改后，形成了《医学人工智能社会治理综合评价指南（送审稿）》，并撰写标准编制说明文件。

三、地方标准主要内容的依据以及与国内外先进标准的对标情况

（一）标准制定的原则

1. 全面系统，兼容并包

《评价指南》查阅和吸纳国内外不同行业关于人工智能治理相关服务标准和指引，展现医学人工智能治理评价的既有共识和前沿发展，全面呈现医学人工智能治理使用者面对不同情境进行不同维度的评估内容。

2. 使用导向，切实可行

《评价指南》以指导实际应用为导向，考量了不同类别医学人工智能产品的特点。在条文的编写上，咨询了医学人工智能开发者、医学人工智能产品使用者、医学人工智能产品监管者等利益相关者的意见，为医学人工智能治理提供了具体指导。

3. 体现特色，适宜推广

《评价指南》在保证体系和内容完善的同时，突出深圳市医学人工智能发展的本土特色，首先深圳市是医学人工智能产品研发的前沿阵地，同时具有良好信息系统建设，深圳

市医疗卫生机构更是医学人工智能产品落实使用的试验田，为此《评价指南》在深圳市推广具有较好的适宜性，可为国内其他地市发展医学人工智能治理工作提供了宝贵的实践经验。

（二）标准制定的依据

本文件严格按照《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》（GB/T 1.1—2020）的要求进行编写。各章主要内容的研制依据情况如下。

本文件第三章 术语与定义，编制依据情况如下表。

内容	依据
3.1 医学人工智能	依据GB/T 41867，关键通用技术相关术语；YY/T 1833.1，定义3.1.2和3.1.3
3.2 医学人工智能治理	依据《Ethics and governance of artificial intelligence for health Guidance on largemulti-modal models》（World Health Organization, 2024）
3.3 医学人工智能治理评价	依据专家研讨达成的专家共识
3.4 训练数据	依据GB/T 41867—2022，定义3.2.34，有修改
3.5 医学人工智能技术内在风险	依据全国网络安全标准化技术委员会发布的《人工智能安全治理框架》编制
3.6 医学人工智能技术应用风险	依据全国网络安全标准化技术委员会发布的《人工智能安全治理框架》编制

本文件第4章 评价原则，参考评价过程需要考虑的事项原则，结合评价过程中需要考虑的原则，经过研究确定了可行性原则、全面性原则、代表性原则、时效性原则。

本文件第 5 章 指标体系，参考《深圳经济特区数据条例》〔深圳市第七届人民代表大会常务委员会公告（第十号）〕、《全国医院信息化建设标准与规范（试行）》（国卫办规划发〔2018〕4 号）、《全国基层医疗卫生机构信息化建设标准与规范（试行）》（国卫规划函〔2019〕87 号）、《生成式人工智能服务管理暂

行办法》（国家互联网信息办公室 中华人民共和国国家发展和改革委员会 中华人民共和国教育部 中华人民共和国科学技术部 中华人民共和国工业和信息化部 中华人民共和国公安部 国家广播电视总局令第15号）、《关于印发医疗机构临床决策支持系统应用管理规范（试行）》（国卫办医政函〔2023〕268号）、《生成式人工智能服务安全基本要求》（信安秘字〔2023〕146号）等政策文件，结合相关标准实际和深圳市对医学人工智能治理评价的需求情况，形成并建立了文件中的评价指标体系。

本文件第6章 指标内涵，编制依据情况如下表。

指标	依据
6.3.1 数据安全	《信息安全技术 个人信息安全规范》（GB/T 35273—2020） 《智能交通 数据安全服务》（GB/T 37373—2019） 《信息安全技术 健康医疗数据安全指南》（GB/T 39725—2020） 《金融数据安全 数据生命周期安全规范》（JR/T 0223—2021） 《电信网和互联网数据安全评估规范》（YD/T 3956—2021） 《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） 《人工智能开发平台通用能力要求 第1部分：功能要求》（YD/T 4392.1—2023） 《信息安全 人工智能数据安全通用要求》（DB11/T 2251—2024）
6.3.2 隐私安全	《信息安全技术 个人信息安全规范》（GB/T 35273—2020） 《信息安全技术 健康医疗数据安全指南》（GB/T 39725—2020） 《人工智能算法金融应用信息披露指南》（JR/T 0287—2023） 《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）
6.3.3 医疗安全	依据《糖尿病视网膜病变人工智能筛查应用规范》（DB52/T 1726—2023）和相关行业专家意见共识编制。
6.3.4 场景渗透、 6.3.5 适用效能、 6.3.6 受众体验	依据《基于人工智能的接入网运维和业务智能化场景与需求》（YD/T 4070 — 2022 ）、《Regulatory consideration sonartificial intelligence for health》（World Health Organization, 2023）和相关行业专家意见共识编制。
6.3.7 算法风险	《信息技术 人工智能 术语》（GB/T 41867—2022） 《人工智能算法金融应用评价规范》（JR/T 0221—2021） 《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） 《人工智能开发平台通用能力要求 第1部分：功能要求》（YD/T 4392.1—2023）

	《人工智能医疗器械 冠状动脉CT影像处理软件 算法性能测试方法》（YD/T 4921—2024）
6.3.8 训练数据风险	《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） 《人工智能开发平台通用能力要求 第1部分：功能要求》（YD/T 4392.1—2023）
6.3.9 生成内容风险	《人工智能医疗器械 质量要求和评价 第4部分：可追溯性》（YY/T 1833.4—2023） 《人工智能开发平台通用能力要求 第1部分：功能要求》（YD/T 4392.1—2023） 《网络安全技术 生成式人工智能服务安全基本要求（征求意见稿）》
6.3.10 社会安全风险、 6.3.11 伦理风险、 6.3.12 社会经济风险	《人工智能算法金融应用信息披露指南》（JR/T 0287—2023） 《基于人工智能的多中心医疗数据协同分析平台参考架构》（YD/T 4043—2022） 《移动智能终端可信人工智能安全指南》（YD/T 4960—2024）《Ethics and Governance of Artificial Intelligence for Health》（World Health Organization,2021） 《Regulatory consideration sonartificial intelligence for health》（World Health Organization,2023） 《Ethics and governance of artificial intelligence for health Guidance on largemulti-modal models 》（World Health Organization,2024）
6.3.13 成本效率、 6.3.14 规模效率、 6.3.15 配置效率	依据《Ethics and Governance of Artificial Intelligence for Health》（World Health Organization,2021）、《Regulatory consideration sonartificial intelligence for health》（World Health Organization,2023）、《Ethics and governance of artificial intelligence for health Guidance on large multi-modalmodels》（World Health Organization, 2024）和相关行业专家意见共识编制。
6.3.16 经济效益、 6.3.17 社会效益、 6.3.18 健康效益	依据《Ethics and Governance of Artificial Intelligence for Health》（World Health Organization,2021）、《Ethics and governance of artificial intelligence for health Guidance on large multi-modal models》（World Health Organization, 2024）和相关行业专家意见共识编制。

（三）与国内外先进标准的对标情况

目前国内外暂无已实行的医学人工智能治理评价的相关标准文件。

四、主要条款的说明以及主要技术指标、参数、试验验

证的论述

《评价指南》现有 6 个章节以及参考文献。包括范围、规范性引用文件、术语与定义、指标选取原则、指标体系、指标内涵（一级指标、二级指标、三级指标）。以下对标准中的主要条款进行简要说明。

（一）范围

本文件提供了开展医学人工智能治理评价的指导，给出了指标选取原则、指标体系的构成、内容、架构图和指标内涵等方面的建议，并给出了相关信息。

本文件适用于评价深圳市行政区域内医学人工智能治理产生的现实或潜在影响的活动。

（二）规范性引用文件

本章节主要包括了标准文本中规范性引用的文件。

（三）术语和定义

本章节主要包括医学人工智能、医学人工智能治理、医学人工智能治理评价、训练数据、医学人工智能技术内在风险、医学人工智能技术应用风险的术语与定义。

（四）指标选取原则

本章介绍了医学人工智能治理评价指标体系中评价指标选取的原则，包括可行性原则、全面性原则、代表性原则、时效性原则。

（五）指标体系

本章介绍了指标体系的构成、指标体系内容及指标体系架构图。医学人工智能治理指标体系的结构分3个层级，其中一级指标3个，二级指标6个，三级指标18个。

（六）指标内涵

本章节给出了医学人工智能治理评价指标体系中一级指标、二级指标、三级指标的内涵。

五、是否涉及专利等知识产权问题

本文件不涉及专利和其他知识产权问题。

六、重大意见分歧的处理依据和结果

无。

七、实施地方标准的措施建议

根据《深圳市地方标准管理办法》（市政府令第345号），市有关行政主管部门在标准发布后，应当组织开展本部门、本行业地方标准的宣传和实施工作。建议主管部门采用以下多种方式开展标准的宣贯与实施工作，积极推动与医学人工智能治理综合评价相关的各利益方理解并实施本文件。

（一）组织标准宣贯与培训工作

首先，由主管部门统筹，通过线上和线下多种渠道和形式，向标准应用相关方宣传和介绍本文件，提高本文件的知晓度；其次，由主管部门组织标准实施的培训工作，要求标准应用的相关利益方参加培训。

（二）结合医学人工智能社会实验试点推广标准

主管部门可以将本文件的实施纳入本市医学人工智能治理实验试点工作的年度计划中，组织标准编制单位制作标

准解读与说明材料，鼓励试点单位结合实际情况开展标准的实施工作。

（三）开展督导检查与交流反馈

由主管部门制定督导检查的工作机制与工作计划，定期组织标准实施的督导检查，记录标准实施情况。同时，组织标准实施交流反馈工作会议。

八、其他需要说明的事项

无。