

DB4403

深 圳 市 地 方 标 准

DB4403/T XXX—XXXX

医学人工智能社会治理综合评价指南

Guidelines for comprehensive evaluation for medical artificial
intelligence social governance

（送审稿）

XXXX-XX-XX 发布

XXXX-XX-XX 实施

深圳市市场监督管理局 发 布

目 次

前言 II

引言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 指标选取原则 1

5 指标体系 2

6 指标内涵 3

7 参考文献..... 7

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由深圳市卫生健康委员会提出并归口。

本文件起草单位：深圳市卫生健康委员会、南方医科大学、深圳市卫生健康发展研究和数据管理中心、中国科学院自动化研究所、南方医科大学深圳医院。

本文件主要起草人：朱春艳、毛燕娜、王冬、汤昊宸、和晓峰、吴旭生、崔书亭、张冬云、吴培凯、丘奂阳、李晨程、陈澄、陈建华、徐青、郭俊毅、李胜利、陈宝颖、帅菲斐。

引 言

随着人工智能技术在医疗领域的广泛应用，医疗卫生行业面临资源配置改革、信息安全、医学伦理、数字鸿沟等诸多挑战，亟需构建一套分类管理、综合评价、贯穿医学人工智能技术全生命周期的人工智能治理评价指标体系，规范医学人工智能治理评价活动，科学研判人工智能技术运用在医疗领域中存在的潜在风险，确保人工智能技术在医疗实践中的安全、可靠、可控，更好地服务于民众健康。

本文件对当前医学人工智能技术在产品研发、使用和推广过程中涉及治理评价需要考虑的因素进行了设计，通过提供医学人工智能治理评价指标体系，为进一步开展医学人工智能治理提供依据和参考。

医学人工智能社会治理综合评价指南

1 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。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

医学人工智能 **medical artificial intelligence**

采用机器学习、逻辑推理、模式识别、自然语言处理等人工智能技术实现预期医学用途的软件系统。

注：软件系统包括智能独立软件和智能设备软件组件。

3.2

医学人工智能治理 **governance of medical artificial intelligence**

通过制定和执行原则、标准、指南与政策法规等，促进医学人工智能发展的过程。

3.3

医学人工智能治理评价 **governance evaluation of medical artificial intelligence**

通过治理评价指标体系考量医学人工智能治理的活动。

注：包括考量医学人工智能的安全与效用现状、风险问题现状、卫生经济问题现状等；不包括医学人工智能技术赋能社会治理活动的评价。

3.4

训练数据 **training data**

用于训练机器学习模型的输入数据样本子集。

[来源：GB/T 41867—2022，定义3.2.34]

3.5

医学人工智能技术内在风险 **technology risk of medical artificial intelligence**

由医学人工智能软件所处系统环境中的算法模型、训练数据、训练场景和算法运行环境等因素而引发的技术安全性风险或社会破坏性风险。

3.6

医学人工智能技术应用风险 **technical application risk of medical artificial intelligence**

因医学人工智能技术使用不当、管理不善或监管不力而造成的技术风险、伦理风险以及社会风险。

4 指标选取原则

医学人工智能治理评价指标选取原则如下：

- 可行性原则**：所选的指标简单、易采集，且成本不宜过高；
- 全面性原则**：需要从多维度、多层级考虑，使选取的指标能够全面反映评价对象的各个方面和特征；
- 代表性原则**：坚持问题导向与目标导向相结合，聚焦医学人工智能治理发展存在的主要问题和短板的典型指标；
- 时效性原则**：结合医学人工智能发展建设进程，对指标体系进行定期评估并调整完善，更好适应建设医学人工智能有效治理的需要。

5 指标体系

5.1 指标体系的构成

医学人工智能治理指标体系的结构分3个层级，一级指标3个，二级指标6个，三级指标18个。

5.2 指标体系的内容

5.2.1 一级指标体系的内容

一级指标体系包括以下内容：

- 安全与效用评价；
- 风险评价；
- 卫生经济评价。

5.2.2 二级指标体系的内容

二级指标体系包括以下内容：

- 安全评价；
- 效用评价；
- 技术内在风险评价；
- 技术应用风险评价；
- 效率评价；
- 效益评价。

5.2.3 三级指标体系的内容

三级指标体系包括以下内容：

- 数据安全；
- 隐私安全；
- 医疗安全；
- 场景渗透；
- 适用效能；
- 受众体验；
- 算法风险；
- 训练数据风险；
- 生成内容风险；
- 社会安全风险；

- 伦理风险；
- 社会经济风险；
- 成本效率；
- 规模效率；
- 配置效率；
- 经济效益；
- 社会效益；
- 健康效益。

5.3 指标体系的架构图

医学人工智能治理评价指标体系架构见图 1。

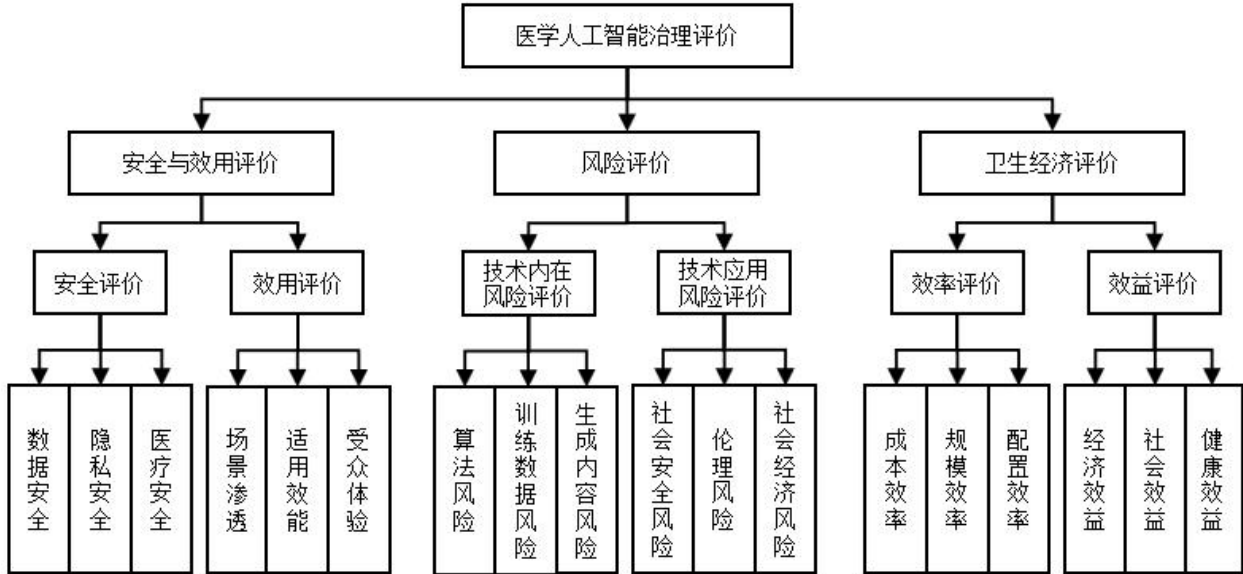


图 1 医学人工智能治理评价指标体系架构图

6 指标内涵

6.1 一级指标内涵

6.1.1 安全与效用评价

安全与效用评价是在评价对象使用过程中，对利益相关方的安全保障与受益程度进行评价的评价指标。

6.1.2 风险评价

风险评价是在评价对象研发、使用、推广的过程中，对其技术内在风险和技术应用风险进行评价的评价指标。

6.1.3 卫生经济评价

卫生经济评价是对评价对象带来的经济成本、社会成本负担与产生的效果、效用和效益进行评价的

评价指标。

6.2 二级指标内涵

6.2.1 安全评价

安全评价是对评价对象在研发、使用、推广的过程中的数据安全、隐私安全以及医疗安全进行的综合评价指标。

6.2.2 效用评价

效用评价是对评价对象在研发、使用、推广的过程中的应用场景渗透广度与深度、场景契合程度、受众体验、硬件环境影响以及适用效能进行评价的评价指标。

6.2.3 技术内在风险评价

技术内在风险评价是对评价对象系统本身所具有的威胁安全、施加影响或逃避监督的可能性进行评价的评价指标。

6.2.4 技术应用风险评价

技术应用风险评价是对评价对象系统本身以外对社会、伦理、法制等其他方面产生的伤害事件预期、系统对人的控制程度和对人工智能系统结果的控制及纠错程度等进行评价的评价指标。

6.2.5 效率评价

效率评价是采用数据包络分析方法对评价对象应用于医疗领域前后对医疗服务供给的技术效率、规模效率、成本效率、配置效率、规模收益状态以及医疗资源配置状态产生的影响进行评价的评价指标。

6.2.6 效益评价

效益评价是对评价对象产生的经济效益、社会效益以及健康效益进行评价的评价指标。

6.3 三级指标内涵

6.3.1 数据安全

数据安全是对评价对象开发与应用过程中涉及健康医疗数据的安全管理的评价指标。

注：健康医疗数据包括个人健康医疗数据以及由个人健康医疗数据加工处理之后得到的健康医疗相关电子数据。例如经过对群体健康医疗数据处理后得到的群体总体医疗数据分析结果、趋势预测、疾病防治统计数据等。

6.3.2 隐私安全

隐私安全是对评价对象保护个人敏感信息的处置方式进行评价的评价指标。

注：个人敏感信息包括身份证件号码、个人生物识别信息、银行账户、通信记录和内容、财产信息、征信信息、行踪轨迹、住宿信息、健康生理信息、交易信息、14 岁以下（含）儿童的个人信息等。

6.3.3 医疗安全

医疗安全是对评价对象在医学场景应用过程中对医疗卫生服务质量带来的影响进行评价的评价指标。

6.3.4 场景渗透

场景渗透是对评价对象在不同等级医院中使用的广度以及在不同临床学科中使用深度进行评价的评价指标。

注：反映该评价对象在社会治理中的难度。

6.3.5 适用效能

适用效能是对评价对象在医学场景应用过程中，与医疗业务情境和业务流程的场景契合程度进行评价的评价指标。

6.3.6 受众体验

受众体验是对评价对象的用户依赖程度进行评价的评价指标。

注：反映该评价对象在社会治理中的优先程度。

6.3.7 算法风险

算法风险是对评价对象的算法模型稳健性、泛化性、鲁棒性、可解释性、可信任性等进行测试，以评价算法的不确定风险的评价指标。

6.3.8 训练数据风险

训练数据风险是对评价对象所使用训练数据的科学性、有效性、合规性以及制定不良训练数据限制规则和防范毒性训练数据的沾染等方面进行评价的评价指标。

6.3.9 生成内容风险

生成内容风险是对评价对象的输出结果是否有包含违反社会主义核心价值观、歧视性、商业违法违规、侵犯他人合法权益的内容和是否满足特定服务类型方面进行评价的评价指标。

6.3.10 社会安全风险

社会安全风险是对评价对象在技术使用中潜在的社会安全风险进行评价的评价指标。

6.3.11 伦理风险

伦理风险是对评价对象在社会伦理方面影响社会公平与稳定的潜在风险进行评价的评价指标。

6.3.12 社会经济风险

社会经济风险是对评价对象在社会经济方面影响社会与对经济发展的潜在风险进行评价的评价指标。

6.3.13 成本效率

成本效率是基于数据包络分析法的COST模型对评价对象应用于医疗领域前后带来医疗资源投入成本变化对医疗服务供给产出的影响进行评价的评价指标。

6.3.14 规模效率

规模效率是基于数据包络分析法的BCC模型对评价对象应用于医疗领域前后医疗资源投入规模变化对医疗服务供给的规模收益状况的影响进行评价的评价指标。

6.3.15 配置效率

配置效率是基于数据包络分析方法COST模型与BCC模型对评价对象应用于医疗领域前后医疗资源投入成本与规模变化对医疗服务供给产出最优状态的影响进行评价的评价指标。

6.3.16 经济效益

经济效益是对评价对象在医疗卫生服务供给方面带来的经济边际效益影响进行评价的评价指标。

6.3.17 社会效益

社会效益是对评价对象在医疗卫生领域推广使用后,对社会人群健康和社会健康发展带来的影响进行评价的评价指标。

6.3.18 健康效益

健康效益是对评价对象在医疗卫生领域推广使用后,对某区域人群、不同年龄结构人群、某疾病患病人群健康状况的影响进行评价的评价指标。

参 考 文 献

[1] GB/T 35273—2020 信息安全技术 个人信息安全规范

[2] GB/T 37373—2019 智能交通 数据安全服务

[3] GB/T 39725—2020 信息安全技术 健康医疗数据安全指南

[4] GB/T 41867—2022 信息技术 人工智能 术语

[5] JR/T 0221—2021 人工智能算法金融应用评价规范

[6] JR/T 0223—2021 金融数据安全 数据生命周期安全规范

[7] JR/T 0197—2020 金融数据安全 数据安全分级指南

[8] JRT 0287—2023 人工智能算法金融应用信息披露指南

[9] MZ/T 165—2020 居民家庭经济状况核对 数据安全要求

[10] NY/T 4261—2022 农业大数据安全管理指南

[11] YY/T 1833（所有部分） 人工智能医疗器械 质量要求和评价

[12] YY/T 1858—2022 人工智能医疗器械 肺部影像辅助分析软件 算法性能测试方法

[13] YD/T 3801—2020 电信网和互联网数据安全风险评估实施方法

[14] YD/T 3865—2021 工业互联网数据安全保护要求

[15] YD/T 3956—2021 电信网和互联网数据安全评估规范

[16] YD/T 4043—2022 基于人工智能的多中心医疗数据协同分析平台参考架构

[17] YD/T 4070—2022 基于人工智能的接入网运维和业务智能化场景与需求

[18] YD/T 4392.1—2023 人工智能开发平台通用能力要求 第1部分：功能要求

[19] YD/T 4921—2024 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

[20] YD/T 4960—2024 移动智能终端可信人工智能安全指南

[21] DB11/T 2251—2024 信息安全 人工智能数据安全通用要求

[22] DB52/T 1726—2023 糖尿病视网膜病变人工智能筛查应用规范

[23] 全国人民代表大会常务委员会. 中华人民共和国个人信息保护法：主席令〔2021〕91号. 2021年

[24] 国家卫生健康委员会规划与信息司，国家卫生健康委员会统计信息中心. 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，2018年

[25] 国家卫生健康委，国家中医药管理局. 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. 2019年

[26] 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. 生成式人工智能服务管理暂行办法：国家广播电视总局令第15号. 2023年

[27] 国家卫生健康委办公厅. 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. 2023年

[28] 深圳市第七届人民代表大会常务委员会. 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）. 2022年

[29] 全国网络安全标准化技术委员会. 生成式人工智能服务安全基本要求. 2024年