ICS 点击此处添加 ICS 号
CCS 点击此处添加 CCS 号

T/ZSMM

浙 江 省 数 理 医 学 学 会 团 体 标 准

T/ZSMM XXXX—2025

医学人工智能治理综合评价指南 第 2 部分：安全评价

Guideline for comprehensive evaluation of medical artificial intelligence social governance —Part 2: Security assessment

（征求意见稿）

（本草案完成时间：2025 年 11 月 28 日）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

浙江省数理医学学会　　发 布

目　　次

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。

《医学人工智能治理综合评价指南》分为以下六个部分：

——第 1 部分：总则；

——第 2 部分：安全评价；

——第 3 部分：风险评价；

——第 4 部分：效用评价；

——第 5 部分：效率评价；

——第 6 部分：效益评价；

本部分为《医学人工智能治理综合评价指南》的第 2 部分。

本文件由浙江省数理医学学会提出并归口。

本文件起草单位：南方医科大学、南方科技大学、深圳市卫生健康委员会、深圳市人民医院、浙江数字内容研究院、深圳市卫生健康发展研究与数据管理中心、中国医学科学院医学信息研究所、阜外华中心血管医院、上海市第六人民医院、南方医科大学第三附属医院、南方医科大学珠江医院、南方医科大学第八附属医院、南方医科大学南方医院赣州医院、南方医科大学中西医结合医院、东莞市石碣医院、广东医科大学附属医院、深圳市第四人民医院、深圳市妇幼保健院、四川大学华西第二医院、香港大学深圳医院、中山市人民医院、珠海市人民医院、佛山市第一人民医院、前海人寿广州总医院、广州市红十字会医院、北京大学深圳医院。

本文件主要起草人：毛燕娜、王冬、姜虹、朱春艳、耿庆山、汤昊宬、丁万夫、郑静、崔书亭、张冬云、李晨程、曹艳林、刘咏梅、许彬彬、陈宝颖、潘鑫、吴超梅、王亚琴、郭洪波、曹蓓、戴辉、杜庆锋、刘仲文、蔡定彬、王诚、李笑天、张少毅、徐小平、黄晓星、郭煜、段光荣、周宏峰、张立贤、赵永胜。

医学人工智能治理综合评价指南
第 2 部分：安全评价

1 范围

本文件提供了开展医学人工智能治理综合评价中关于安全维度评价的指导，给出了安全维度涉及的权利、需要考虑的评价要素，以及评价执行的相关信息。

本文件适用于开展医学人工智能对安全治理产生的现实或潜在影响的综合评价活动。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 39725 信息安全技术 健康医疗数据安全指南

DB4403/T 634—2025 医学人工智能治理综合评价指标体系

T/ZSMM XXXX—XXXX 医学人工智能治理综合评价指南 第1部分：总则

3 术语和定义

T/ZSMM XXXX—XXXX界定的术语和定义适用于本文件。

4 安全评价指标体系内容

4.1 指标体系架构图

安全评价根据T/ZSMM XXXX—XXXX提供的建议，作为医学人工智能治理综合评价指标体系的一级指标，可再下设两个层级的评价指标，包括二级评价指标3个，三级评价指标19个，见图1。

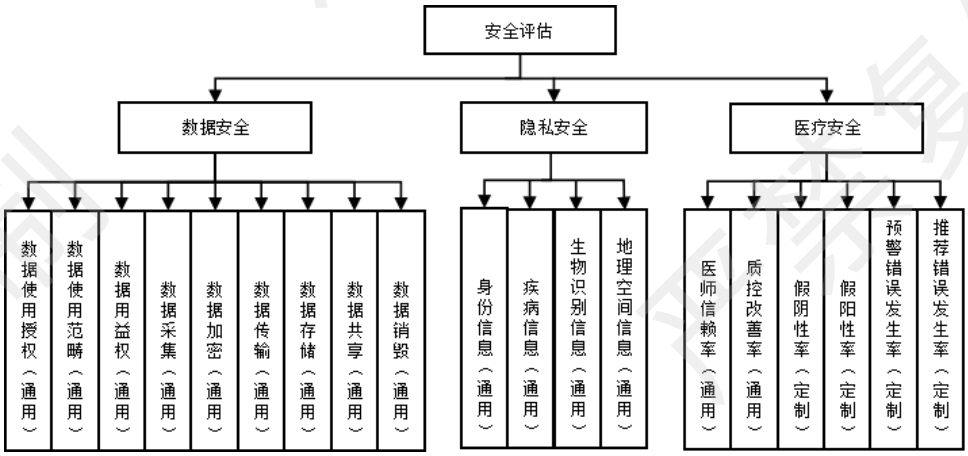

图1 医学人工智能治理综合评价中安全评价指标体系架构图

4.2 二级指标体系的内容

二级指标体系宜包括以下内容：

——数据安全；

——隐私安全；
——医疗安全；

4.3 三级指标体系的内容

三级指标体系可包括以下内容：
——数据使用授权；
——数据使用范畴；
——数据用益权；
——数据采集；
——数据加密；
——数据传输；
——数据存储；
——数据共享；
——数据销毁；
——身份信息；
——疾病信息；
——生物识别信息；
——地理空间信息；
——医师信赖率；
——质控改善率；
——假阴性率；
——假阳性率；
——预警错误发生率；
——推荐错误发生率。

5 安全评价指标内涵

5.1 二级指标

5.1.1 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。

5.1.2 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。

5.1.3 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。

5.2 三级指标

5.2.1 数据使用授权

描述数据所有权代理人或代理机构在数据来源符合法律规定的前提下，将数据的使用权合法授予数据使用方。属于通用性因素，宜涵盖以下内容：
——该软件算法模型所使用数据是否具备数据所有权代理人或代理机构的合法授权证明；
——该软件算法模型所使用数据的授权证明的授权过程是否合理合规；
——该软件算法模型所使用数据是否存在扩大数据使用权限范畴的不合理情况。

授权使用的数据量宜满足国家与行业主管部门要求，同时在不改变该数据相关权利与义务的前提下进行使用授权。

5.2.2 数据使用范畴

描述医学人工智能软件开发、使用以及测试的过程中所使用的相关数据要符合我国《数据安全法》、《网络安全法》以及《个人信息保护法》所规定的使用范畴。属于通用性因素，宜涵盖以下内容：

——该软件算法模型数据使用范畴协议是否满足相关政策法规规范；

——在追溯和监管该软件算法模型数据使用情况时是否存在超范畴使用情况；

——该软件算法模型数据使用过程中数据使用者是否采取有效的数据安全保护与监管措施。

5.2.3 数据用益权

描述数据所有权代理人在转让数据使用权后通过签订数据用益分配的协议或合同促进在医学人工智能软件开发中对数据进行积极开发的同时商定数据复利的收益分配规则，侧重于监管数据使用方之间开展数据使用权和收益权相互转让的合法性。属于通用性因素，宜涵盖以下内容：

——该软件算法模型里的数据使用前是否签订用益权合同，详细说明数据使用与权益分配情况、明确利益关系与数据使用权力转让的条件；

——该软件算法模型跨部门使用数据时是否征求医疗机构同意并签订数据用益权转让合同。

5.2.4 数据采集

描述医学人工智能软件开发、使用以及测试过程中在进行数据收集的过程。属于通用性因素，宜涵盖以下内容：

——该软件算法模型使用过程中，个人信息主体主动提供个人信息的数据量说明；

——该软件算法模型使用过程中，通过个人信息主体交互或记录个人信息主体行为等自动采集的数据量说明；

——该软件算法模型使用过程中，通过共享、转让、搜集公开信息等间接获取个人信息的数据量说明；

——如果产品或服务的提供者提供工具供个人信息主体使用，提供者不对个人信息进行访问的，则不属于本标准所称的收集。

5.2.5 数据加密

描述医学人工智能软件开发、使用以及测试过程中在进行数据传输、运算、测试或存储前时使用加密技术或隐私计算技术进行处理。属于通用性因素，宜涵盖以下内容：

——该软件算法模型使用过程中数据存储使用的加密技术说明；

——该软件算法模型使用过程中数据传输使用的加密技术说明；

——该软件算法模型使用过程中数据运算的隐私计算技术说明。

5.2.6 数据传输

描述医学人工智能软件开发、使用、测试过程中涉及医疗健康数据从一个实体发送至另一个实体的过程，存在数据传输中断、篡改、伪造及窃取等安全风险。属于通用性因素，宜涵盖以下内容：

——该软件算法模型开发安全管理，保障数据传输工具的安全性，开展必要的源码安全审计、三方组件安全评审、渗透测试、支持库漏洞查找等工作；

——采用防火墙、入侵检测等安全技术或设备，确保数据传输网络安全性；

——不同网络区域或者安全域之间宜进行安全隔离和访问控制；

——终端宜采取准入控制、终端鉴别等技术措施，防止非法或未授权终端接入内部网络；

——宜对通信双方进行身份确认，确保数据传输双方是可信任的；

——宜采用数字签名、时间戳等方式，确保数据传输的抗抵赖性；

——宜采用密码技术或非密码技术等方式，确保数据完整性；

——宜选用安全的密码算法，可使用国密序列 SM3、SM4 等，不宜使用如 MD5、DES-CBC、SHAI 等不安全的算法。

5.2.7 数据存储

描述使用医学人工智能软件的医疗卫生机构在提供医疗服务、医疗业务运营等活动中，将医疗健康数据进行持久化保存的过程，包括但不限于采用磁盘、磁带、云存储服务、网络存储设备等载体存储数据。数据存储过程，可能存在数据泄漏、篡改、丢失、不可用等安全风险。属于通用性因素，宜涵盖以下内容：

——宜根据数据安全级别、重要性、量级、使用频率等因素，将数据分域分级存储；
——宜定期对数据存储过程中可能产生的影响进行风险评价，并采取相应安全防护措施；
——脱敏后的数据宜与用于还原数据的恢复文件隔离存储，使用恢复原始数据的技术宜经过严格审批，并留存相关审批及操作记录；
——宜采取一定措施确保数据存储的完整性。存储 3 级及以上数据时，宜采用密码技术、权限控制等技术措施保证数据完整性，3 级及以上数据可按照 GB/T 39725 中的 6.2 划分数据分级。

5.2.8 数据共享

描述医学人工智能软件开发、使用以及测试过程中对数据集进行完全公开共享、受控公开共享以及领地公开共享的情况。属于通用性因素，宜涵盖以下内容：

——该软件算法模型使用过程中数据集通过互联网直接公开发布的数据量的情况说明；
——该软件算法模型使用过程中数据集通过数据使用协议对数据的使用进行约束的数据量的情况说明；
——该软件算法模型使用过程中数据集在物理或者虚拟的领地范围内共享，数据不能流出到领地范围外的数据量的情况说明。

5.2.9 数据销毁

描述医学人工智能软件开发、使用以及测试后，依据《个人信息保护法》的相关规定，在实现日常业务功能所涉及的系统中去除个人信息的行为，使其保持不可被检索与访问的状态。属于通用性因素，宜涵盖以下内容：

——该软件算法模型使用后，个人信息数据销毁的技术说明；
——该软件算法模型使用后，个人信息数据销毁的周期说明；
——该软件算法模型使用后，个人信息数据销毁的合规说明。

5.2.10 身份信息

描述社会身份信息，如身份证号、医保识别号、职业、单位、家庭住址、等，根据《个人信息保护法》，需对以上信息进行保护。属于通用性因素，宜涵盖以下内容：

——该软件在收集身份信息前，宜通过合同协议等方式，明确双方在数据安全方面的责任与义务；
——该软件在收集身份信息前，宜通过合同协议等方式，明确数据采集范围、频度、类型、用途等；
——该软件在收集身份信息前，宜提供相关个人身份信息主体的授权；
——宜对身份信息数据访问权限和实际访问控制情况进行审计，宜每半年 1 次对访问权限规则和已授权清单进行复核，及时清理已失效的账号和授权；
——身份信息数据不宜导出，确需导出宜使用加密、脱敏等技术手段防止数据泄漏，同时宜经卫生行政机构高级管理层批准，并配套数据跟踪溯源机制；
——通过数据溯源方法从算法模型中推断出使用的数据的身份信息的情况同样适用以上的规定。

5.2.11 疾病信息

描述患者疾病诊疗相关信息，根据《个人信息保护法》，需对该信息进行保护。属于通用性因素，宜涵盖以下内容：

——该软件在收集疾病信息前，宜通过合同协议等方式，明确双方在数据安全方面的责任与义务；

——该软件在收集疾病信息前，宜通过合同协议等方式，明确数据采集范围、频度、类型、用途等；该软件在收集疾病信息前，宜制定数据供应方约束机制，并事前开展数据安全影响评价，针对数据处理活动，检验其合法合规程度，判断其对相关方合法权益造成损害的各种风险，以及评价相关保护措施有效性的过程；
——宜对疾病信息数据访问权限和实际访问控制情况进行审计，宜每半年 1 次对访问权限规则和已授权清单进行复核，及时清理已失效的账号和授权；
——疾病信息数据不宜导出，确需导出宜使用加密、脱敏等技术手段防止数据泄漏，同时宜经卫生行政机构高级管理层批准，并配套数据跟踪溯源机制。

5.2.12 生物识别信息

描述个人基因、指纹、声纹、掌纹、耳廓、虹膜、面部识别特征等，根据《个人信息保护法》，需对以上信息进行保护。属于通用性因素，宜涵盖以下内容：
——该软件在收集生物识别信息前，宜通过合同协议等方式，明确双方在数据安全方面的责任与义务；
——该软件在收集生物识别信息前，宜通过合同协议等方式，明确数据采集范围、频度、类型、用途等；
——该软件在收集生物识别信息前，宜制定数据供应方约束机制，并事前开展数据安全影响评价，针对数据处理活动，检验其合法合规程度，判断其对相关方合法权益造成损害的各种风险，以及评价相关保护措施有效性的过程；
——宜对生物识别信息数据访问权限和实际访问控制情况进行审计，宜每半年 1 次对访问权限规则和已授权清单进行复核，及时清理已失效的账号和授权；
——生物识别信息数据不宜导出，确需导出宜使用加密、脱敏等技术手段防止数据泄漏，同时宜经卫生行政机构高级管理层批准，并配套数据跟踪溯源机制。

5.2.13 地理空间信息

描述GPS定位地理信息、居住地址、工作单位地址、出生地地址等信息，根据《个人信息保护法》，需对该信息进行保护。属于通用性因素，宜涵盖以下内容：
——该软件在收集地理空间信息前，宜通过合同协议等方式，明确双方在数据安全方面的责任与义务；
——该软件在收集地理空间信息前，宜通过合同协议等方式，明确数据采集范围、频度、类型、用途等；
——该软件在收集地理空间信息前，宜制定数据供应方约束机制，并事前开展数据安全影响评价，针对数据处理活动，检验其合法合规程度，判断其对相关方合法权益造成损害的各种风险，以及评价相关保护措施有效性的过程；
——宜对地理空间信息数据访问权限和实际访问控制情况进行审计，宜每半年 1 次对访问权限规则和已授权清单进行复核，及时清理已失效的账号和授权。

5.2.14 医师信赖率

描述医学人工智能软件介入医疗服务供给过程后，医师支持使用并不愿意放弃使用该软件的百分比。属于通用性因素，宜涵盖以下内容：
——该软件介入后，不同临床学科医师支持使用的情况；
——该软件介入后，不同临床学科医师放弃使用的情况；
——该软件介入后，不同年资医师支持使用的情况；
——该软件介入后，不同年资医师放弃使用的情况。

5.2.15 质控改善率

描述医学人工智能软件介入医疗服务供给过程后，病历质量改善的情况。属于通用性因素，宜涵盖以下内容：
——该软件介入后，某临床学科科室病历质量的错误发生减少情况；
——该软件介入后，某临床学科科室医疗失误发生减少情况。

5.2.16 假阴性率

描述在医学人工智能软件介入疾病诊断辅助决策过程，患者实际患病，但根据辅助决策系统被判为无病的百分比，又称误诊率或第Ⅰ类错误。属于疾病诊断辅助决策场景定制性因素，宜涵盖以下内容：

——假阴性人数与金标准阳性的比率；

——在医学人工智能软件介入疾病诊断辅助决策下，某科室某类疾病出现假阴性率的情况。

5.2.17 假阳性率

描述在医学人工智能软件介入疾病诊断辅助决策过程，患者实际无病，但根据辅助决策系统被判为患病的百分比，又称漏诊率或第Ⅱ类错误。属于疾病诊断辅助决策场景定制性因素，宜涵盖以下内容：

——假阳性人数与金标准阴性人数的比率；

——在医学人工智能软件介入疾病诊断决策过程下，某科室某类疾病出现假阳性率的情况。

5.2.18 预警错误发生率

描述在医学人工智能软件介入患者健康管理决策过程，患者实际健康状况平稳，但根据辅助决策系统被判为需医疗处置状况的百分比。属于健康管理决策场景定制性因素，宜涵盖以下内容：

——该软件介入后，患者健康管理过程中错误预判的情况；

——该软件介入后，患者疾病诊疗处置错误预判的情况。

5.2.19 推荐错误发生率

描述在医学人工智能软件介入医疗服务供给过程，根据辅助决策系统推荐错误诊断、错误治疗处置、错误用药等情况的百分比。属于疾病诊断辅助决策场景定制性因素，宜涵盖以下内容：

——该软件介入后，患者疾病诊疗过程中错误预判疾病诊断的情况；

——该软件介入后，患者疾病诊疗过程中错误预判疾病治疗处置的情况；

——该软件介入后，患者疾病诊疗过程中错误预判用药方案的情况。

参 考 文 献

[1] GB/T 20001.8—2023 标准起草规则 第8部分：评价标准
[2] GB/T 35273—2020 信息安全技术 个人信息安全规范
[3] GB/T 37373—2019 智能交通 数据安全服务
[4] GB/T 39725—2020 信息安全技术 健康医疗数据安全指南
[5] GB/T 41867—2022 信息技术 人工智能 术语
[6] GB/T 44811—2024 物联网 数据质量评价方法
[7] GB/T 45288.2—2025 人工智能 大模型 第2部分：评测指标与方法
[8] GB/T 45288.3—2025 人工智能 大模型 第3部分:服务能力成熟度评估
[9] GB/T 45654—2025 网络安全技术 生成式人工智能服务安全基本要求
[10] GB/T 45674—2025 网络安全技术 生成式人工智能数据标注安全规范
[11] GB/T 45652—2025 网络安全技术 生成式人工智能预训练和优化训练数据安全规范
[12] GB/T 45438—2025 网络安全技术 人工智能生成合成内容标识方法
[13] GB/T 46071—2025 数据安全技术 数据安全和个人信息保护社会责任指南
[14] GB/T 45577—2025 数据安全技术 数据安全风险评估方法
[15] GB/T 45574—2025 数据安全技术 敏感个人信息处理安全要求
[16] GB/T 46347—2025 人工智能 风险管理能力评估
[17] JR/T 0197—2020 金融数据安全 数据安全分级指南
[18] JR/T 0221—2021 人工智能算法金融应用评价规范
[19] JR/T 0223—2021 金融数据安全 数据生命周期安全规范
[20] JR/T 0287—2023 人工智能算法金融应用信息披露指南
[21] MZ/T 165—2020 居民家庭经济状况核对 数据安全管理要求
[22] NY/T 4261—2022 农业大数据安全管理指南
[23] YD/T 3801—2020 电信网和互联网数据安全风险评估实施方法
[24] YD/T 3865—2021 工业互联网数据安全保护要求
[25] YD/T 3956—2021 电信网和互联网数据安全评估规范
[26] 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
[27] YD/T 4960—2024 移动智能终端可信人工智能安全指南
[28] YY/T 1833（所有部分） 人工智能医疗器械 质量要求和评价
[29] 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
[30] DB11/T 2251—2024 信息安全 人工智能数据安全通用要求
[31] DB37/T 4845—2025 人工智能技术应用伦理风险的治理要求
[32] DB52/T 1726—2023 糖尿病视网膜病变人工智能筛查应用规范
[33] DB4403/T 643—2025 医学人工智能治理综合评价指标体系
[34] European Union.General Data Protection Regulation[Z].Geneva:EU,2018
[35] 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
[36] 全国人民代表大会常务委员会.中华人民共和国数据安全法：主席令〔2021〕84号.2021年
[37] 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
[38] 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
[39] 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
[40] 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

[41] 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