《医学人工智能治理综合评价指南 第 1 部分：总则》（征求意见稿）编制说明

《医学人工智能治理综合评价指南　第 1 部分：总则》

标准编制组

二〇二五年十一月

目 录

一、 编制的目的和意义.....1
（一）研究背景.....1
（二）编制目的.....3
二、 任务来源及编制原则和依据.....4
（一）任务来源.....4
（二）编制原则.....4
（三）编制依据.....4
三、 编制过程.....5
四、 主要内容的确定.....7
（一）范围.....7
（二）规范性引用文件.....7
（三）术语和定义.....7
（四）基本原则.....8
（五）评价机构和人员.....8
（六）评价流程.....9
（七）评价指标体系.....9
五、 采标情况.....11
六、 重大分歧意见的处理.....11
七、 与国家法律法规和强制性标准的关系.....11
八、 标准实施的建议.....12
九、 其他应予说明的事项.....12

一、 编制的目的和意义

（一）研究背景

近年来，人工智能技术在全球范围内迅速发展，尤其在医学领域展现出广泛的应用前景与深刻的治理需求。2017 年，国务院印发《新一代人工智能发展规划》，明确提出分阶段推进人工智能技术与产业融合的战略目标。2021 年，《新一代人工智能伦理规范》发布，进一步将伦理治理贯穿于人工智能全生命周期，为我国人工智能的负责任发展提供了制度保障。2023 年，习近平主席在第三届“一带一路”国际合作高峰论坛上提出《全球人工智能治理倡议》，倡导构建“以人为本、智能向善、公平普惠”的全球治理体系，标志着我国在人工智能全球治理中发挥日益重要的引领作用。至 2025 年，“人工智能+”被写入《政府工作报告》，成为推动高质量发展的重要路径；同期发布的《人工智能示范法 3.0》及 2025 世界人工智能大会形成的《人工智能全球治理行动计划》，进一步强化了法治保障与国际协作共识，为我国医学人工智能的治理实践提供了顶层设计依据。

世界卫生组织于 2024 年发布的《Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models》，系统阐释了

医疗人工智能在伦理与治理方面的关键挑战与应对原则。该指南强调，必须通过建立系统化治理机制，落实包括保护自主性、促进人类福祉、强化透明与问责等六大核心原则，以实现人工智能在医疗场景中的可信、可控与可持续应用。这为我国构建本土化医学人工智能治理体系提供了重要参考。

我国医学人工智能实践层面，相关技术已在疾病预测、辅助诊疗、药物研发等环节实现广泛应用，同时也带来了数据安全、算法偏见、隐私泄露、权责界定不清等新型风险。为应对上述挑战，我国已初步构建涵盖国家标准、行业标准与地方标准的多层次治理体系。国家标准有《网络安全技术 人工智能生成合成内容标识方法》（GB 45438-2025）、《网络安全技术 生成式人工智能数据标注安全规范》（GB/T 45674-2025）及《网络安全技术 生成式人工智能服务安全基本要求》（GB/T 45654-2025)等标准规范，与地方标准《医学人工智能治理综合评价指标体系》（DB4403/T 634—2025）、《信息安全 人工智能数据安全通用要求》（DB11/T 2251-2024）及行业标准《移动智能终端可信人工智能安全指南》（YD/T 4960-2024）互为补充，共同在数据安全、模型治理、终端部署等方面形成了一定的制度基础。

然而，现有标准体系在系统性、协同性与场景适配性方面仍存在不足，尤其在面向医疗机构的人工智能系统影响综合评价方面，尚未形成统一、可操作的实施指南。因此，亟需制定一部专门针对医学人工智能领域的治理综合评价指南，建立科学、系统、可推广的评价框架，明确评价的基本原则、组织机制、评价流程与核心内容，为实现医学人工智能的规范化管理、防范医疗安全风险、促进技术负责任落地提供标准化支撑，也为全国范围内相关机构的治理实践提供参考范式与实施路径。

（二）编制目的

为促进医学人工智能安全、稳定及可持续发展，贯彻落实国家《新一代人工智能发展规划》、《全球人工智能治理倡议》等战略部署，响应“人工智能+”行动与全球治理合作号召，南方医科大学牵头多家单位，依托浙江省数理医学学会平台，共同发起《医学人工智能治理综合评价指南 第1部分：总则》的编制工作。在于构建一套科学、系统、可操作的综合评价指南，明确医学人工智能治理评价的基本原则、组织机制、工作流程与核心内容，为各类医疗卫生机构、技术开发主体和监管单位提供统一、规范的评价指引。通过医学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价结果为规范建设人工智能治理体系提供决策性参考，助力浙江省医学人工智能产业健康发展。

二、 任务来源及编制原则和依据

（一）任务来源

本标准编制任务来源于浙江省数理医学学会于 2025 年 5 月 2 日下达的浙数医 [2025]11 号关于批准《医学人工智能治理综合评价指南　第 1 部分：总则》等两项团体标准立项的通知，归口单位为浙江省数理医学学会，标准名称为《医学人工智能治理综合评价指南　第 1 部分：总则》，项目编号:ZSMM—2025—004。

（二）编制原则

本标准的制定工作遵循“统一性、协调性、适用性、一致性、规范性”原则，本着先进性、科学性、合理性和可操作性的原则，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》给出的规则编写。

（三）编制依据

本文件的编制主要参考与依据以下文件：

1.《标准起草规则　第 8 部分：评价标准》（GB/T 20001.8—2023 ）

2.《信息安全技术　健康医疗数据安全指南》（GB/T 39725—2020）

3.《医学人工智能治理综合评价指标体系 》（DB4403/T 643—2025）

4.《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）

5.《网络安全技术 生成式人工智能服务安全基本要求》（GB/T 45654—2025)

6.《信息技术服务 从业人员能力评价要求》（GB/T 37696—2019）

7.《网络安全技术 人工智能生成合成内容标识方法》（GB/T 45438—2025）

8.《数据安全技术 数据安全和个人信息保护社会责任指南》（GB/T 46071—2025）

9.《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）

三、 编制过程

1、实地调研阶段,2025 年 1 月至 2 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。

2、规划准备阶段,2025 年 2 月 21 日编制组正式成立，由医学人工智能领域、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、数据安全领域的学者专家、行政管理者、卫生技术人员、工程师等组成。编制组制定了详细的编制计划方案，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，分析适宜医学人工智能治理的规范性要素、技术要点和框架结构。

3、标准起草阶段,2025 年 3 月至 4 月，在充分调研和理论研究的基础上，编制组开始标准文本的起草工作，形成了《医学人工智能治理综合评价指南　第 1 部分：总则》工作组讨论稿。

4、申请立项阶段，2025 年 4 月 5 日，标准编制组向浙江省数理医学学会递交团体标准立项申请表，于 2025 年 4 月 9 日收到受理通知书。

5、立项论证阶段，2025 年 4 月 24 日，浙江省数理医学学会标准化工作委员会组织召开立项论证会，《医学人工智能治理综合评价指南　第 1 部分：总则》通过立项论证评审，经公示，于 2025 年 5 月 2 日成功获批立项。

6、标准研制阶段，2025 年 5 月至 11 月，标准编制组根据专家意见，经过多轮研讨，对《医学人工智能治理综合评价指南　第 1 部分：总则》工作组讨论稿进行修改，于

2025 年 11 月 28 日完成《医学人工智能治理综合评价指南 第 1 部分：总则》征求意见稿与编制说明。

四、 主要内容的确定

本文件的重要技术内容系基于对国家相关政策法规的系统研究、对国内外相关标准的参考借鉴，并经由起草组专家多轮专题研讨后最终确定。《医学人工智能治理综合评价指南 第 1 部分：总则》有七个章节和参考文献。其中主要内容包括范围、规范性引用文件、术语和定义、基本原则、评价机构和人员、评价流程、评价指标体系。

（一）范围

本文件确立了开展医学人工智能治理综合评价的基本原则，提供了评价机构和人员、评价流程和评价内容等方面的指导与建议。

本文件适用于组织开展医学人工智能技术对社会活动产生的现实或潜在影响进行综合评价。

（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的术语与定义。

本文件第三章 术语与定义，编制依据情况如下表。

术语与定义	依据
3.1 医学人工智能治理	依据《医学人工智能治理综合评价指标体系》(DB4403/T 634—2025)，定义 3.1、《Ethics and governance of artificial Intelligence for health Guidance on large multi-modal models》（World Health Organization，2024）以及专家研讨达成的共识
3.2 医学人工智能治理评价	依据《医学人工智能治理综合评价指标体系》(DB4403/T 634—2025)，定义 3.2 和专家研讨达成的共识

（四）基本原则

本章所确立的医学人工智能治理综合评价基本原则，包括安全有益、动态评价、短期与长期结合以及整体与部分结合原则。其形成主要依据《医学人工智能治理综合评价指标体系》（DB4403/T 634—2025）、《网络安全技术 生成式人工智能服务安全基本要求》（GB/T 45654—2025）及《医学人工智能治理综合评价指标体系》（DB4403/T 643—2025）等相关文件，并在系统分析评价工作内在需求的基础上，经由起草组专家研讨审议后达成共识。

（五）评价机构和人员

本章评价机构和人员，根据医学人工智能治理综合评价过程中所涉及到的相关机构和人员和专家研讨，经过研究确定评价机构一般由相关利益方和相关监管与治理部门共同组成。评价人员部分依据《中华人民共和国职业分类大典（2025 版）》、《信息技术服务 从业人员能力评价

要求》（GB/T 37696—2019)建立“职业种类”和“职业等级分级方法”二维度的评价人员能力模型，对医学人工智能治理综合评价人员的职业种类、等级及要求进行细化与量化。

（六）评价流程

本章评价流程，根据医学人工智能治理综合评价过程中所涉及的流程，经专家研讨确定医学人工智能治理综合评价可按照以下程序进行：成立评价组，描述评价对象符合性评估，按评价内容模块分类评价，评价材料收集，形成评价意见与结论，资料归档，评价结果管理与应用。

（七）评价指标体系

本章节给出了医学人工智能治理综合评价指标体系，医学人工智能治理指标体系由两个层级的评价指标构成，一级评价指标 5 个，二级评价指标 18 个。其中一级指标包括安全评价、风险评价、效用评价、效率评价及效益评价，并且医学人工智能技术发展的不同时期，形成事前、事中、事后的动态评估医学人工智能社会治理综合评价体系。

本文件第七章 评价指标体系，编制依据情况如下表。

一级指标	二级指标	依据
7.1安全评价	数据安全	《数据安全技术 数据安全风险评估方法》（GB/T 45577—2025） 《信息安全 人工智能数据安全通用要求》(DB11/T 2251—2024） 《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》（GB/T 45652—2025） 《网络安全技术 生成式人工智能数据标注安全规范》（GB/T

		45674—2025） 《物联网　数据质量评价方法》（GB/T 44811—2024） 《医学人工智能治理综合评价指标体系》(DB4403/T 634—2025) 《信息安全技术个人信息安全规范》（GB/T35273—2020） 《智能交通数据安全服务》（GB/T37373—2019） 《信息安全技术健康医疗数据安全指南》（GB/T39725—2020） 《金融数据安全数据生命周期安全规范》（JR/T0223—2021） 《电信网和互联网数据安全评估规范》（YD/T3956—2021） 《人工智能开发平台通用能力要求第 1 部分：功能要求》（YD/T4392.1—2023）
	隐私安全	《数据安全技术 敏感个人信息处理安全要求》（GB/T 45574—2025） 《数据安全技术 数据安全和个人信息保护社会责任指南》(GB/T 46071—202） 《医学人工智能治理综合评价指标体系》(DB4403/T 634—2025) 《信息安全技术个人信息安全规范》（GB/T35273—2020） 《信息安全技术健康医疗数据安全指南》（GB/T39725—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/T1833.3—2022）
	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。
7.3风险评价	社会安全风险 伦理风险 社会经济风险	《医学人工智能治理综合评价指标体系》(DB4403/T 634—2025) 《人工智能技术应用伦理风险的治理要求》（DB37/T 4845—2025） 《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/T4043—2022） 《移动智能终端可信人工智能安全指南》（YD/T4960—2024） 《Ethics and Governance of Artificial Intelligence for Health》（World Health Organization,2021） 《Regulatory considerations on artificial intelligence for health》(World Health Organization,2023)
	模型风险 训练数据风险 生成内容风险	《医学人工智能治理综合评价指标体系》(DB4403/T 634—2025) 《人工智能　大模型　第 2 部分：评测指标与方法》（GB/T 45288.2—2025） 《人工智能　大模型　第 3 部分:服务能力成熟度评估》（GB/T 45288.3—2025） 《信息技术人工智能术语》（GB/T41867—2022） 《人工智能算法金融应用评价规范》（JR/T0221—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/T1833.3—2022） 《人工智能开发平台通用能力要求第 1 部分：功能要求》（YD/T4392.1—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/T4921—2024） 《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》（GB/T 45652—2025） 《网络安全技术　人工智能生成合成内容标识方法》（GB/T 45438—2025 ）
7.4效用评价	场景渗透 受众体验 适能效用	依据《医学人工智能治理综合评价指标体系》(DB4403/T 634—2025)、《基于人工智能的接入网运维和业务智能化场景与需求》（YD/T4070 — 2022 ）、《Regulatory considerations on artificial intelligence for health》（World Health Organization，2023）和相关行业专家意见共识编制。
7.5效率评价	规模效率 成本效率 配置效率	依据《Ethics and Governance of Artificial Intelligence for Health》(World Health Organization,2021)、《Regulatory considerations on artificial intelligence for health》（World Health Organization，2023）和相关行业专家意见共识编制。
7.6效益评价	经济效益 社会效益 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。

五、 采标情况

无

六、 重大分歧意见的处理

本标准制定过程中无重大分歧。

七、 与国家法律法规和强制性标准的关系

本标准为指南类团体标准，与有关的现行法律、法规和强制性国家/行业标准无抵触。

八、 标准实施的建议

标准发布后视各方反映情况，可以举办培训班来指导标准的实施。

九、 其他应予说明的事项

无

《医学人工智能治理综合评价指南　第 1 部分：总则》

团体标准编制组

2025 年 11 月 28 日