

《儿童医学大模型评测规范》（征求意见稿）

编制说明

一、工作简况

（一）任务来源

本标准由相关单位提出，经中国技术市场协会批准，于 2025 年 5 月正式列入 2025 年团体标准制修订计划，标准名称为《儿童医学大模型评测规范》。

（二）项目背景

当前大模型技术在医疗健康领域快速渗透，以深度学习为核心的儿童医学大模型已成为赋能儿科诊疗、健康管理、公共卫生服务的关键技术载体。儿童生理发育具有特殊性、疾病诊疗精细化要求高，儿童医学大模型的专业性、安全性、准确性直接关系到医疗质量与儿童健康。

行业内长期缺乏针对儿科场景的大模型专项评测标准，模型能力评价无统一依据、应用效果参差不齐、安全伦理规范缺失，制约了大模型在儿科领域的可信落地与规模化应用。为落实国家人工智能与医疗健康深度融合发展部署，补齐儿科领域大模型评测体系短板，以标准规范引领技术创新与产业升级，特制定本标准。

（三）目的意义

本标准聚焦儿童医学大模型全生命周期评测需求，构建覆盖基础能力、儿科专业知识、医疗应用、科研教学、安全与伦理五

大维度的系统化评价体系，明确大模型的技术门槛、评测方法与等级划分，为大模型研发、验证、部署、监管提供权威依据。

通过标准化评测，保障儿童医学大模型输出精准、可靠、合规，筑牢临床应用安全底线；规范大模型研发方向，推动技术聚焦儿科真实场景持续迭代升级；促进优质大模型产品分级应用，提升儿科医疗服务智能化水平；强化行业自律与良性竞争，以标准支撑大模型技术在儿科领域健康有序发展，助力医疗人工智能产业高质量创新。

（四）起草单位及起草人名单

本标准起草单位：首都医科大学附属北京儿童医院、××××××× 等。本标准主要起草人：××××、××××、××××、×××× 等。

（五）主要起草过程

1. 文本调研：2025 年 3 月启动标准文本调研工作，系统梳理儿科医学、医疗大模型、人工智能评测相关政策、文献与技术资料，2025 年 4 月完成资料收集与分析。

2. 标准立项：2025 年 5 月向中国技术市场协会提交立项申请，2025 年 5 月获正式批准立项。

3. 组建标准起草工作组：2025 年 8 月召开项目启动会，同步成立标准起草工作组，明确分工并部署标准调研与框架设计事项。

4. 形成标准草案：2025 年 10 月起草组汇报资料调研成果，开展线上研讨；7 月中旬经组内论证确定标准整体框架与核心内

容；12 月上旬对标准初稿进行现场研讨并征集修改意见；12 月底完成意见整合与文本修订，形成标准草案。

5. 形成征求意见稿：2026 年 3 月起草组对标准草案进行全面审议与完善，最终形成《儿童医学大模型评测规范》（征求意见稿）。

二、确定标准主要内容的论据

（一）编制原则

本标准按照 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》以及《中国技术市场协会团体标准工作程序》的规定起草。

（二）标准主要内容及适用范围

1. 标准主要内容

本标准规定了儿童医学大模型评测的体系框架、核心能力要求、评测内容、评价方法及分级规范。此外，本标准还详细规定了基于选择题、判断题和案例题的综合评分规则，以及根据得分情况划分的 A 至 E 五个评价等级。

2. 标准适用范围

本标准适用于以下相关主体开展儿童医学大模型的评估与评价工作：

医疗及研究机构：用于对引入或研发的模型进行专业性能核验。

企业与服务提供商：指导开发单位优化模型算法并提升儿科专科适配性。

监管部门与行业协会：作为对儿童医疗人工智能产品进行质

量监测和准入评估的技术参考依据。

（三）确定标准主要内容的论据

1. 评测维度构建的论据

本标准确定的五大评测维度（模型基础能力、儿科专业知识能力、医疗应用能力、科研教学能力、安全与伦理能力）是基于对当前大语言模型（LLM）技术特性与儿科临床需求的深度融合分析：

儿科专科化需求： 考虑到儿童生长发育的动态性、罕见病占比高及表达能力受限等特点，本标准特设“儿童生长发育”和“抗干扰与容错”指标，论据源于儿科临床诊疗中对发育里程碑评估和家属代诉信息的依赖性。

多维度权重分配： 权重设置（专业知识与医疗应用共占60%）参考了《医疗人工智能评估指南》及专家咨询结果，确保模型评价重心落在实际临床价值与知识准确性上。

2. 技术指标与参数确定的论据

知识覆盖度论据： 标准要求评测集覆盖《诸福棠实用儿科学》95%以上的病种。该教材作为中国儿科医学的权威典籍，其病种覆盖率是衡量模型是否具备全面专业能力的公认物理基准。

多模态处理能力指标： 论据基于儿科高度依赖影像（如骨龄片）、心音、视频（如癫痫发作观察）等非文本数据的临床现实，确立了图像、音频、视频问答的评测细则。

安全与伦理的一票否决制： 规定“安全与伦理单项低于 80

分则评定为 E 级”。论据源于国家网信办《生成式人工智能服务管理暂行办法》及医学伦理“不伤害”原则，强调医疗大模型的底线红线。

3. 实验方法与打分规则的论据

评测集组合逻辑： 推荐按照“选择题（40%）+ 判断题（40%）+ 案例题（20%）”进行组合。

论据： 客观题用于验证基础医学常识的检索准确度，案例题用于考察临床思维逻辑与因果推理能力，模仿了医学资格考试的结构模型。

双盲实验验证： 提出模型与不同职称医生评分优劣性的对比实验。

论据： 引入专业人工审阅（具备高级职称）作为金标准，确保自动化评测结果不偏离临床实际，解决了 LLM 评价中常见的“幻觉”识别难题。

4. 等级划分与推荐适用范围的论据

分级准入逻辑： 标准将评测等级与医疗机构级别挂钩（A 级推荐三级医院使用，B 级推荐基层机构）。

论据： 遵循“风险等级与应用场景匹配”原则。三级医院涉及更多危重症和复杂决策，对模型 A 级（综合得分 ≥ 80 ）的高可靠性有刚性要求；基层机构则侧重常见病筛查，B 级性能即可提供辅助价值。

三、主要试验[或验证]情况分析、技术经济论证、预期经济效果

（一）主要试验（验证）情况分析

本标准在编制过程中，依托北京儿童医院及相关医疗机构开展了多轮验证性测试，重点对“评测维度”和“评测方法”的科学性进行了实证分析。

评测集构建验证：参照本标准 7.1 节要求，课题组构建了涵盖 0-18 岁各年龄段、覆盖常见儿科病种的多模态专科评价数据库。验证结果显示，以《诸福棠实用儿科学》为基准的题库能有效覆盖 95% 以上的儿科临床场景，确保了评测的全面性。

评测方法一致性验证：采用“自动化评测与人工评测相结合”的方式，邀请具备高级职称的专家对模型输出的案例题进行双盲评分。分析表明，本标准 7.3 节制定的打分规则与专家临床判断的一致性高达 90% 以上，证明了评测体系的有效性。

安全与伦理阈值验证：通过对 3 个不同参数规模的儿科大模型进行压力测试，验证了“安全与伦理单项低于 80 分一票否决”制度的必要性，该设定能准确拦截存在数据脱敏不彻底或伦理偏见的模型进入临床推荐范围。

（二）技术经济论证

技术先进性：本标准首次针对儿科领域提出了“多模态交互”、“抗干扰与容错”及“儿童生长发育”等专项评测指标。相较于通用医疗大模型评测，更契合儿童疾病谱广、表达受限、依赖影像和生长监测的临床特征，填补了儿科医学大模型评价体系的空白。

应用可行性：标准采用百分制雷达图展示单项能力，并根据综合得分将模型划分为 A-E 五个等级。这种分级准入机制为

三级医院到基层医疗机构提供了明确的适用指南，降低了各级机构引入 AI 技术时的选型成本。

（三）预期经济与社会效果

提升诊疗效率与质量：通过本标准评测的高质量模型，可支撑临床诊疗全流程，包括自动生成病历、智能追问病史及辅助鉴别诊断。预期可缩短门诊患者排队等待时间，降低误诊、漏诊率，特别是在基层医疗机构的应用可显著提升当地儿科诊疗水平。

优化医疗资源分配：模型具备的医保政策解读及费用优化建议能力，能帮助家庭推荐高性价比的诊疗方案，减轻经济负担。

规范行业健康发展：通过统一的评测规范，可引导开发商加强数据安全与伦理建设，有效防止儿童敏感健康信息的泄露，为儿科人工智能产业的规模化、规范化应用奠定坚实的制度基础。

四、采用国际标准和国外先进标准的程度

经全面文献检索与行业调研，未发现与本标准完全对应的国际标准或国外先进标准，本标准为自主制定，旨在填补国内外儿科医学大模型评测领域的标准空白。

五、重大分歧意见处理经过及依据

本标准在制定全过程中，经多轮研讨与专家论证，未出现重大分歧意见，各项内容达成一致。

六、与现行相关法律、法规及相关标准的协调性

（一）与现行法律法规的符合性

严格遵循《中华人民共和国个人信息保护法》《中华人民共和国数据安全法》《生成式人工智能服务管理暂行办法》及医疗人工智能临床应用相关规范，保障数据安全、隐私保护与伦理合规。

（二）与强制性国家标准的协调性

本标准 of 推荐性团体标准，不涉及强制性国家标准冲突；编写完全符合 GB/T 1.1—2020 要求，术语引用遵循 GB/T 5271.1-2000、GB/T 25069-2010 等国标。

（三）与国内相关标准的关系

本标准与 GB/T 41867-2022、T/SAIAS 020-2024 等医疗 AI 标准保持一致，在通用标准基础上聚焦儿科垂直领域细化，填补国内儿科医学大模型评测规范空白，形成互补协同。

七、知识产权情况说明

1. 知识产权归属：本标准著作权由各起草单位共同所有。
2. 必要专利识别：经知识产权检索，截至本征求意见稿发布，未识别出标准技术内容涉及必要专利，核心评测框架、维度与规则基于公有领域技术原理与行业共识编制。
3. 引用文件版权：本标准引用的国家标准、团体标准均获合法授权，无版权侵权风险。

八、其他应予说明的事项

无。

《儿童医学大模型评测规范》

团体标准起草组

2026 年 3 月