

# 团体标准《医疗大模型应用测试方法》（征求意见稿）

## 编制说明

### 一、工作简况

#### （一）任务来源

医疗人工智能作为现代产业体系中成长性、关联性和带动性较强的战略新兴产业，是人工智能技术赋能实体经济的重要方向。当前，在技术进步与政策支持的双重驱动下，该产业发展态势良好，产业生态基本形成，应用场景持续创新，商业进程不断加速，发展环境逐渐优化。但整体来看，产业仍处于发展初期，在技术成熟度、标准体系建设等方面面临诸多挑战。

“十四五”时期是我国医疗数字化、智能化转型升级的关键期，也是医疗人工智能产业发展的重大机遇期。近年来，我国医疗产业高速发展，市场规模迅速扩大，但产业发展不平衡不充分的问题依然突出，尤其在高端医疗器械领域，存在关键工艺依赖进口、整机制造水平较低等“卡脖子”问题。因此，借助人工智能、大数据、云计算等新一代信息技术推动产业升级转型，成为加快产品迭代、提升性能、实现跨越式发展的必然选择。为积极响应国家对医疗领域技术创新和标准化建设的号召，推动医疗大模型技术的健康发展，促进其在医疗行业的广泛应用，制定本团体标准，旨在规范和指导医疗大模型应用测评的框架体系。

本标准由上海人工智能创新中心提出，上海市人工智能行业协会批准立项。《医疗大模型应用测试方法》主要起草单位：上海人工智能创新中心，计划应完成时间为2026年2月。

#### （二）主要参与起草单位

本文件的主要参与起草单位：上海人工智能创新中心、复旦大学附属中山医院、上海交通大学医学院附属瑞金医院、广东省人民医院、上海交通大学医学院附属第九人民医院、讯飞医疗科技股份有限公司、蚂蚁科技集团股份有限公司、上海库帕思科技有限公司等单位共同负责起草。

#### （三）主要工作过程与主要起草人所做工作

本标准编制过程到目前主要经历了标准工作组成立、标准调研、立项阶段和标准研制四个阶段。各阶段主要工作总结如下：

### 1. 标准工作组成立

2024年11月，为保证标准编写质量，面向行业征集标准起草单位，成立了标准起草工作组，吸纳了国内从事医疗大模型研发与应用领域的企业、高校和医疗行业的专家，共同开展标准研制工作。

### 2. 标准调研

2024年11月 - 2025年4月，标准工作组对医疗大模型评测存在的问题、标准发展情况、建设与应用需求等进行资料收集、查阅、分析和整理。对团体标准政策制度进行梳理，确定标准主要内容和标准框架结构。对照《标准化法》、《团体标准管理规定》和GB/T 1.1-2020等标准，编写标准草案。期间为提高标准关于医疗大模型评测的普适性和可操作性，标准工作组通过召开研讨会以及书面反馈意见等多种形式，对标准文本进行修改完善。

### 3. 立项阶段

2025年6月，标准工作组组织各参编单位代表开展了《医疗大模型应用测试方法》团体标准立项会，会上标准工作组汇报了标准编制的必要性、目的和意义，以及标准草案的框架和主要内容，与会专家一致同意标准立项。

### 4. 标准研制

2025年6月 - 2025年9月，标准工作组组织本标准参编单位代表开展了2轮《医疗大模型应用测试方法》团体标准的标准研制会，进一步针对标准框架和技术内容展开技术研讨，征求各相关方意见。期间标准工作组共处理内部意见120余条，形成标准征求意见稿及编制说明。

标准工作组主要成员及其分工如下：

标准工作组主要成员：徐捷、丁金如、卞谋啸、李佳稭、柏志安、尤梦祥、梁会营、金文忠等。

主要分工：徐捷为工作组组长，主持全面协调工作。丁金如、卞谋啸为本标准主要持笔人，负责本标准的起草、编写。柏志安、尤梦祥、梁会营等为组员负责对医疗大模型应用测评框架及测评方法进行全面调研、研究分析和资料汇总整理等工作，提炼总结医疗大模型应用测评指南，并进行了反复验证、修改完善。

## 二、标准编制原则和确定主要内容的论据及解决的主要问题

### （一）原则

本文件编制遵循“统一性、适用性、一致性、规范性”的原则，注重标准的可操作性。本文件的编写符合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》的编制要求。

标准内容充分考虑了模型提供者、应用服务者和用户等的评测需求，制定了切实可行、合理有效的以医疗大模型评测方法。

### （二）确定主要内容的论据

本文件规定了应用于医疗服务的医疗大模型的测试方法，包括场景应用能力测试、性能测试、安全性测试、模型基础能力测试、模型服务能力测试。

本文件适用于专业测试机构开展医疗大模型的测试工作。企业、科研院所等医疗大模型开发应用机构也可参照执行。

主要技术内容如下：

——第6章 场景应用能力测试：给出医疗大模型开展的辅助诊断决策、辅助治疗方案推荐、病历生成、病历质控、病历信息管理等面向医院侧的医疗服务能力的测试方法；给出医疗大模型开展的疾病预防、诊前指导、用药指导、诊中指引、诊后康复、在线咨询、便民服务等面向患者侧医疗服务能力的测试方法。本章节内容参考《卫生健康行业人工智能应用场景参考指引》，同时结合上海人工智能创新中心提出的医疗大模型测评体系设计完成。

——第9章 模型基础能力测试：从模态支持度、任务支持度、语言处理能力、图像分析能力等几个方面，给出测试模型基础能力的指标和方法。该章节技术内容参考GB/T 45288.1-2025《人工智能 大模型 第1部分：通用要求》和GB/T 45288.2-2025《人工智能 大模型 第2部分：评测指标与方法》标准，经过标准工作组讨论确认。

——第10章 模型服务能力测试：给出围绕个性化服务能力、服务可靠性、服务配套性等方面给出测试内容和方法。

### （三）解决的主要问题

本标准明确了医疗大模型在功能、性能及安全测试上的基本要求、评测指标、测试方法与计算方式，适用于多种医疗大模型的评估，包括但不限于语言模型、

病理诊断、健康管理、药物研发及临床决策支持等类型。本标准旨在填补医疗健康领域内标准化评测体系的空白。立足于医院和患者的使用需求，本标准从医疗服务场景出发，将能力需求分为医院侧和患者侧，并详细定义了各自应具备的服务能力，提供了具体的测评指标与测试方法，确保模型的实际应用价值。此外，针对模型的技术服务能力与安全性，制定了多维度的评估体系，以促进医疗大模型的健康有序发展，推进医疗行业的数字化和标准化进程。

## 二、主要试验情况分析

在《医疗大模型应用测试方法》编制过程中，标准起草工作组紧密围绕《关于加快场景创新以人工智能高水平应用促进经济发展的指导意见》等国家政策导向，严格遵循人工智能标准化建设要求与团体标准管理规范，开展多轮次、系统性的实验验证工作。工作组选取部分较为先进的大模型，根据标准大纲和草案中的评测维度与方法等，面向模型提供者 and 使用者开展评测试验，统计并分析数据，验证标准草案的评测维度、评测方法的设计，调整修改评测题库设计、评测人员筛选。通过验证结果，优化标准内容，确保技术可行性与行业适用性。

## 四、知识产权情况说明

本文件中没有涉及专利和相关知识产权问题。

## 五、产业化情况、推广应用论证和预期达到的经济效果

该标准的制定和实施将极大推动医疗健康人工智能大模型的应用与发展，有效扩大市场边界。通过统一测试方法与评估标准，可构建市场信任机制，吸引更多主体参与医疗AI领域投资与研发，形成“标准引领-技术突破-产业升级”的良性循环。随着市场规模扩张，产业链将趋于完善，带动上下游企业协同发展，为经济增长注入新动能。

标准既为技术创新提供明确方向，也通过规范测试流程加速技术迭代。企业可依据标准体系开展针对性研发，提升产品性能与可靠性，增强上海在医疗AI领域的核心竞争力。更重要的是，该标准将促进医学与计算机科学的深度交叉融合，催生跨领域创新成果和多元应用场景，形成技术创新的乘数效应。

该标准的发布和实施将打破区域技术壁垒，促进医疗AI资源优化配置。对医疗资源薄弱地区而言，标准化产品和服务的引入可快速提升服务效能，缩小城乡医疗水平差距，助力区域经济均衡发展。这种标准化普及不仅体现技术普惠价值，

更通过资源共享机制推动区域产业协同升级。通过提高医疗健康服务的效率与质量，并减少不必要的医疗资源浪费，该标准不仅降低了社会医疗成本，还支持了经济的可持续发展。同时，标准的实施有利于推动医疗健康行业的绿色、低碳发展，减少环境污染与资源浪费，实现经济与环境的和谐共生。

#### **六、转化国际标准和国外先进标准情况**

无。

#### **七、与现行相关法律、法规、规章及相关标准的协调性**

当前国内人工智能领域的强制性标准与法律法规主要聚焦于数据安全、隐私保护等议题，包括：

1) 《中华人民共和国数据安全》：规定数据全生命周期安全管理要求，覆盖收集、存储、使用、传输环节。

2) 《中华人民共和国个人信息保护法》：规范个人信息处理规则，强调合法性、正当性与必要性。

在标准方面，GB/T 40629 - 2021《人工智能 术语》等基础标准，对人工智能领域的专业术语进行了统一规范，避免行业内因术语理解不一致而产生的沟通障碍与误解，为产业的有序发展奠定基础。

#### **八、重大分歧意见的处理经过和依据**

无。

#### **九、标准性质的建议**

本标准批准后作为推荐性团体标准使用。

#### **十、贯彻标准的要求和措施建议**

建议该标准发布即实施，发布后，建议项目提出单位组织开展标准宣贯，出台相关文件推动该标准的落地应用。

#### **十一、替代或废止现行相关标准的建议**

无。

#### **十二、其它应予说明的事项**

《医疗大模型应用测试方法》团体标准编制起草组

2025-11-3