

《人工智能 智能健康术语歧义消解测试要求》团体标准编制说明

一、任务来源

随着医疗健康数字化转型加速，智能交互系统、电子健康档案平台等应用场景中，健康术语歧义问题导致的服务错配、数据互通困难等问题日益突出。据行业调研，健康术语歧义引发的智能服务误判率高达 15%-20%，跨平台数据语义对齐成本占系统开发成本的 30%以上。为解决行业痛点，推动健康术语处理标准化发展，本团体标准由北京健康有益科技有限公司提出并归口，联合协和医院、301 医院、中国医学科学院信息所、北京康健数字化健康管理研究院、中国人民大学、上海师范大学天华学院、天津大学、北京工业大学、河北大学、北京减脂时代科技有限公司、北京数果科技有限公司、健康有益（河北）科技有限公司。

二、制定规范的必要性和意义

2.1 必要性

2.1.1 解决术语处理乱象问题：当前健康术语存在“一词多义”“多词一义”现象，如“运动”在减肥、康复场景下语义差异显著，各企业自定义处理规则导致产品兼容性差，亟需统一技术规范。

2.1.2 保障健康服务精准性：术语歧义易引发智能诊断、用药建议等服务偏差，如将“无菌性炎症”误判为“细菌感染”可能导致治疗方案错误，需通过标准明确歧义消解流程与质量要求。

2.1.3 支撑健康数据互联互通：电子健康档案、医疗数据分析等场景需统一术语语义，现有系统术语体系碎片化，跨机构数据共享需重复映射，标准制定可降低数据整合成本。

2.2 意义

2.2.1 行业层面： 填补健康术语歧义消解与标准化领域的标准空白，建立统一技术框架，引导行业从“各自为战”向“协同发展”转型，提升整体技术应用水平。

2.2.2 企业层面： 为企业提供清晰的技术研发依据，减少重复开发成本，降低服务合规风险，促进跨企业、跨平台的技术协同与数据互通。

2.2.3 用户层面： 确保用户获得精准、一致的健康服务，避免因术语歧义导致的信息误解，提升用户对医疗健康数字化产品的信任度与使用体验。

三、主要编制过程

1、启动阶段（2025年11月-2025年12月）： 成立由医疗信息化专家、NLP技术专家、临床医师组成的编制工作组，明确任务分工；开展行业调研，收集国内外相关技术资料、企业案例及用户需求，形成《标准编制工作方案》。

2、草案阶段（2026年1月-2026年2月）： 结合调研结果，梳理术语定义、技术流程等核心内容，参考GB/T 35273-2022、WS/T 447-2014等现行标准，完成标准草案初稿；组织3轮内部研讨，修改完善后形成草案征求意见稿。

四、制定规范的原则和依据，与现行法律、法规、标准的关系

4.1 制定原则

4.1.1 科学性原则： 基于多模态语义嵌入、对比学习等前沿技术，结合医疗健康领域知识特点，确保技术要求与指标的科学性、合理性。

4.1.2 实用性原则： 聚焦行业实际痛点，内容覆盖术语采集、歧义消解、标准化映射全流程，指标可量化、流程可操作，满足企业落

地需求。

4.1.3 合规性原则：严格遵循数据安全、医疗健康相关法律法规，确保标准内容合法合规。

4.1.4 前瞻性原则：预留新兴术语增量嵌入接口，支持多模态数据融合技术升级，适应行业发展趋势。

4.2 制定依据

《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》；

《“十四五”医疗健康信息化发展规划》《关于加快推进电子健康档案普及应用的指导意见》等政策文件；

GB/T 35273-2022《信息安全技术 个人信息安全规范》、WS/T 447-2014《电子健康档案基本架构与数据标准》、GB/T 38667-2020《人工智能 术语》、WS 363-2011《卫生信息数据元目录》等现行标准；行业内主流企业的技术实践、科研机构的多模态语义处理研究成果及临床需求调研数据。

4.3 与现行法律、法规、标准的关系

本标准与现行法律、法规、标准无冲突，是对现有医疗健康信息化标准体系的补充与细化：

在数据安全方面，遵循 GB/T 35273-2022 要求，进一步明确健康术语关联敏感数据的加密存储、传输规范；

在术语体系方面，衔接 WS/T 447-2014、WS 363-2011，补充歧义术语的标准化映射规则，提升术语体系的实用性；

在人工智能技术应用方面，参考 GB/T 38667-2020，定义多模态语义嵌入等领域特有术语，完善 AI 医疗应用的术语体系。

五、主要条款的说明，主要技术指标、参数、试验验证的论述

5.1 主要条款说明

术语和定义：明确健康术语、歧义消解、多模态语义嵌入等 5 个核心术语，统一行业认知，为后续技术内容奠定基础。

技术流程：规范术语采集预处理、多模态特征融合、歧义检测、消解、标准化映射及结果反馈六步流程，指导系统开发逻辑。

核心要求：从术语采集、语义嵌入、歧义消解、动态术语库四个维度提出量化指标，确保处理质量。

质量评估：建立自动化测试、人工评估、用户调研三维评估体系，全面验证标准执行效果。

5.2 主要技术指标及试验验证

技术指标	指标要求	试验验证论述
高频歧义术语消解准确率	$\geq 96\%$	选取“运动”“调理”等 20 个高频歧义术语，构建包含 5000 条标注样本的测试集，在 3 家企业的原型系统上测试，平均消解准确率达 96.8%，满足指标要求。
语义向量余弦相似度均值	≥ 0.85	使用标准术语测试集（含 1000 条标准术语），通过 FGCrossNet 模型生成语义向量，计算同类术语相似度均值为 0.88，异类术语相似度均值为 0.23，语义区分度良好。
术语库覆盖率	$\geq 90\%$	采集 1000 条真实用户健康咨询文本，统计可匹配到动态术语库的术语数量为 923 条，覆盖率达 92.3%，符合要求。
歧义消解响应时间	≤ 1 秒	模拟 100 用户并发请求场景，测试歧义消解平均响应时间为 0.82 秒，峰值响应时间为 0.95 秒，满足实时性要求。

六、重大意见分歧的处理依据和结果的说明

在标准征求意见阶段，主要存在以下重大意见分歧及处理结果：

分歧 1：关于“多模态数据融合种类”，部分企业认为融合 3

种异构数据成本过高，建议简化为 2 种；科研机构则坚持 3 种数据融合能提升语义准确性。

处理依据：试验验证显示，融合 3 种数据的歧义消解准确率比 2 种数据高 4.2%，且当前主流硬件设备可支撑融合成本。

处理结果：保留“融合至少 3 种异构数据”的要求，同时在附录 B 提供轻量化融合方案备选，平衡准确性与成本。

分歧 2：关于“术语库更新频率”，有专家建议每月更新 1 次，企业则认为每季度更新更符合实际运维节奏。

处理依据：统计近 2 年权威医学指南及标准更新频次，平均每季度新增/修订术语约 50 条，每季度更新可满足需求。

处理结果：确定“术语库更新频率不低于每季度 1 次”，同时要求对重大公共卫生事件相关术语启动紧急更新通道。

七、采用国际标准或国外先进标准程度的说明，以及国内、外同类产品或标准的对比情况

7.1 采用国际标准或国外先进标准程度

目前国际上尚无健康术语歧义消解与标准化处理的专项标准，国际标准化组织（ISO）、世界卫生组织（WHO）的相关标准（如 ISO 18104、ICD-11）主要聚焦健康数据格式、疾病分类等基础内容。本标准在制定过程中，参考了 ICD-11 的疾病术语分类逻辑及欧盟 GDPR 的数据隐私保护原则，但未直接采用特定国际标准，属于结合国内行业需求自主制定的原创性团体标准。

7.2 国内、外同类产品或标准对比

对比维度	国外情况	国内情况	本标准
标准现状	无专项标准，侧重通用术语体系构建	无专项标准，现有标准聚焦数据元与档案格式	首个健康术语歧义消解与标准化专项标准
技术方法	以单文本语义嵌入为主，多模态	多模态技术探索中，	明确 FGCrossNet 多模态融合架构，提

	融合应用少	缺乏统一融合框架	供详细训练指南
动态更新	术语库更新周期长（6-12个月）	更新机制不健全，新兴术语响应慢	24小时增量嵌入+季度更新，支持紧急更新通道
评估体系	侧重技术指标，缺乏用户体验评估	评估维度单一，多为企业自定标准	技术+人工+用户三维评估，指标量化明确

八、实施标准的措施建议

宣贯培训：由归口单位联合行业协会组织标准宣贯会，面向企业、医疗机构开展专题培训，解读技术要点与实施案例。

试点示范：选取5家不同类型的健康科技企业（智能问诊、慢病管理、健康科普）开展试点应用，总结实施经验并形成白皮书。

工具研发：鼓励第三方机构基于标准研发开源的术语处理工具包，降低企业应用门槛，推动标准快速落地。

动态维护：成立标准维护工作组，每2年对标准进行复审，根据技术发展与行业需求适时修订，确保标准时效性。

九、其他说明

本标准推荐性团体标准，不具有强制性，但鼓励医疗健康信息化领域相关企业、机构自愿采用。标准实施过程中，如遇重大技术突破或政策调整，编制工作组将及时收集反馈意见，启动标准修订程序。本标准的附录A为规范性附录，附录B为资料性附录，均为标准内容的重要补充，供使用者参考执行。