ICS 35.020
CCS L70

T/ISC

团 体 标 准

T/ISC 0170—2026

医疗认知大模型系统技术要求

Technical requirement for Medical cognitive large model system

（发布稿）

2026－10－28 发布 2026－11－28 实施

中国互联网协会 发 布

目　　次

1 范围 1
2 规范性引用文件 1
3 术语和定义 1
4 缩略语 1
5 总体要求 2
6 功能需求 2
6.1 医疗基座大模型要求 2
6.2 思维链推理要求 3
6.3 检索增强生成功能要求 3
6.4 个性化医疗信息提取要求 3
6.5 医疗上下文编辑能力要求 4
7 准确性要求 4
7.1 医疗基座大模型准确性指标 4
7.2 推理框架准确性指标 4
7.3 检索增强生成准确性指标 5
7.4 个性化医疗信息提取准确性指标 5
7.5 医疗上下文编辑准确性指标 6
8 性能效率要求 6
8.1 时延响应指标 6
8.2 吞吐量指标 6
9 易用性要求 6
9.1 可理解性 6
9.2 易学性 6
9.3 易操作性 6
10 安全性要求 7
10.1 基础设备安全性 7
10.2 数据安全 7
10.3 应用安全 7

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。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国互联网协会归口。

本文件起草单位：安徽省合数智医科技有限公司、讯飞医疗科技股份有限公司、中国信息通信研究院、安徽医科大学第一附属医院、安徽省立医院（中国科学技术大学附属第一医院）、安徽省胸科医院、蚌埠市第三人民医院（蚌埠市中心医院）、上海市第六人民医院安徽医院、瑶海区胜利明光街道社区卫生服务中心、中国科学技术大学、合肥市第四人民医院（合肥市精神病医院、合肥市精神卫生防治中心）、蚌埠医科大学第一附属医院（蚌埠医科大学附属肿瘤医院）

本文件主要起草人：韩苹、陈良、鲁舜毅、张志、朱颜渝、张金良、葛健聪、白晶晶、陶晓东、鹿晓亮、彭跃辉、汪雪豆、雒靖元、张勇、周兆旦、姜磊、宋时德、尤梦祥、史惠婷、张晓燕、王琪、杜雨暄、任九选、贾斐、相识、张卓然、黄涂半特、刘泊宁、陈冠衣、王慧莹、张洁、朱德刚、杜明超、张岩、王曦、查晓丽、程静、田仰光、郑昌成、陈玉俊、姜东兴、吕华毅、谢丹、朱懿、韩永生、陈荣、刘和福、周杨、杨启晗、徐小童、王振杰

医疗认知大模型系统技术要求

1 范围

本文件规定了医疗认知大模型系统涉及的技术能力要求，从功能要求、准确性要求、性能效率要求、易用性要求和安全性要求等维度对医疗认知大模型系统的能力提出要求。

本文件适用于医疗认知大模型系统提供方评估自身技术与应用能力，管理方对医疗认知大模型系统服务提供方的服务能力进行要求，第三方评估医疗认知大模型系统提供方的能力。

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 35273-2020 信息安全技术 个人信息安全规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

大模型 large model

指可学习参数量大的深度学习模型，本文件中特指至少具备 1亿参数的深度学习模型。

3.2

医疗认知大模型系统

以经医疗领域数据后训练的大语言模型为推理核心，集成医学检索增强生成、医疗信息检索、医疗信息抽取和医疗上下文管理等技术，面向医疗场景提供医学知识理解、推理、检索与信息处理辅助功能的软件系统。

3.3

思维链 chain of thought

指通过令大模型在输出答案前额外输出文本作为辅助，提升大模型医疗复杂问题推理与决策准确性的推理框架。

4 缩略语

下列缩略语适用于本文件

CPU：中央处理器（Central Processing Unit）

GPU：图形处理器（Graph Processing Unit）

ICD-10：国际疾病分类第十版（International Classification of Disease, 10th Revision）

JSON：JavaScript 对象表示法（JavaScript Object Notation）

KG：知识图谱（Knowledge Graph）

GraphRAG：图检索增强生成（Graph Retrieval-Augmented Generation）

RAG：检索增强生成（Retrieval-Augmented Generation）

ROC：接受者操作特性曲线（Receiver Operating Characteristic）

SNOMED CT：医学系统命名法 —— 临床术语（Systematized Nomenclature of Medicine – Clinical Terms）

XML：可扩展标记语言（Extensible Markup Language）

EMR：电子病历（Electronic Medical Record）

HIS：医院信息系统（Hospital Information System）

MTBF：平均无故障时间（Mean Time Between Failure）

5 总体要求

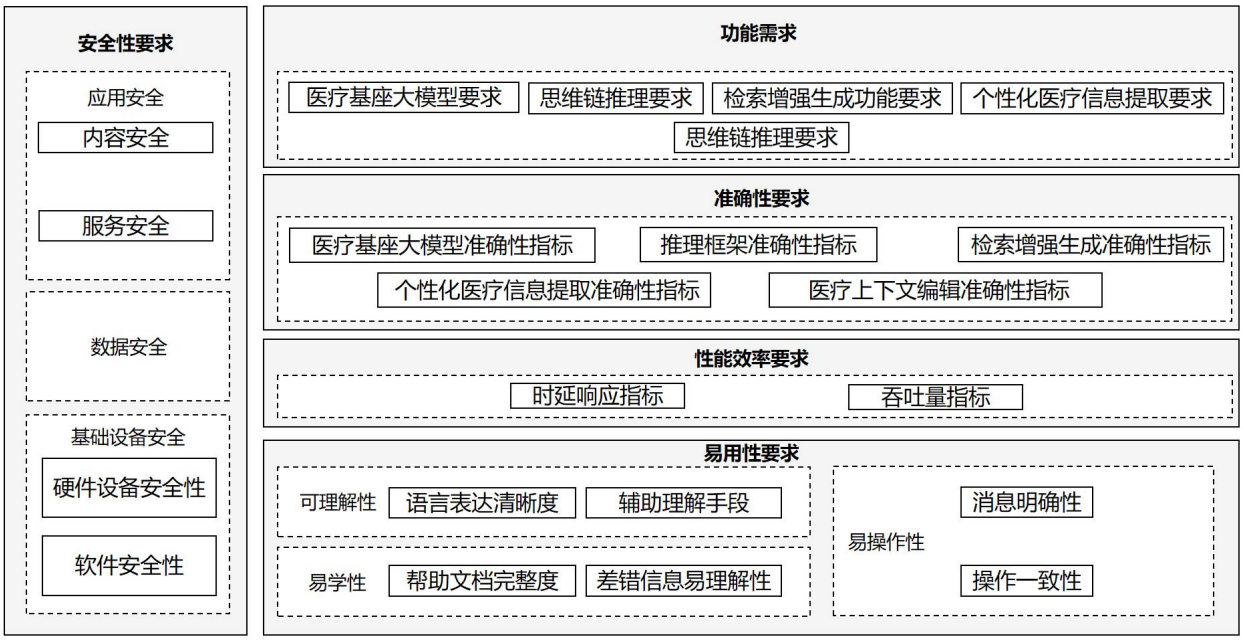

图 1 医疗认知大模型技术要求架构图

医疗认知大模型系统应基于医疗信息完成基座大模型后训练，集成思维链推理、医学 RAG（纯向量、KG、多层次 GraphRAG）、个性化医疗信息检索、医疗信息抽取、医疗上下文编辑核心能力，覆盖医疗认知、推理、检索、信息处理全流程需求。

医疗认知大模型系统应在以下方面满足要求：功能完备性要求、准确性要求、性能效率要求、易用性要求、安全性要求。

6 功能需求

6.1 医疗基座大模型要求

医疗认知大模型系统的基座模型应完成医疗领域专项后训练，适配医疗场景认知需求：

a) 医疗语料适配能力：应支持医学教材、临床指南、病历、医学文献等专业语料的后训练与知识对齐，覆盖基础医学、临床医学等多领域核心内容，应通过 SFT（有监督微调）等技术对语料进行精细化处理，确保模型所学知识与临床实际应用匹配；
b) 医学术语规范能力：应支持 ICD-10、SNOMED CT 等医疗编码体系的知识融合，输出的 ICD-10 编码应符合 GB/T 14396-2016；
c) 医疗常识对齐能力：应通过后训练技术使模型实现临床诊疗逻辑、伦理安全、医学禁忌、用药规范等基础常识的学习与固化；
d) 实体抽取能力：应支持疾病、药品、症状、检查、手术等医疗实体的抽取，覆盖临床核心实体类型；
e) 关系抽取能力：应支持实体间诊疗关系、用药关系、因果关系等抽取，挖掘实体间的内在关联；
f) 事件抽取能力：应支持入院、手术、复查、不良反应等医疗事件的抽取；
g) 结构化输出能力：应支持将抽取结果输出为 JSON、XML 等标准结构化格式，便于后续数据处理与系统对接；
h) 非标文本适配能力：应支持手写病历、口述病历、不规则医疗文档的信息抽取，适配各类非结构化医疗文本，通过 OCR 识别、语音转写等技术处理手写、口述内容，精准抽取其中的医疗实体与事件，打破文本格式限制；

i） 抽取校验能力：应支持对抽取结果的医学逻辑校验，纠正错误抽取信息，提升抽取准确性（如校验“青霉素过敏”患者的抽取结果中是否存在“使用青霉素”的用药关系，及时纠正逻辑矛盾的错误信息）；
j） 知识更新能力：宜支持医疗新指南、新疗法、新药品的增量后训练与知识迭代，跟上医疗行业发展节奏。可通过增量 SFT 技术实现新知识的快速融入，无需对模型进行全量重训，降低知识更新成本。

6.2 思维链推理要求

医疗认知大模型系统应具备思维链推理能力，通过显式的逐步推理过程提升复杂医疗问题的可解释性与准确性，并支持基于强化学习算法的推理策略优化：

a） 逐步推理能力：应支持将医疗问题分解为连续的推理步骤，每一步基于前一步结论进行推导，形成完整的推理链条（如从患者主诉、现病史、体格检查结果逐步分析以生成初步诊断、鉴别诊断及检查建议）；
b） 推理格式规范：应支持结构化的推理输出格式，使用特定词元（如〈think〉... 〈/think〉）包裹完整的中间推理过程，最终答案置于标签外部以便于系统在交互界面上对推理内容进行折叠或缩略显示，使临床人员可快速获取结论，同时按需展开详细推理步骤进行核验；
c） 强化学习优化能力：宜采用 GRPO（Group Relative Policy Optimization）等强化学习算法对思维链推理策略进行训练与优化，通过组内相对比较奖励机制，促使模型生成逻辑一致、医学合理的推理路径，减少幻觉与跳跃性推理，提升多步推理的稳定性和准确性；
d） 推理摘要能力：应支持从完整思维链中自动提取关键推理节点（如核心证据、决策转折点、排除诊断依据等），生成简化的推理流程摘要或流程图，便于临床人员快速把握推理过程。

6.3 检索增强生成功能要求

医疗认知大模型系统应支持多层次医学 RAG 检索，实现精准知识增强：

a） 向量检索能力：应支持医疗文本、病历、文献的切片、嵌入生成、向量库构成，快速匹配相关知识，提升检索响应速度，宜支持 ANN 算法（如 NSW、HNSW 等）以快速检索相关的知识片段，如特定疾病的诊疗文献；
b） 重排序能力：应支持使用基于双向编码器对问题与召回文本配对进行重新排序以提高检索内容质量；
c） 医学知识图谱检索能力：应支持对医疗实体与关系三元组的图数据库检索，精准定位结构化医学知识；
d） 多层次 GraphRAG 能力：应支持文档级、段落级、实体级多层图谱检索，提升不同颗粒度内容检索精度；
e） 知识库对接能力：应支持临床指南库、药品库、疾病库等权威医疗知识库的对接，保障知识来源的权威性；
f） 过滤能力：应支持基于检索文本元数据字段对检索信息进行二次过滤以减少过时信息对大模型推理的影响；
g） 检索修正能力：应支持对低质量检索信息评估并重写问题重新召回文本的能力；
h） 问答匹配能力：应支持对文本生成潜在问题并对潜在问题嵌入以增强对用户问题匹配程度的能力；
i） 知识融合能力：应支持基于检索内容进行推理并附带引用以减少大模型幻觉影响。

6.4 个性化医疗信息提取要求

医疗认知大模型系统应支持个性化医疗信息检索，适配不同用户与场景需求：

a） 患者画像适配能力：应结合患者病史、过敏史、基础病生成个性化检索策略，适配不同患者的诊疗需求（如针对有高血压基础病的患者，检索时优先关联高血压相关的诊疗知识与用药注意事项）；
b） 场景适配能力：应支持门诊、住院、急诊、慢病管理等不同诊疗场景的检索偏好适配，贴合场景需求（如急诊场景优先检索快速急救方案，慢病管理场景优先检索长期干预与随访相关知识）；

c） 权限分级能力：应支持医师、护士、患者等不同角色的检索权限控制，保障医疗信息安全（医师可检索完整病历与诊疗指南，患者仅可检索与自身相关的健康知识，避免信息泄露）；
d） 跨源检索能力：应支持 EMR、HIS、医学文献等多源医疗数据的统一检索，实现数据互通；
e） 结果过滤能力：应支持对低价值检索结果的自动过滤，提升检索有效性与用户体验。

6.5 医疗上下文编辑能力要求

医疗认知大模型系统应支持医疗上下文编辑，优化信息交互与文档处理：
a） 上下文管理能力：应支持长程医疗对话、多轮检索结果的上下文存储与管理，保障交互的连贯性，可持久化存储多轮医疗对话内容与检索结果，后续交互时无需重复输入，如连续对话中可关联历史诊疗建议，提升交互效率；
b） 错误修正能力：应支持对上下文内的医学术语、数据、逻辑错误自动修正，确保信息的准确性（如自动修正“糖尿症”为“糖尿病”）；
c） 信息补充能力：应支持对缺失的关键医疗信息进行智能补全，完善上下文内容（如基于患者医疗画像补全病历中缺失的患者年龄、既往病史等关键信息，或补全诊疗建议中缺失的用药剂量、复查时间等内容）；
d） 结构化整理能力：应支持将零散医疗信息整理为规范病历、诊疗记录等格式，提升文档规范性；

7 准确性要求

7.1 医疗基座大模型准确性指标

评估医疗认知大模型系统中医疗基座大模型处理医疗领域自然语言处理任务准确性：
a） 医疗三元组抽取正确率：综合衡量系统抽取医疗实体与关系三元组（SPO）的整体准确性，计算公式如下：

$$Acc_{SPO} = \frac{N_{SPO-correct}}{N_{SPO}}$$

式中：
Acc_{SPO}——医疗三元组抽取正确率；
$N_{SPO\text{-}correct}$——正确抽取三元组数量；
N_{SPO}——文本三元组总数。
b） 医学术语规范率：衡量系统输出属于符合 ICD-10、SNOMED CT 等编码的比例，计算公式如下：

$$Acc_{term} = \frac{N_{term-correct}}{N_{term}}$$

式中：
Acc_{term}——医学术语规范率；
$N_{term\text{-}correct}$——输出中医学术语被正确编码的数量；
N_{term}——输出中医学术语总数。
c） 医疗专业词汇词元覆盖率：计算分词器（tokenizer）对医学专业词汇的覆盖比例，反应模型对医疗领域的适配度，计算公式如下：

$$P_{coverage} = \frac{N_{covered}}{N_{total-term}}$$

式中：
$P_{coverage}$——医疗专业词汇词元覆盖率；
$N_{correct}$——分词器包含的测试医学专业词汇数量；
$N_{total\text{-}term}$——测试医学术语总数。

7.2 推理框架准确性指标

评估医疗认知大模型系统使用推理框架后进行医疗相关决策准确性指标：
a） 检查推荐准确率：衡量系统对辅助检查推荐的准确度，计算公式如下：

$$Acc_{examination} = \frac{N_{examination-correct}}{N_{examination}}$$

式中：

$Acc_{examination}$——检查推荐准确率；

$N_{examination\text{-}correct}$——正确推荐的检查数量；

$N_{examination}$——推荐检查总数。

b） 诊断准确率：衡量系统对疾病诊断的准确度，计算公式如下：

$$Acc_{diag} = \frac{N_{diag-correct}}{N_{diag}}$$

式中：

Acc_{diag}——诊断准确率；

$N_{diag\text{-}correct}$——正确诊断案例数量；

N_{diag}——诊断案例总数。

c） 推理路径正确率：衡量系统诊断正确案例中推理路径正确率，计算公式如下：

$$Acc_{path} = \frac{N_{path-correct}}{N_{diag-correct}}$$

式中：

Acc_{path}——推理路径正确率；

$N_{path\text{-}correct}$——正确诊断案例中推理路径正确的案例数量；

$N_{diag\text{-}correct}$——正确诊断案例总数。

7.3 检索增强生成准确性指标

评估医疗认知大模型系统检索增强生成模块的检索准确度：

a） 检索精确度：评估向量、知识图谱、GraphRAG 检索到的 k 个结果的精确率，k 值选取取决于系统召回文本数，计算公式如下：

$$Precision@k = \frac{TP@k}{TP@k+FP@k}$$

式中：

Precision@k——检索精确度；

TP@k——前k个文本块中与问题相关的文本块数量；

FP@k——前k个文本块中与问题无关的文本块数量。

b） 忠实度：评估系统基于检索结果生成的文本是否被检索内容支撑，计算公式如下：

$$Faithfulness = \frac{N_{supported}}{N_{claim}}$$

式中：

Faithfulness——忠实度；

$N_{supported}$——生成结果中被支持的论点；

N_{claim}——生成结果中论点数量。

c） 最低余弦相似度：检索文本中与原问题或改写后问题的最低余弦相似度，衡量召回文本和问题的相关性，计算公式如下：

$$Min_{cossim} = min_{i=1,2\ldots k}[cossim(Q, R_i)]$$

式中：

Q——问题字符串；

R_i——第i个检索文本块；

cossim——余弦相似度；

7.4 个性化医疗信息提取准确性指标

评估医疗认知大模型系统提取个体医疗信息的准确度，使用疾病相关信息抽取正确率评估，衡量系统抽取当前患者主诉相关的健康信息的比例，计算公式如下：

$$Acc_{extract} = \frac{N_{extract-correct}}{N_{extract}}$$

式中：

$Acc_{extract}$——疾病相关信息抽取正确率；

$N_{extract\text{-}correct}$——正确的抽取的疾病相关信息数量；

$N_{extract}$——应被抽取的疾病相关信息数量。

7.5 医疗上下文编辑准确性指标

评估医疗认知大模型系统对于上下文中不准确信息纠错能力：

a) 词元缩减数：对上下文进行缩略后上下文词元数减少的数量；

b) 隐私脱敏成功率：对于上下文中患者敏感信息脱敏且不破坏推理需要的比例。

8 性能效率要求

8.1 时延响应指标

时延响应指标衡量医疗认知大模型系统推理响应速度：

a) 首令牌输出时延：收到标准化输入请求后，系统流式输出第一个令牌的间隔时长，如具备推理能力，则计算输出第一个非思维链令牌的间隔时长；

b) 推理平均时延：具备推理能力的模型完成单次推理的平均耗时；

c) 99 分位推理时延：统计周期内位于 99%的首令牌输出时延，通过排除极端长尾时延，衡量模型时延稳定性。

8.2 吞吐量指标

吞吐量指标衡量医疗认知大模型系统流式输出吞吐能力：

a) 每秒输出量：模型在稳定推理状态下每秒输出的令牌总数量；

b) 单日最大输出量：模型在运行状态下单日可处理与输出的令牌总量，衡量模型长期运行稳定性。

9 易用性要求

9.1 可理解性

9.1.1 语言表达清晰程度

医疗认知大模型系统应用界面文字、提示及交互内容应简洁准确并以标准医学术语表示，避免口语化内容导致的歧义。

9.1.2 辅助理解手段

医疗认知大模型系统涉及召回信息时应显式引用相关来源辅助医师理解。

9.2 易学性

9.2.1 帮助文档完整性

医疗认知大模型系统应配备结构化帮助文档，含功能说明、操作指南及常见问题解答，支持关键词检索，内容随平台更新同步修订。

9.2.2 差错信息易理解性

医疗认知大模型系统系统异常时，差错信息应明确原因并提供解决方案，不应以技术代码表述。

9.3 易操作性

9.3.1 操作一致性

医疗认知大模型系统各功能模块操作逻辑、交互样式应保持统一，降低用户学习成本。

9.3.2 消息明确性

医疗认知大模型系统进行图表理解与回答输出时，输出文本应简洁明了，不应使用冗长复杂的表述。

10 安全性要求

10.1 基础设备安全性

10.1.1 硬件设备安全性

医疗认知大模型系统涉及的硬件设备（如网络设备、存储设备、计算设备等）的安全防护能力应包含：

a) 通用安全要求：
   1) 应满足物理安全保障要求，包含防火、防雷、防水、灾备等；
   2) 应满足功能安全保障要求，包含设备标签、硬件接口安全、固件安全、驱动程序安全等；
   3) 应满足管理安全保障要求，包含管理机制、管理人员、授权等；
b) 计算设备安全专用要求：
   1) 应具备保障人工智能加速芯片应具备通用安全保障能力，包含 AI 加速芯片信息窃取防护、架构安全漏洞防护等；
   2) 应具备保障人工智能加速芯片在异构场景下稳定运行的能力，包含 CPU 与 GPU 相结合的场景；
   3) 应具备保障人工智能加速芯片运行环境安全的能力。

10.1.2 软件安全性

医疗认知大模型系统应支持多种设施如依赖库、AI框架、向量数据库、中间件、接口等具备安全防护能力，包含：

a) 漏洞管理：软件设施应定期进行漏洞扫描和修复，具备完善的漏洞响应机制；
b) 安全更新：软件设施应及时更新安全补丁，以防止新出现的安全威胁。

10.2 数据安全

除应符合 GB/T 35273-2020 外，基于医疗认知大模型还应满足以下数据安全要求：

a) 生物特征数据保护：
——大模型采集的可用于识别特定自然人身份的生物特征数据，应参照 GB/T 35273-2020 中个人生物识别信息的保护要求进行同等防护；
——原始数据应在终端侧完成特征提取，应上传脱敏后的特征。
b) 数据留存与注销隔离：
——患者健康档案的存储期限应遵循国家关于病历与健康档案保存期限的相关规定，应在存储期满后对数据进行不可逆匿名化处理；
——患者注销账户后，因法律法规要求需留存的数据，应在完成去标识化处理后与日常运营数据库进行物理隔离，仅可用于法定的公共卫生统计或科研目的。

10.3 应用安全

10.3.1 内容安全

医疗认知大模型系统输出内容（含生成内容、决策内容）应符合全人类普适的道德伦理及医学伦理要求。

a) 应支持尊重人权，包括医疗认知大模型系统输出内容（含生成内容、决策内容）应遵循人权的普遍性和不可侵犯性的原则，尊重人类平等、尊严和自由的权利；
b) 应支持无偏见歧视性，包括医疗认知大模型系统输出内容（含生成内容、决策内容）避免产生偏见及歧视性结果的程度；
c) 应符合科技伦理原则，包括增进人类福祉、坚持公平公正、推动透明可释、确保可控可信等；
d) 应遵循科技伦理指标，包括公平性、透明可释性、数据隐私、可控可靠性、内容向善、责任可追溯、可持续性等。

10.3.2 服务安全

医疗认知大模型系统应支持服务安全可信、内容安全可信等应用相关内容具备安全防护能力，包含：

服务安全：医疗认知大模型系统涉及的模型安全性应满足模型安全保障要求，包含MTBF、服务安全性、服务合规性、反馈处置机制等。