

团 体 标 准

T/ISC 0096—2025

医疗健康行业智能体 健康咨询技术要求

Intelligent agent in the healthcare industry - Technical requirements for
health consultation

（发布稿）

2025 - 12 - 26 发布

2026 - 01 - 26 实施

中 国 互 联 网 协 会 发 布

目 次

前 言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 总体要求 1

5 功能完备性技术要求 2

 5.1 健康档案 2

 5.2 健康问答 3

 5.3 药物查询 3

 5.4 报告解读 3

 5.5 专病管理路径 4

6 准确性要求 6

 6.1 二分类任务 4

 6.2 文书生成类任务 5

 6.3 多分类任务 6

 6.4 图像分割任务 7

7 智能体能力要求 8

 7.1 感知能力 8

 7.2 规划能力 9

 7.3 记忆能力 10

 7.4 执行能力 11

8 易用性要求 12

 8.1 可理解性 12

 8.2 易学性 12

 8.3 易操作性 12

9 安全性要求 13

 9.1 基础设施安全 13

 9.2 数据安全 13

 9.3 应用安全 14

参 考 文 献 15

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提出并归口。

本文件起草单位：讯飞医疗科技股份有限公司、中国信息通信研究院、北京协和医院、上海交通大学医学院附属瑞金医院、北京安定医院、中国科学技术大学附属第一医院、国家卫生健康委人口文化与基层健康中心、国家卫生健康委信息统计中心、南京鼓楼医院、东南大学医学院附属中大医院、盐城市第一人民医院、盐城市第三人民医院、呼和浩特市第一医院、安徽医科大学第一附属医院、安庆市立医院、安徽中医药大学第一附属医院、太和县人民医院、安徽医科大学第五附属医院、合肥市第二人民医院、安徽医科大学附属口腔医院、北京市朝阳区医院、山东第一医科大学第一附属医院（山东省千佛山医院）、深圳市宝安区中心医院、中国民族卫生协会、呼和浩特市第二医院、北京民海生物科技有限公司、首都医科大学附属北京安贞医院、新疆医科大学第二附属医院、昌吉州人民医院、鄂尔多斯市中心医院、烟台市中医医院、山东大学附属儿童医院、肥城市人民医院、内蒙古自治区妇幼保健院、内蒙古医科大学第二附属医院、哈尔滨医科大学附属第一医院、鄂尔多斯市乌审旗卫生健康委员会、新疆医科大学第一附属医院、新疆维吾尔自治区人民医院、南通大学附属医院、江南大学附属医院、北京大学深圳医院、河南省人民医院、河南科技大学第一附属医院、河南大学第一附属医院、许昌市中心医院、郑州市中心医院、郑州市第九人民医院、阜外华中心血管病医院、宁波大学附属第一医院、深圳市宝安区妇幼保健院、临沂市兰山区卫生健康局、烟台毓璜顶医院、海南省人民医院、河北大学附属医院、深圳市南方科技大学附属龙华医院、唐山市妇幼保健院、河北医科大学第二医院、中国中医科学院广安门医院保定医院、邢台市人民医院、石家庄市第四医院、保定市第二中心医院、河北医科大学第一医院、白城市医院、沧州市中心医院、中南大学湘雅三医院、保定市第一中心医院、无锡新瑞医院（上海瑞金医院无锡分院）、福建医科大学附属第一医院、广西医科大学附属武鸣医院、阜新市中心医院、锦州医科大学附属第一医院、沈阳医学院附属中心医院、锦州医科大学附属第三医院、厦门大学附属第一医院、海南省卫生健康委员会统计信息中心、鞍钢集团公司总医院、海南省肿瘤医院、营口市中心医院、吉林大学中日联谊医院、河北医科大学第四医院、三亚市人民医院（四川大学华西三亚医院）、吉林大学白求恩第一医院、来宾市人民医院、重庆医科大学附属两江医院、新疆生产建设兵团第六师医院、梧州市人民医院、沈阳市第四人民医院、大连医科大学附属第二医院、辽宁省人民医院、中国医科大学附属盛京医院、大连医科大学附属第一医院、大连市皮肤病医院、大连市妇女儿童医疗中心（集团）、大连市第二人民医院、陆军军医大学西南医院、重庆大坪医院、重庆医科大学附属第一医院、重庆市中医院、徐州医科大学附属医院、海南省第五人民医院、张家港市第一人民医院、杭州师范大学附属医院、湖州市中心医院、合肥市口腔医院、蚌埠医科大学第一附属医院、济宁医学院附属医院、山东大学齐鲁医院、山东大学齐鲁医院德州医院、合肥美亚光电技术股份有限公司、北京中兴医院

本文件主要起草人：陶晓东、鹿晓亮、贺志阳、陈祖吉、何忠芮、刘学学、程美、尤梦祥、王慧莹、贾斐、张卓然、杜雨暄、陈冠衣、任九选、周翔、徐宇辰、王刚、丰雷、李楠茜、贺雪梅、黄程、毕婷婷、杨爱平、张并立、李腾、朱岩、苏逸飞、焦蕴、倪红波、王吟吟、张跃、云琳、朱德刚、杜明超、张洁、夏祺霖、梁栋、于大江、苗杰、杨恒康、陈香梓、徐建光、张梅、杨旗、李锋、孙兆伟、贾旭、田丰、潘晓雄、何怡华、王静怡、张善泉、邢瀚文、乔向东、方建军、刘玉祥、赵侃、张文斌、刘翔宇、彭宇、郭小兵、韩雷、杨毅宁、孙刚、江海林、刘丽、徐荣、许昌、刘广芝、罗慧中、郭柯磊、许峰波、庞艺芊、刘晓今、范军朝、李得江、朱一彬、荆琳、李晓莉、金迟、牛勇勇、韩金丹、徐波雷、汪君、龚维龙、徐培国、廖旺、董斌、杨国杰、纪晨、张海涛、潘文森、霍海洋、刘高金、郑中强、刘朝霞、张强、赵鹏、庞珈琦、马辉、于姗姗、郑欣、翁锦阳、郑炜、张镇洋、项俊铮、刘永伟、刘振坤、刘博、孙中海、王存库、姜红辉、张瑞、刘德志、崔永亮、魏君、张楠、谢君、杨洋、谭诚、黄荣、杨博洋、李华明、辛涛、孙岩国、朱立夫、邵尉、张滨、曹中法、刘洋、吴洋、汪鹏、王飞、黄昊、胡磊、秦涵书、李浩、郑海源、周龙、马行凯、屈陈江、

田昕、王静华、潘奂泉、朱军梅、杨闽、彭欣元、罗永钱、周昂、王清亮、郭吉显、邵宏伟、李晖、王竞波、李贺

医疗健康行业智能体 健康咨询技术要求

1 范围

本文件规定了医疗健康行业智能体 健康咨询在应用过程中涉及的技术能力，从功能要求、智能体能力要求、易用性要求和安全性要求等维度对智能体技术在健康咨询场景中应用的能力提出要求。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1 智能体 AI Agent

又称人工智能代理，是指驻留在某一环境下，能持续自主地发挥作用，具备驻留性、反应性、社会性、主动性等特征的计算实体。

3.2 医疗健康行业智能体 healthcare ai agent

在通用智能体的基础上，结合医疗健康行业特点设计的智能体，与医疗健康相关任务的适配度较高。

3.3 健康咨询 mobile internet health consultation

健康咨询是指医疗机构或其他健康平台通过移动互联网技术和设备，如智能手机、平板电脑等，向用户提供健康和医疗相关信息咨询服务，健康咨询服务包括但不限于疾病预防、营养和健康饮食指导、常见疾病症状分析、医学检查解读、进一步检查建议，健康咨询非诊断、治疗服务，不出具处方。

3.4 健康档案 health records

健康档案是指按照一定的规则建立的关于用户身体健康的一系列指标数据变化的记录，如身高、体重、血压、血氧、心率、血糖、体脂等在一个时间阶段的变化情况以反映一个人的身体健康状态。

4 总体要求

本文件规定了医疗健康行业智能体 健康咨询技术要求，包括功能完备性技术要求、准确性要求、智能体能力要求、易用性要求和安全性要求，主要分为以下内容：功能完备性技术要求部分包括健康档案、健康问答、药物查询、体检报告解读、专病管理路径；准确性要求部分包括二分类、多分类、文书生成及图像分割四项任务；智能体能力要求部分包括感知能力、规划能力、记忆能力和执行能力；易用性要求部分包括可理解性、易学性和易操作性；安全性要求部分包括基础设施安全、数据安全、应用安全。

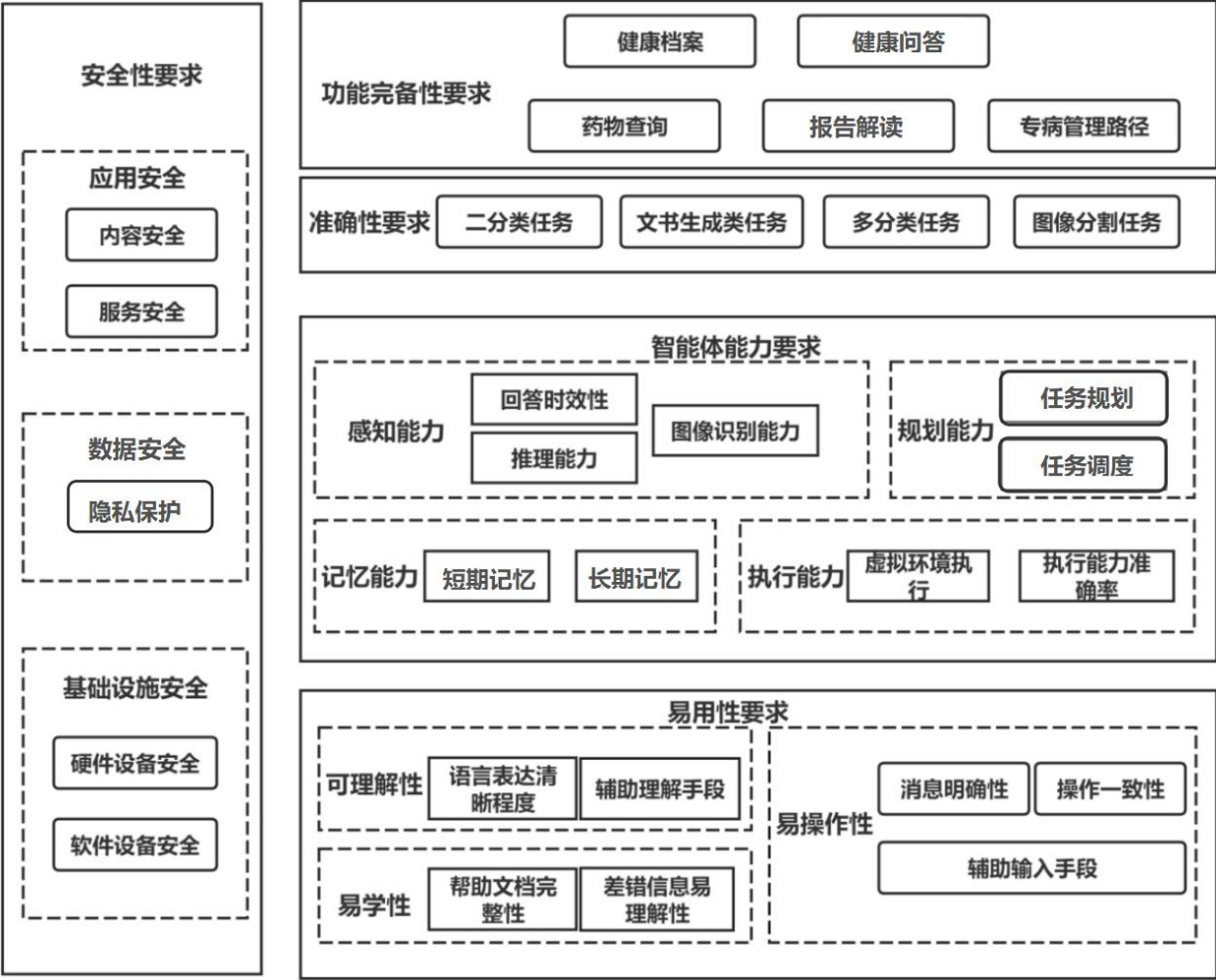


图 1 架构图

5 功能完备性技术要求

5.1 健康档案

医疗健康行业智能体 健康咨询应支持健康档案查询。

- a) 建立健康档案：应支持处理多源异构数据，包括但不限于个人及家人的健康史：既往史、个人史、家族史等，用药计划、复查计划、饮食建议、运动建议等；应支持从各类文件格式中提取患者数据，包括文本、图片、语音等，并支持上传病历、报告单、药物清单、体检报告等资料至健康档案资料夹，应支持将各类异构数据（含提取及上传的多源数据）进行有效整合，形成完整、统一的健康档案，实现多源数据的集中管理，为后续针对性咨询提供全面数据支撑。
- b) 全面指标管理：应按照医学领域具有专业性、权威性、科学性和实用性的指南，如《中国高血压防治指南(2024 年)》，支持健康数据管理，包括血糖、血压、血脂、肝功能、肾功能、心电图等基础生理指标，同时支持通过智能手表、血压计等硬件设备导入数据，并对指标变化进行动态分析。应重点关注与慢性疾病相关的核心指标，如高血压（对应血压指标）、糖尿病（对应血糖指标）、高脂血症（对应血脂指标）、高尿酸血症（对应尿酸指标）、贫血（对应血常规相关指标）等，实现全面覆盖、重点监控的管理效果。
- c) 风险评估：应通过分析个人健康指标数据、家族史等信息，进行各项指标解读，给出全面的健康评估报告。基于风险评估结果，提供针对性的健康干预建议，为用户提供个性化的健康管理指导，帮助用户及早采取预防措施，降低疾病发生的风险。

- d) 健康风险可视化呈现与追踪：应通过图表（如风险等级雷达图、指标异常热力图）等可视化方式直观呈现健康风险，清晰展示风险类别（如心血管疾病风险、代谢疾病风险等）及风险等级；针对每一条健康风险，需在档案中关联呈现疾病科普内容（如疾病定义、发病原因、常见症状）、病程记录追踪（如风险首次发现时间、历次评估结果变化）、病程指标记录及分析（如与该风险相关的指标历史数据、指标变化对风险的影响分析），实现健康风险从识别、追踪到管理的全流程记录，为 AI 咨询服务提供可追溯的风险数据支撑。
- e) 使用干预建议：应支持根据个人健康评估结果，匹配对应的检测建议，提供个性化的健康管理建议，包括饮食建议、运动指导、生活方式指导等，应提供疾病预防指南。
- f) 信息分类与存储：应确保信息条理清晰，便于快速检索与查看，同时要保障数据存储的安全性，遵循相关法律法规并进行加密等保护措施。
- g) 数据更新与维护：应支持实时或定期更新健康档案中的数据，当有新的诊疗结果或健康状况变化时，能及时将相关信息录入并更新档案内容，保证档案始终反映患者最新的健康状况，且在更新过程中要确保数据的准确性和完整性。

5.2 健康问答

医疗健康行业智能体 健康咨询应支持个人看病前的健康问答功能。功能应包括：

- a) 建立咨询者信息：应支持收集咨询者信息应包括性别、年龄，也可包括姓名、身高、体重等；
- b) 咨询问题描述：应支持采用文字、图文、语音描述待咨询的健康问题及想要获得的帮助；
- c) 健康问答小结：系统应支持咨询结束后生成 AI 咨询小结的功能，结合咨询者问题内容解读分析与个人健康档案，输出个性化答复，该小结需涵盖咨询者信息、健康问题、原因分析、医学循证、处置建议、就医建议等核心内容。最终应形成就医小贴士，包括病情描述、既往史、家族史、检查记录、疑似风险、推荐检查等，并支持导医导诊到医院科室。。
- d) 健康问答记录：应支持记录采用文字、图文进行的自查功能，需明确记录咨询时间、自查过程中的问答交互内容，包括补充的症状细节、系统的引导性提问。支持记录症状的具体特征，自动关联系统生成的病情分析结论，为后续就医沟通或健康管理提供全维度参考。

5.3 药物查询

医疗健康行业智能体 健康咨询应支持用药时的查询功能。

- a) 药物信息多渠道查询：用户可通过拍照识别或手动输入药品名称、成分等关键信息进行查询，系统应提供详细的药物信息，包括适应症、用法用量、禁忌反应、副作用等，帮助用户全面了解药品特性。
- b) 个性化用药建议：结合用户的健康档案，包括既往病史、用药记录等信息，为用户提供个性化的用药建议，如药物使用的注意事项、潜在的药物相互作用风险等，同时应支持导入用药计划，并设置定时用药提醒，辅助用户合理规划、规范执行用药方案。

5.4 报告解读

医疗健康行业智能体 健康咨询应支持检查检验报告单、体检报告解读功能。

- a) 上传与信息采集：应支持图片、文件等多格式、多类型的体检报告上传，如历史体检报告、近期复查报告、专项检查报告等。借助 OCR 等技术精准提取报告内文字、数值等信息，确保基础数据采集全面且准确，为后续解读提供可靠素材。
- b) 健康管理空间：应包含原始报告、各类检查数据展示，如各科检查详情，包括一般检查、内科检查、外科检查、眼科、耳鼻喉检查、血常规等原始检查数据；
- c) 提取健康信息：应支持从体检报告中提取个人健康基本信息，如身高、体重、性别、年龄、血压、血糖等，加入健康档案。
- d) 智能解读与健康信息整合：应运用医疗知识图谱与 AI 算法，对报告指标进行智能解读，清晰呈现指标是否异常、异常程度等；同时提取身高、体重、血压等基础健康信息，并支持与既往健康史整合，形成更完整的健康数据维度。
- e) 疾病预测与风险分层：基于解读后的指标及整合的健康信息，结合临床指南与大数据模型，对疾病发生风险进行预测，且明确划分高、中、低等风险等级，让用户直观知晓健康隐患程度。

- f) 个性化解决方案推荐：应针对不同风险等级，匹配差异化解决方案。高风险时，对接就医资源，提供导医导诊服务，助力及时诊疗；中风险则制定定期复查计划或推荐体检套餐，持续监测健康；低风险或需日常关注的情况，推送健康知识、康复建议等内容，辅助日常健康管理。
- g) 服务闭环构建：应结合复查、健康管理等环节，在推荐解决方案后，持续推进就医资源对接、复查计划落实、健康知识触达等后续服务，形成从报告解读到健康干预的完整闭环，保障用户健康需求得到持续响应。
- h) 安全与隐私保护：应可通过多项权威安全认证，如 DSMC 国家级数据安全领域认证、ISO27001 信息安全管理体系国际标准认证、可信云认证、国家信息安全登记保护三级认证等，保障用户体检数据及健康信息的隐私安全，构建权威、准确、易懂、安全、个性化的健康报告。
- i) 交互体验便捷性与持续性：应支持用户以语音、文字等多种方式快速查询体检报告解读相关内容，给出详细且通俗易懂的回答，界面设计应符合用户使用习惯，可快速定位到自己关心的报告解读信息。应及时收集用户对体检报告解读服务的反馈意见，用于持续优化解读内容、改进交互流程，不断提升用户体验。
- j) 健康风险提示准确性：应以专业医学研究成果和临床数据为参考依据，对体检报告解读中疾病预测的准确性进行评估。

5.5 专病管理路径

医疗健康行业智能体 健康咨询应支持专病管理路径查询功能。

- a) 构建专病管理路径：应严格依据国内外临床指南、高影响因子论文及专家共识来制定。对疾病的诊疗、康复等各阶段关键节点进行明确规范，确保路径具备权威的医学理论支撑，能为后续个性化管理提供科学、通用的框架指引，涵盖疾病从确诊到康复全周期的核心干预要点与流程规范。
- b) AI 抽取患者画像：应依托病历、检验、检查等多维度医疗数据，运用先进的人工智能技术进行深度分析与提取。要精准识别患者的疾病类型、严重程度、基础健康状况、既往病史、用药情况等关键信息，生成全面且精准的患者个体特征画像，为后续个性化管理路径生成提供精准的个体数据基础。
- c) 自动执行康复计划：可借助外呼机器人、微信小程序等数字化工具，按照生成的个性化康复计划，自动向患者推送康复目标、要点、用药指导、运动指导、饮食建议、居家护理等内容。确保信息传递的及时性与准确性，能高效、规范地将康复计划落地执行，保障患者在院外也能按计划开展康复相关活动。
- d) 患者咨询：可为患者提供便捷的咨询渠道，支持患者与医生进行实时互动交流。患者可就康复过程中遇到的疑问、身体出现的异常状况等进行咨询，医生需及时、专业地予以解答，为患者提供针对性的指导，助力患者更好地遵循康复计划，解决康复过程中的各类问题。
- e) 效果总结：可全面收集患者在康复过程中的各项数据，包括体征监测结果、康复目标达成情况、疾病评估数据等。对康复计划的执行效果进行系统、科学的总结与评估，分析康复过程中的优势与不足，为后续优化专病管理路径、改进康复计划提供依据，同时也能清晰呈现患者的康复成果，让患者和医护人员都能直观了解康复效果。

6 性能评估指标

6.1 二分类任务

医疗健康行业智能体 健康咨询涉及的二分类任务包括健康指标（正常、异常）、健康风险（有、无）等。将通过以下指标衡量智能体准确性。

二分类混淆矩阵

分类	人工智能分类	
	阳性	阴性

参考标准分类	阳性	真阳性（TP）	假阴性（FN）
	阴性	假阳性（FP）	真阴性（TN）

说明：TP为真阳性（患病/异常且正确识别），TN为真阴性（健康/正常且正确识别），FP为假阳性（健康/正常但误判为异常），FN为假阴性（患病/异常但误判为正常）。

a) 灵敏度（Sen）

$$\text{Sen} = \text{TP} / (\text{TP} + \text{FN}) \times 100\%$$

说明：衡量实际异常样本中被正确识别的比例，核心聚焦“不漏诊”，如肺结节“有结节/无结节”判断、眼底“病变/正常”识别，确保潜在疾病不被遗漏。

b) 特异度（Spe）

$$\text{Spe} = \text{TN} / (\text{TN} + \text{FP}) \times 100\%$$

说明：衡量实际正常样本中被正确识别的比例，核心聚焦“不误诊”，如甲状腺超声“良性/恶性”初步筛查，避免假阳性导致用户过度焦虑和不必要的医疗干预。

c) 准确率（Acc）

$$\text{Acc} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \times 100\%$$

说明：整体正确识别的样本占总样本的比例，反映模型综合识别效果，如皮肤中“良性痣/恶性黑色素瘤”初步判断，给出整体可靠的识别结果。

6.2 文书生成类任务

医疗健康行业智能体 健康咨询涉及的文书生成类包括咨询问题描述、健康问答小结、使用干预建议（根据个人健康评估结果，匹配对应的检测建议）、生成的个性化用药建议、个性化解决方案推荐等，针对此任务，将通过以下指标衡量生成内容与标准答案的区别。

a) ROUGE-N：对摘要任务，计算客观指标ROUGE-N，其计算公式如下：

$$\text{ROUGE} - N = \frac{\sum S \in \{\text{ReferenceSummaries}\} \sum \text{gram}_n \in \text{sCount}_{\text{match}}(\text{gram}_n)}{\sum S \in \{\text{ReferenceSummaries}\} \sum \text{gram}_n \in \text{sCount}(\text{gram}_n)}$$

式中：

N——即 n-gram，文本内容滑动窗口字节数，参考值为 2；

$\text{Count}_{\text{match}}(\text{gram}_n)$ ——参考摘要和机器生成摘要中共有的 n-gram 的数量；

$\text{Count}(\text{gram}_n)$ ——参考摘要中 n-gram 的数量；

说明：衡量机器生成文本与参考文本的重叠程度，反映生成内容的完整性和准确性，用于健康咨询中影像相关文书生成，如医学影像解读报告、影像相关健康问答小结，确保生成内容与专业参考文本高度契合。

b) 关键词命中率

$$\text{Hit_Rate} = \frac{\text{Count}_{\text{Hit}}}{\text{len}(\text{keyword})}$$

式中：

$\text{Count}_{\text{Hit}}$ ——机器生成文本中命中关键词的数量

$\text{len}(\text{keyword})$ ——关键词的数量

说明：衡量生成文本对核心信息的覆盖程度，确保关键医学信息不缺失，用于个性化建议生成，确保关键词准确命中。

c) BERTScore: 对生成任务, 计算客观指标BERTScore, 计算公式如下:

$$\begin{aligned} sim(x_i, y_i) &= \frac{Emb(x_i) \cdot Emb(y_i)}{\|Emb(x_i)\| \|Emb(y_i)\|} \\ Precision &= \frac{1}{|x|} \sum_{x_i \in x} \max_{y_i \in y} sim(x_i, y_i) \\ Recall &= \frac{1}{|y|} \sum_{y_i \in y} \max_{x_i \in x} sim(x_i, y_i) \\ BERTScore &= \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \end{aligned}$$

式中:

$Emb(x_i)$ ——句子 x 中词语在经过编码器后的嵌入向量;

$Emb(y_i)$ ——句子 y 中词语在经过编码器后的嵌入向量;

$sim(x_i, y_i)$ ——两词语嵌入向量的余弦相似度;

$Precision$ ——精确率;

$Recall$ ——召回率;

说明: 通过预训练语言模型, 将生成文本与参考文本(标准答案, 如专业医生撰写的健康咨询文书、临床指南推荐的规范表述)映射至语义空间, 计算两者 token 级别的语义相似度, 最终输出综合得分。如在健康问答小结生成: 参考文本为“含咨询者信息、症状原因分析(符合临床指南)、医学循证依据、处置建议(如休息 + 用药提醒)、就医建议(如持续发热需就诊)”的规范小结; 生成文本为智能体输出的问答小结, 通过 BERTScore 衡量两者在“症状归因逻辑”、“医学建议语义一致性”上的匹配度, 确保生成小结语义符合专业表述, 避免因字符相似但语义偏差导致的误导(如“避免辛辣饮食”与“可少量食用辛辣”的语义区分)。

6.3 多分类任务

医疗健康行业智能体 健康咨询涉及的多分类任务为全面指标管理(如血糖、血脂、肝功能等, 对指标异常类型进行多类别判断, 例如“血糖异常”可分为“正常、轻度升高、中度升高、重度升高”4类)、疾病风险评估(如低风险、中风险、高风险)、识别患者的疾病类型、健康干预方案等, 针对此任务, 我们将通过以下指标衡量智能体的准确性。

a) Macro-Precision: 宏精准率

$$Macro - P = \frac{1}{K} \sum_{i=1}^K Precision_i$$

式中:

K ——分类任务的总类别数(如三分类时 $K=3$)

$Precision_i$ ——第 i 类精准率, 计算公式为 $Precision_i = \frac{TP_i}{TP_i + \sum_{j \neq i} FP_{j \leftarrow i}}$ (TP_i 为 i 类真阳性数, $FP_{j \leftarrow i}$ 为实际非 i 类但预测为 i 类的假阳性数, $j \neq i$ 时属于 i 类的误判)

说明: 计算各类别精准率的平均值, 不考虑类别样本不均衡, 平等对待每一类, 用于健康咨询中多级别异常判断, 如“血糖指标异常”分为正常、轻度升高、中度升高、重度升高类别, 确保各类别预测结果的可信度。

b) Macro-Recall: 宏召回率++

$$\text{Macro-R} = \frac{1}{K} \sum_{i=1}^K \text{Recall}_i$$

式中：

K——分类任务的总类别数（如三分类时 K=3）

Recall_i ——第 i 类精准率，计算公式为 $\text{Recall}_i = \frac{TP_i}{TP_i + \sum_{j \neq i} FN_{i \rightarrow j}}$ （ TP_i 为 i 类真阳性数， $FN_{i \rightarrow j}$ 为实际 i 类但预测为 j 类的假阴性数， $j \neq i$ 时属于 i 类的漏判）

说明：计算各类别召回率的平均值，保障每一类别的样本都能被有效识别，可用于疾病风险多等级评估，如肺结节“低风险、中风险、高风险”分类，确保不同风险等级的结节都能被准确识别，避免高风险漏判。

c) Macro-F1：宏精准率与宏召回率的调和平均

$$\text{Macro-F1} = \frac{2 \times \text{Macro-P} \times \text{Macro-R}}{\text{Macro-P} + \text{Macro-R}}$$

说明：调和宏精准率与宏召回率，平衡“不误判”和“不漏判”，解决多分类场景下类别不均衡问题。用于健康咨询中复杂影像多分类任务，如肝脏“正常、轻度脂肪肝、中度脂肪肝、重度脂肪肝、肝癌”分类，综合保障各类别识别的准确性和全面性。

6.4 图像分割任务

医疗健康行业智能体 健康咨询涉及的图像分割任务为医学影像识别、健康档案中生理指标图像显示、患者画像抽取等，针对此任务，我们将通过以下指标衡量智能体的准确性。

a) 准确率(Accuracy)

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \times 100\%$$

说明：实际分割结果中正确的像素占总像素的比例，反映整体分割准确性，适用于健康咨询中各类医学影像病灶分割，如肿瘤影像分割、骨折区域分割，快速评估整体分割效果，为后续健康评估提供基础数据。

b) 交并比 (IOU)

$$\text{IOU} = TP / (TP + FP + FN)$$

说明：衡量模型分割结果与医生标注金标准的重叠比例，直观反映分割精准度，用于健康咨询中精准病灶分割任务，如肝癌影像中肿瘤边界分割，确保分割区域与实际病灶高度吻合，支撑精准病情评估。

c) DICE 系数 (DICE)

$$\text{DICE} = 2 * TP / (TP + FP + TP + FN)$$

说明：与 IOU 高度相关，更侧重衡量两个区域的重叠程度，取值 0-1，越接近 1 表示分割效果越好，用于小病灶或不规则病灶分割评估，如肺小结节影像分割、脑血管狭窄区域分割，精准衡量分割结果与金标准的契合度，保障小病灶不被遗漏。

d) 精确率 (Precision)

$$\text{Precision} = TP / (TP + FP)$$

说明：衡量分割结果中实际为病灶的比例，避免过度分割导致的病情误判，用于需严格控制假阳性的分割任务，如影像钙化点分割，避免将正常钙化误判为病变钙化，减少不必要的进一步检查。

e) 召回率 (Recall)

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN})$$

说明：衡量实际病灶中被成功分割的比例，确保病灶不被遗漏，用于关键病灶分割任务，如肺癌影像中肿瘤浸润区域分割，确保所有病灶区域都能被完整分割，为疾病分期和治疗建议提供全面依据。

7 智能体能力要求

7.1 感知能力

7.1.1 回答时效性

医疗健康行业智能体应具备一定的回答时效性。

通过用户从发起请求到系统或应用程序返回结果的时间计算响应实时性，计算方式参见式：

$$\text{ES_T} = \text{R_T}_{\text{finish}} - \text{R_T}_{\text{start}}$$

式中：

ES_T ——响应时间；

R_T_{finish} ——医疗健康行业智能体返回结果的时间；

R_T_{start} ——用户发起请求的开始时间。

说明：计算用户从发起咨询请求到智能体返回结果的耗时，反映响应速度，可用于紧急咨询，如急诊快速解读、突发症状分析，确保在短时间内给出响应，满足紧急健康咨询需求。

7.1.2 图像识别能力

a) 准确率（Acc）

$$\text{Acc} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

说明：整体正确识别的影像样本占总样本的比例，反映图像识别综合效果，可用于健康咨询中各类通用影像识别，如对用户上传的X光影像进行“是否存在肺纹理增粗”识别时，用准确率衡量整体分类正确性（如准确率≥92%）。若准确率达标，可快速输出“有肺纹理增粗，提示支气管炎风险”或“无异常”的结论，为健康评估提供基础。

b) 灵敏度（Sen）

$$\text{Sen} = \text{TP} / (\text{TP} + \text{FN})$$

说明：实际异常影像中被正确识别的比例，聚焦“不漏诊”，如疾病筛查类影像识别的肺癌CT影像筛查、宫颈癌HPV检测等影像识别，确保潜在疾病影像不被遗漏，及早发现健康隐患。

c) 特异度（Spe）

$$\text{Spe} = \text{TN} / (\text{TN} + \text{FP})$$

说明：实际正常影像中被正确识别的比例，聚焦“不误诊”，需避免过度医疗的影像识别，如普通体检中的腹部超声影像判断，避免将正常影像误判为异常，减少用户不必要的担忧和检查。

d) 精确率（Pre）

$$\text{Pre} = \text{TP} / (\text{TP} + \text{FP})$$

说明：预测为异常的影像中实际为异常的比例，反映阳性预测结果的可信度，可用于高风险疾病影像初步诊断，如脑梗死影像识别，确保预测为异常的结果具有高可信度，为后续就医指导提供可靠依据。

e) F1分数

$$\text{F1} = 2 \times (\text{Pre} \times \text{Sen}) / (\text{Pre} + \text{Sen})$$

说明：综合衡量图像识别结果的精准性、全面性，调和精确率与灵敏度，平衡“不漏诊”和“不误判”，解决类别不平衡问题，可用于健康咨询中复杂影像识别任务，如多发性病变影像识别（如多部位结节、多器官异常），综合保障识别的全面性和准确性。

7.1.3 推理能力

7.1.3.1 指代消解

医疗健康行业智能体在指代消解能力上应具备一定的准确率。

计算对话系统的指代消解准确率，即多轮对话中某个轮次代词或名词可能指代多种不同事物情况下识别正确。计算公式如下：

$$P_M = \frac{m}{M} \times 100\%$$

式中：

P_M ——本轮指代消歧平均准确率；

m ——本轮中每个代词或名词短语被正确识别的次数；

M ——本轮中所有代词或名词短语数量。

可用于影像相关多轮健康咨询，如用户先提及“肺部CT显示有阴影”，后续询问“它需要进一步检查吗”，智能体需准确识别“它”指代“肺部阴影”，确保问答逻辑连贯。

7.1.3.2 知识推理

医疗健康行业智能体在知识推理能力上应具备一定的准确率。

根据推理总数和推理正确数，计算F1值：

$$F1 = \frac{2 \times P \times R}{P + R}$$

式中：

P ——预测正确的数量/预测出的总数量；

R ——预测正确的数量/实际总数量。

可用于基于影像的健康推理咨询，如结合用户“肺部结节影像”、“吸烟史”、“家族肺癌史”，推理结节恶变风险，给出个性化健康建议。

7.2 规划能力

7.2.1 任务规划

7.2.1.1 目标拆解

医疗健康行业智能体在目标拆解能力上具备一定的性能优越度。

- 目标识别认知：医疗健康行业智能体应支持对目标进行深入认知，包括但不限于关键信息和潜在障碍，如用户目标为“解读肺部CT报告并评估风险”，需明确核心信息为“结节大小、形态、密度”，潜在障碍为“报告缺乏关键参数”。
- 目标分析预测：医疗健康行业智能体应支持对目标进行分析预测，包括但不限于目标概念、结构、复杂性、层次、目标间关系等，如“健康管理”目标可拆解为“健康解读、风险评估、干预建议、随访规划”等子目标。
- 拆解关联度：医疗健康行业智能体应确保子目标间高度关联，如“甲状腺超声解读”拆解为“特征识别、良性/恶性初步判断、后续检查建议”，子目标层层递进；
- 拆解合理性：医疗健康行业智能体拆解目标时宜参考拆解子目标可行性、依赖关系和优先级，保障拆解目标及可操作性，如“体检综合解读”需先完成各单项（如血常规、血压）解读，再进行综合健康评估；
- 拆解可解释性与可视化：医疗健康行业智能体应支持提供拆解规划方案的详细解释和可视化展示，帮助用户或开发者理解方案的生成过程和结果，如通过流程图向用户呈现“报告上传→信息提取→智能解读→风险评估→建议生成”的完整流程。

7.2.1.2 规划策略

医疗健康行业智能体应支持任务内或任务间的组织规划。

- 规划结构性：支持按照一定的结构的任务结构进行任务规划，如线性、分层、并行、树状、网状、条件、迭代等；
- 规划逻辑性：医疗健康行业智能体拆解任务的合理性，确保子任务之间有明确的逻辑关系；
- 规划一致性：医疗健康行业智能体在任务间或任务内部组织规划时，应具备规划一致性和协调性，避免重复任务、死循环任务、冲突任务及无效任务等不一致问题；

- a) 冲突解决预案：智能体应具备对内部策略冲突进行预案的能力。在规划阶段应对可能出现的策略冲突节点进行预判并准备合理冲突解决预案；
- d) 规划策略准确率：即智能体为了完成指定任务给出的规划步骤中，有多少步骤是必要的有效步骤。计算公式如下：

$$E_p = \frac{R_1}{R} \times 100\%$$

式中：

E_p ——规划策略准确率；

R_1 ——完成指定任务所提供的规划策略中有效的操作数量；

R ——完成指定任务提供的规划策略中总的操作数量；

7.2.2 任务调度

7.2.2.1 调度机制

医疗健康行业智能体应支持多种任务调度机制，具备一定鲁棒性。

- a) 调度机制：医疗健康行业智能体调度机制的可选度，如先来先服务、短作业优先、轮转调度机制、优先级调度机制、最早截止时间优先等；
- b) 鲁棒性：医疗健康行业智能体在面对异常情况时应能够迅速适应并重新规划任务调度，如自动干预及手动干预；
- c) 自主性：系统应支持自动调度医疗健康行业智能体工作和协同。

7.2.2.2 组织协调

医疗健康行业智能体执行任务时应具有各项组织协调能力。

- a) 资源协调：医疗健康行业智能体对时间分配、计算资源管理、数据访问与存储、多源信息整合的支持度；
- b) 任务分配：医疗健康行业智能体应在并发请求场景下，智能体应能依据实时资源状况，将任务动态分配到不同的处理单元，实现负载均衡；
- c) 进度监控：医疗健康行业智能体应支持监控流程执行进度，并对异常情况进行报警；
- d) 应急处置：当紧急事件发生，医疗健康行业智能体应支持灵活调整任务分配策略，具备应急能力。

7.3 记忆能力

7.3.1 短期记忆能力

7.3.1.1 痕迹检索

医疗健康行业智能体应支持痕迹检索功能。

- a) 检索精确度：检索到的痕迹是否准确，是否与原始输入匹配，计算公式如下：

$$P_A = \frac{A_1}{A} \times 100\%$$

式中：

P_A ——短期痕迹检索精确度；

A_1 ——正确检索的数量；

A ——总检索数量；

- b) 检索完整性：检索到的痕迹是否完整，是否包含所有相关的信息，计算公式如下：

$$P_B = \frac{B_1}{B} \times 100\%$$

式中：

P_B ——短期痕迹检索完整性；

B₁ ——完整检索的数量；

B ——总检索数量；

c) 检索速率：完成检索的平均时间，计算公式如下：

$$P_c = \frac{C}{D}$$

式中：

PC ——平均检索时间；

C ——总检索时间；

D ——检索次数。

7.3.1.2 提示词管理

医疗健康行业智能体应具备提示词管理相关功能。

- a) 模板丰富度：医疗健康行业智能体应具备多种预制的提示词模板，如文本生成类、知识问答类、逻辑推理类等；
- b) 框架丰富度：医疗健康行业智能体应支持的提示词框架丰富度，即在不同框架提问下效果稳定，如ICIO 框架、CRISPE 框架、BROKE 框架等；
- c) 模板管理：医疗健康行业智能体应具备提示词模板管理功能，如创建、修改、删除等；

7.3.2 长期记忆能力

7.3.2.1 记忆存储

医疗健康行业智能体应支持记忆尽量多轮次的历史对话。

- a) 历史对话轮次：计算在模型性能没有明显下降的情况下，医疗健康行业智能体最长可以支持的历史对话轮次；
- b) 知识更新：医疗健康行业智能体知识更新频次与质量；
- c) 存储容量：能够记住和存储的长期痕迹字符数量。

7.3.2.2 快速检索

医疗健康行业智能体应支持快速检索功能。

- a) 检索速度：医疗健康行业智能体从接收到查询请求到返回检索结果所需的时间；
- b) 检索准确性：医疗健康行业智能体返回的检索结果与用户查询意图的匹配程度；
- c) 检索覆盖范围：医疗健康行业智能体能够检索到的信息来源和类型。

7.4 执行能力

7.4.1 虚拟环境执行能力

医疗健康行业智能体在虚拟环境下应具备虚拟环境执行能力。

- a) 交互积极性：医疗健康行业智能体的交互积极性，应可以从被动服务向主动服务转变；
- b) 交互对象多样性：医疗健康行业智能体与软件环境中的其他实体进行交互的支持度，其他实体包括其他agent、系统、环境本身等；
- c) 数据格式多样性：医疗健康行业智能体需要对接收到的软件环境信息进行理解和解码的能力，

环境数据包括文本及多模态数据；

- d) 工具丰富度：医疗健康行业智能体可以调用外部工具的数量，如文档解析、语音识别、数据库访问、图像识别等。
- e) 执行容错与回退机制：智能体在执行过程中应具备处理操作失败等异常情况的能力，能进行错误提示、启动备选方案或安全回退，并记录故障信息。

7.4.2 执行能力准确率

- e) 计算执行能力准确率，即智能体为了完成指定任务给出的规划步骤中，有多少步骤执行后得到了正确的结果。计算公式如下：

$$P_p = \frac{C_1}{C} \times 100\%$$

式中：

P_p ——执行能力准确率；

C_1 ——完成指定任务所提供的规划策略中得到正确结果的操作数量；

C ——完成指定任务提供的规划策略中总的操作数量。

8 易用性要求

8.1 可理解性

8.1.1 语言表达清晰程度

医疗健康行业智能体界面文字、提示及交互内容应简洁准确，避免生僻医学术语。必要术语需附通俗释义，患者端信息表述需无歧义。

8.1.2 辅助理解手段

医疗健康行业智能体涉及复杂医学知识、操作流程等操作宜辅以图文、动画或视频展示；关键环节宜设引导提示，提升信息理解效率。

8.2 易学性

8.2.1 帮助文档完整性

医疗健康行业智能体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。

8.2.2 差错信息易理解性

医疗健康行业智能体操作错误或系统异常时，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。

8.3 易操作性

8.3.1 操作一致性

医疗健康行业智能体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。

8.3.2 消息明确性

- a) 医疗健康行业智能体向用户推送的各类消息，如检查提醒、复诊通知、用药提示等，内容应明

确具体，包含关键信息，如时间、地点、注意事项等。

- b) 医疗健康行业智能体向用户推送的各类消息的标题和正文应简洁明了，不应使用冗长复杂的表述。
- c) 医疗健康行业智能体消息推送应具备合理的频率和时机，不应过度打扰用户。

8.3.3 辅助输入手段

医疗健康行业智能体应支持智能联想、语音、手写等多种输入方式。

9 安全性要求

9.1 基础设施安全

9.1.1 硬件设备安全性

医疗健康行业智能体涉及的硬件设备（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、授权等；
 - （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、管理人员等；
- b) 网络设备安全专用要求：分布式训练、推理时应满足组网安全保障要求，包含网络带宽、网络时延、网络丢包率、网络抖动等；
- c) 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应具备稳定运行的能力，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。

9.1.2 软件设备安全性

医疗健康行业智能体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。

9.2 数据安全

医疗健康行业智能体应支持数据采集、数据预处理、数据使用等数据相关内容具备安全防护能力，包含：存储安全、隐私保护、过程安全、销毁安全等。

9.2.1 隐私保护

应遵循最小必要原则采集数据，对个人信息特别是医疗健康个人敏感信息，需在采集前明确告知用

户采集目的、范围及使用方式，获取用户明同意；数据使用过程中需采取去标识化或匿名化处理（如删除姓名、身份证号等直接标识符，对关键医疗指标进行脱敏转换），避免敏感信息直接暴露；同时禁止将医疗健康个人敏感信息用于与健康咨询无关的场景，禁止向未授权第三方共享。

9.3 应用安全

9.3.1 内容安全

医疗健康行业智能体输出内容（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，包括医疗健康行业智能体输出内容（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，包括医疗健康行业智能体输出内容（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。

9.3.2 服务安全

医疗健康行业智能体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：医疗健康行业智能体涉及的模型安全性应满足模型安全保障要求，包含 MTTF、服务安全性、服务合规性、反馈处置机制等。

参 考 文 献

[1]ISO/IEC TS 4213:2023, Artificial intelligence-Performance evaluation methods for image segmentation[S]
