ICS 35.020
CCS L70

团 体 标 准

T/ISC 0172—2026

医疗语音大模型技术要求

Technical requirements for medical speech large model

（发布稿）

2026 - 10 - 28 发布 2026 - 11 - 28 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 缩略语 2
5 总体要求 3
6 模型能力要求 4
6.1 语义理解 4
6.2 推理应用 4
6.3 语音识别 5
6.4 语音合成 5
6.5 语音交互 6
7 模型训练要求 6
7.1 数据集构建 6
7.2 部署应用 7
8 评估指标要求 7
8.1 客观指标 7
8.2 主观指标 12
9 易用性要求 12
9.1 可理解性 12
9.2 易学性 12
9.3 易操作性 13
10 安全性要求 13
10.1 基础设施安全 13
10.2 数据安全 13
10.3 应用安全 13

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、安徽省第二人民医院（安徽医学高等专科学校附属医院、安徽省职业病防治院）、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）、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）、合肥市蜀山区卫生健康委员会、合肥市庐阳区四里河街道社区卫生服务中心

本文件主要起草人：韩苹、陈良、鲁舜毅、张志、朱颜渝、白晶晶、彭黎君、邵天辰、尹大海、贺志阳、倪金红、程美、彭跃辉、周兆旦、张清池、尤梦祥、史惠婷、张晓燕、王慧莹、贾斐、任九选、张卓然、杜雨暄、陈冠衣、相识、刘泊宁、黄涂半特、黄亚、王真明、陆姚、李飞、王曦、宋荣、巩天宁、吴淑华、吴娅莉、刘树瀚、汤小铭、周文静、陈笑、瞿其梅

医疗语音大模型技术要求

1 范围

本文件规定了医疗语音大模型涉及的技术能力要求，从模型能力要求、模型训练要求、评估指标要求、易用性要求和安全性要求等维度对医疗大模型技术在语音大模型场景中应用的能力提出要求。

本文件适用于医疗语音大模型提供方对自身技术与应用能力的评估、管理方对医疗语音大模型服务提供方服务能力的要求，以及第三方对医疗语音大模型提供方能力的评估等场景。

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/T41813.1-2022信息技术智能语音交互测试方法第1部分:语音识别

GB/T41813.2-2022信息技术智能语音交互测试方法第2部分:语义理解

GB/T41867-2022信息技术 人工智能术语

GB/T45288.1-2025人工智能 大模型第1部分:通用要求

GB/T45288.2-2025人工智能 大模型第7部分:语音大模型

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

语音大模型 speech large model

能按要求理解或生成语音数据，模型输入和输出至少有一方包含语音数据的大模型。

注 1：“要求”一般以自然语言、符号或语音的形式描述。

注 2：语音数据指人类语音交流中产生的数据，包含话语内容、环境噪声、语气、情感以及非言语的声效等，及其自然语言或符号说明信息。

注 3：语音大模型的输入与输出由任务决定，包含语音、标识符号、自然语言等的一种或多种。

注 4：语音大模型一般具备语音识别、语音合成等一种或多种能力。

[来源：GB/T45288.2-2025人工智能 大模型第7部分:语音大模型]

3.3

医疗语音大模型

基于大规模医疗语音数据和医学知识进行预训练，具备医疗场景下语音识别、语音合成、语音理解、语音交互等核心能力，能满足医疗健康领域特定语音相关需求的大模型。

3.4

医疗语音识别

利用功能单元进行的，从医疗场景语音信号到语音内容（含医学术语、诊疗信息等）的转换过程。

3.5

语义理解 semantic understanding

使医疗语音大模型理解医疗场景中人说话的意图，包括诊疗需求、咨询诉求、指令传达等。

3.6

语音合成 speech synthesis

通过机械的、电子的方法合成符合医疗场景规范和沟通需求的人类语言的过程。

3.7

全双工交互

模型在进行语音输出时，可实时接收用户的语音输入（即用户可随时打断模型输出），模型需立即终止当前输出，且不丢失此前已交互的上下文信息（如用户打断模型讲解用药指导后，模型可衔接之前的对话逻辑，无需重新开始），适配医疗场景中用户可能随时补充症状、追问诊疗相关问题的需求，提升交互连贯性与实用性。

3.8

公有云部署

将医疗语音大模型部署在第三方云服务商提供的公共云计算基础设施上，通过网络向多个用户或机构提供模型服务的方式。

3.9

私有化部署

将医疗语音大模型部署在用户自建或专属的计算基础设施上，数据与服务完全独立运行、不与其他用户共享资源的部署方式。

3.10

鲁棒性 robustness

人工智能模型的核心性能指标，在医疗语音大模型场景中，指模型面对医疗场景复杂声学环境（如诊室背景噪音、不同设备录音、远距离发音）、多样化说话特征（如医护/患者不同口音、语速、声调）、专业术语口语化表达等非理想输入时，仍能保持稳定、准确的语音识别和语义理解能力，是模型适配临床实际应用的关键特性。

4 缩略语

下列缩略语适用于本文件。

AAC：高级音频编码（Advanced Audio Coding）

API：应用程序编程接口（Application Programming Interface）

BLEU：双语评估替换（Bilingual Evaluation Understudy）

CDA：临床文档架构（Clinical Document Architecture）

CER：字错误率（Character Error Rate）

DER：分离错误率（Diarization Error Rate）

DICOM：医学数字成像和通信（Digital Imaging and Communications in Medicine）

EMR：电子病历系统（Electronic Medical Record System）

FAR：错误接受率（False Acceptance Rate）

FLAC：无损音频压缩编码（Free Lossless Audio Codec）

FN：假负例的数量（False Negative）

FP：假正例的数量（False Positive）

HDFS：分布式文件系统（Hadoop Distributed File System）

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

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

LIS：实验室信息系统（Laboratory Information System）

M4A：音频格式（MPEG 4 Audio）

MOS：平均意见得分（Mean Opinion Score）

MP3：动态影像专家压缩标准音频层面（MPEG Audio Layer 3）

MTBF：平均无故障工作时间（Mean Time Between Failure）

OGG：开源音频压缩格式（Ogg Vorbis）

RESTful：表述性状态转移接口（Representational State Transfer）

SCR：句识别正确率（Sentence Correct Rate）

SDK：软件开发工具包（Software Development Kit）

TN：真负例的数量（True Negative）

TP：真正例的数量（True Positive）

WAV：波形音频文件格式（Waveform Audio File Format）

WER：词错误率（Word Error Rate）

WS 363-2011：卫生信息数据元目录（卫生行业标准）

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

5 总体要求

医疗语音大模型是人工智能大模型技术的重要应用形态，通过语音识别、语音合成、语义理解与语音交互等能力，在门诊、病房等医疗场景中实时生成电子病历、病程记录、护理记录与检查报告，可开展健康宣教、用药指导与出院随访，辅助智能问诊与远程会诊等，减轻医护人员文书录入负担，为患者提供便捷的语音交互服务，全面提升医疗服务效率、质量与可及性。

本文件规定了医疗语音大模型技术要求，包括模型能力要求、模型训练要求、评估指标要求、易用性要求和安全性要求，主要分为以下内容：模型能力要求部分包括语义理解、推理应用、语音识别、语音合成、语音交互；模型训练要求部分包括数据集构建和部署应用；评估指标要求部分包括客观指标和主观指标；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。

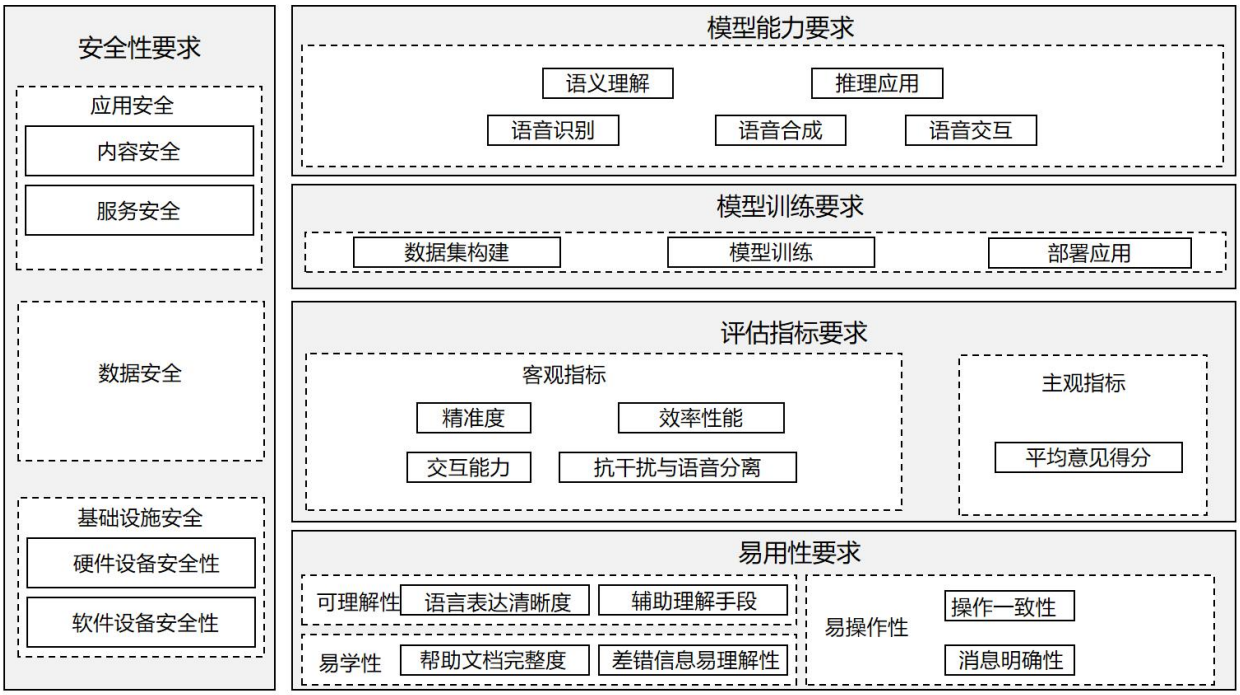

图 1　医疗语音大模型技术要求示意图

6　模型能力要求

6.1　语义理解

医疗语音大模型的语义理解与推理应用能力要求如下：

a）基础语义理解能力：应具备信息抽取、逻辑推理、任务分解、摘要总结、命名实体识别、敏感信息辨识等基础语义理解能力；

b）医学知识融合能力：应支持深度融合多源医学知识与语音数据，实现多源知识统一表示，确保语音内容与对应医学知识精准关联，提升模型临床专业性；

c）医疗场景诉求识别能力：应精准识别各类医疗场景的核心诉求，涵盖诊疗咨询、病因询问、用药指导、检查解读、健康管理等场景，能明确区分问题的主次关系与潜在需求，完整捕捉复杂医学问题的关键信息（如多症状并发咨询、跨系统疾病疑问），无核心诉求遗漏或理解偏差，确保问题理解与用户真实意图一致；

d）多语境医疗表述适配能力：应适配不同用户、不同语境的医疗表述习惯，能准确理解医学问题中的专业表述、通俗表达、模糊描述（如“身上不舒服”、“没力气”），同时可理解不同文化、地域、民族背景下的医疗相关特定表达（如地方通俗症状描述、民族特有健康认知相关表达），避免因表述或文化语境差异导致的语义理解偏差；

e）医疗意图识别能力：应具备医疗场景意图识别能力，以准确理解用户在语音交互中的真实意图；

f）多轮上下文理解能力：应具备上下文相关的多轮对话理解能力，有效记忆并利用历史交互信息，保持语义连贯；

g）指代消解与省略句补全能力：应具备处理指代消解与省略句补全的能力，可正确处理对话中的代词、指代性表述及省略句式（如“他”、“她”、“该检查项目”、“明天做？”等）；

h）多语种多方言语义理解能力：应具备多语种、多方言的自由混说语义理解能力，不会因语言切换导致语义中断或理解偏差；

i）用户情感感知能力：宜具备感知用户情感的能力，可识别对话中的情绪倾向，为交互回复提供依据。

6.2　推理应用

6.2.1 推理过程

医疗语音大模型的推理应用的推理过程能力要求如下：

a) 应以医学知识（诊疗指南、药品说明书、临床路径、医学共识等）为依据，按照“信息整合-逻辑推导-结论形成”的流程，符合临床诊疗逻辑；
b) 应支持多维度信息融合推理，综合患者症状、病史、个体情况（年龄、性别、过敏史等）、医疗检查结果等多源信息，避免单一信息维度的片面推理；
c) 应具备因果关联性，明确输入信息-医学推理依据-结论的对应关系，确保推理链条完整可追溯、可验证；
d) 应具备高可信度，针对高风险医疗场景（如危急重症咨询、药物相互作用查询、禁忌症判断），明确标注适用范围与限制条件。

6.2.2 推理结果

医疗语音大模型的推理应用的推理结果能力要求如下：

a) 应符合医疗规范与临床标准，无违背诊疗原则、用药禁忌、医疗伦理的内容，降低医疗风险；
b) 应表述清晰、准确，避免模糊歧义，确保医护人员、患者能清晰理解，同时提供推理依据说明（如引用的指南名称、条款编号），支持结果验证；
c) 应具备医疗场景结构化文本填充能力，包括但不限于症状名称、发病时间、用药剂量、检查项目、患者年龄、既往病史、过敏史、诊疗操作史等，确保核心医疗信息无遗漏。

6.3 语音识别

医疗语音大模型的语音识别应具备高准确率、多语种适配、复杂场景适应与鲁棒性，要求如下：

a) 应具备医疗场景专业语音识别能力，依托医学专业知识库与上下文语义理解技术，有效识别医疗场景下易混淆的同音字、多音字，高准确率识别各类临床、药学、医技等医疗专业术语，同时有效识别医患沟通中出现的网络用语（如新冠相关的“阳了、二阳、阳康、转阴”等），避免识别错误造成医疗文书偏差，保障诊疗信息记录的精准性与规范性；
b) 应具备普通话医疗语音精准转写能力，能将接收到的普通话语音数据转写为与语音内容相符的文字结果，支持医疗术语、诊疗流程表述、症状描述，以及临床常用简单专业英文单词等专业内容的精准转写；
c) 应具备普通话与至少一种方言混合转写能力，且支持普通话与该方言的混合输入转写（如普通话 + 粤语、普通话 + 四川话）；
d) 应具备至少一种常用少数民族语言语音转写能力（如蒙语、藏语等）；
e) 宜具备非中文语种及多语种混合语音识别能力，可将至少一种非中文的常用语种的语音转写为对应的文本（如英语、法语、阿拉伯语等），并支持两种及以上语种的混合语音识别，能在同一句话或对话中无缝处理混合语音（如汉语+英语）；
f) 应具备多口音医疗语音识别能力（如地域口音、非母语口音、特殊人群口音等）；
g) 应具备多发音人重叠语音转写能力，能将包含两个及以上发音人的重叠医疗语音精准区分说话人并转写为对应的文本；
h) 应具备特殊发音人群语音转写能力，可对构音障碍患者、老年发音不清患者、术后发音异常患者等群体的语音完成精准转写。
i) 应具备不同语速的医疗语音识别能力；
j) 应具备低信噪比下的医疗语音识别能力。

6.4 语音合成

医疗语音大模型的语音合成能力要求如下：

a) 应具备将汉语文本（含医疗术语、医嘱、诊疗建议等）转换为相应的语音输出的能力；
b) 应具备将至少一种常用的少数民族语言文本转换成相应的医疗语音输出的能力，（如蒙语、藏语等）；
c) 应具备将汉语文本转换成多种方言的医疗语音输出的能力，（如粤语、闽南语等）；

d) 应具备将至少一种非中文语种的文本转换成相应的医疗语音输出，（如英语、法语、阿拉伯语等）；
e) 应具备合成包含各种指定副语言信息的医疗语音的能力，（如指定风格、语气、性别、情感、笑声、 拟声词等）；
f) 应具备将包含至少两种语言的文本转换为相应语音输出的能力，如汉语和英语混合；
g) 应具备合成具有饱满情感、自然停顿、舒缓韵律等的拟人化特征的医疗语音的能力；
h) 应具备多角色对话的医疗语音合成能力，能为至少两位说话人分配差异化语音特征（如音色、语调、性别）；
i) 应具备生成清晰、流畅、自然的，接近人类发音、符合人类口语习惯的语音的能力，无明显的机械感、杂音、书面化表达等。

6.5 语音交互

医疗语音大模型的语音交互能力要求如下：
a) 多轮连续对话能力：应具备流畅、自然的多轮连续医疗语音对话能力，基于语义理解结果生成符合人类预期、语法正确的医疗语音回复；
b) 全双工交互能力：应支持用户实时打断并调整输出内容，打断后模型需立即停止输出，且不丢失已交互的上下文信息；
c) 语言风格自适应能力：应具备根据对话语境自动调整语言风格的能力，如正式、随意、幽默等；
d) 回复多样化语音回复配置能力：应具备设定多样化的语音回复的能力，如不同年龄、性别、音色、语言风格等；
e) 情感化回复能力：宜具备情感化语音表达能力，可根据用户情绪快速响应，生成带有特定韵律、语调、节奏和措辞的语音回复，以表达相应的情感，如安慰、祝贺等；
f) 交互日志留存能力：应支持交互日志留存，完整记录语音交互的时间戳、说话人信息、交互内容、模型响应结果，如日志留存时间不低于 6 个月，支持日志加密存储和合规导出，满足医疗场景追溯需求。

7 模型训练要求

7.1 数据集构建

医疗语音大模型的数据集构建要求如下：
a) 应具备多样性，包含多语种、口音、年龄、性别、健康状况的说话人，确保数据代表性；
b) 应具备专业性，语音内容符合医疗规范，医疗术语准确，场景对话逻辑合规，无虚假、误导性内容；
c) 应具备数据采集规定，数据采集应需获得用户书面知情同意，明确数据使用范围与期限；
d) 应覆盖不同场景，包含门诊问诊、手术指令、住院查房、远程会诊、基层常见病咨询、健康宣教、慢病管理等语音数据；
e) 应具备基础标注（如音频路径、时长、语种、方言类型、噪声类型、文本转录、说话人属性）、医学标注（医学关键词、语义标签、意图类别、知识关联标签（如关联疾病、药物、诊疗指南）、临床逻辑关系）、基层标注（场景类型、用户属性、方言变体类型、通俗化表达映射等），确保表达准确性与语音采集覆盖的完整性；
f) 应支持多种数据格式，如音频格式应支持 WAV、MP3、FLAC、M4A、OGG、AAC 等主流格式，标注格式：采用 JSON、XML 标准格式，字段定义符合 WS 363-2011 等 规范，包含必要字段（id、wav_path、text、medical_tags、scene_type 等），存储格式应支支持分布式存储（如 HDFS）；
g) 应规范统一数据文件命名（如 “scene-language-accent-id.wav”），目录结构清晰，便于检索与使用；
h) 应支持数据的上传、下载、删除、替换等功能；
i) 应控制数据及质量，精准还原医疗语音信息，医疗术语、症状描述等无偏差；
j) 应进行数据清洗，严格剔除噪声过高、无医疗意义、标注错误及音频失真严重的数据，确保数据集整体质量满足模型训练与测试核心需求；

k) 应具备隐私安全保护，全面覆盖患者身份、病史、诊疗记录等敏感数据，采用匿名化等技术全量脱敏，实现语音特征与个人信息解绑，符合隐私保护法规；
l) 应建立质量评估机制，建立“客观指标+人工审核”双重体系，客观维度涵盖音频质量、标注准确性等；人工抽样核查医疗术语标注、脱敏效果及核心信息完整性。

7.2 部署应用

医疗语音大模型的部署应用要求如下：

a) 多部署模式：应兼容公有云部署与私有化部署两种核心模式，满足不同医疗机构的合规与业务需求；其中私有化部署需支持本地数据中心、专属云等基础设施部署，公有云部署需支持主流云服务商的标准化部署。
b) 设备兼容：应适配医疗场景主流设备，包括但不限于医院诊疗一体机、基层医疗机构专用终端、医生工作站、移动护理终端、可穿戴医疗设备、普通计算机、智能手机（含千元机等低配置机型），支持 Windows、Linux、Android、iOS 等主流操作系统。
c) 系统兼容：应支持与医疗核心业务系统对接，包括但不限于医院信息系统（HIS）、电子病历系统（EMR）、实验室信息系统（LIS）、影像归档和通信系统（PACS），兼容 HL7FHIR、DICOM、CDA 等医疗数据交换标准，实现语音识别结果、语义理解数据与业务系统无缝同步。
d) 接口兼容：应提供标准化 API、SDK 及 WebSocket 接口，兼容 RESTful 等通用接口规范，支持与第三方医疗软件、定制化业务系统快速集成，降低集成开发复杂度。
e) 运行稳定：
   1) 公有云部署，应具备多地域高可用能力，支持弹性扩容，无明显卡顿或中断；
   2) 私有化部署，应具备独享计算资源，避免多租户干扰，连续运行无故障，核心业务场景（如门诊问诊、手术指令识别）无服务中断。
f) 性能稳定：应满足医疗场景端到端推理延时交互需求，支持弱网环境（带宽≤1Mbps）下的稳定运行，基层场景离线部署时核心功能（语音识别、本地语义处理）正常可用，网络恢复后数据自动同步。
g) 容错稳定：应具备异常场景容错能力，支持音频格式错误、设备断连、网络中断等异常的自动检测与恢复，异常恢复后可无缝接续服务，针对医疗设备突发故障，支持数据本地缓存与断点续传，避免核心医疗信息丢失。

8 评估指标要求

8.1 客观指标

8.1.1 精准度

8.1.1.1 准确率

准确率是正确分类的样本数与样本总数之间的比例：

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

式中：

Acc——准确率；

TP——真正例的数量，即模型正确预测为正类的实例数量；

FP——假正例的数量，即模型错误预测为正类的负例数量；

TN——真负例的数量，即模型正确预测为负类的负例数量；

FN—— 假负例的数量，即模型错误预测为负类的正例数量。

8.1.1.2 召回率

召回率是实际为正类的样本中被正确预测为正类的比例，公式：

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

式中：

Recall——召回率；

TP——真正例的数量，即模型正确预测为正类的实例数量；

FN—— 假负例的数量，即模型错误预测为负类的正例数量。

8.1.1.3 精确率

精确率是被预测为正例的样本中，实际为正例的比例，公式：

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

式中：

Precision——精确率；

TP—— 真正例的数量，即模型正确预测为正类的实例数量；

FP—— 假正例的数量，即模型错误预测为正类的负例数量。

8.1.1.4 字错误率

字错误率是识别结果中错误字符（汉字、字母、符号等）数量占参考文本总字符数量的比例，公式：

$$CER = \frac{S+I+D}{N} \quad (4)$$

式中：

CER——字错误率；

S—— 替代错误字数；

I—— 插入错误字数；

D—— 删除错误字数；

N——参考文本的总字数。

8.1.1.5 词错误率

词错误率是识别结果中错误词汇数量占参考文本总词汇数量的比例，公式：

$$WER = \frac{S+I+D}{N} \quad (5)$$

式中：

WER——词错误率；

S——替代错误词数；

I——插入错误词数；

D——删除错误词数；

N——参考文本的总词数。

8.1.1.6 句识别正确率

句识别正确率是被正确识别的句子数量占标注的总句子数量的比例，公式：

$$SCR = \frac{S}{N} \quad (6)$$

式中：

SCR——句识别正确率；

S——正确识别的句子数；

N——标注的总句子数。

8.1.1.7 错误拒绝率

错误拒绝率是错误拒绝的数目占测试集合中应被接受的测试数据的比例，公式：

$$FRR(\theta) = \frac{S(\theta)}{N} \cdots\cdots (7)$$

式中：

FRR(θ)—— 阈值θ下的错误拒绝率；

θ ——阈值；

S(θ) ——得分低于阈值θ的假阴性数，即错误拒绝的正类样本数；

N——正类样本总数。

8.1.1.8 错误接受率

错误接受率是错误接受的数目占测试集合中应被拒绝的测试数据的比例，公式：

$$FAR(\theta) = \frac{T(\theta)}{N} \cdots\cdots (8)$$

式中：

FAR(θ)——阈值θ下的错误接受率；

θ——阈值；

T(θ)——得分高于阈值θ的假阳性数，即错误接受的负类样本数；

N——负类样本总数。

8.1.1.9 双语评估替换

双语评估替换 (BLEU)是一种用于机器翻译任务的评测指标，公式：

$$BLEU = BP \cdot exp(\sum_{n=1}^{N} W_n \, log(p_n) \cdots\cdots (9)$$

式中：

BLEU——n-gram(n 元组）的 BLEU 分数；

BP—— 长度惩罚因子；

N—— 最大 n-gram 的阶数；

W_n——触发的 n-gram 的权重；

P_n——n-gram 的准确率。

8.1.2 效率性能

8.1.2.1 端到端推理延时

端到端推理延时是指从用户发起请求开始，到模型输出完整响应之间的时间间隔，公式：

$$T = t_2 - t_1 \cdots\cdots (10)$$

式中：

T——端到端推理延时；

t_1——用户发起请求时间；

t_2—— 模型输出完整响应时间。

8.1.2.2 吞吐率

吞吐率是指在单位时间内能够处理的语音数据时长，通常以样本数/秒(samples/s)或音频时长/秒 (seconds of audio/s)为单位，公式：

$$TR = \frac{N}{T} \quad (11)$$

式中：

TR——吞吐率；

N—— 语音数据时长；

T——处理语音数据所需的时间。

8.1.3 交互能力

8.1.3.1 打断成功率

打断成功率是指某段时间内，语音打断操作被正确响应的次数占总次数的比例，公式：

$$P_i = \frac{N_i}{N} \quad (12)$$

式中：

P_i—— 打断成功率；

N_i——语音打断正确响应的次数；

N——语音对话中需要执行打断操作的总次数。

8.1.3.2 打断延时

语音对话的打断延时是指从用户尝试中断当前会话的时间点，到模型停止当前响应的时间点之间 的时间间隔，公式：

$$T = t_2 - t_1 \quad (13)$$

式中：

T——打断延时；

t_1——用户中断当前会话时间；

t_2—— 模型停止当前响应时间。

8.1.3.3 对话成功率

对话成功率是指一定的时间内，成功的语音对话总数占有效的语音对话总数的比例。成功的语音对 话指获取到完整的语音服务结果，期间未产生差错的语音对话；有效的语音对话指全部的语音对话去除 由于用户终端故障或用户行为、参数错误导致的失败对话。对话成功率，公式：

$$P = \frac{S}{S+F} \quad (14)$$

式中：

P—— 对话成功率；

S——对话成功的次数；

F—— 对话失败的次数。

8.1.3.4 响应延时

语音对话的响应延时是指从用户会话的结束时间，到模型响应会话的开始时间之间的时间间隔，公式：

$$T = t_2 - t_1 \quad (15)$$

式中：

T——响应延时；

t_1——用户会话结束时间；

t_2——模型响应开始时间。

8.1.3.5 误触发率

误触发率是指系统在聆听状态下，无正常交互输入，且周围存在环境噪声或非交互人声时，以及正 常交互输入后，系统开始回复播报过程中，周围存在环境噪声或非交互人声时，系统被触发响应的概率，公式：

$$P = \frac{\mathrm{N}}{\mathrm{T}} \quad (16)$$

式中：

P—— 误触发率；

N——测试用例总时长内的误触发总次数；

T——测试用例总时长。

8.1.4 抗干扰与语音分离

8.1.4.1 信噪比改善

信噪比改善为输出语音信噪比与输入语音信噪比的比值。

8.1.4.2 噪声抑制量

噪声抑制量，公式：

$$D_{NR} = \frac{\Sigma_{\{n=0\}}^{N-1} \left| v_{in(n)} \right|^2}{\Sigma_{n=0}^{N-1} \left| v_{out(n)} \right|^2} \quad (17)$$

式中：

D_{NR}—— 噪声抑制量，单位为分贝 (dB)；

$v_{in(n)}$——输入信号中第 n 个噪声信号的振幅；

$v_{out(n)}$——输出信号中第 n 个噪声信号的振幅；

N——输入信号频谱频率分量的总数量。

8.1.4.3 分离错误率

分离错误率是分离错误的语音片段时长占整个有效语音片段时长的比例，公式：

$$DER = \frac{\Sigma_{S=1}^{S} dur(s) \times \{\max[N_{ref(s)}, N_{hyp(s)}] - N_{correct(s)}\}}{\Sigma_{S=1}^{S} dur(s) \times N_{ref}} \quad (18)$$

式中：

式中：

DER——分离错误率；

S——实际结果和输出结果都包含同一个说话人(对)的说话人片段数

量；

$dur(s)$——片段s的时长；

$N_{ref(s)}$——片段 s 中实际结果的数量；

$N_{hyp(s)}$——片段 s 中输出结果的数量；

$N_{correct(s)}$——片段 s 中输出结果与实际结果正确对应的数量。

8.2 主观指标

8.2.1 平均意见得分

主观评测中，通常采用人工评测指标 MOS 来评测语音大模型的效果。依据针对不同任务设定的人工评测的指标维度，由选定的参与者以分数的形式来进行评分。参与者可结合给定的指标维度给出总体的评测得分，也可分别对每个指标维度进行评测。通常，将参与者的评测结果分为 5 个等级（5 分优秀、4 分良好、3 分一般、2 分较差、1 分很差，分值越高表示语音质量与听觉体验越好。），最终以总体和每个维度的平均得分来评测大模型对应的能力 。

a） 总体平均得分：对每条数据的总体评分取平均值，得出一个总体评测得分，来评测该内容的优秀程度；

b） 每个维度的平均得分：按指标维度对每条数据评分取平均值，分别得出各个维度的平均得分，公式：

$$MOS = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}S_{ij} \quad (19)$$

式中：

MOS—— 平均意见得分；

N——评测样本数；

M—— 参与者人数；

S_{ij}——第 j 个参与者对第 i 个样本的评分。

9 易用性要求

9.1 可理解性

9.1.1 语言表达清晰程度

医疗语音大模型界面文字、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。

9.1.2 辅助理解手段

医疗语音大模型涉及复杂医学知识、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。

9.2 易学性

9.2.1 帮助文档完整性

医疗语音大模型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

医疗语音大模型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。

9.3 易操作性

9.3.1 操作一致性

医疗语音大模型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

a) 医疗语音大模型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) 医疗语音大模型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) 医疗语音大模型消息推送应具备合理的频率和时机，不应过度打扰用户。

10 安全性要求

10.1 基础设施安全

10.1.1 硬件设备安全性

医疗语音大模型涉及的硬件设备（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

医疗语音大模型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

医疗语音大模型应支持数据采集、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。

10.3 应用安全

10.3.1 内容安全

医疗语音大模型输出内容（含生成内容、决策内容）应符合应保障人权平等、杜绝歧视，并符合科技伦理、普适道德伦理、医学伦理相关要求。

a) 应支持尊重人权，包括医疗语音大模型输出内容（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，包括医疗语音大模型输出内容（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) 应遵循科技伦理管控要求，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；

e) 普适道德伦理原则：应遵循普适道德伦理原则，包括公序良俗、人格尊严与人权平等、诚信向善、生命健康优先等；

f) 普适道德伦理管控要求：应满足普适道德伦理管控要求，包括平等无歧视、尊重人格尊严、内容正向向善、不危害生命身心健康、不违背公共秩序、不传播违法不良信息、表述真实诚信等；

g) 医学伦理原则：应符合医学伦理核心原则，包括尊重自主、不伤害、行善有利、公平公正等；

h) 医学伦理管控要求：应满足医学伦理管控要求，包括知情告知清晰、风险最小化、患者健康获益优先、人群公平无医疗偏见、患者隐私保护、弱势群体特殊保护、诊疗建议合规可追溯、明确模型仅作辅助参考等。

10.3.2 服务安全

医疗语音大模型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：医疗语音大模型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。