|  |  |
| --- | --- |
| ICS |  |
| CCS | |  | | --- | | D:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T.pngD:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T后面的反斜杠.png |   点击此处添加CCS号 |

     团体标准

T/XXX XXXX—XXXX

医疗大模型语料一体机应用指南

Requirements Guide of Medical Large Model Integrated Machine

XXXX - XX - XX发布

XXXX - XX - XX实施

       发布

**版权声明**

目录

[前言 III](#_Toc13972)

[1 范围 1](#_Toc9022)

[2 规范性引用文件 1](#_Toc31822)

[3 术语和定义 1](#_Toc20074)

[4 缩略语 1](#_Toc20705)

[5 系统框架 2](#_Toc27915)

[6 应用分类和能力要求 4](#_Toc17421)

[6.1 一类应用 4](#_Toc14838)

[6.2 二类应用 4](#_Toc12064)

[6.3 三类应用 5](#_Toc13281)

[7 模型要求 5](#_Toc32271)

[7.1 基础模型选型要求 5](#_Toc18083)

[7.2 垂类模型选型要求 6](#_Toc9795)

[8 智能体要求 6](#_Toc3900)

[9 语料要求 6](#_Toc24815)

[9.1 语料对象与范畴 7](#_Toc6508)

[9.2 医疗通识语料要求 7](#_Toc11494)

[9.3 医疗行业知识语料要求 7](#_Toc10905)

[9.4 医疗临床数据要求 7](#_Toc31815)

[9.5 语料管理要求 8](#_Toc3416)

[9.6 语料工具链选型要求 8](#_Toc21635)

[9.7 向量数据库选型要求 9](#_Toc10715)

[10 算力要求 10](#_Toc7039)

[10.1 智能算力配置要求 10](#_Toc23296)

[10.2 普通算力配置要求 10](#_Toc25317)

[10.3 一体机操作系统要求 10](#_Toc11500)

[11 部署环境要求 11](#_Toc19170)

[11.1 机房及类机房环境部署要求 11](#_Toc22859)

[11.2 非机房及非类机房环境部署要求 11](#_Toc8159)

[参考文献 12](#_Toc13991)

[附 录 A 13](#_Toc13047)

[附 录 B 14](#_Toc15229)

[附 录 C 15](#_Toc11069)

1. 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这些专利的责任。

本文件由上海人工智能协会提出并归口。

本文件起草单位:

本文件起草人：

本标准首次制定。

首期执行单位:

本文件版权归所有。未经许可，不得擅自复制、转载、抄袭、改编、汇编、翻译或将本标准用于其他任何商业目的。

医疗大模型语料一体机应用指南

* 1. 范围

 本文件规定了医疗大模型语料一体机的能力要求，包括系统框架、应用场景能力、软硬件配置、语料配置等内容。

本文件适用于医疗大模型语料一体机的研发、配置、维护、采购等工作。

* 1. 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 41867-2022 信息技术 人工智能 术语

* 1. 术语和定义

GB/T 41867-2022界定的术语和定义适用于本文件。

医疗大模型语料一体机medical large model integrated machine

专门为医疗领域设计的高性能计算设备，以一体机服务器形式存在。

医疗大模型 large medical model

基于深度学习等AI技术，针对医疗领域特定问题（如疾病诊断、药物研发等）训练而成的大型神经网络模型，具备处理和分析复杂医疗数据，提供精准医疗决策支持等能力。

医疗语料库 medical Corpus

医疗领域相关文本、图片、音频、视频等数据的集合，可用于训练和优化医疗大模型，以提升其在医疗领域的应用效果。

临床决策支持clinical decision support

基于大量患者数据，使用医疗大模型，为医生提供诊断和治疗建议的智能代理，旨在辅助医生作出更加精准和科学的医疗决策。

算力动态分配 Dynamic Computing Power Allocation

根据医疗大模型的实时需求和系统负载情况，动态调整算力资源分配的智能化策略，以确保系统在高负载情况下仍能稳定运行，并可优先处理关键任务。

* 1. 缩略语

AEs：医疗不良事件(Adverse Events)

AI：人工智能 （Artificial Intelligence）

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

ASR ：自动语音识别(Automatic Speech Recognition)

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

FPS：每秒帧数（Frames Per Second）

GPU：图形处理单元（Graphics Processing Unit）

IDE：集成开发环境（Integrated Development Environment）

NIfTI：神经影像信息学技术倡议（Neuroimaging Informatics Technology Initiative）

NPU：神经网络处理单元（Neural Processing Unit）

SAEs:严重不良事件（Severity Adverse Events）

TOPS：每秒万亿次操作（Tera Operations Per Second）

TPU：张量处理单元（Tensor Processing Unit）

* 1. 系统框架

医疗大模型语料一体机通过集成医疗应用区、医疗模型区、算力区、医疗语料区、智能体引擎区以及医疗一体机操作系统等关键功能模块，实现医疗资源的优化配置和医疗服务的高效提供。系统遵循实用性、可靠性、安全性、可扩展性原则，确保设备安全、网络安全及数据安全，并通过定时数据备份保障系统稳定运行。

一体机总体功能宜按照“1+5”架构部署，包括1个一体机操作系统，医疗应用区、模型区、智能体区、医疗语料区、算力区等5个功能区，具体内容如下图所示：



图1 医疗大模型语料一体机架构图

一体机功能概述具体如下：

1. 医疗应用区：提供一系列与医疗相关的功能模块，包括但不限于常见疾病/慢性疾病诊断、医学知识科普、康复养老辅助、复杂疾病辅助诊断、诊疗流程优化、药物作用分析、疾病监控与分析、医学知识图谱、文献综述生成等。
2. 模型区：包含基础模型、垂类模型。
3. 算力区：配置多种AI加速卡及其适配引擎。
4. 智能体区：集成规划引擎、工具引擎、执行引擎、记忆引擎等关键组件。
5. 医疗语料区：包括多个医疗语料库、向量数据库以及语料清洗、标注工具。
6. 医疗一体机操作系统：包括医疗模型管理、引擎调度、智能体管理、应用管理、语料管理、算力调度管理及集群管理等模块。
   1. 应用分类和能力要求

参考国家卫健委《卫生健康行业人工智能应用场景参考指引》的内容，以及医疗应用本身复杂程度，将医疗大模型语料一体机的应用分为一类应用、二类应用以及三类应用，本章对应用的能力要求进行解释。

* + 1. 一类应用

6.1.1诊疗辅助

应具备根据患者症状、体检结果和实验室数据提供初步诊断的能力；

应具备辅助医生制定个性化治疗方案的能力；

应具备实时更新医学知识，以提供最新的诊疗建议。

6.1.2 医院管理

应具备辅助优化医院资源分配，提高医疗服务效率的能力；

应具备根据医院运营数据辅助决策制定的能力；

应具备提供改善患者就医流程建议的能力。

6.1.3 健康管理

应具备跟踪个人健康数据，提供健康风险评估的能力；

应具备根据个人健康状况提供定制化的健康建议和干预措施。

6.1.4 养老托育服务

应具备为老年人和儿童提供个性化护理计划的能力；

应具备根据老年人和儿童的健康状况数据，及时预警健康风险；

应具备提供心理支持和社交互动，增强老年人和儿童的生活质量。

* + 1. 二类应用

二类应用包括临床专病辅助决策、基层全科医生辅助决策、公共卫生服务、临床用药辅助、药物研发、中医中药辅助等应用。

6.2.1 临床专病辅助决策

应具备针对特定疾病提供深入的临床知识支持的能力；

应具备辅助医生进行复杂病例分析和治疗方案选择的能力。

6.2.2 基层全科医生辅助决策

应具备提供全面的基层医疗服务知识的能力；

应具备辅助全科医生进行常见病和多发病的诊断和治疗建议。

6.2.3 公共卫生服务

应具备根据公共卫生数据预测疾病流行趋势的能力；

应具备制定和实施公共卫生干预措施的能力；

应具备提供公共卫生教育和健康促进知识的能力。

6.2.4 临床用药辅助

应具备药物信息查询和药物相互作用分析的能力；

应具备辅助医生进行药物剂量调整和用药指导的能力；

应具备根据患者用药反应及时调整治疗方案的能力。

6.2.5 药物研发

应具备药物设计和筛选的辅助能力；

应具备药物效果预测和副作用评估的能力；

应具备临床试验设计和数据分析的能力。

6.2.6 中医中药辅助

应具备中医理论学习和应用的能力；

应具备中药方剂配伍和调整建议的能力；

应具备中医诊断和治疗建议的能力。

* + 1. 三类应用

三类应用包括医学影像辅助诊断、手术辅助规划、医学教学、医学科研等应用。

6.3.1 医学影像辅助诊断

应具备医学影像的辅助识别和分析能力；

应具备辅助医生进行疾病诊断和病情评估的能力；

6.3.2 手术辅助规划

应具备辅助手术方案设计和风险评估的能力；

应具备手术过程中的实时辅助和建议能力；

应具备制定术后恢复和并发症预防规划的能力。

6.3.3 医学教学

应具备提供医学教育资源和模拟训练的能力；

应具备辅助医学生和医生进行专业知识学习的能力；

应具备评估学习效果和提供个性化教学建议的能力。

6.3.4 医学科研

应具备医学文献检索和分析的能力。

应具备辅助科研设计和实验方法建议的能力。

* 1. 模型要求
     1. 基础模型选型要求

本节对医疗大模型语料一体机大模型选型进行要求，包括性能与指标、自然语言处理、知识储备、价值对齐、安全与可靠性等方面。

7.1.1模型基础能力要求

模型应具备自然语言理解、知识推理、文本生成、跨语言处理、跨模态理解、跨模态生成等方面能力，其能力宜通过行业公认的测评集测试；

7.1.2长文本处理能力要求

模型应能处理8k以上的文本输入，宜能够处理32k的连续文本输入。

7.1.3多模态数据处理能力要求

模型应能同时处理图像数据（X射线图像、CT图像、MRI图像、超声图像、病理切片图像等）、文本数据（电子病历、医生诊断报告、医学文献等）、实验室检查数据、基因数据、音频数据（医患沟通音频、心音、肺音等）、视频数据等多模态数据。

7.1.4参数规模要求

一类应用宜使用不大于13B参数的稠密模型或使用混合专家模型激活总参数在13B及以下；

二类应用宜使用13B-30B参数的稠密模型或使用混合专家模型激活总参数在13B-30B；

三类应用宜使用30B-72B参数的稠密模型或使用混合专家模型激活总参数在30B-72B；

7.1.5性能与指标要求

在标准硬件配置下平均响应时间应小于500ms；

在标准硬件配置下文本生成速度不小于10tokens/s；

* + 1. 垂类模型选型要求

7.2.1专业性要求

行业大模型应根据医疗领域设计，具备医疗知识问答、医学文本分类、医学信息抽取、医学自然语言推理、医学内容生成、医学多模态数据处理等能力，其能力宜通过行业公认的测评集测试；

7.2.2精准性要求

针对于医疗场景下的问题，行业大模型应提供更精准的问题解决方案，准确率不低于95%；

在医疗任务上的处理效率应高于基础大模型；

应能理解医疗行业特有术语和上下文；

7.2.3知识库更新要求

应支持根据具体业务需求进行调整和优化；

应该支持持续学习，随行业发展不断更新知识库；

7.2.4价值对齐要求

生成的内容应符合公认的医疗道德伦理规范和法律规定；

生成的内容应符合社会主流价值观，不得对某些社会群体产生不利影响；

* 1. 智能体要求

智能体区应具备智能体引擎和智能体开发功能，要求如下：

1. 智能体引擎：应具备规划引擎、工具引擎、执行引擎、记忆引擎功能。记忆引擎应支持向量数据库及外部数据访问；
2. 智能体开发：应支持低代码或无代码开发智能体功能。
   1. 语料要求
      1. 语料对象与范畴

医疗大模型语料一体机语料应包含医疗通识语料、医疗行业知识语料、医疗临床数据三部分，补充内容可参见附录。

* + 1. 医疗通识语料要求

医疗通识语料应包括医疗领域的相关知识，包括基础医学、临床医学、保健医学、康复医学、理疗学、麻醉学、内科学等，具体医疗领域详见附录A。医疗通识语料应包含但不限于如下部分内容，部分数据源详见附录B：

1. 书籍：包括国内外权威出版社获取的医疗健康领域教科书，及其他高质量生命健康相关书籍。
2. 期刊论文：包括国内外知名期刊或者从权威数字出版平台获取的医疗健康领域的学术论文。
3. 数据库：包括从医学数据库获取的临床试验、人类疾病、基因序列、蛋白质序列和功能信息、药物药理药效等数据。
4. 知识库：包括从医学知识库中获取的生物、疾病和基因等多维度之间的关联知识数据。
5. 专利：包括从专利搜索引擎检索得到的医疗健康领域的专利正文。
6. 循证医学知识库：包括用于支持临床决策、疾病预防、治疗方案选择和健康管理等信息，通过整合和分析大量的医学研究数据和临床实践经验构建的循证医学结构化知识库。
7. 临床指南：包括官方机构发布的关于某个疾病和专科领域内容有关诊断、管理和治疗的决策标准的文档。
8. 药品数据库：包括药品的名称、成分、适应症、用法用量、不良反应、药物间相互作用、注意事项等信息的数据库。
9. 医疗器械资料：如医疗设备的使用说明书、操作手册等，包括常见的诊断设备（如心电图机、血糖仪等）和治疗设备（如呼吸机、除颤仪等）的相关资料。
10. 专家共识：包括由特定领域具有专业知识和经验的专家共同讨论、协商达成的统一意见或建议，一般发表于高水平的期刊论文中。
11. 中医学典籍：包括中医领域的经典著作。
12. 药材药方：包括经典中药书籍中记录的历代中医方剂，包括方剂的组成、用法、主治病症等。
13. 相关法律法规：收集国家和地方关于社区医疗的政策、指南和标准，尤其是一些涉及医疗报销方面的信息。
14. 健康宣教材料：涵盖各类健康教育和预防知识，如疾病预防、健康生活方式等。
15. 其他结构化数据：包括疾病对应的症状列表、疾病对应的药品列表、药品通用名说明书、症状的通俗叫法，同义词等。
    * 1. 医疗行业知识语料要求

医疗行业知识语料应包含但不限于如下部分内容：

1. 专科专病数据：结合医疗管理部门重点监测的单病种（单病种清单详见附录C）形成的人工智能基地的行业专科专病语料集，赋能后续的临床研究、临床决策支持、医院运营管理和患者管理等方面。
2. 医学数据库：包含了海量的医学文献资源的权威医学数据库。
3. 医学知识图谱：医学知识图谱以图结构的形式表示医学领域的知识，包括疾病、症状、药物、治疗等实体以及它们之间的关系。
   * 1. 医疗临床数据要求

医疗临床数据应包含但不限于如下部分内容：

1. 电子病历：包括针对个体的医院门诊、住院形成的电子病历，含诊断、实验室检验数据、影像检查报告、治疗方案等内容。
2. 医疗住院数据：包含入出院业务、转科记录业务、药品医嘱业务、住院非药品医嘱业务、住院费用业务、住院费用分类业务、住院发票业务、住院退药记录等内容。
3. 研发阶段临床试验数据：记录受试者在接受药物治疗期间发生的任何不良事件（AEs），以及严重不良事件（SAEs）。评估药物达到预定疗效指标的情况，如缓解率、生存期延长等。药物在体内吸收、分布、代谢和排泄的过程，以及药物对机体的作用强度随时间变化的关系。
   * 1. 语料管理要求

医疗通识语料应至少每年更新一次；

医疗行业知识语料至少半年更新一次；

医疗临床语料至少每月更新一次；

语料库应存在定期审核和更新机制，可移除已被证伪的信息，保证语料库的准确性和完整性；

语料数据应以知识块、知识对、三元组、知识图谱等的方式存在；

语料数据应具有规范性、一致性、完整性、及时性、准确性、有效性和及时性；

语料数据应遵从数据安全管理的相关法律法规、标准以及合同约定等，并不得损害国家利益、社会公共利益和他人合法权益；

语料数据应具有正确的价值导向，符合科技伦理和社会道德。

* + 1. 语料工具链选型要求

语料工具链选型要根据工具功能的不同属性，分为清洗工具选型规范与标注工具选型规范两部分。

9.6.1清洗工具选型要求

9.6.1.1文本数据清洗工具

文本数据清洗工具功能要求如下：

1. 应具备对医疗文档进行格式标准化处理的能力；
2. 应具备获取医疗文档基本信息，并进行标签记录的能力；
3. 应具备识别医疗文档语言类型并进行标记的功能；
4. 应具备清除转换后医疗文本中无关信息的能力；
5. 应具备检测医疗文档敏感数据，并确保其符合隐私保护标准的能力。

9.6.1.2视频数据清洗工具

视频数据清洗工具功能要求如下：

1. 应具备自动获取医疗视频文件的帧率（fps）、时长（duration）等信息的功能；
2. 应具备使用算法模型自动根据参数切分医疗视频镜头的功能；
3. 应具备使用算法与模型抽取医疗视频文字信息，并提取音频内容转换为文字信息的功能；
4. 应支持自动抽取医疗视频中的关键帧，或根据逻辑判定抽取关键帧的功能；
5. 应支持自动识别医疗视频中的单帧水印，并在相应的帧数上进行标记记录的功能。

9.6.1.3图片数据清洗工具

图片数据清洗工具功能要求如下：

1. 应具备使用算法与模型抽取医疗图片文字信息，或对图片进行医疗领域描述；
2. 应具备使用深度学习模型对医疗图像进行分类，以区分不同的疾病类型；
3. 应具备使用算法与模型，在医学图像中识别并分割出所需区域的功能。

9.6.1.4音频数据清洗工具

音频数据清洗工具功能要求如下：

1. 应具备对医疗音频进行ASR处理，并生成文本数据的能力；
2. 应具备获取医疗音频基本信息，并进行标签记录的能力；
3. 应具备识别医疗音频语言类型并进行标记的功能；
4. 应具备清除转换后医疗音频中无关信息的能力；
5. 应具备检测医疗文档敏感数据，并确保其符合隐私保护标准的能力。

9.6.2标注工具能力要求

标注工具应当具备以下人工标注能力：

1. 应提供针对各个数据模态的数据标注功能，包括文本、图片、音频、视频等；
2. 宜提供标注工具支持医疗领域的特殊数据要求，包括DICOM、NIfTI等格式。

9.6.2.1文本数据标注工具

文本数据标注工具应具备文选区、工具区、标签区。

标注工具功能要求如下：：

1. 支持针对医疗文本的主题分类预标注、情感倾向预标注、关键词提取预标注、实体链接预标注、语义角色预标注等功能；
2. 支持多语种；
3. 数据导出格式多样；
4. 支持基于插件的扩展性（可选项）。

9.6.2.2图像数据标注工具

图像数据标注工具应具备画选区、工具区、标签区；

标注工具应支持针对医疗图片的目标检测预标注、场景识别预标注、图像质量评估预标注等功能。

9.6.2.3视频数据标注工具

视频数据标注工具应具备视频选区、工具区、标签区；

标注工具应支持针对医疗视频的镜头分割预标注、关键帧提取预标注、行为识别预标注等功能。

9.6.2.4音频数据标注工具

音频数据标注工具应具备音频选区、工具区、标签区

标注工具应具备支持针对医疗音频的声音事件检测预标注、语音识别预标注（方言）、语音识别预标注（外语）、音频信号源分离预标注、语调模式识别预标注等功能。

* + 1. 向量数据库选型要求

向量数据库提供语料向量形式存储，要求如下：

1. 应支持多种向量类型，能够存储和处理多种类型的向量数据；
2. 应具备向量索引机制，应支持快速的相似性检索和范围查询；
3. 应提供查询功能；
4. 应提供API接口，支持多种编程语言和框架；
5. 应提供数据加密功能与访问控制机制；
6. 数据库应提供数据备份和恢复机制。
7. 应支持不少于100W向量规模；
8. 应提供不多于300ms返回速度。
   1. 算力要求

一体机硬件配置要求可分为智能算力配置要求、普通算力配置要求以及算力分配和调度要求。具体分类如下图所示：

* + 1. 智能算力配置要求

智能算力提供AI加速能力，包括GPU、NPU、TPU等架构，其配置要求如下：

（a）使用稠密模型参数在13B及以下或使用混合专家模型激活总参数在13B及以下，并发请求数不大于5，FP16浮点总算力宜不低于280（TFlops/s），AI加速卡总内存宜不低于64GB，配置4张AI加速卡情况下，单卡FP16浮点总算力配置宜不低于 70（TFlops/s），单卡内存宜不低于16GB。

（b）使用稠密模型参数在13B以上30B及以下或使用混合专家模型激活总参数在13B以上30B及以下，并发请求数不大于10，FP16浮点总算力宜不低于640（TFlops/s），AI加速卡总内存宜不低于96GB，配置4张AI加速卡情况下，单卡FP16浮点总算力配置宜不低于 160（TFlops/s），单卡内存宜不低于24GB。

（c）使用稠密模型参数在30B以上72B及以下或使用混合专家模型激活总参数在30B以上72B及以下，并发请求数不大于15，FP16浮点总算力宜不低于1600（TFlops/s），AI加速卡总内存宜不低于192GB，配置8张AI加速卡情况下，单卡FP16浮点总算力配置宜不低于 200（TFlops/s），单卡内存宜不低于24GB。

* + 1. 普通算力配置要求

普通算力提供通用的CPU算力，其配置要求如下：

1. 当AI加速卡在4卡及以下时，CPU核心数不宜少于24，内存宜按照CPU核心数进行1:4适配。
2. 当AI加速卡在8卡及以上时，CPU核心数宜不低于64，内存宜按照CPU核心数进行1:4适配。

表1 普通算力配置要求表

|  |  |  |
| --- | --- | --- |
| 一体机卡数  配置要求 | 4卡及以下 | 8卡及以上 |
| CPU核心数 | 宜不少于24 | 宜不少于64 |
| 内存 | 按照核心数进行1:4适配 | |

* + 1. 一体机操作系统要求

一体机操作系统应配置提供对整体资源管理和调度。具体参考要求如下：

1. 应用管理：应支持应用的安装部署、启用、停止、升级、卸载。
2. 智能体管理：应支持智能体及引擎的安装、升级、启用、停用、卸载。
3. 模型管理：应支持模型、推理引擎和微调引擎的装载、加载、卸载。
4. 语料管理：应支持提示词、微调数据集、知识库数据集、测评数据集的存储、启用、更新、备份、删除功能，应支持向量数据库的安装部署、启用、停止、升级、卸载、备份。
5. 算力管理：应支持对AI加速卡的资源分配、回收、抢占等调度功能，可依据医疗大模型的实时需求进行动态分配，以适应工作负载的变化和不同的服务优先级。
6. 高可用性能力：医疗大模型的算力分配和调度系统应采用高可用性架构。
7. 容错能力：系统应集成容错机制。
8. 故障处理能力：系统应在AI加速卡单点故障发生时，对应资源隔离。
9. 监控能力：系统应配备监控系统，可实时跟踪算力资源的使用情况和系统性能指标。
10. 运维能力：应包含定期维护和更新机制。
    1. 部署环境要求
       1. 机房及类机房环境部署要求

工作温度应保持在5℃至35℃之间。

应控制环境温度变化率不超过20℃/小时。

工作湿度应维持在8%至90%的相对湿度范围内。

宜控制环境湿度变化率不超过20%/小时。

应保持机房内空气清洁，无腐蚀性气体和过多尘埃。

宜定期清洁机房，使用空气净化设备。

应提供稳定、符合设备规格的电源供应，并确保良好接地。

宜配备UPS和/或发电机作为备用电源。

* + 1. 非机房及非类机房环境部署要求

工作温度应保持在5℃至35℃之间。

应控制环境温度变化率不超过20℃/小时。

工作湿度应维持在8%至90%的相对湿度范围内。

宜控制环境湿度变化率不超过20%/小时。

应保持环境通风良好，减少有害物质积聚。

应定期清洁设备周围区域，减少尘埃和污垢。

应提供稳定、可靠的电源供应，并确保良好接地。

参考文献

1. T/SAIAS 015—2024 语料库建设导则
2. 《卫生健康行业人工智能应用场景参考指引》



附 录 A

（资料性附录）

医疗通识语料医疗领域

医疗通识语料应包括医疗领域的相关知识，包括但不限于如下领域知识：

基础医学、临床医学、保健医学、康复医学、理疗学、麻醉学、内科学、外科学、妇产科学、儿科学、眼科学、耳鼻咽喉科学、口腔医学、皮肤病学、性医学、神经病学、精神病学、急诊医学、核医学、肿瘤学、护理学、预防医学与卫生学、药学、中医学、民族医学、中西医结合医学、生物学、生理学、遗传学、放射生物学、分子生物学、生态学、神经生物学、植物学、昆虫学、动物学、微生物学、病毒学、人类学、生物工程、心理学；

附 录 B

（资料性附录）

语料数据源范围

医疗通识语料部分数据源选取参考如下：

1. 书籍包括但不限于从人民卫生出版社、高等教育出版社、科学出版社、Elsevier、Springer等国内外出版社获取的医疗健康领域教科书，及其他高质量生命健康相关书籍；
2. 期刊论文包括但不限于《中华医学杂志》《中国药学杂志》《中国公共卫生》《中国中医药杂志》《Nature》《Cell》《The Lancet》《JAMA》等国内外期刊或者知网、Elsevier、Springer等数字出版平台；
3. 数据库包括但不限于中国临床试验数据库、人类疾病数据库、GenBank、UniProt、DrugBank等；
4. 知识库：包括包括但不限于KEGG（生物系统知识库）、OMIM（人类遗传疾病知识库）等知识库；
5. 专利搜索引擎包括但不限于从PatentScope、Google Patents等。
6. 临床指南发布机构包括WHO Guidelines、National Guideline Clearinghouse (NGC)、国家卫生健康委员会、中华医学会等；
7. 中医学典籍包括但不限于《黄帝内经》、《伤寒杂病论》、《本草纲目》、《中药大辞典》等中医经典；
8. 药材药方经典著作包括《方剂学》、《普济方》等书籍；

附 录 C

（资料性附录）

国家卫健委重点监控的单病种目录

国家卫健委重点监控的单病种包括（51种）：急性心肌梗死（ST段抬高型，首次住院）（STEMI）、 心力衰竭（HF）、冠状动脉旁路移植术（CABG）、脑梗死（首次住院）（STK）、短暂性脑缺血发作（TIA）、社区获得性肺炎（成人，首次住院）（CAP）、社区获得性肺炎（儿童，首次住院）（CAP2）、慢性阻塞性肺疾病急性发作（住院）（AECOPD）、髋关节置换术（THR）、膝关节置换术（TKR）、剖宫产（CS）、肺癌（手术治疗）（LC）、甲状腺癌（手术治疗）（TC）、乳腺癌（手术治疗）（BC）、围手术期预防感染（PIP）、围手术期预防深静脉血栓栓塞（DVT）、 住院精神疾病（HBIPS）、房颤（AF）、主动脉瓣置换术（AVR）、二尖瓣置换术（MVR）、房间隔缺损手术（ASD）、室间隔缺损手 术（VSD）、出血性卒中（ICH）、脑膜瘤（初发手术治疗）（MEN）、胶质瘤（初发，手术治疗）（GLI）、垂体腺瘤（初发，手术治疗）（PA）、急性动脉瘤性蛛网膜下腔出血（aSAH）、惊厥性癫痫持续状态（CSE）、帕金森病（PD）、哮喘（成人，急性发作，住院）（CAC）、异位妊娠（手术治疗）（EP）、子宫肌瘤（手术治疗）（UM）、胃癌（手术治疗）（GC）、甲状腺结节（手术治疗）（TN）、中高危风险患者预防静脉血栓栓塞症(VTE)、严重脓毒症和脓毒症休克（SEP）早期治疗。