T/JNBDA

济 南 市 大 数 据 协 会 团 体 标 准

T/JNBDA 0008-2025

医药标准化人工智能体构建指南

Guide for Building a Medical Standardization Artificial Intelligence System

2025 - 11- 13 发布　　2025 - 11 - 13 实施

济南市大数据协会发布

前　言

随着医疗数据的快速增长和人工智能技术的成熟，药事服务工作对智能化、精准化的需求日益迫切。本标准填补了当前处方审核和患者用药指导智能化的空白，具有显著的社会效益和经济效益。

本标准旨在指导RAG（检索增强生成）技术的药物相互作用智能分析系统，为医院药学提供智能化处方审核服务和个体化的患者用药指导服务，确保患者用药的安全性和合理性。

本文件由济南市大数据协会提出并归口，按照GB/T 1.1-2020 《标准化工作导则》要求编写。

本规范由山东未来互联科技有限公司、山东大学齐鲁第二医院主编。

本规范主要参与单位：山大地纬软件股份有限公司、国控大健康科技（山东）有限公司、山东省数字化应用科学研究院有限公司、泰山财产保险股份有限公司。

本规范主要起草人孟洋、孙超、陈晨、张章、王金喜、杨培。

目录

医药标准化人工智能体构建指南

1 总则

本规范适用于RAG（检索增强生成）技术和确定性可信数据网络空间的药物相互作用智能分析系统。

系统设计应统筹规划、联合建设、资源共享，合理利用已有网络设施和服务器装备，满足建设资源节约型、环境友好型社会的要求。

系统设计应以保证分析质量为基础，进行多方案比较，提高经济效益，降低系统造价。

当本规范与国家有关标准规范相矛盾时，应按国家标准规范的相关规定执行。在系统设计中，采用本规范，但在特殊情况下执行本规范个别条款确有困难时，应充分阐述理由，提出解决方案，并报请主管部门审批。

2 规范性引用文件

下列文件对于本规范的应用是必不可少的。凡是注日期的引用文件,仅所注日期的版本适用于本规范。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本规范。

GB 50174《数据中心设计规范》

GB 50009《建筑结构荷载规范》

GB 50611《电子工程防静电设计规范》

GB 50370《气体灭火系统设计规范》

GB 50011 《建筑抗震设计规范》

GB 50343《建筑物电子信息系统防雷技术规范》

GB 50052《供配电系统设计规范》

YD/T 4418-2023《电信网络的确定性IP网络的总体架构和技术要求》

3 术语与定义

下面术语和定义适用于本规范。

3.1 RAG（检索增强生成）

RAG是一种将信息检索（IR）系统与大语言模型（LLM）相结合的技术架构。它的核心思想是：在让大模型生成答案之前，先从外部知识库中检索相关信息，然后将这些信息作为上下文提供给模型，从而让模型生成更准确、更可信、且更少幻觉的答案。

3.2 AI医疗智能体（AI Agent）

AI医疗智能体是具备感知环境信息、自主理解任务目标、规划决策执行路径，并能通过学习持续优化行为以达成特定目标的人工智能系统。

3.3 向量数据库

向量数据库是一种专门用于存储和高效查询高维向量数据，并支持快速相似性搜索的数据库系统。

3.4 LLM大语言模型(Large Language Model）

大语言模型是基于深度学习技术，通过海量文本数据训练，能够理解和生成自然语言的人工智能模型。

3.5 重排序（Rerank）

重排序（Rerank）是对初步检索结果进行深度语义分析并重新排序，以提升相关性和准确性的技。

3.6 Embedding模型

Embedding模型是将离散数据（如文本、图像）映射为低维稠密向量，以捕捉语义关系并支持高效计算的技术。

3.7 确定性网络

确定性网络是一种通过新一代通信技术构建的新型网络，具有大带宽、低时延、低抖动、高可靠等特点，提供精准控制所需的差异化服务，有效解决传统网络数据传输中的拥堵、延迟、抖动等痛点问题。

4 缩略语

下列缩略语适用于本规范。

CPU：中央处理器（Central Processing Unit）

GPU：图形处理器（Graphics Processing Unit）

UPS:不间断电源(Uninterruptible Power Supply）

IR:信息检索(Information Retrieval）

OCR:光学字符识别(Optical Character Recognition）

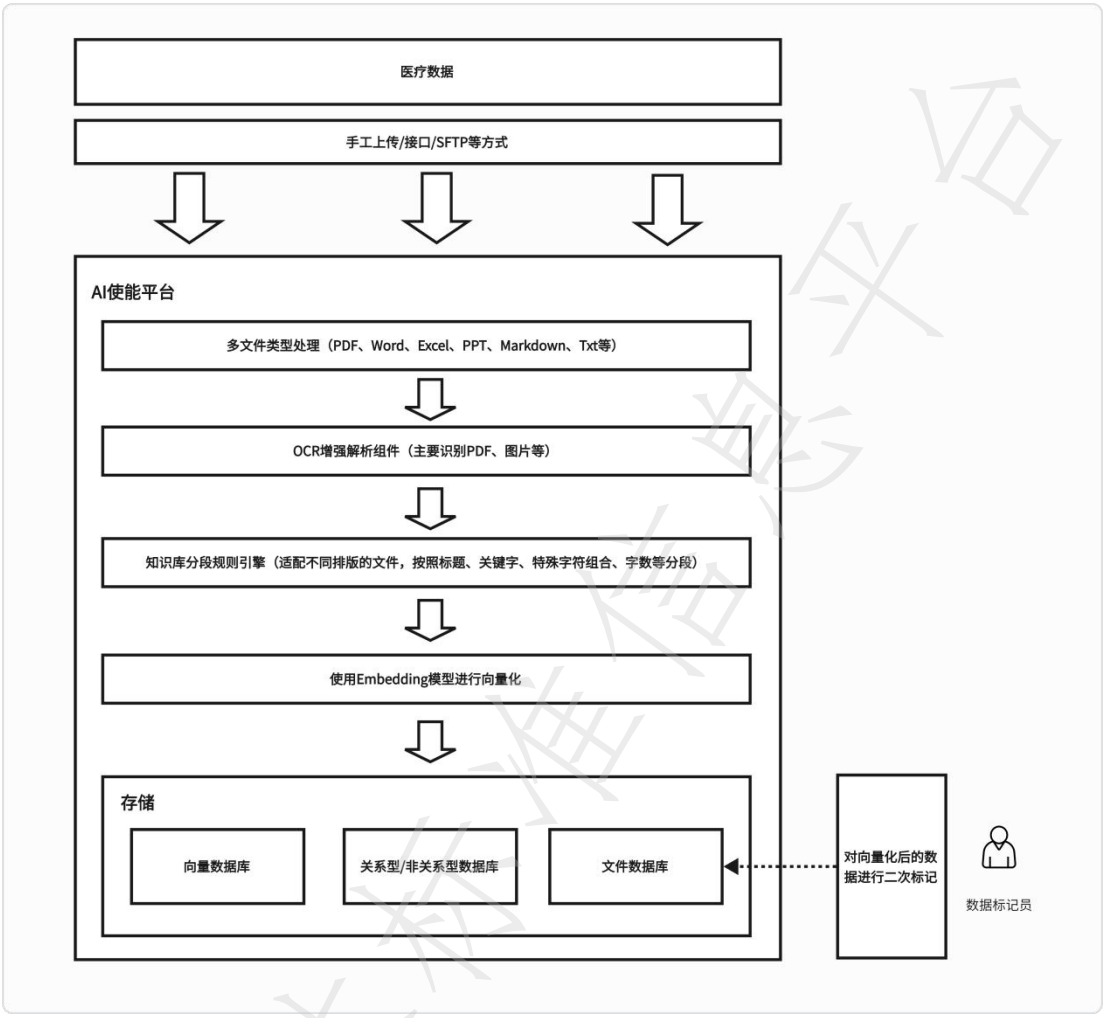
SSD：固态硬盘（Solid State Drive）

5 系统架构及功能

5.1 系统架构

构建基于AI技术的智能审方辅助系统。系统需集成多源权威医疗数据，构建医疗知识库，并基于检索增强生成（Retrieval-Augmented Generation, RAG）技术实现高效信息提取。采用大参数预训练语言模型与智能体（Agent）编排技术构建核心审核决策引擎，为临床医生提供实时、精准的用药建议，提升医疗质量与用药安全性。

5.2 知识库搭建流程

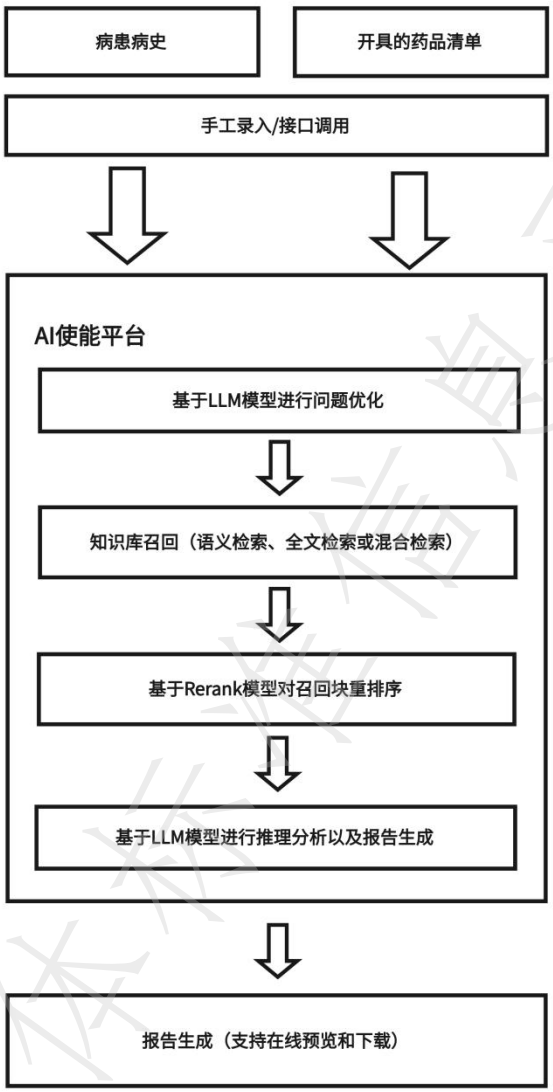


1）数据接入与预处理：多源异构医疗数据通过AI使能平台接入，依次经由OCR增强解析引擎进行文本识别与结构化处理，并通过知识库分段规则引擎进行语义划分与标准化。

2）向量化与持久化存储：对预处理后的数据，应用Embedding模型进行向量化表征，并将生成的向量数据存储至向量数据库。同时，关联的结构化/半结构化数据及原始文档分别持久化存储于关系型数据库、非关系型数据库及文档数据库中。

3）质量控制与标注验证：医学标注专员对知识库中预处理完成的数据样本进行召回率测试与质量评估。基于测试结果，执行二次数据标注流程，以优化标注质量并迭代知识库内容。

5.3 智能体搭建流程



1）患者信息与处方数据接入：目标患者的诊疗信息及待审核药品清单，通过手工录入或系统接口调用的方式，输入至AI医疗智能体。

2）知识检索与关联分析：AI医疗智能体使用大语言模型对输入信息进行语义解析与查询优化，并基于优化后的查询，从构建的医疗知识库中进行信息检索。

3）结果重排序：重排序（Rerank）模型对检索返回的语义关联片段进行相关性排序。

4）推理分析与报告生成：经重排序后的高相关性知识片段，连同原始输入信息，一并提交给推理分析模块（基于大语言模型）进行综合判断与决策推理，最终生成结构化用药审核报告。

5）报告输出与管理：生成的审核报告需支持用户在线实时预览及标准格式文件（如PDF）下载功能。

5.4 数据采集表

数据采集表

一般信息	**住院号/门诊号**			**性别**	
	年龄			**体重**	
	诊断				
	既往用药史及过敏史				
药物治疗情况	药品名称	用法	用量	开始时间	停止时间
辅助检查	项目	检查时间	结果		
目前表现					

5.5 标准化报告产出

产出的报告分为药物重整、用药指导、药学会诊等类型。可根据需要选择一种或多种报告。

1.药物重整报告：重点审查药物治疗情况中的配伍禁忌、药物相互作用评价、调整建议。并附证据依据（参考资料来源）。

2.用药指导报告：重点给出特殊用药交代/治疗过程中特殊注意事项、治疗过程中需监测的指标。并附证据依据（参考资料来源）。

3. 药学会诊报告：重点给出治疗建议、具体用药方案及备选方案。并附证据依据（参考资料来源）。

6 系统设备的要求

医疗人工智能系统的核心价值在于高效执行复杂任务（如实时数据推理、多模态交互、自主决策等），需要硬件设备提供稳定、高效的算力支撑、数据传输能力与环境保障，所需设备需满足智能体系统运行。

医疗人工智能系统的设备符合以下标准:

1 医疗人工智能应选择高可靠性能计算机和其他标准化的配套设备。设备系统规模容量除满足当期工程外，宜满足近期业务发展的需要；其接口应能根据需要随时扩容。

2 系统所需服务器设备，需选择高性能多核CPU、国产/进口先进GPU卡、大型模型内存配置、高速SSD存储等、多个热插拔冗余电源等。

3 系统所需网络要满足高带宽、低时延、低抖动、零丢包、高安全性能，已满足数据快速传输、数据低时延传输、数据稳定可靠的交互以及数据在传输过程中的安全。

4 根据需求配置外部存储设备，如磁盘阵列等，以提供更大的存储容量和数据备份功能。

5 系统设备需要不间断电源（UPS）支撑，以防止突然停电对设备和数据造成损害，确保在停电后能够继续为设备供电一段时间，以便进行数据保存和系统关闭等操作。

7 系统的网络连接

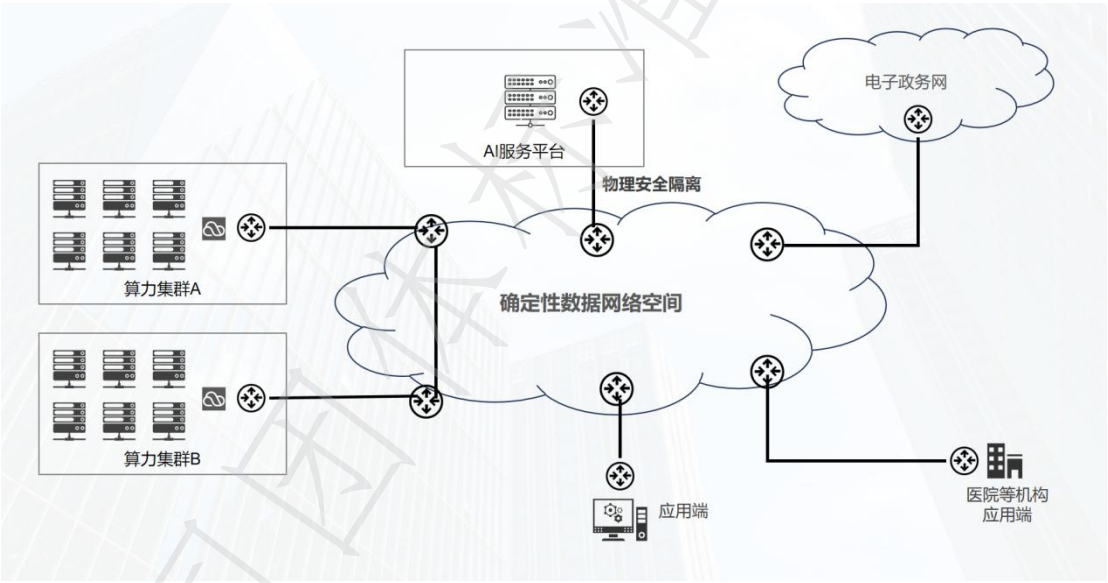
7.1 网络连接的方式

根据服务对象、范围和工程实际情况，系统的数据传输可采用确定性DIP网络、确定性SDN数据专线等连接。

采用确定性网络传送时，应遵循YD/T 4418-2023《电信网络的确定性IP网络的总体架构和技术要求》所规定的网络架构和通信协议。

7.2 系统网络连接的架构

通过确定性网络，连接电子政务网、各算力集群、AI服务平台、医疗机构等，提供可信数据空间内的AI应用服务。具体的网络拓扑如下所示：



8 设备配置

8.1 设备硬件配置要求

AI医疗智能体服务器设备应符合以下配置要求：

1 配置两颗CPU，每颗CPU内核数≥48核，基本频率≥2.1GHz

2 配置八卡GPU，单块GPU卡显存≥90GB

3 内存配置≥1500GB，速率≥5600MT/s

4 系统盘配置两块960GB SSD

5 数据盘配置≥2*3.84T NVMe SSD

6 网卡配置为单块双口25G网卡或两块单口25G网卡

8.2 设备软件配置要求

1 软件应采用图形用户界面，并基于模块化设计，确保各模块高内聚、低耦合。单个模块的修改或升级不应影响其他模块运行。软件应具备系统备份、安全管理、容错处理及性能监控能力。此外，软件系统应满足用户友好、易于维护与升级、模块可测性与可修改性、全部软件及其模块易于升级、开放性等性能。

2 系统安全性，软件系统应具备保护机制，防止因过载导致的错误。程序代码及只读数据应受到保护。系统应支持相关数据的导入与导出功能，并具备自检与自复位能力。

3 软件安装，软件系统的安装对环境有相应的要求，主要包括操作系统、硬件配置、网络环境、依赖库支持等多个方面。

4 软件更新与升级，系统应支持本地或远程进行软件更新与升级。升级或修改操作应在不影响系统正常运行的前提下完成。

9 设备安装

9.1 机房条件

1 AI医疗智能体机房通用环境要求按应用场景应符合GB 50174《数据中心设计规范》的相关规定。

2 AI医疗智能体机房楼面均布活荷载值应符合GB 50009《建筑结构荷载规范》的相关规定。

3 AI医疗智能体机房地板、墙面、吊顶等的防静电应符合标准GB 50611《电子工程防静电设计规范》的相关规定

4 AI医疗智能体机房必须有消防措施，应符合标准GB 50370《气体灭火系统设计规范》的相关规定

5 AI医疗智能体机房及相关设备的防雷接地应符合标准GB 50343《建筑物电子信息系统防雷技术规范》的相关规定。

6 AI医疗智能体机房及相关设备供电应符合标准GB 50052《供配电系统设计规范》的相关规定。

7 AI医疗智能体机房及相关设备应满足抗震要求，应符合GB 50011《建筑抗震设计规范》的相关规定。

9.2 设备硬件安装布置

设备布置应符合高性能智算中心机房的统一规划，在有利于提高机房空间规划、环境适配基础上适当考虑机房整齐美观。具体应满足以下要求：

1 选择恒温、恒湿、防尘、隔音、降噪条件良好的机柜环境.

2 安装位置应预留一定空间，便于通风散热、便于维护施工，并预留设备扩容位置。

3 安装区域应远离高电磁辐射源，若无法避免，需对安装场所进行电磁屏蔽处理。

4 安装环境需有充足且均匀的照明，方便设备操作与维护

5 机房电源线、光纤连接线、通信电缆线宜分开布放。线缆布放位置应合理，不得妨碍或影响日常维护、测试工作的进行。布线距离应尽量短而整齐，且应考虑今后扩容时设备安装及线缆布放。线缆两端粘贴清晰标识，标注用途、连接设备、信号类型等。