采购需求

1 建设方案

1.1 结核病高风险人群精准识别数据大模型及智能体研究支撑服务

1.1.1 结核病防治垂类大模型训练

本部分目标是构建一个对结核病防治领域知识具有深刻理解、能够进行专业推理的垂类大语言模型。该模型是本项目智能应用的核心引擎，其性能直接决定后续智能体的识别精准度与决策辅助能力。

1.1.1.1 训练数据集准备

为确保模型具备坚实的领域知识基础，需构建高质量、多模态、符合伦理与隐私保护要求的训练数据集。数据来源须在符合国家法律法规及双方签订的数据安全协议框架下获取，主要包含：

（1）结构化数据

从上海市疾控中心传染病报告系统、结核病专病管理系统、以及合作医疗机构的医院信息系统（HIS）、实验室信息系统（LIS）中，经严格脱敏处理后，提取近 5-8 年的历史病例数据、诊断信息、检验检查结果（如痰涂片、培养、分子诊断、CD4 细胞计数、病毒载量）、治疗方案及转归信息。

（2）非结构化文本数据

包括临床电子病历（EMR）中的主诉、现病史、既往史、诊疗计划文本；国内外最新的结核病防治指南、专家共识、诊疗规范；权威医学期刊相关文献摘要；公共卫生培训教材及科普材料。

（3）其他数据

经授权的、已脱敏的医学影像报告文本、流行病学调查报告文本等。工作量体现在对上述多源异构数据的协调、获取、格式标准化、整合与初步归类，形成原始数据池，为后续精细加工奠定基础。

1.1.1.2 数据筛选与清洗

针对原始数据池进行严格的质量控制，是保障模型训练有效性的关键前置步骤。具体工作包括：

（1）去重与无效数据剔除：清除完全重复的记录、测试数据、空白记录及明显错误的条目。

（2）不一致性处理：识别并修正逻辑矛盾的数据（如诊断结果与用药方案矛盾、性别与特定疾病矛盾等），需制定明确的业务规则库进行自动化与人工核对结合的处理。

（3）缺失值处理：分析关键字段（如诊断、关键检验指标）的缺失模式，采用合理的填充策略（如基于同类病例的统计值填充、标记为“未知”）或制定策略性舍弃标准，确保用于监督学习的数据完整性。

（4）标准化与归一化：将来自不同系统的诊断名称、药品名称、检验项目等术语，统一映射至标准医学术语体系（如 ICD-10、SNOMED CT 核心子集、国内临床术语标准），并对数值型指标进行标准化处理，消除量纲影响。此阶段交付物为一份高质量、可用于标注的洁净数据集及详细的数据清洗报告。

1.1.1.3 数据标注

组织临床、公卫专家对训练样本进行多维度标注，需组建由临床医学（呼吸科、感染科）、公共卫生、疾病预防控制专家组成的标注团队，在项目组制定的统一标注规范下进行工作。

（1）标注内容

-实体与关系标注：在病历文本中标注疾病实体（如“肺结核”、“肺外结核”、“HIV 感染”）、症状体征（如“咳嗽”、“午后低热”）、检查检验、药物、手术、流行病学因素（如“密切接触史”）等，并标注其间的关系（如“患者-患有-疾病”，“药物-治疗-疾病”）。

-风险等级标注：根据病例信息，结合《中国结核病预防控制工作技术规范》（2020 年版）及防控要求，对病例进行“高风险”、“中风险”、“低风险”的等级标注，并注明核心判断依据。

-意图与问答对标注：针对模拟的医生、公卫人员查询场景，构建高质量的问答对，例如“针对一名有糖尿病史的 HIV 感染者，出现咳嗽症状，应重点考虑哪些鉴别诊断？”及其标准答案。

-诊疗路径标注：对典型病例的诊疗过程进行阶段划分和意图标注，用于训练模型的逻辑推理能力。

-偏好数据标注：在多次问答数据的基础上，人工标注符合业务专家偏好的问答对，并且给出偏好的原因。

-工具使用数据标注：在基线模型的基础上，针对性的合成和人工标注模型调用药品说明书数据库、访问各类医学计算器、各类感染风险评分问卷工具的对话数据。

（2）质量控制

标注过程需实行“双人背靠背标注-仲裁校验”制度，确保标注一致性率不低于98%。需开发或采用高效的标注平台，支持专家在线协作与质量监控。

1.1.1.4 数据增强

针对结核病高风险人群样本在真实数据中可能相对稀少的问题，采用技术手段扩充相关样本的多样性，以提升模型对“长尾”风险模式的识别鲁棒性。

（1）文本数据增强：对已标注的文本数据，采用回译（中英互译）、同义词替换、句式变换、实体替换（在保护隐私前提下）等方法，生成语义一致但表达多样的新样本。

（2）结构化数据增强：对于结构化风险特征数据，在医学合理性边界内，使用SMOTE（合成少数类过采样技术）或条件生成对抗网络（cGAN）生成符合高风险人群特征分布的合成数据样本。

（3）对抗性样本生成：构造具有轻微扰动、可能迷惑模型的边界样本，用于后续增强模型的抗干扰能力。

（4）多轮对话样本合成：采用超大模型针对性的合成多学科符合结核病防治业务场景的多轮对话数据。

1.1.1.5 模型架构设计

设计适合本领域任务的模型架构方案。基于成熟的开源或可商用的通用大语言模型（如特定版本的ChatGLM、Qwen、Kimi等）进行领域化适配设计。

（1）基座模型选型与评估：评估不同基座模型在医学理解、逻辑推理、中文处理方面的能力，结合计算资源、授权许可等因素，选择最合适的1-2个基座模型。

（2）领域适配架构设计：设计高效的领域知识注入与能力激发架构（如稠密模型和混合专家模型等）。重点规划如何将清洗标注后的领域数据用于模型的不同训练阶段（预训练、微调），并设计模型外挂知识库/工具调用的接口方案。

（3）轻量化与部署考量：在架构设计中需考虑最终模型的部署场景（如本地服务器、云端），为可能的模型量化、剪枝、蒸馏等轻量化方案预留设计空间，确保模型在可用算力下能达到实用级响应速度。

交付详细的模型架构设计说明书。

1.1.1.6 模型训练

整体训练流程需分步骤、渐进式地进行。

（1）继续预训练（Continued Pre-training）：使用本项目准备的大规模领域文本数据（指南、文献、清洗后的病历文本等），在选定的基座模型上进行继续预训练。此阶段目标是让模型“广泛阅读”专业资料，将结核病防治知识深度融合到其参数中，形成领域语言和理解能力。需采用大规模分布式训练技术，管理训练过程中的灾难性遗忘问题。

（2）有监督微调（Supervised Fine-Tuning, SFT）：使用高质量标注的指令数据（问答对、风险标注对、诊疗路径数据等）对模型进行有监督微调。此阶段目标是教会模型遵循人类指令，完成具体的风险评估、问答、报告生成等任务。需要精心设计指令模板和损失函数。

（3）强化学习（Reinforcement Learning from Human Feedback, RLHF）或基于AI反馈的强化学习（RLAIF）：构建一个奖励模型，用于评估SFT后模型输出结果的质量（如准确性、全面性、安全性、符合指南程度）。通过强化学习算法（如PPO、DPO、GRPO等）进一步优化模型，使其输出更符合专家偏好和临床安全规范。此阶段是提升模型决策质量与可靠性的关键。

（4）工具使用学习（Tool Learning）：训练模型学会调用外部工具，例如：查询最新的药品说明书数据库、访问计算某个体重指数（BMI）或感染风险评分的小工具、检索内部知识库中的最新防控政策。这能极大扩展模型的能力边界和事实准确性。此阶段需详细记录训练超参数、损失曲线、评估指标变化，并产出多个中间检查点模型。

1.1.1.7 模型评估与优化

构建多层次、多维度的评估体系，对训练出的模型进行严格评测与迭代优化。

（1）构建综合评估集：包含“封闭集”和“开放集”。封闭集来自标注数据的留出

部分；开放集为项目组新构造的、模型未见过但符合真实场景的复杂案例和问题。

（2）评估维度：

知识与事实准确性：通过专业考试题、事实性问答进行评测。

风险评估性能：采用精准率、召回率、F1-score、AUC-ROC 曲线等指标，评估模型识别高风险人群的能力，核心目标是在保证高特异度（>90%）的前提下，追求高灵敏度（>85%）。

推理与决策合理性：邀请领域专家对模型给出的诊断思路、防控建议进行盲评打分。

安全性与合规性：评估模型输出是否存在偏见、是否产生有害建议、是否符合医疗伦理和患者隐私保护原则。

实用性能：测试模型的响应延迟、吞吐量等。

（3）迭代优化：

根据评估结果，分析模型缺陷，可能回溯至数据增强、SFT 数据补充或 RLHF 奖励模型调整等环节，进行多轮迭代优化，直至模型在核心评估集上达到预定性能标准。

交付最终评估报告和优化后的可交付模型。

1.1.2 结核防控高风险人群精准识别智能体构建支撑服务

本部分旨在将训练好的垂类大模型封装成一个具备自主任务执行能力、可集成到现有工作流程的实用化智能体系统。

1.1.2.1 智能体架构设计

设计一个模块化、可扩展的智能体系统架构。

（1）核心引擎：集成优化后的垂类大模型作为“大脑”。

（2）规划模块：负责解析用户指令（如“筛查今日门诊患者中的结核高风险者”），将其分解为可执行的子任务序列（数据查询、模型推理、结果汇总）。

（3）记忆模块：设计短期对话记忆和长期知识缓存机制，使智能体能在多轮交互中保持上下文连贯。

（4）工具集成模块：提供标准接口，使智能体能够安全、可控地调用外部 API、数据库查询接口、计算公式等。

（5）安全与审核模块：设计输出过滤、风险内容识别和关键决策建议人工复核触

发机制。

（6）接口层：设计面向用户（Web 界面、API）和面向其他系统（与 HIS、公卫平台对接）的接口规范。交付智能体系统架构设计图及接口文档。

1.1.2.2 智能体模块开发

根据智能体架构设计，实现以下核心功能模块：

（1）风险识别模块：能够手动识别或者按照预设周期或触发条件，从对接的医疗信息系统中抽取新入患者/就诊者数据，调用模型进行批量风险评估，并生成疑似高风险名单及简要依据。

（2）交互式咨询模块：提供自然语言对话界面，供医生或公卫人员主动询问个案或群体风险。智能体能理解复杂查询，调用模型进行分析，并以结构化报告或清晰文本形式返回结果，支持多轮追问。

（3）系统对接适配器：开发与上海市疾控中心指定信息系统、以及典型医院信息系统（HIS）进行安全数据交换的适配器，实现数据的自动化流入（脱敏后）和结果的自动化流出（至医生工作站或管理后台）。

（4）管理后台：提供配置管理（如风险阈值调整）、任务监控、日志审计、用户权限管理等功能。

1.1.2.3 智能体测试与优化

在模拟环境和可控的真实环境中进行系统化测试。

（1）功能测试：验证所有设计功能是否正确实现，包括数据对接、自动筛查、交互问答、管理配置等。

（2）性能测试：测试智能体在高并发请求下的响应时间、吞吐量及稳定性，确保满足日常业务峰值需求。单次风险评估 API 响应时间应在秒级（如<5 秒）。

（3）集成测试：与模拟的或测试环境中的业务系统进行对接联调，测试数据流和业务流的完整性与正确性。

（4）安全测试：进行渗透测试、漏洞扫描，确保系统无高危安全漏洞；测试数据脱敏、传输加密、访问控制的有效性。

（5）用户验收测试（UAT）：邀请最终用户（疾控人员、临床医生）进行实际操作

测试，收集可用性反馈，并基于反馈对用户界面、交互流程、报告格式等进行多轮优化，提升用户体验和接受度。

1.1.2.4 智能体整体封装

将完整的智能体系统、其所有依赖（模型、运行时环境、配置文件）打包封装为标准化的 Docker 镜像。提供详细的镜像部署手册，包括环境要求、启动命令、配置说明、健康检查方式等。确保该镜像可在疾控中心指定的云平台或本地服务器上通过容器化技术（如 Docker，Kubernetes）一键部署、平滑扩展，实现高效、稳定的服务交付。

交付最终可运行的 Docker 镜像文件及完整的部署文档。

2 其他招标需求

2.1 项目实施要求

投标人需提供详细的项目实施方案，包括项目的组织及人员配备，项目实施过程，系统测试及验收方案等。

投标人需要保证投入到本项目的开发人员能够准确、全面地理解业务需求并在项目中进行实现，能够满足本项目需求涉及面广、专业性强的要求。

同时，在保证实施团队技术力量的前提下，需要对项目实施人员的职责进行明确规定、对项目实施过程进行有效监督，加强需求的提供方与系统开发方之间的沟通效率，确保项目的顺利实施。

投标人需提供合理可行的设备到货计划方案，并严格按照业主要求配送到指定地点，配合业主及其指定的人员完成设备的到货验收。

2.2 项目工期要求

合同签订后 12 个月内完成全部建设内容并通过验收，包含 1 个月试运行期。

供应商需在合同生效后立即组织项目实施人员开展具体工作。

2.3 质量保证要求

投标人需提供规范的软件工程实施和项目管理方法，提供完整的质量保证体系，对项目实施的全过程进行监控和管理，对项目进度和工程实施进度进行控制。

本项目模型通过上海市医疗大模型应用检测验证中心测评。测评通过结果将作为项目验收的先决条件。

2.4 系统验收及交付要求

按照招标文件的要求，实施方需要配合业主进行系统的整体验收，完成本项目的验收和交付，并提交全部相关文档资料等。实施方需要根据本项目建设需求和系统设计方案，提供详细的系统验收手册和验收流程、验收通过标准，经业主方认可后，由双方共同组织人员进行系统验收。

一、模型基础能力验收要求

1、模型参数规模：交付的基座模型参数量应严格控制在 14B（140 亿）至 70B（700 亿）范围内。

2、模型版本与精度：根据项目实际需求，模型可交付量化版本或非量化版本。若交付量化版本，供应商须明确说明所采用的量化技术（如 GPTQ、AWQ、INT8/INT4 等），并提供量化前后在关键医学任务上的精度损失评估报告，确保量化后的模型性能仍能满足业务要求。

3、API 接口与文档：必须提供标准化的 HTTP API 接口，并配套详细的技术文档。文档内容需涵盖接口地址、请求方式、Headers、请求体与响应体示例、错误码说明以及鉴权方式（如 AK/SK 或 Token），便于后续系统集成与二次开发。

4、医学方向性验证：模型需经过医学语料的针对性预训练或指令微调，具备明确的医学方向性。验收时将采用基础医学知识问答、医学推理等测试集进行评测，验证模型在医学语义理解方面的能力，排除通用对话模型直接用于医疗场景的情况。

二、智能体性能与业务能力验收要求

1、系统响应性能：模型 API 的端到端响应时间应达到秒级。验收时将在标准生产环境下，使用压力测试工具模拟并发请求，统计不同负载下的响应时间分布，确保系统具备良好的实时响应能力。

2、高风险结核病识别业务指标：针对高风险结核病筛查这一核心任务，模型需同时满足以下三项关键指标：

（1）精准率（Precision） > 85%：在所有被模型判定为高风险的患者中，实际患病者的比例应高于 85%。

（2）特异度（Specificity） > 90%：在所有实际未患病的患者中，被模型正确判定为低风险的比例应高于 90%。

（3）灵敏度/召回率（Sensitivity/Recall） > 85%：在所有实际患病的患者中，被模型成功识别为高风险的比例应高于 85%。

3、业务接口与文档：业务层面提供的 API 接口及相关文档需清晰易懂，验收时将安排开发人员依据文档进行联调测试，验证接口的可用性和文档的准确性。

三、安全合规性要求

模型通过上海市医疗大模型应用检测验证中心测评：模型整体方案及测评报告必须符合上海市医疗大模型应用检测验证中心的测评要求。供应商需提供该中心出具的检测报告或通过其测评的证明文件，作为安全合规性验收的前置依据。

四、总体验收要求

本项目的模型基础能力、智能体性能与业务能力、安全合规性都满足要求，且交付物完整，即交付物包含完整的模型文件、配置文件、运行环境依赖说明，以及模型部署与使用的操作手册，由实施方和采购人协商验收时间，由采购人组织进行项目验收。

2.5 培训要求

本项目要求供应商对采购人技术人员进行全面技术培训，培训内容应包括：供应商所提供系统的性能、操作使用方法，维护管理技术，并能实际上机操作。

供应商对采购方人员的培训，应使采购方人员可以独立操作、维护、管理，从而使采购方人员能独立进行管理、故障处理、日常测试维护等工作，确保系统能正常安全运行。

投标人需提供详细的项目培训方案。

2.6 技术文档要求

系统验收前需提交各类成果文档(电子文档是成果不可分割的部分)及系统源代码，电子文档可不限于以下几类：

1、需求分析说明书

2、系统设计说明书

3、数据库设计

4、用户操作手册

5、安装维护手册

6、系统测试报告

2.7 售后服务要求

投标人需针对本项目软件部分提供不少于一年的免费维保服务。投标人应根据招标要求，承诺相应质保服务，并在免费维护期中，提供详细的维护服务和技术支持计划，若遇到与软硬件产品相关的技术问题，能够及时响应，委派技术工程师提供现场技术支持和服务。

投标人承诺，维护期内的技术服务免费，并至少提供以下服务：

1、提供固定的售后服务联系人员 1 名，提供人员名单、联系电话和地址。

2、对系统故障应满足 2 小时内响应，及时安排工程师根据情况进行排除,包括远程调试和现场支援,根据用户需求 8 个小时到达故障现场，24 小时内解决问题。

3、提供 7*24 小时本地语言（中文）支持。

4、无论由于哪一方产生的问题导致系统发生不正常情况时，供应商需立即派工程师到场全力配合解决，使系统尽快恢复正常。

5、以上所有服务的相关费用均需计入投标总价。

2.8 安全保密要求

1、对于采购人提供给各供应商的招标文件和其他业务需求说明文件，供应商有为采购人保密的义务；

2、供应商应在采购人指定的工作机房内对数据进行处理，不得对用户的数据进行拷贝、备份并带出工作机房，同时不得对外泄露用户数据资料（包括自然人、法人和其他共享数据）的具体内容。否则由此引起的用户损失、相关社会和法律责任，由供应商负责承担；

3、供应商需协助用户采取各种管理和技术的手段，确保用户的数据不外传和泄露；

4、协助用户进行技术方案审理优化、进度质量控制和验收管理的一方，有为用户和包件供应商提供的项目有关文件和技术信息保密的义务。

2.9 项目建设团队要求

投标人应为本项目合理配置项目实施团队，提供项目人员组织方案，包括组织结构、项目负责人、组成人员及分工职责，核心技术人员需具备类似项目实施经验。

项目团队不少于 5 人，包括项目经理 1 人，具有计算机或软件类专业中级及以上职称。项目经理具有大型信息系统规划设计、项目管理经验，担任过类似项目的项目经理；有较好的文字表达和沟通协调能力，具备良好的沟通和组织能力，能带领和指导团队顺利完成分析工作。项目经理提供相关证书、学历证书及工作简历。

项目组其他成员 4 人，包括：产品设计工程师 1 人，负责需求调研及产品设计；数据分析工程师 1 人，负责数据治理分析服务；技术开发及测试工程师 2 人，相关软件开发及测试等。主要成员提供相关证书、学历证书及工作简历。

2.10 第三方接口要求

涉及到本项目的建设内容，需与甲方现有信息系统进行对接，接口开发涉及的费用包含于本项目报价中，甲方不再另行支付接口费用。