

# **上海医保经办服务人工智能应用子系统 建设项目采购需求**

上海市医疗保险事业管理中心

# 目录

一、 项目概况 .....	1
(一) 项目背景 .....	1
(二) 建设目标 .....	1
(三) 项目建设内容 .....	2
二、 项目建设要求 .....	3
(一) 总体要求 .....	3
1. 技术要求 .....	3
2. 性能要求 .....	3
(二) 应用系统建设要求 .....	5
(三) 医保智算中心建设要求 .....	8
(四) 与市医保信息平台双向同步要求 .....	17
(五) 商业密码应用需求 .....	17
(六) 项目部署要求 .....	17
三、 项目实施、培训、售后服务等要求 .....	18
(一) 系统软硬件产品供货要求 .....	18
(二) 项目实施要求 .....	19
(三) 项目工期要求 .....	20
(四) 项目团队及驻场人员要求 .....	20
(五) 项目培训要求 .....	21
(六) 项目验收要求 .....	21
(七) 售后服务要求 .....	23
(八) 应急响应要求 .....	23
(九) 投标人综合能力要求 .....	24
(十) 技术文件要求 .....	24
(十一) 知识产权及保密要求 .....	25

## **一、项目概况**

### **(一) 项目背景**

为深入贯彻落实党中央、国务院关于“人工智能+”战略部署、国家医疗保障局办公室《关于安全稳妥有序推进人工智能大模型在医保领域应用的通知》（医保办函〔2025〕45号）的相关要求，根据市委、市政府主要领导对《上海市医疗保障经办领域人工智能应用规划》(2025-2030)的批示要求，全力推进人工智能与医疗保障领域的深度融合，以医保数据作为资源基础，以应用场景为驱动力，全面推进医保数据治理和数据画像标签化应用、多模态数据集和专业语料库建设、支撑平台搭建及各应用场景的拓展，形成“技术赋能监管、数据驱动决策、智能优化服务”的智慧医保新生态，全面提升医保治理的精准性、服务的便捷性和制度的可持续性，为超大城市医疗保障体系现代化建设提供强劲动力，努力打造全国智慧医保创新发展的示范标杆。

### **(二) 建设目标**

市医保深入实施创新驱动战略，聚焦民众对医疗保障服务的新期盼，以人工智能技术革新为引领，旨在提升医保服务质量和效率，以医保数据为资源基础，以参保人异常就医行为识别、两定机构资格证书复核、零星报销智能辅助审核、数据质量智能监测、医保政策知识检索等场景应用为驱动引擎，统筹推进医保数据治理及数据画像标签化应用、多模态数据集和垂类语料库建设、支撑平台搭建及各场景应用拓展，充分激发人工智能与医疗保障经办深度融合的创新潜能，推动医保经办管理向智能化、服务向精准化、决策向科学化转型升级，全面构建智慧医保服

务新格局，助力医疗保障改革向更深层次、更广领域迈进。

### (三) 项目建设内容

本项目建设内容包括：

#### 1. 构建市医保人工智能底座，推进智慧医保大模型应用

基于市医保专用网内部私有化部署，包含数据层、计算层、应用层等硬件建设和医保知识库、规则库等软件建设于一体的基础底座平台，建立人工智能驱动的医疗保障服务新模式，开展智能算法训练，推进智慧医保大模型应用，为医保政策落地实施和优化提供精准数据支撑。

#### 2. 夯实市医保数据基础支撑，提升人工智能数据应用能效

推动医保数据资源与人工智能技术的深度融合，建设多模态数据集和语料库，优化算法模型，提升数据应用能效，为智慧医保提供坚实的数据支撑。

#### 3. 深化市医保创新场景，推动医保服务迈入智能化

统筹推进人工智能与医保业务的深度融合，建设第一阶段规划的应用场景，借助持续的技术迭代和业务优化，推动医保服务迈入智能化和精准化。

本次计划建设的智能应用场景包括：参保人异常就医行为识别、医保零星报销智能辅助审核、智能纳保、医保政策知识检索、医保数据质量智能监测。

## **二、项目建设要求**

### **(一) 总体要求**

#### **1. 技术要求**

(1) 投标人提供的大模型需支持私有化部署模式，适配主流算力芯片及加速卡，确保模型推理性能稳定高效。

(2) 大模型需支持分层模块化架构，支持不同业务场景下的模型微调与功能扩展，单个模块的迭代升级不影响整体服务可用性。

(3) 大模型版本需支持向下兼容，新版本迭代时需提供完整的接口适配方案，确保在模型升级过程中业务系统调用不受影响，推理服务中断时间控制在规定阈值内。

(4) 配套的大模型训练、推理及监控工具需为主流成熟产品，提供完整的模型生命周期管理功能，包括数据清洗、模型训练、评估、部署、监控等，并能与已建成的市医保信息平台实现数据互通与功能对接。

#### **2. 性能要求**

性能指标涵盖知识库、智能体、业务系统 3 个维度：

##### **(1) 知识库**

知识库性能指标主要包括单个文件入库平均时间、文件更新平均时间、文件删除更新平均时间、信息检索响应时间、信息检索准确率、混合检索响应时间、混合检索准确率等。具体如下：

- 1) 单个文件入库平均时间  $\leqslant$  50 秒
- 2) 文件更新平均时间  $\leqslant$  50 秒
- 3) 文件删除平均时间  $\leqslant$  1 秒

- 4) 信息检索响应时间  $\leq 2$  秒
- 5) 信息检索准确率  $\geq 90\%$
- 6) 混合检索响应时间  $\leq 8$  秒
- 7) 混合检索准确率  $\geq 90\%$

### (2) 智能体

智能体性能指标主要包括文档图像识别率、关键信息提取准确率、篡改检测误报率、日处理文件数量、单个文件平均处理时间等。具体如下：

- 1) 文档图像识别率  $\geq 90\%$
- 2) 关键信息提取准确率  $\geq 90\%$
- 3) 篡改检测误报率  $\leq 10\%$
- 4) 日处理文件数量  $\geq 1000$  份
- 5) 单个文件平均处理时间  $\leq 5$  分钟

### (3) 业务系统

主要从系统响应时间给出系统性能指标。系统响应时间指完成目标系统中的交互或批量业务处理所需的响应时间。根据业务处理类型的不同，划分为两类：交互类业务、查询类业务，分别给出响应时间要求的参考值，包括峰值响应时间、平均响应时间。

#### 1) 交互类业务

交互类业务是指平时工作中在系统中进行的业务处理，如录入、修改或删除一条记录、发布一条信息等操作。

- 平均响应时间： $\leq 1$  秒；

- 峰值响应时间：≤ 3 秒。

批量前台经办业务数据导入（按一次 2000 条评估）。

- 平均响应时间：≤ 5 秒；
- 峰值响应时间：≤ 10 秒。

## 2) 查询类业务

查询业务由于受到查询的复杂程度、查询的数据量大小等因素的影响，需要根据具体情况而定，在此给出一个参考范围。

- 简单查询平均响应时间：≤ 1 秒；
- 复杂查询平均响应时间：≤ 3 秒。

## (二) 应用系统建设要求

序号	功能名称		功能描述
1	大模型 底座建 设	大模型知识库工具	需要实现知识采集、知识存储、知识处理、知 识管理、知识服务功能
2		大模型智能体工具	需要实现智能体构建、智能体要素、智能体运 营功能
3		基础模型部署	需要实现与本项目相关的医保大语言模型部署 功能
4	参保人 异常就 医行为 识别	数据预处理	需要实现数据处理、数据增强功能
5		医保参保人就医识别 审核知识库建设	需要实现知识语料管理与标准化、知识标记与 数据集构建、知识图谱构建与维护、知识间关 系组织与管理、知识库更新与优化功能

6	参保人异常就医用法用量识别智能体建设 参保人异常就医配伍禁忌识别智能体建设 配套业务改造	参保人异常就医用法用量识别智能体建设	需要实现药品说明书质量评估、信息定位与字段识别、语义补全与规则推理、单位换算、异常识识别与审核输出功能
7		参保人异常就医配伍禁忌识别智能体建设	需要实现信息定位与字段识别、医学知识激活与禁忌识别触发、配伍校验与合规识别、禁忌判定解释与交互反馈、审核报告生成功能
8		配套业务改造	需要实现药品说明书审核管理、“药品-诊疗行为-费用”三维知识图谱功能
9	零星报销智能辅助审核 医保零星报销审核知识库建设 零星报销智能辅助审核智能体建设 配套业务改造	数据预处理	需要实现数据处理、数据增强功能
10		医保零星报销审核知识库建设	需要实现知识语料管理与标准化、知识抽取与归类、知识图谱构建与维护、知识增强与共享、知识库更新与优化功能
11		零星报销智能辅助审核智能体建设	需要实现报销材料质量评估、关键字段识别与结构提取、风险检测与篡改识别、审核判断与反馈检测、结构化结果与可信评分功能
12		配套业务改造	需要实现医保零星报销登记管理、医保零星报销审核管理、长三角电子票据共享接口开发功能
13	智能纳保 医保两定机构纳保审核知识库建设	数据预处理	需要实现数据处理、数据增强功能
14		医保两定机构纳保审核知识库建设	需要实现知识语料管理与标准化、知识归类与关键信息提取、知识图谱构建与维护、知识增强与共享、知识库更新与优化功能

15		两定机构资格证书复核智能体建设	需要实现资格证书图像质量评分、材料自处理与文件结构解析、图像篡改检测与内容真实性评估、字段一致性比对与语义补全、智能复核与结构化输出、字段识别及语义复核功能
16		配套业务改造	需要实现两定机构资格证书复核管理、两定机构纳保管理、医保支付资格管理功能
17		数据预处理	需要实现数据清洗、数据转换、数据去噪、数据标准化、数据脱敏功能
18	数据质量智能监测	医保数据监测报表知识库建设	需要实现知识语料管理与标准化、知识标记与数据集构建、知识质量评估与映射、知识间关系组织与管理、知识库更新与优化功能
19		医保数据报表智能查询及展示智能体建设	需要实现智能问答交互、查询请求智能处理、SQL语句生成优化、数据查询与结果分析、动态查询优化与自学习功能
20		配套业务改造	需要实现数据指标管理、规则模板管理、医保数据质量预警管理、数据分级分类管理、数据语义智能分析管理、数据质量监测展示功能
21	医保政策知识检索	数据预处理	需要实现数据处理、数据增强功能
22		医保政策知识库建设	需要实现知识语料管理与标准化、知识切分与向量化、知识归类与元数据管理、知识检索与溯源、知识库更新与优化功能

23		医保政策知识分析检索智能体建设	需要实现政策文件质量评估、文档结构识别与处理、智能检索以及语境构建、知识聚合与推理分析、回复响应生成与交互反馈回溯功能
24		配套业务改造	需要实现政策文件标签管理、政策文件质量评估、政策文件管理、政策文件检索、政策文件解读、政策文件关联性分析、政策文件归集分析、政策文件发文单位分析功能
25	密码应用功能	/	需要实现用户身份认证机制模块、业务重要数据安全传输模块、服务器虚拟机设备日志/访问控制信息完整性模块、重要可执行程序签名验签模块、用户访问控制信息签名验签模块、应用系统重要数据加解密模块、应用系统重要数据签名验签模块功能

### (三) 医保智算中心建设要求

医保智算中心拟部署并承载大量医保领域垂类人工智能模型的训练任务，其中部分大模型训练任务的持续时长可达数小时至数天，且训练过程中涉及海量模型中间检查点（checkpoints）的数据读写操作。若系统发生故障，将可能造成重大损失，因此，对平台的高可靠性与高安全性提出了明确要求。

本项目需采用成熟、稳定、可靠的算力产品和技术，充分复用医保信息平台现有的物理环境、网络环境以及安全体系来构建医保智算中心，

以保证后续长期可靠的运行。

## 1. 基本要求

- (1) 要求投标人充分理解项目建设需求、建设内容，设计的医保智算中心需要充分与现有物理环境、网络环境以及安全体系进行无缝衔接。方案需要以独立的章节进行详细描述。
- (2) 投标人提供的配套软硬件产品，需是主流成熟的产品，所有配件需配齐，以构成一套完整的系统，并能够完全兼容已建成投用的市医保信息平台。
- (3) 投标人选择的产品需满足招标技术参数要求，其中“▲”指标为关键性参数，不得虚假应答，并提供相关证明材料，材料需清晰可辨识。
- (4) 系统软硬件采购清单

序号	名称	质量保证期	数量
1.	高速交换设备	三年	2
2.	数据预处理节点	三年	3
3.	小规模训练节点	三年	2
4.	大规模训练节点	三年	2
5.	多模态训练节点	三年	2
6.	算力监控平台	一年	1
7.	大模型管理平台	一年	1

## 2. 技术参数指标要求

- (1) 高速交换设备

序号	功能项	技术规格要求
----	-----	--------

1.	配置要求	配置冗余主控， $\geq 48$ 个 SFP28 口， $\geq 8$ 个 QSFP28 口，冗余电源模块，配置 $\geq 16$ 个模块。
2.	基础性能	交换容量 $\geq 100\text{Tbps}$ ，包转发率 $\geq 76800\text{Mpps}$
3.	业务插槽数	业务插槽数 $\geq 6$
4.	电源冗余	电源模块冗余
5.	安全性	▲要求设备支持防火墙业务插卡，提供第三方测试报告截图证明
6.	控制引擎	主控引擎模块 $\geq 2$ ，满足 1+1 冗余
7.	ACL	支持双向 ACL， $\text{ACL} \geq 4\text{K}$ 支持端口 ACL 支持 VLAN ACL
8.	QoS	每端口支持 8 个优先级队列，3 个丢弃优先级，支持 SP、WRR、SP+WRR 三种队列调度算法 支持精细化的流量监管，粒度可达 8K 支持流量整形 Shapping 支持 WRED 拥塞避免 支持 802.1p、TOS、DSCP、EXP 优先级映射
9.	可靠性	双引擎快速倒换，主备切换时候板内转发无丢包 支持 NSF/GR for OSPF/BGP/IS-IS 支持热补丁功能，可在线进行补丁升级 支持 BFD，BFD for VRRP/BGP/IS-IS/OSPF/RSVP/LDP/RIP/静态路由。
10.	软件规格	▲MAC 表 $\geq 1\text{M}$ ，IPv4 FIB $\geq 3\text{M}$ ，IPv6 FIB $\geq 1\text{M}$ ，ARP 表 $\geq 256\text{K}$ ，提供第三方测试报告截图证明
11.	可视化	支持 Telemetry 流量可视化功能

12.	MPLS	支持 L3 VPN、支持 VLL、支持 VLPS、支持 MCE
13.	跨三层互联技术	支持主流的 MAC in IP 技术，如 EVI，实现跨三层网络的二层互联，提供工信部或下属实验室出具的第三方测试报告，提供官网公开配置手册和截图证明
14.	售后服务	3 年原厂维保，提供售后服务承诺函

## (2) 数据预处理节点

序号	功能项	技术规格要求
1.	设备外型	$\geq 2U$ 机架式
2.	处理器	支持多达 2 颗第三代可扩展处理器，单颗 $\geq 16$ 个核心，主频 $\geq 2.4GHz$ ，合计 $\geq 32$ 个核心
3.	内存容量	支持 $\geq 32$ 个 DDR4 内存插槽，速率可高达 3200MT/s，支持 RDIMM 或 LRDIMM，单颗处理器最高支持 $\geq 6TB$
4.	磁盘扩展	支持 $\geq 12$ 盘 2.5/3.5 英寸 SAS/SATA/NVMe
5.	存储控制器	支持 RAID0/1/10/5
6.	PCIe 插槽	支持 $\geq 14$ 个 PCIe 4.0 标准插槽 支持 $\geq 1$ 个 OCP 3.0 插槽
7.	GPU 计算模块	支持 $\geq 4$ 个双宽 GPU， 单模块基础频率 $\geq 2,280MHz$ ，CUDA 核心 $\geq 14,592$ 个，Tensor 核心 $\geq 456$ 个，RT 核心 $\geq 114$ 个
8.	风扇	支持热插拔冗余风扇
9.	网络	支持 $\geq 1/10/25/40/50/100/200GE$ PCIe 标准网卡
10.	电源	支持 $\geq 2$ 个 1+1 白金电源，支持 N+N 冗余

11.	配置要求	$\geq 2$ 颗处理器, $\geq 384G$ 内存, $\geq 1TB$ SSD 盘, $\geq 5TB$ SAS 盘, $\geq 2$ 张 $24G$ 缓存 GPU 计算模块, $\geq 4$ 口千兆电口, $\geq$ 冗余电源、风扇, 3 年原厂维保
-----	------	--

### (3) 小规模训练节点

序号	功能项	技术规格要求
1.	设备外型	$\geq 4U$ 机架式
2.	处理器	支持多达 2 颗第三代可扩展处理器, 单颗 $\geq 12$ 个核心, 主频 $\geq 2.1GHz$ , 合计 $\geq 24$ 个核心
3.	内存容量	支持 $\geq 32$ 个 DDR4 内存插槽, 速率可高达 $3200MT/s$ , 支持 RDIMM 或 LRDIMM, 最高支持 $\geq 12TB$ , 支持内存高级 ECC, 内存镜像模式, 在线内存备份模式
4.	磁盘扩展	支持 $\geq 24$ 盘 2.5/3.5 英寸 SAS/SATA/NVMe 或 12 盘 2.5/3.5 英寸 SAS/SATA/NVMe
5.	存储控制器	支持 RAID0/1/10/5/6/50/60
6.	PCIe 插槽	支持 $\geq 8$ 个双宽 x16 插槽 支持 $\geq 4$ 个单宽 X16 插槽 支持 $\geq 1$ 个 OCP 3.0 网卡专用插槽 支持 $\geq 1$ 个阵列卡专用插槽
7.	GPU 计算模块	支持 $\geq 8$ 个双宽 GPU, 单模块基础频率 $\geq 2,235$ MHz, CUDA 核心 $\geq 16,384$ 个, Tensor 核心 $\geq 512$ 个, RT 核心 $\geq 128$ 个
8.	风扇	支持 $\geq 8$ 个风扇, 支持 N+1 热插拔冗余风扇
9.	网络	支持 $\geq 1/10/25/40/50/100/200GE$ PCIe 标准网卡
10.	电源	支持 $\geq 4$ 个 2000W/2400W 白金电源, 支持 N+N 冗余
11.	配置要求	$\geq 2$ 颗处理器, $\geq 512GB$ DDR 内存, $\geq 5TB$ SSD 硬盘, $\geq 4$ 端口千兆网卡, $\geq 8$ 张 $24G$ 缓存 GPU 计算模块, $\geq 4$ 个冗余电源, 3 年原厂维保

#### (4) 大规模训练节点

序号	功能项	技术规格要求
1.	设备外型	$\geq 4U$ 机架式
2.	处理器	支持多达 2 颗第三代可扩展处理器，单颗 $\geq 12$ 个核心，主频 $\geq 2.1GHz$ ，合计 $\geq 24$ 个核心
3.	内存容量	支持 $\geq 32$ 个 DDR4 内存插槽，速率可高达 3200MT/s，支持 RDIMM 或 LRDIMM，最高支持 $\geq 12TB$ ，支持内存高级 ECC，内存镜像模式，在线内存备份模式
4.	磁盘扩展	支持 $\geq 24$ 盘 2.5/3.5 英寸 SAS/SATA/NVMe 或 12 盘 2.5/3.5 英寸 SAS/SATA/NVMe
5.	存储控制器	支持 RAID0/1/10/5/6/50/60
6.	PCIe 插槽	支持 $\geq 8$ 个双宽 x16 插槽 支持 $\geq 4$ 个单宽 X16 插槽 支持 $\geq 1$ 个 OCP 3.0 网卡专用插槽 支持 $\geq 1$ 个阵列卡专用插槽
7.	GPU 计算模块	支持 $\geq 8$ 个双宽 GPU， 单模块 FP32 算力 $\geq 59.35$ TFLOPS，INT8 算力 $\geq 239$ TOPS，FP8 算力 $\geq 239$ TOPS，显存 $\geq 48GB$ ，显存位宽 $\geq 384\text{-bit}$
8.	风扇	支持 $\geq 8$ 个风扇，支持 N+1 热插拔冗余风扇
9.	网络	支持 $\geq 1/10/25/40/50/100/200GE$ PCIe 标准网卡
10.	电源	支持 $\geq 4$ 个 2000W/2400W 白金电源，支持 N+N 冗余
11.	配置要求	$\geq 2$ 颗处理器， $\geq 512GB$ DDR 内存， $\geq 5TB$ SSD 硬盘， $\geq 1$ 块 2GB 缓存阵列卡， $\geq 4$ 端口千兆网卡， $\geq 8$ 张 48G 缓存 GPU 计算模块， $\geq 4$ 个冗余电源，3 年原厂维保

## (5) 多模态训练节点

序号	功能项	技术规格要求
1.	设备外型	$\geq 4U$ 机架式
2.	处理器	支持多达 2 颗第三代可扩展处理器，单颗 $\geq 12$ 个核心，主频 $\geq 2.1GHz$ ，合计 $\geq 24$ 个核心
3.	内存容量	支持 $\geq 32$ 个 DDR4 内存插槽，速率可高达 3200MT/s，支持 RDIMM 或 LRDIMM，最高支持 $\geq 12TB$ ，支持内存高级 ECC，内存镜像模式，在线内存备份模式
4.	磁盘扩展	支持 $\geq 24$ 盘 2.5/3.5 英寸 SAS/SATA/NVMe 或 12 盘 2.5/3.5 英寸 SAS/SATA/NVMe
5.	存储控制器	支持 RAID0/1/10/5/6/50/60
6.	PCIe 插槽	支持 $\geq 8$ 个双宽 x16 插槽 支持 $\geq 4$ 个单宽 X16 插槽 支持 $\geq 1$ 个 OCP 3.0 网卡专用插槽 支持 $\geq 1$ 个阵列卡专用插槽
7.	GPU 计算模块	支持 $\geq 8$ 个双宽 GPU， 单模块 FP32 算力 $\geq 59.35$ TFLOPS，INT8 算力 $\geq 239$ TOPS，FP8 算力 $\geq 239$ TOPS，显存 $\geq 48GB$ ，显存位宽 $\geq 384\text{-bit}$
8.	风扇	支持 $\geq 8$ 个风扇，支持 N+1 热插拔冗余风扇
9.	网络	支持 $\geq 1/10/25/40/50/100/200GE$ PCIe 标准网卡
10.	电源	支持 $\geq 4$ 个 2000W/2400W 白金电源，支持 N+N 冗余
11.	配置要求	$\geq 2$ 颗处理器， $\geq 512GB$ DDR 内存， $\geq 5TB$ SSD 硬盘， $\geq 1$ 块 2GB 缓存阵列卡， $\geq 4$ 端口千兆网卡， $\geq 8$ 张 48G 缓存 GPU 计算模块， $\geq 4$ 个冗余电源，3 年原厂维保

## (6) 算力监控平台

序号	功能项	技术规格要求
1.	系统平台	算力监控平台采用 B/S 架构，系统平台和管理对象模块分离。
2.	管理平台	▲ 基于医保现有运维平台扩容算力监控模块，提供日维度的运维工作总览视图，在一张图上至少包含当前重要事件、近期重要事件、昨日设备性能变化汇总情况，提供相关证明材料。
3.	并发用户数	支持不少于 10 个并发用户同时登录系统并进行操作；
4.	分布式采集器	▲ 满足 1 个分布式采集器具备上千台设备的数据采集能力；采集器可被现有运维平台纳管，提供相关证明材料。
5.		支持跨广域网的分布式采集，单套产品至少具备支持 5 个及以上的分布式采集器；
6.	告警信息输出	平台支持屏幕颜色告警、告警邮件提醒、短信猫告警、企业微信告警及钉钉告警等方式。
7.		支持根据用户需求自定义外发告警信息格式，自定义内容包含但不限于：告警状态、告警项、告警时间、告警发生节点、告警项取值、告警模板名称及告警等级等。
8.	算力节点监控要求	支持对于主流国产操作系统和开源算力操作系统的管理，支持对于 CPU、内存、进程、日志、网络流量、磁盘性能的监控；
9.		支持对 GPU 计算模块运行情况进行监控；
10.		支持 IPMI 方式监控算力节点的硬件，监控内容包括温度、电源、风扇、电压等硬件状态；提供按照故障分类以及开机时间的分布查询，为全局快速分析硬件提供信息支持；

11.		提供算力节点可用率、性能报表、系统告警统计报表等多套基于采集和管理数据生成的客观统计报表
12.	算力业务运维管理要求	支持医保各大模型运行情况状态监控，提供延迟情况，请求吞吐量，Token 吞吐量等监控指标，并提供同期历史数据对比。
13.	售后服务	1 年原厂维保，提供售后服务承诺函
14.	配置要求	满足医保算力底座 $\geq 11$ 个算力节点监控授权， $\geq 58$ 个 GPU 计算模块监控授权， $\geq 2$ 个网络设备复杂信息监控授权，1 年原厂维保

## (7) 大模型管理平台

序号	功能项	技术规格要求
1.	基础服务管理	支持多元异构算力的纳管及调度； 支持高性能存储及高品质网络；
2.	数据服务管理	支持微调数据、回流数据的一站式端到端治理；基于 AI 语义理解实现数据内容优化、自动分类、自动生成、自动标注等；
3.	模型训练管理	支持各种资源管理功能； 支持预制算法； 支持智能调度、优化资源等；
4.	模型推理管理	支持多元模型推理功能； 推理服务全生命周期管理；
5.	模型评估管理	支持模型评估工具； 支持模型迭代优化等；
6.	模型资产管理	支持模型镜像、插件、知识库、提示词等统一管理；

7.	模型仓库管理	支持各模型版本迭代版本管理；
8.	服务要求	1年原厂维保，提供售后服务承诺函
9.	配置要求	配置模型资产管理和通用大模型模块仓库管理模块授权，1年原厂维保

#### (四) 与市医保信息平台双向同步要求

本项目系统运行所需的基础数据依赖市医保信息平台，投标人需结合市医保信息平台和本项目的实际情况，通过与市医保信息平台对接，形成完整、准确的数据双向同步机制，实时双向同步两定机构纳保数据。投标人需提供可行的技术方案，符合市医保信息平台技术标准的对接机制，涵盖两定机构纳保数据加密传输方案，并基于医保业务编码规范制定增量同步策略以及异常数据熔断回滚机制。

#### (五) 商业密码应用需求

为满足国密使用要求，投标人需对上海医保经办服务人工智能应用子系统提出详细、可行的改造方案，相关改造内容包括：网络和通信、设备和计算、应用和数据等层面的密码应用改造，对用户身份认证机制模块、业务重要数据安全传输模块、服务器虚拟机设备日志/访问控制信息完整性模块、重要可执行程序签名验签模块、用户访问控制信息签名验签模块、应用系统重要数据加解密模块、应用系统重要数据签名验签模块等开发改造适配。

#### (六) 项目部署要求

本项目大模型相关核心模块的建设部署依托医保智算中心完成。该智算中心作为项目核心基础设施支撑平台，包含各类大模型训练节点和

产品软件，构建软硬件协同适配的专业化部署环境。依托该部署架构，医保智算中心为项目大模型的稳定运行、知识库的安全存储与高效调用，以及智能体的协同交互等关键环节，提供了合规、可靠的底层技术支撑，确保相关核心模块在医保专属机房的专属部署环境下，实现安全可控、高效协同的运转效能。

相关配套业务系统依托上海市电子政务云建设，包括云服务器、存储、网络等基础设施由市电子政务云统一提供。系统需依据国产化要求实现，按照上海市电子政务云相关管理要求，以及项目安全及密码应用需要完成环境及应用部署。

投标人需要根据项目要求并结合当前本市医保经办服务人工智能应用相关业务量的估算对本项目所需要的资源进行设计，并配置配套数据库、操作系统、中间件以及密码测评所需安全认证网关服务、签名验签服务、可信密码服务、数字证书服务、时间戳服务。投标人需提供系统的详细部署方案。智算中心相关资源情况参考本节第三点“医保智算中心建设要求”，云资源相关情况如下：

序号	资源类型	资源名称	vCPU (核)	内存 (GB)	存储数量 (GB)
1	虚拟机	数据库服务器	64	256	6400
2	虚拟机	应用服务器	256	512	6400

### 三、项目实施、培训、售后服务等要求

#### (一) 系统软硬件产品供货要求

- 投标人提供的所有产品需具有在中国境内的合法使用权，且产品授权为本次项目招标人和最终用户。

2. 投标人提供的产品需是市场现有设备，招标人不接受厂商已停产或即将停产（自投标日起1年内）的设备。
3. 投标人需在合同签订后的45天内完成设备采购及到货。
4. 投标人提供的产品，应保证包装完好、数量与供货清单相符。
5. 投标人需将其供货设备运送到招标人指定现场，并提供完整清晰的供货清单，双方进行开箱点货验收。点货验收后，双方签字确认，货物的保管权移交给招标人，招标人在保管期内只对货物的数量和外观原状负责。
6. 投标人提供的产品若具有固有的设计缺陷，投标人应承担对该设备的修复和更换责任，且不以质保期为期限，直至问题得到彻底解决。

## **(二) 项目实施要求**

1. 投标人有固定的售后服务团队。
2. 投标人应充分理解本项目的建设要求及内容，提出完整且详细的项目实施、培训、管理、验收、售后服务方案等。
3. 投标人应根据对项目的理解做出项目的人员配置管理计划，包括组织结构、项目负责人、组成人员及分工职责。项目经理及项目成员的更换需书面通知招标人，经招标人同意后方可更换。
4. 招标人有权监督和管理此项目的开发、故障排错、测试及系统验收等各项工作，投标人应接受并服从招标人和监理方、测评方的监督和管理要求。
5. 本项目开发地点在上海，集中开发，中标人需承担开发设备环境、

开发人员的开发场地和食宿等相关内容和费用。

6. 投标人需提供针对本项目的风险控制方案。

### (三) 项目工期要求

投标人应根据建设内容，分阶段制定合理的时间进度，并且应根据招标人要求进行调整和细化。

本项目要求自合同签订之日起 10 个月内完成项目建设并通过竣工验收，需包含一个月试运行，要求提供详细可行的项目实施进度及人员安排。

### (四) 项目团队及驻场人员要求

1. 投标人需具有稳定的技术保障团队，能提供及时的技术支持服务，针对本项目投标人需提供不少于 40 人的项目服务团队（包括项目经理、技术负责人、研发等），其中至少 32 人驻场。投标人提供的服务人员需具备相应的服务能力，需提供相关证明。

角色	主要职责	人员数量	人员要求	驻场要求
项目经理	负责项目质量和进度控制	1 人	具备高级工程师职称（计算机、信息类专业）的；15 年以上从业经验，管理过不少于 5 个相关项目	驻场
技术负责人	负责技术架构和重大技术问题解决	1 人	具备系统分析师证书及信息系统项目管理师证书；10 年以上从业经验，在不少于 5 个相关项目担任技术负责人	驻场

研发	负责项目具体 开发与实施	至少 38 人	具备系统分析师、软件设计师、注册 信息安全专业人员（CISP）等证书	至少 30 人 驻场
----	-----------------	------------	---------------------------------------	------------------

2. 本项目质保期一年内投标人应提供不少于 10 人的驻场服务团队，其中项目经理 1 人，技术工程师不少于 9 人。

### **(五) 项目培训要求**

投标人应提供业务操作培训，应提供详细培训方案。

1. 项目验收前提供至少 1 次系统操作培训。在 12 个月的质量保证期内，至少提供 1 次与项目相关的必要培训。

2. 投标人需要开展分层次的人员培训工作，每次培训后应提供培训签到表、培训记录和培训报告。培训应具有培训教材、培训环境和高水平的培训讲师。

3. 投标人应提供对操作人员的基础操作培训和系统管理员的日常应用维护培训，确保培训对象能够掌握对应的操作技能。

### **(六) 项目验收要求**

本项目按下列方式开展验收。

1. 验收分初验和终验。

2. 初验前，投标人需完成软件开发和信息系统的调试等，并对本项目进行功能和运行检测，确保所有信息系统功能模块能够正常运行且已达到本项目约定的各类标准要求。投标人应以书面形式向招标方递交初验申请，招标方应当在接到通知后的 5 个工作日内确定初验的具体日期，由双方按照本项目的约定完成本项目的初验。招标方有权委托第三方检

测机构进行验收，对此投标人应当配合。

3. 初验时，投标人需提供软件文档（包括《需求规格说明书》、《概要设计说明书》、《详细设计说明书》、《三方功能需求确认单》、《测试报告》、《用户使用手册》、《系统部署手册》等）及可安装的程序运行文件。所交付的文档与文件应当是可供自然人阅读的书面和电子文档。项目通过招标方组织的初验评审，并且所有初验材料通过验收后，视为初验通过。如有缺陷，招标方应向投标人出具书面意见，陈述需要改进的缺陷。投标人应立即严格依照招标方的书面报告中的要求改进缺陷，并再次进行初验。

4. 自初验通过之日起，投标人应当按照招标方的要求提供现场技术支持服务，解决信息系统试运行期间可能出现的各类问题，或进一步提高与完善信息系统运行水平。

5. 初验通过且信息系统试运行期已经达到本项目约定的时间，经投标人确认信息系统具备正常运行条件，且信息系统通过运行测试，投标人应以书面形式通知招标方信息系统已准备就绪等待终验。招标方在收到终验通知后发起组织专家验收会。

6. 如果属于投标人原因致使本项目未能通过终验，投标人应当排除缺陷，直至本项目完全符合验收标准，由上述情形而产生的相关费用应由投标人自行承担。

7. 如果由于招标方原因导致本项目在终验通过前出现故障或问题，投标人应及时配合排除该方面的故障或问题。

8. 如本项目连续 3 次验收未通过（含初验未通过或终验未通过），

招标方有权解除项目，并有权依照本项目约定的违约条款追究投标人的违约责任。

## **(七) 售后服务要求**

1. 本项目整体从验收通过之日起 1 年内提供免费技术支持和售后服务。
2. 在质量保证期内，投标人将按照售后服务的承诺提供保修和运行维护服务。
3. 在质量保证期内，投标人负责信息系统的运行维护工作，确保信息系统安全、稳定、可靠地运行。本项目涉及的运行维护工作范围包括但不限于：远程技术支持、预防性维护、故障修复、重大活动现场保障、技术交流与培训、性能优化等。

## **(八) 应急响应要求**

投标人对系统故障应能够实时响应，若系统发生故障，在接到通知后 30 分钟之内响应，专业工程师 2 小时内到达现场。特殊故障与客户沟通协商后，按照协商的方式制定解决方案并进行处理。

具体故障级别及对应的应急响应要求如下：

一级故障：在 1 小时内确诊，总故障解决时间不超过 4 小时。

二级故障：在 2 小时内确诊，并在 4 小时内由专家到达现场确诊并解决，总故障解决时间不超过 8 小时；

三、四级故障：在 4 小时内确诊故障，总故障解决时间不超过 16 小时。

## **(九) 投标人综合能力要求**

投标人需具有 ISO/IEC27001 信息安全管理体系建设证书、ISO/IEC20000 信息技术服务管理体系认证证书、数据管理能力成熟度 DCMM 证书、CCRC 信息安全服务资质认证证书、ITSS 信息技术运行维护标准证书的优先考虑。

## **(十) 技术文件要求**

投标人提供的书面技术资料应能确保系统正常运行所需的管理、运营及维护有关的全套文件。技术文件应该全面、完整、详细。投标人提供的技术文件至少应包括：

1. 系统说明文件；
2. 技术手册(安装、测试、操作、维护、故障排除等)；
3. 项目文档，应该包括：
  - (1) 软件需求说明书
  - (2) 系统总体设计说明书
  - (3) 应用软件功能清单
  - (4) 安装配置报告
  - (5) 测试方案及报告
  - (6) 试运行方案及报告
  - (7) 验收方案及报告
4. 招标人提出的其他项目验收文档。

根据招标人项目验收要求，提供全套技术文件纸介质 1 套以及电子文件 1 套。

## (十一) 知识产权及保密要求

### 1. 知识产权要求

投标人拥有上海医保经办服务人工智能应用子系统第三方产品外所有工作件、交付品文档、定制软件、应用系统软件、数据资源及其他附属产品的知识产权，未经招标人许可，投标人不得用于第三方。

### 2. 保密要求

投标人同招标人签订保密协议，保证在此次项目建设中：由招标人向投标人提供的用户需求书、图纸、技术文档等所有资料，投标人获得后，应对其保密。除非招标人同意，投标人不得以任何形式向第三方透露或将其用于本次项目以外的任何用途。全部实施及验收工作完成后，所有资料应交回招标人。投标人驻场工作人员应签订保密承诺书，承诺内容需得到招标人认可，并与合同、保密协议一并提交给招标人。