

政府采购合同

甲方（需方）：河南省公平竞争审查事务中心

乙方（供方）：浪潮软件股份有限公司

依据 河南省公平竞争审查事务中心人工智能融合训推一体机采购项目 采购（招标/项目编号：豫财招标采购-2026-164）的招标采购结果，现依照《中华人民共和国民法典》及有关法律法规、规章规定的内容，为明确供、需双方责任，双方达成如下协议：

1、合同标的与价格

品名	制造商	规格型号	质保期	单价(元)	数量	金额(元)
深度学习验证平台（服务器）	北京神州鲲泰信息技术有限公司	KunTai A924	3年	895000.00	1	895000.00
深度学习验证平台（调试终端）	联想(北京)有限公司	ThinkStation P3 Tower、N2739Q（HF22270QK0）	3年	33000.00	2	66000.00
深度学习验证平台（可移动调试终端）	联想(北京)有限公司	ThinkPad T14p Gen3-041	3年	17000.00	2	34000.00
深度学习验证平台（软件）	浪潮软件股份有限公司	灵犀有言平台	1年	265000.00	1	265000.00
技术驻场服务	浪潮软件股份有限公司	定制	1年	0	2	0.00

合同总价（小写）：人民币 1,260,000.00

合同总价（大写）：人民币壹佰贰拾陆万元整

1.合同总价包括但不限于设备费、运至甲方指定地点的运输费、保险费、伴随服务费、安装调试费、质保期内的维修维护费（人为损坏的除外）、操作人员培训费、国家强制要求检验费用、税费等所产生的一切费用，甲方不再另行支付任何费用。

2.合同货物的技术参数等详见合同附件。



2、质量要求和技术标准

2.1 质量要求：乙方应保证所供货物是全新的，未使用过的，并必须达到或高于招标（谈判）要求及投标（报价）承诺。

2.2 技术标准：本合同货物以《附件：设备技术参数》为准，且符合产品说明所述的技术规格和标准。如果没有提及适用标准，则应符合货物来源国适用的国家标准，这些标准必须是有关机构发布的最新版本的标准。

2.3 乙方保证，其提供的合同货物具备相关资质证明（证书），并且满足国内行业管理的有关规定。

3、交货

3.1 交货方式：乙方负责送货到交货地点完成安装调试，并承担运输过程中发生的一切费用及风险。

3.2 交货期：合同签订后、接到甲方通知之日起 180 个日历日内完成交货及安装调试。

3.3 交货地点：甲方指定地点。

4、供货清单及包装、运输要求

4.1 供货清单：包括产品主机、随机备品备件、专用工具的名称及数量。

4.2 包装及运输要求：

4.2.1 乙方所提供的全部货物是厂家出厂的原包装。

4.2.2 乙方提供的全部货物须采用相应标准及保护措施进行包装，这种包装方式适用于相应的运输方式，并有良好的防潮、防震、防锈和防野蛮装卸等保护措施，以便保证货物安全运抵现场。货物在运输过程中所发生锈、损坏和丢失及其他任何损失由乙方承担责任和费用。

4.2.3 每件包装应附有详细装箱清单和质量合格证书。

5、验收

5.1 乙方按照招投标文件要求提供货物到达交货地点，完成安装、调试等工作且运维人员正式入驻后，由甲方组织项目验收，验收结果以甲方意见为准。

5.2 乙方应积极配合甲方建立确保货物安全运行的工作环境，并对完善相应的操作规范等工作制度提出专业性的意见和建议。

5.3 合同货物验收时，由甲方签署货物验收单。

5.4 乙方应派代表参与验收过程，乙方未派代表参与或对验收意见有异议但未在3个工作日内书面提出的，视为乙方对验收意见无异议。如乙方在验收完成后3个工作日内书面提出异议，以甲方委托的第三方验收意见为准。

5.5 验收合格后，乙方应在甲方要求的时间内直接交付甲方使用。合同货物交付使用前由乙方负责保管，合同货物的毁损或灭失风险由乙方承担。

5.6 甲方根据本合同约定提出换货、退货或解除合同的，乙方应在收到甲方通知后3个工作日内自行收回不符合合同约定的货物，并承担因退换货或解除合同所产生的一切费用。

6、售后服务

6.1 质保期为货物经验收合格之日起3年。质保期内，乙方向甲方提供3年软件系统免费升级服务。

6.2 故障响应时间：在质保期内接到甲方通知后，乙方需在1小时内到达，2小时内修复；2小时内无法修复的，乙方提供相应配置的代用设备或更换新设备，以保证甲方工作生产不中断，其中发生一切费用由乙方承担。特殊情况下，由乙方与甲方协商，并经甲方同意后在双方约定的时间内完成设备的修复或更换。

6.3 质保期结束后，乙方仍应负责提供终身维修服务，但只能收取零配件费，零配件价格不得高于市场同类产品价格。乙方保证能长期提供维修配件，具体的维修服务协议待质保期满后另行签订。

6.4 回访及不定期维修：乙方承诺对所有维修服务工作进行定期回访（3个月一次），乙方应每3个月向甲方提供维修服务，维修报告应包括每次维修或保养到场时间、维修持续时间、故障地方、更换的配件等，并接受甲方的监督和检查。甲方可根据合同货物的使用情况要求乙方在规定时间内免费为合同货物进行检修、日常维护及保养服务，以保证合同货物的长期正常使用。

6.5 技术培训：乙方应向甲方免费提供合同货物的操作使用及维护的培训，直至使用单位的技术人员能完全掌握设备操作技能。

6.6 技术资料：乙方应向甲方提供完整的中文技术资料，包括：产品验收标准，技术说明书，使用说明书，操作手册，设备安装调试材料，安装维修手册，

维修线路原理图及其维修资料，零部件目录，备品备件易耗件清单（含价格）及专用工具清单（如有的话），代理商与厂家之间的维保合同（如乙方为设备代理商）等文件资料。

7、付款条件与方式

7.1 付款条件

乙方同意，甲方按 7.2 约定的付款进度将相应款项汇入乙方指定账户。乙方账户信息如下：

户 名：浪潮软件股份有限公司。

开户银行：中国建设银行股份有限公司济南济东支行。

账 号：37050161555200000281。

7.2 付款方式

签订合同后付合同总价的 50%，到货后支付合同总价的 40%，验收完成后支付合同总价的 10%。到货后乙方需向甲方开具合同总价 10%的质量保函（有效期一年）。

甲方付款前，乙方应向甲方开具足额、合法、有效的 13%增值税专用发票。若乙方开具的发票不符合甲方要求，甲方有权拒绝或延迟付款，且无需承担任何责任。

8.知识产权

乙方须保障甲方在使用该货物或其任何一部分时不受到第三方关于侵犯专利权、商标权或工业设计权等知识产权的指控。如果任何第三方提出侵权指控与甲方无关，乙方须与第三方交涉并承担可能发生的全部责任与一切费用。如甲方因此而遭致损失的，乙方应赔偿该损失。

9、违约责任

9.1 乙方未能按时交货或未能按时交付使用的，每逾期一日，乙方应支付逾期交货货款1%违约金。逾期超过30个日历日，甲方有权单方解除本合同，乙方应另外支付合同总价5%的违约金。

9.2 合同货物验收不合格的，甲方有权选择解除合同或换货。如甲方选择换货，乙方重新供货导致交货延期的，按 9.1 条处理；若甲方选择单方解除合同的，乙方支付合同总价5%的违约金。

9.3 乙方提供的货物不符合 2.3 条规定,在合同货物最终验收合格前发现的,按 9.2 条处理;在合同货物最终验收合格后发现的,甲方有权退货,如甲方已支付货物价款,乙方应在甲方规定的时间内予以返还,此外,乙方应另外向甲方支付合同总价 5% 的违约金(如验收合格后发现货物不合格,由甲方委托的第三方鉴定确认)。

9.4 乙方的投标(报价)资料有弄虚作假、隐瞒事实内容等情形,在合同货物最终验收合格前发现的,按 9.2 条处理;在合同货物最终验收合格后发现的,甲方有权退货,如甲方已支付货物价款,乙方应在甲方规定的时间内予以返还,此外,乙方应另外向甲方支付合同总价 5% 的违约金。(如验收合格后发现货物不合格,由甲方委托的第三方鉴定确认)。

9.5 因乙方原因导致退换货的,乙方应承担退换货所需的一切费用。如乙方未在规定的时间内收回不合格的货物,甲方不对上述货物的灭失或损坏承担任何责任。如乙方逾期超过 30 个日历日仍未收回的,甲方有权自行处理上述货物。

9.6 质保期内,若乙方实际的维修响应(到达现场)时间不满足本合同要求的,每次应支付合同总价 1% 违约金,甲方有权另聘第三方对设备提供技术维修服务,由此产生的维修费用由乙方承担;乙方未按照本合同其他要求及投标(报价)文件售后服务承诺书的条款履行义务的,每次应支付合同总价 1% 违约金。

9.7 除本合同另有约定外,在补救违约而采取的任何其他措施未能实现的情况下,即在甲方发出违约通知后 10 个日历日内乙方仍未纠正其任何一种违约行为,甲方有权单方解除本合同,乙方除应退还甲方已支付的款项外,还应向甲方支付合同总价 5% 的违约金。

9.8 本合同约定的违约金无法弥补甲方损失的(损失包括但不限于诉讼费、保全费、律师费、鉴定费、执行费、直接损失、可期待利益损失等),乙方应继续承担相应的赔偿责任。甲方有权直接从未付的款项中扣除乙方根据本合同应付未付的违约金,赔偿金等。

10、不可抗力

10.1 因不可抗力造成违约的,遭受不可抗力一方应及时向对方通报不能履行或不能完全履行的理由,并在随后取得有关主管部门证明后的 15 个日历日

内向另一方提供不可抗力发生以及持续期间的充分证据。基于以上行为，允许不可抗力一方延期履行、部分履行或不履行合同，并根据情况可部分或全部免于承担违约责任。

10.2 本合同中的不可抗力指不能预见、不能避免并不能克服的客观情况，包括但不限于：自然灾害如地震、台风、洪水、火灾；政府行为、法律规定或其适用的变化或者其他任何无法预见、避免或者控制的事件。

10.3 当事人一方因不可抗力的原因不能履行合同的，应及时通知对方，以减轻可能给对方造成的损失，并应当在合理期限内提供证明。

11、保密及廉洁条款

11.1 保密条款：双方应对本协议的内容(包括补充协议)及在本协议的签订、履行过程中获悉的对方所有商业信息（秘密信息）和相关资料承担保密义务，未经对方的事先书面同意，不得向第三方透露或以履行本合同以外的目的使用相关秘密信息，造成损失的应向对方承担赔偿责任。

11.2 廉洁条款：双方员工不得以任何形式向对方相关人员提供回扣或返利。对于一方员工未经授权擅自向另一方做出的承诺，双方一概不予承认，由此造成的损失，由过错方自行承担。

12、合同的转让

乙方不得擅自部分或全部转让其应履行的合同义务。

13、合同纠纷处理方式及送达

13.1 因本合同或与本合同有关的一切事项发生争议，由双方友好协商解决。协商不成的，任何一方均可向甲方所在地有管辖权的人民法院提起诉讼。

13.2 本协议履行过程中发生任何纠纷，则任何一方当事人的工商登记公示地址视为司法机关邮寄法律文书及另一方邮寄催收函、律师函、通知等文件的送达地址。司法机关及本协议当事人向上述送达地址送达的，视为有效送达。本协议任何一方当事人送达地址需要变更的，应提前 3 天以书面形式告知协议其他方和司法机关（若争议已经进入司法程序解决）。因载明的地址有误或未及时告知变更后的地址，导致相关文书及诉讼文书未能实际被接收的，邮寄送达的，相关文书及诉讼文书退回之日即视为送达之日。

14、其他约定

14.1 磋商文件、投标文件和招标现场谈判补充的条款是本合同的有效组成部分，具有与本合同同等的法律效力。

14.2 上述条款如有未尽事宜，应经过双方协商一致后以书面补充，作为附件，具有与本合同同等的法律效力。

14.3 本合同一式柒份，甲方执伍份，乙方执贰份，具有同等法律效力。

14.4 本合同自双方签订并加盖公章之日起生效。

7
月
)
0

甲 方：河南省公平竞争审查事务中心

乙 方：浪潮软件股份有限公司

单位地址：

单位地址：山东省泰安市岱岳区东岳大街
527号

法定代表人：

法定代表人：

授权代表：

授权代表：霍苗苗

签订时间：

签订时间：

电 话：

电 话：

开户银行：

开户银行：中国建设银行股份有限公司济
南济南支行

账 号：

账 号：37050161555200000281

附件：

1、设备技术参数；

1、设备技术参数

设备名称	设备参数
深度学习验证平台（服务器）	<p>品牌型号：神州鲲泰 KunTai A924 一台</p> <p>(1)▲国产品牌,采用风冷散热方式,整机提供的稳定峰值 AI 总算力为 2.24PFLOPS@FP16; (已提供产品彩页和技术白皮书证明材料)</p> <p>(2)规格:标准机架式服务器,机箱高度为 4U;</p> <p>(3)▲处理器:单机配置 4 颗国产自主可控 ARM 架构鲲鹏 920 处理器,单颗 CPU 核数为 48 核,主频为 2.6GHz; (已提供产品彩页证明材料)</p> <p>(4)▲AI 芯片:配置 8 颗 AI 芯片,8 路 NPU 双向互联带宽为 392GB/s,整机 AI 芯片内存规格为 512GB HBM; (已提供技术白皮书证明材料)</p> <p>(5)内存:配置 1024GB DDR4 内存,内存频率为 3200MHz,最大支持 32 个内存插槽;</p> <p>(6)存储:本次配置 2 块 960GB SATA SSD 硬盘,2 块 3.84TB SATA SSD 硬盘;</p> <p>(7)RAID 卡:配置 1 块 RAID 标卡,缓存为 2GB,支持 RAID0/1/5/6/10,配置超级电容掉电保护;</p> <p>(8)网卡:配置 2 张双口 25GE 网卡,配置 4 块双口 200GE 光网卡;</p> <p>(9)电源:配置 4 块 2600W 热插拔交流电源,2+2 冗余备份,支持热插拔;</p> <p>(10)可维护性:设备集成服务器管理系统,支持远程监控图形界面,可实现与操作系统无关的远程对服务器的完全控制,包括远程的开机、关机、重启、虚拟软驱、虚拟光驱等操作;支持中文 BIOS 界面;支持国产主流操作系统;</p> <p>(11)产品资质:服务器产品具备中国国家强制性产品认证、中国节能产品认证、中国环境标志产品认证,已提供相关认证证书证明材料;</p> <p>(12)提供 3 年原厂维保服务。已提供原厂授权书和原厂三年质保函。</p> <p>详见 9.4.1.1 服务器配置参数。</p>
深度学习验证平台（调连试终端）	<p>品牌型号：联想 ThinkStation P3 Tower、联想 N2739Q（HF22270QK0）</p> <p>调试终端两台，每台具体参数如下：</p> <p>处理器：16 核 24 线程，基础频率 3.4GHz，最大睿频 5.4GHz，缓存 33MB</p> <p>内存：64GBDDR5 4800MHz 存储：1TBPCIe4.0NVMeSSD(读取 3500MB/s)+4TB7200rpmHDD</p> <p>显卡：16GBGDDR7 显存，CUDA 核心 8950，浮点性能 40TFLOPS，支持 PCIe5.0 和独显直连</p> <p>PCIe 插槽：1 个 PCIe4.0(x16)插槽、1 个 PCIe 4.0 插槽</p> <p>USB 接口：3 个 USB3.2Gen2 接口（含 Type-A 和 Type-C）、2 个 USB3.2Gen1 接口</p> <p>电源：750W80Plus 金牌认证</p> <p>屏幕尺寸：27 英寸，2K 分辨率（2560×1440），刷新率 210Hz</p> <p>屏幕面板：IPS，色域 99%sRGB，接口 HDMI2.0+DP1.2</p>

<p>深度学习验证平台（可移动调试终端）</p>	<p>品牌型号：联想 ThinkPad T14p Gen3-041 可移动调试终端两台，每台具体参数如下： 处理器：16 核（含性能核+能效核）16 线程，最大睿频 5.1GHz，NPU 算力 13TOPS 内存：32GB DDR5 5600MHz 存储：1TBPCIe4.0NVMeSSD 显卡：独立显卡，显存 8GB GDDR7，显卡功耗 55W 屏幕：14 英寸，分辨率 3072×1920（3K），刷新率 120Hz，色域覆盖率 100%DCI-P3，亮度 500nits 电池：容量 75Wh 散热系统：支持智能降噪散热技术、双出风口 其他要求：接口含两个 USB3.2、1 个 HDMI2.1、1 个 Type-C</p>
<p>深度学习验证平台（软件）</p>	<p>品牌型号：浪潮灵犀有言平台 一套 深度学习验证平台功能要求 1 知识库管理模块 知识管理是知识空间的核心领域，涵盖知识文件的新建、解析、内容摘要、问答生成、下载导出与检索、附件的上传与引用等环节。平台以多模态解析能力为基础，结合字段结构化管理与语义处理能力，实现知识内容的全生命周期数字化治理。 1.1 空间创建与管理 ▲知识空间可设置空间名称、AI 生成空间描述信息、封面、可见范围、空间管理员、空间知识维护人员。（提供界面截图并加盖软件开发商公章） ▲空间及目录权限配置支持组织权限分配、角色权限分配、人员组权限分配的功能。（提供已通过“具备 CMA 或 CNAS 认证标志的第三方检测机构”的检测报告） ▲每个知识空间下支持构建独立目录树，允许用户创建多级目录结构，并可根据业务需求对目录进行重命名、顺序调整、位置移动、逻辑归并与删除等操作。（提供界面截图并加盖软件开发商公章） 目录支持层级化管理，可对任一层级目录单独设置可见范围与知识官，并支持目录下快速新建或接入知识文件。通过灵活的目录划分机制，用户可对复杂的知识体系实现清晰化组织，提升内容的查找效率与维护便利性。 1.2 知识导入 支持原始内容的导入、解析工具选择、导入预览等关键环节。 1.3 数据清洗 内置数据清洗模块，自动处理知识中的 HTML 格式符号、无效字符、页眉页脚等信息，将清洗的后的知识存储入库。 1.4 知识管理 1.4.1 知识新建 平台支持文档、表格、图片、视频、音频等多类型知识文件的上传，支持多种解析工具，对于多模态文档，可预览文件分段情况；对于半结构化数据，支持配置字段映射。 1.4.2 知识详情 用户可在知识文件详情页查看原文件、文档解析内容、知识摘要与元数据等。 1.4.3 解析块维护 平台支持对解析工具生成的解析块进行预览，支持按索引快速跳转解析块，支持编辑解析块内容、上传或引用附件、添加标签等，确保知识内容具备明确边界与统一表达，利于智能服务的高效调用。 1.4.4 知识标注维护</p>

为知识文件配置标签信息，用于增强知识文件的语义标注能力。系统支持标签的创建、删除操作。标签体系与目录结构互为补充，构成平台双重知识组织结构。

1.4.5 问答萃取

平台集成问答生成模型，可对结构化内容或段落内容进行自动问答对抽取，形成标准化的“问题-答案”知识单元。用户可对生成的问答对进行重新生成、采用、编辑和删除操作。

1.4.6 内容摘要

知识文件支持自动生成段落摘要，系统结合结构识别模型与自然语言生成模型生成高质量摘要信息。

1.5 附件管理

支持为知识资产挂载多类型附件，包括原文 PDF、扫描件、参考图、视频片段、配套表格等，所有素材具备统一预览功能

1.6 知识提取

▲ 支持批量上传多格式文件，包括.pdf、.md、.xlsx、.csv、.docx、.txt、.mp3、.mp4、.wmv、.avi、.png、.jpg、.jpeg、.wps、.et、.ett、.pptx、.ofd 格式，用户可通过本地上传、服务器直传等方式将内容接入平台。（提供已通过“具备 CMA 或 CNAS 认证标志的第三方检测机构”的检测报告）

▲支持结构化数据库或表格数据的直连导入，面向 Excel、CSV、本地数据库，通过统一导入流程完成字段映射与内容解析。（提供界面截图并加盖软件开发商公章）

1.7 多模态知识管理能力

▲按多模态数据上传、多模态数据自动切片、多模态数据切片预览一体化构建流程；多模态数据上传包含图、文、音、视四种模态数据的上传；多模态数据自动切片包含图、文、音、视四种模态数据上传后通过多模态大模型自动提取多模态数据知识并按知识内在逻辑进行切分；多模态数据切片预览在数据自动切片后可查看切片后数据情况。（提供界面截图并加盖软件开发商公章）

1.8 知识问答与多模态附件管理能力

▲针对上传的多模态数据进行附件补充后进行知识问答；附件补充功能需包含针对切片后的数据进行图、文、音、视多模态附件补充附件个数；知识问答可选择特定模型问答、可选择制定知识空间进行问答、针对单次上传的知识进行问答、针对全部知识空间进行问答、流式输出的答案中展示针对某条知识补充的多模态附件数据且可预览。（提供界面截图并加盖软件开发商公章）

1.9 多级用户系统管理能力

▲具备超级管理员、首席知识官、知识官、人员组多级管理能力。（提供已通过“具备 CMA 或 CNAS 认证标志的第三方检测机构”的检测报告）

1.10 知识空间配置

支持模型配置、存储配置，统一管理此空间内的模型资源。

▲(1)在启动 4 张算力卡（单卡 280TFlops,24GB）情况下，100 并发下知识问答首词元时间，平均响应时间小于 2 秒”的性能指标要求。（提供已通过“具备 CMA 或 CNAS 认证标志的第三方检测机构”的检测报告）

▲(2)在启动 4 张算力卡（单卡 280TFlops,24GB）情况下，100 并发对知识问答词元间隔下平均词元间隔响应时间小于 0.5 秒。（提供已通过“具备 CMA 或 CNAS 认证标志的第三方检测机构”的检测报告）

▲（3）支持向量库混合查询，应支持常用的运算符，包括比较运算符（=,!=,>,>=,<,<=）和逻辑运算符（与、或、非）等。（提供已通过“具备 CMA 或 CNAS 认证标志的第三方检测机构”的检测报告）。

1.11 角色体系定义

平台预设四类核心角色：首席知识官（CKO）、知识官（KO）、普通用户与平台管理员，各类角色权限边界清晰、互不重叠。CKO 具备全局资源管理权限，可创建空间、新建知识官、监管内容质量；KO 拥有分空间管理权限，承担知识资产的维护与解析职责；普通用户以阅读、搜索和问答应用为主，具备基础使用权限；平台管理员主要负责系统层级配置与账号安全治理。角色体系支持扩展，便于适配不同组织结构与知识治理职责分配方案。包括可见范围配置和集成策略功能。

2 模型训练平台

模型训练平台包括全局看板、模型编排、模型微调训练、模型部署、模型评测、应用管理和算力资源管理模块。

2.1 全局看板

支持为用户提供训练平台的全局看板，方便用户了解平台内部的各项功能的运行状况。支持对通知、应用统计、模型统计、训练概览、算力统计、节点概览等模块的运行情况进行展示。

2.2 模型编排

模型编排支持提供处理模型运行前置、后置工作的机制及其快捷入口，支持输入参数控制、Python 脚本支持、输出内容控制等功能，以便于让模型更好的适应各种运行场景。

2.2.1 输入参数控制功能

模型编排需具备强大且灵活的输入参数控制能力。可通过可视化界面，对模型运行所需的各类输入参数进行精确设置。支持数值型参数的范围限定，避免因参数取值不当导致模型运行异常；同时，系统还支持参数的分组管理与批量设置，方便用户在处理复杂模型或多个相似任务时，高效配置参数，大幅提升工作效率。

2.2.2 Python 脚本支持功能

模型编排深度集成 Python 脚本功能。用户可直接在系统提供的脚本编辑区域编写 Python 代码，实现对模型运行流程的个性化控制。支持对输入数据进行复杂的预处理逻辑；支持在模型运行过程中添加自定义的中间计算步骤；支持对输出结果进行特定规则的加工处理。

2.2.3 输出内容控制功能

在输出内容控制方面，模型编排需赋予用户充分的自主性。用户可以根据实际业务需求，灵活选择输出内容的格式，如常见的 JSON、CSV 等格式。同时，能够对输出内容进行筛选和过滤，只保留关键信息，去除冗余数据，使输出结果更加简洁、精准。

2.3 模型微调训练

模型微调训练应支持通过选定可用的算力、训练数据、基座模型进行微调训练任务的设定。需支持通过获取微调训练数据，选择训练算力，使用基础模型，微调得到专用任务模型。通过模型微调训练模块可以清晰的查看训练进度、训练剩余时间等情况。

2.3.1 微调训练任务设定

可基于训练任务的规模、复杂度及时间要求，从平台的 CPU、GPU 算力资源池中，便捷选定合适的算力资源。支持选用通用的基座模型开启训练，支持基于往期微调训练过的模型继续优化。系统提供模型对比分析工具，从模型结构、性能指标、适用场景等维度进行评估，辅助用户挑选最契合任务的模型。

支持对微调参数进行配置，包括训练轮次（epoch）、训练批次（per_device_train_batch_size）、梯度下降批次（gradient_accumulation_steps）的设置。

2.3.2 训练微调过程管理与展示

需支持以可视化界面实时呈现训练进度。用户可清晰掌握当前训练所处阶段，如剩余训练时长等信息。

2.4 模型部署

模型部署应支持将训练好的模型开放给业务系统，包括部署任务创建、部署模型选择、部署算力选择。整个过程可视化的方式进行配置，部署成功后可进行可视化测试。

2.4.1 部署算力选择

平台需支持多样化的算力资源供用户在模型部署时选择。用户可根据模型的规模、复杂度以及预期的推理请求量，灵活挑选合适的 CPU 或 GPU 算力。

2.4.2 部署模型选择

在模型选择环节，平台需提供便捷的操作界面，用户可以轻松浏览已训练完成的模型列表。用户可根据业务需求，选择适合部署的模型。

2.4.3 模型发布

当完成算力和模型的选择后，用户可通过简单的操作步骤实现模型发布。平台提供一键式发布功能，系统会自动完成模型格式转换、依赖环境配置等一系列准备工作，确保模型能够在目标环境中正常运行。

2.4.4 API 调用

模型部署完成后，需支持通过平台开放的 API 便捷地调用模型的推理能力。API 接口设计遵循标准化规范，具备高兼容性和稳定性，支持多种常见的编程语言。使用方只需按照接口文档提供的参数格式和调用方式，传入待推理的数据，即可快速获取模型的推理结果。

2.5 模型评测

模型评测应支持不同类型的模型，采用不同指标进行测评，测评结果应支持反馈至前端页面，为模型优化与选择提供强有力的数据支持，进一步推动智能化水平的提升。模型训练完毕后需要对训练好的模型使用评测数据进行指标评测，如精确率、召回率、F1 指标等。通过指标与业务系统要求进一步对模型的训练提出指导。

2.5.1 文本分类模型评测

对于文本分类模型，重点关注精确率、召回率、F1 值等核心指标。精确率体现模型预测为正类样本中真正正类的比例，召回率衡量实际正类样本被正确预测的比例，F1 值则综合两者反映模型的平衡性能。

2.5.2 命名实体识别模型评测

在命名实体识别模型评测中，以实体识别准确率、召回率和 F1 值为关键指标。不仅关注模型识别出的实体是否准确，还考量是否完整识别出文本中的所有实体。

2.6 应用管理

应支持展示已有的应用数量（总量、较上月增长量、调用量）、应用调用总量排行。并且应支持通过网关模块，通过颁发认证密钥的方式来控制外部对运行模型的访问控制。

2.7 算力资源管理

提供算力资源管理，支持 CPU、GPU 等不同类型算力节点管理、算力资源管理。

3 智能体平台

3.1 创建应用

支持在应用工作室中创建应用；支持通过应用模板创建，支持导入文件或者直接创建空白应用；支持基础编排和工作流编排两种方式。

3.2 流程编排

提供流程编排，支持多种类型能力节点编排接入。

3.3 应用调试和预览

支持通过应用的调试和预览，在工作流编排页面，输入测试的问题，进行过程化调试。输入运行后，整个的运行过程应支持在流程编排页面进行展示，并应支持查看当前问题行走的路径、节点。

3.4 应用发布

	<p>经过编排和预览后的应用，应支持对外发布功能。发布后的应用应支持独立运行提供服务、嵌入到第三方网页提供服务、独立接口形式提供服务。</p> <p>3.5 知识集创建</p> <p>应支持知识集的创建，也应支持通过接口调用的形式创建知识集。知识集应支持名称、描述、索引模式、检索设置、文件分段等设置。索引模式应支持高质量和经济两种模式；检索模式应支持倒排索引、向量召回、混合索引三种模式，用于不同的场景；文件分段设置应支持分段的分隔符配置、每段的最大长度设置等。</p> <p>3.6 知识库接口</p> <p>应提供文本创建文档、通过文件创建文档、创建知识空间、获取知识库列表等接口，应支持第三方系统通过获取访问密钥进行接口调用。</p> <p>3.7 知识库应用</p> <p>知识库的数据应支持通过应用 workflow 编排中的知识检索节点来进行调用。新建知识检索节点后，选择一个或者多个知识库进行召回。随后将知识库中召回的知识作为模型的入参输入模型，提高模型输出的专业性和准确性。</p>
<p>技术驻场服务</p>	<p>驻场保障：提供 2 人 1 年的技术支撑（1 人驻场、1 人远程），工程师配合相关工作，遵守招标方规章制度，签订保密协议，保障服务响应及时。</p> <p>专项培训：开展智能体编排、训练平台使用培训，覆盖操作流程、基础配置等核心内容，提供实操指导和问题答疑，助力相关人员快速上手。</p> <p>知识库支持：提供知识库系统知识编目技术支持，协助梳理编目规则、解决编目过程中的技术问题，保障知识入库规范高效。</p> <p>智能体研究配合：结合河南省市场监督管理局现有业务场景，协助开展智能体的相关研究工作。</p>