第五部分 采购内容要求

商务要求

1、项目成果交付期：乙方应在接到甲方通知后120个日历日内完成服务内容并通过甲方的检验。

2、质保期：自终验合格之日起，软件系统免费升级维护5年，硬件设备质保3年。

3、付款方式：

3-1合同签订后60个日历天内，甲方支付合同总价的40%，作为预付款项；乙方完成全部服务及相关配套内容并达到约定，经甲方初验合格后60个日历天内，甲方支付合同总价的30% ；项目经甲方终验合格后60个日历天内支付合同总价的30%。

3-2支付方式：银行转帐。

3-3结算方式：成交单位开具合同全额发票给采购人。

3-4履约保证金：合同签订前中标单位应向甲方缴纳合同价款的5%作为履约保证金，在乙方全部履行完毕合同约定的服务及其他义务，且经甲方使用单位确认乙方服务合格或对服务质量无争议后，甲方一次性无息退还乙方。

4、项目验收要求：

4-1初验：所有硬件到位，语言智能与语言能力测试平台基础版（含6个主要模块的开发）研发测试完成，可申请初验，按照软件功能验收表进行功能测试；功能全部实现可视为初验通过。

4-2终验：所有功能及技术参数实现，软件系统试运行30日后，提供完整功能版本；提供相关接口说明、用户运维手册、培训文档；完成所有交付清单内容，完成系统培训，可申请终验。终验验收所产生的验收费由中标单位承担。

5、售后服务要求

5-1须有售后服务承诺,规定明确的服务范围和内容。

5-2售后服务从系统软件终验通过之日起计算。

5-3免费维护期内采用电话、eMail、远程、现场技术支持等方式解决日常维护中的问题，包括性能维护、操作使用、技术讲解和问题排故等。免费维护期内对影响生产的软件故障应提供5\*8小时，2小时内响应，24小时内提供解决方案。免费维护期外应提供运行、功能故障的解决。

5-4在免费维护期内，在面临重要日期、重大事件和严重问题时，必须采用现场技术支援的形式到场服务。

5-5在质保期内，为学校免费提供各种技术服务，包括在线答疑、定期回访、适应性修改、少量增量性修改、版本升级与技术培训。

5-6提供完善的产品开发、管理、运维、应急等文档资料。

5-7质保期内，中标方应免费进行系统跨版本升级，保障产品最新功能与中标方官方网站发布内容一致；免费进行系统维护，负责保障系统安全运行。

5-8培训服务应包括产品使用操作、简单维护等培训内容。中标方需按采购人要求的时间为采购人免费培训技术人员若干名，培训服务以受培训人员熟练掌握相应技能为原则。在产品投入使用初期进行必要的跟踪指导，保障产品的稳定运行。软件产品需在甲方现场培训的，供应商应按要求履行，供应商培训人员在培训过程产生的交通费、食宿费、培训费等均由供应商承担。

二、采购内容

1、本平台主要面向三类用户，包括语言学习者、教授者及研究者：

（1）语言学习者（学生）：交互式语言学习，智能语言百科等。

（2）语言教授者（教师）：翻译教学实例数据生成（课堂内容扩展），口语语音评测，翻译能力评测，评测数据分析。

（3）语言研究者：多模态数据导出（文本、图像及语音），教学科研数据支持，语言学习者数据获取，评测数据分析。

2、主要基本模块介绍如下：

（1）机器翻译-为翻译教学提供翻译实例，延伸教学内容。



图：机器翻译模块原理示意图.

2）智能语言百科-语言学百科全书，为用户提供相关知识快速查询接口



图4：智能语言百科原理示意图.

3）自动摘要-为翻译学，跨语言、跨文化交流等提供辅助工具



图5：自动摘要模块原理示意图.

4）语音识别-为口语评测及语言学习、语言研究提供支撑及评价依据



图6：语音识别模块原理示意图.

5）语音翻译-为口语学习、翻译实例学习提供接口



图7：语音翻译模块原理示意图.

6）图文转换-为语言学习、翻译学习提供辅助手段

图

8：图文转换模块原理示意图.

系统集成任务

（1）★中标方须按照学校要求提供所需接口并永久免费开放所有接口。

（2） a) 要求所有系统实现与学校统一身份认证对接，实现单点登录。

b) 要求按照学校的统一要求和标准规范免费完成与我校各公共平台的深度对接集成和适配工作，包括但不限于于我校移动校园、掌上西外、网上办事大厅、统一消息中心、统一短信网关的数据集成、消息集成、服务集成、应用集成，实现统一信息管理。

c) 本项目内涉及到的移动端应用均要求采用H5开发，并集成到学校移动校园微信企业号、掌上西外app和小程序，H5页面需遵循学校移动应用视觉设计规范和交互设计规范。

（3）★要求完成与学校公共数据中心的数据采集集成，包括系统的所有结果数据，提供系统数据库的只读账号、数据字典、代码表、API 接口、集成文档等相关文档。

（4）★本项目涉及到与学校相关业务系统的应用集成和数据对接，如涉及费用，均由中标方承担。

（5）将提供本项目正常运转所必须的全部系统软件、中间件、插件等，确保系统正常稳定运行。

（6）提供本项目正常运转所必须的全部系统软件、中间件、插件等基础环境，确保系统正常稳定运行。

四、安全性要求

为保证本系统在使用过程中的安全性，具体实现要求参考如下：

(1)实现https安全传输协议，实现兼容TLSv1.2以上版本使用。

(2)中标单位与学校签署保密协议，完成网络安全等级保护测评（二级）工作，并协助学校完成等级保护测评备案，费用由中标单位承担。

(3)系统具备数据保密机制，数据存储经过密码保护，针对项目中的数据信息安全等内容会签署保密协议，防止信息泄露。

(4)系统具有数据访问日志和数据审计策略，系统日志实现留存6个月以上。

(5)系统具备数据的本地备份机制，每日备份至本地；系统中存在的核心关键数据，具备异地数据备份功能，可备份至指定的服务器或存储系统上；每周制作数据的备份在指定服务器，确保系统一旦发生故障时能够快速恢复，备份数据不得更改；业务数据实现定期、完整、真实、准确地转储到不可更改的介质上，并实现集中保存，保存期限2年。

(6)提供数据对接、提供和第三方数据交换、集成规范和标准、开放并提供数据接口。

(7)中标单位确保系统运行环境的安全运维和防护。

(8)平台数据来源保证合法合规。

五、技术要求

1.★自主可控的国产软件，翻译引擎采用国产自主研发的深度学习平台。

2.系统为B/S架构，实现系统部署于服务器端，无需安装客户端，用户只需通过浏览器即可在线访问系统，即可执行翻译项目、术语库、语料库、人员管理等操作，不限制访问设备的类型。

3.系统实现市面主流浏览器登录，包括谷歌Chrome、360浏览器、IE浏览器等。

4.SDK：实现多种平台的终端接入：Android、iOS、Windows、Linux、WebSDK、WebAPI、微信小程序。

5.Cloud Services：语音评测云平台，包含负载均衡服务、授权服务、评测计算服务集群、存储服务等。

六、功能要求

（一） 翻译引擎

1.★实现英中、中英翻译，可局域网部署的机器翻译引擎。

2.实现翻译引擎再训练能力，保证翻译效果持续更新。

3.翻译引擎采用国产自主研发的深度学习平台，非采用国外开源深度学习平台。

4.★实现基于对齐语料，可选择训练集、测试集和干预集，一键进行训练学习。

5.自动评估，针对模型训练由机器给出评估BLEU值。

6.可以看到过往训练的版本，查看训练日志、部署相应版本。

（二）文本和文档翻译

1.在源语端输入文本，目标语端自动输出译文、具有原文语种自动识别、原文清空、译文复制、双语高亮等功能，具备每次翻译5000字符能力。

2.★ 具备术语干预和记忆库匹配。

3.具备在翻译界面将术语和翻译句对添加到术语库或记忆库。

4.格式要求：pdf，docx、txt等。

5.翻译前可指定术语库、记忆库配置。

6.下载要求：实现多种格式选择下载，如双语对照格式、docx格式下载等。

7.实现原文和译文对照预览。

8.★ 实现在预览同时进行校对。

（三）语音翻译

1.实现中文和英文的流式语音相互翻译。

2.语音识别准确率：中文≥95%，英文≥90%。

3.文本翻译准确率：中译英≥90%，英译中≥90%。

4.★中英互译端到端准确率达80%。

5.实现流式语音翻译，实时语音翻译api采用websocket协议的连接方式，接收实时的音频流，输出带有标点的识别结果以及翻译结果。实现pcm格式的原始音频数据，支持但不限于16000采样率，单声道，16bits，小端序。字符编码统一使用utf-8编码。

6.支持 pcm s16le 16k, pcm s16le 8k, alaw 16k, alaw 8k, ulaw 16k, ulaw 8k, vok 8k, vox 6k 等采样率和格式。

（四）智能语言百科

1.多渠道接入：实现与网站、微信公众号、企业号渠道、APP、小程序、实体终端等多渠道接入方式。

2.★智能交互:

1）基础寒暄。具备全面的自然语言库，能够回应对话过程中的的基础寒暄常用语。

2)热门问题推送。在用户发起会话的同时，实现推送热门问题列表，引导用户点选。

3)上下文语义理解。系统应具有上下文语义理解能力，师生提问时缺省关键信息，系统可以自动分析关联上文内容，定位用户需求，给出准确回复。

4)动态问答交互。同一个问题，实现根据用户提供的信息（教职工，学生等），回复与其信息相匹配的不同答案。

5)推荐回复。根据师生提问关键词，智能给用户推荐给用户相关性最大的问题列表。

6)富文本式回复。智能客服机器人的回复内容实现文字、图片、超链接、音频、视频、文件、富文本等多种形式。

7)用户反馈。在答复师生提问后，用户可以选答案“有用”或“无用”。

8）问题解决调查。在对话结束后，系统自动发起问题是否解决的调查。

9）用户评价。用户可以主动对整个会话解决过程的满意度进行评价。

3.★知识库管理平台：系统实现知识库管理维护工具，实现知识库分类、分级、分权限管理。业务知识库包含单轮问答库、动态问答库和多轮问答库，根据不同场景灵活使用。知识管理实现一点维护和多点发布，具备多渠道集中管理功能。

知识维护。提供业务知识库的增删改查和导入导出及实效管理；知识库导入时，系统实现重复内容检测。

知识库内容管理。知识库可以逐条添加，也可批量导入，维护简单方便。批量导入时，系统内置导入模板。回答内容素材实现图片、文档、音频、视频、链接、接口调用等。

知识展示。实现不同类型的问题分类展示，可显示知识点的使用频次。

知识点应用。对同一个问题，可自定义生效渠道、生效时间，实现自定义回复答案，同时可关联多个问题，也可关联相关知识链接。

未识别问题管理。对于机器人无法回复的问题，系统可以统一聚类管理和查看，实现通过问题追溯到原聊天记录；在处理新问题时，系统具备相似知识点智能推荐功能。

4.训练学习平台

系统实现自训练，用户可手动添加样本训练，提升识别能力。

实现统一收集未识别问题，可进行问题忽略、添加为新问题、学习到已有问题、追溯聊天记录进行人工回复等操作。

训练后，可随时测试训练效果，且对错误的答案，实现随时纠正。

实现已训练问题、未训练问题、已训练样本、待训练样本数量的查看。

5.智能统计：实现多维度统计，实现多种方式的查询、导出，能够为不断完善系统和优化决策提供充分可量化的直观数据和统计报表。

总体数据概况。实现按时间维度查看各渠道的总体数据情况，包含机器人会话量、人工会话量、满意度评价比、系统消息收发总量的数据情况，可直观查看会话量趋势，并实现数据导出。

访客分析。实现按时间维度查看各渠道的用户访问情况分析，包含地域来源分析、访问渠道占比分析、首次访问占比分析。

热点咨询统计。实现按时间维度对用户的咨询内容进行分析，包含热词排行统计、各部门知识库热门业务排行统计、热门问题排行统计，实现统计报表的可视化呈现及统计数据的导出。

机器人回复效果统计。实现按时间维度对各渠道师生的提问效果进行分析，包含问题准确匹配率、满意度评价，实现统计报表的可视化呈现及统计数据的导出。

人工工作量统计。每个系统维护人员处理未识别问题总数需清晰可见；实现按时间维度对各维护人员的工作质量情况进行分析，包含在线时长、回复问题条数、回复字数、人工回复师生次数、满意度、一次性问题解决率等不少于5个维度的数据呈现，并实现统计数据的导出。

6.操作日志

实现所有会话记录永久留存，至少可以一次性筛选出5个月的会话记录并导出会话内容。

实现操作日志记录，实现对所有管理账号登录、知识库增删改查、未识别问题处理等操作行为进行日志记录。

7.性能

知识库容量不少于1000条。

年度会话调用量不少于10万次。

网络情况良好情况下，会话响应时间小于1秒，支持每秒不少于500人次同时访问对话。

（五）文本摘要

结合传统语义特征和深度学习模型，充分考虑段落分布和篇章结构，准确计算新闻语句的重要性，对文本内容进行全面的语义理解与分析。

★能根据需求灵活控制摘要长度，自动抽取关键信息，形成摘要结果。可用于内容理解、内容分发、智能写作等多种应用，为翻译学，跨语言、跨文化交流等提供辅助工具。

实现将文本按句切分，实现分词、词性分析和词频统计。

（六）口语评测

★英文单词评测：总分、音标得分、实现音节重音检测、实现传音标评分、实现英美式发音区分、实现松紧度调节。

英文单词纠音评测：总分、检测音素级漏读、多读、错读、识别学生发音音素、实现传音标评分、实现英美式发音区分、实现松紧度调节。

★英文句子评测:总分、流利度、完整度、准确度、韵律度、各单词发音得分、各音节发音得分、检测重读、停顿、句末升降调、实现失去爆破、连读检测、实现传音标评分、实现英美式发音区分、实现松紧度调节。

英文句子纠音评测:总分、检测单词级漏读、多读、错读、检测重读、停顿、句末升降调、识别学生发音单词、实现传音标评分、实现英美式发音区分、实现松紧度调节。

★英文段落评测：总分、流利度、完整度、准确度、韵律度、各句子发音得分、各单词发音得分、检测重读、句末升降调、检测单词级漏读、多读、错读、实现传音标评分、实现英美式发音区分、实现松紧度调节。

英文AI Talk：识别分支、置信度得分、流利度、完整度、准确度、实现关键词检测、实现传音标评分、实现英美式发音区分。

英文口语选择评测：总分、发音得分、流利度得分、实现单选、多选。

英文开放题型评测：总分、流利度、语法得分、内容得分、发音得分、实现关键词命中。

英文自由识别评测：总分、流利度、各单词发音得分。

（七）图文转换

图文转换模块具有物体识别、场景名称、通用文字识别、图片文字识别、平台还可提供定制图像分类、物体检测等模型，自定义识别标签，实现定制物体和场景识别等功能，实现获取图片识别结果对应的百科信息，接口返回百科词条URL、图片和摘要描述，可选择是否需要返回百科信息。具体特性如下：

识别动物、植物、商品、建筑、风景、动漫、食材、公众人物等10万个常见物体及场景，接口返回大类及细分类的名称结果。

可提供基于图像内容安全的色情识别、违禁识别、公众人物识别、恶心图像识别。

可对接百度百科，实现获取图片识别结果对应的百科信息，接口返回百科词条URL、图片和摘要描述，可选择是否需要返回百科信息。

可根据用户拍摄照片，识别图片中物体名称及百科信息，提高用户交互体验，广泛应用于智能手机厂商、拍照识图及科普类app中。

支持不少于1000QPS并发量。

实现识别引擎增量训练，保证识别转换效果持续更新。

识别引擎采用国产自主研发的深度学习平台，非采用国外开源深度学习平台。

具备识别术语干预和记忆库匹配。

下载要求：实现多种格式选择下载，如双语对照格式、docx格式下载等。

 (八）API接口

文本翻译API接口：基于HTTP协议的翻译API接口，可以便捷的集成嵌入到第三方的业务平台或其他应用系统中。

语种识别API接口：自动识别原文语言类型。

文档翻译API接口翻译：文本翻译，译文中可以保留格式。

术语API接口：提供用户术语词典操作的相关API接口，保证机器翻译译文中术语翻译的准确性和一致性。

翻译记忆API接口：提供用户翻译记忆操作的相关接口，通过API调用将人工翻译的正确译文实时添加到翻译系统中的翻译记忆库中，避免机器翻译重复犯错。

★增量训练API接口：便于和现有系统对接开发，提供训练能力。

七、项目成果交付清单

（1）语言智能与语言能力测试平台1套。

（2）文本语料库（单语/双语）-双语：不少于1,000万句对；单语：不少于1亿词次。

（3）语音语料库-不少于300小时\*2种语言。

（4）可执行代码和源代码：保证系统正常运行的所有可执行代码，以及项目开发过程中为用户方定制功能的源代码。

（5）在本项目的开发过程中和交付使用后，交付各个阶段的成果和文档资料并提供明确的交付清单。成果和文档资料必须符合软件工程的相关要求。要交付的成果和文档资料主要包括但不限于以下部分：

提供《信息系统安全等级保护项目计划书》、《系统等级测评方案》、《系统等级测评报告》、《信息系统安全等级保护定级备案证明》和《信息系统整改建议书》。

（6）项目文档：项目开发中的各种技术文档，包括用户手册、管理员手册、开发环境配置说明、需求分析说明、方案设计、参数设定、软件说明书、数据库结构、数据结构（数据字典）、代码清单、软件工具清单、需求及实施变更说明、需求跟踪矩阵、API接口手册以及有关系统集成接口的技术说明文档、元数据模型说明、系统设计说明、系统维护说明、系统培训资料，系统测试报告、试运行报告等。

（7）管理文档：包括项目开发中的一些工作文档，如实施计划、阶段报告、讨论纲要、会议记录，培训资料、竣工文档（含项目实施总结、系统配置情况等）、验收测试报告（测试计划、测试用例和测试结果等）等。

（8）提供项目实施过程中的过程记录，如文档管理、版本管理、变更记录等。

（9）项目完工后提供验收报告、实施方案、进度安排、项目周报等。交付的所有成果应汇集成册交付给学校，并提供成果的电子化版本。

八、技术支持与服务

（1） 符合国家有关规定：系统建设必须符合我国相关部门制订的标准，安全策略、密码与安全设备选用、网络互联、安全管理等方面必须符合我国信息安全法律法规。

（2） 统一信息标准：实现学校智慧校园统筹规划的应用需求和未来发展，符合并遵守学校制定的教育信息化技术规范、软件设计与开发规范、软件设计开发标准等标准规范要求，系统的设计和开发遵循学校统一的信息标准。

（3） 集成对接要求：与学校已建成或待建的公共数据中心、统一身份认证、移动校园等平台对接集成；提供和第三方数据交换、集成规范和标准、开放并提供数据接口。

（4） 具备良好的兼容性：须跨主流浏览器运行，支持IE10以上系列、Google chrome、360浏览器等主流浏览器，并保证各浏览器网页显示效果的一致性；兼容学校其他业务系统各项服务功能。

（5）所提供的软件平台需支持IPV6网络协议。

（6）系统须部署在学校的虚拟化平台上，采用分布式高可用性部署；系统用户端和管理端均应采用完全的B/S架构。

注：为保证本项目顺利进行并投入使用，以上采购内容不允许负偏离。