# AISERVERS实时互动数字人一体机操作手册

## 概述

爱微思服数字人一体机基于大语言模型和 RAG 的开源知识库问答系统。RAG 技术赋能大语言模型，使其在生成内容时精准参考内部知识库，有效攻克 LLM 模型处理特定领域请求时易出现的幻觉、知识泛化不佳、可解释性差等难题。这意味着一体机在面对各类复杂业务咨询时，能够给出高度精准、逻辑严谨的回应，为企业决策与客户服务提供坚实可靠的信息支撑。

凭借先进的问题解析与知识库匹配算法，数字人一体机精准捕捉用户需求。无论是在学术研究场景下查找专业文献资料，还是企业内部员工查询流程规范，亦或是客服应对客户售后疑问，只需输入自然语言问题，就能迅速定位最相关知识，节省宝贵时间，提升办事效率。

充分考虑企业数据安全诉求，支持本地部署与调用本地大模型。这一举措严密管控越级访问风险，杜绝公有模型数据传输中的安全隐患，确保企业核心知识资产牢牢掌握在自己手中，让企业安心畅享智能服务。

## 配置要求

### 2.1 硬件配置要求

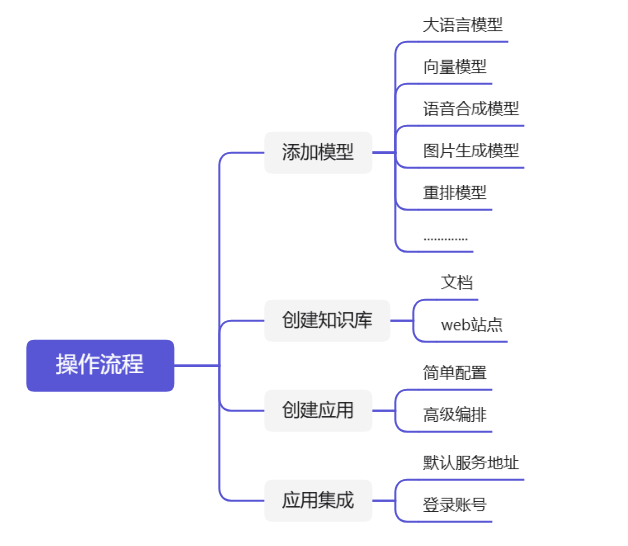
|  |  |  |  |
| --- | --- | --- | --- |
| 设备名称 | CPU | 内存 | 显卡 |
| 建议配置 | i7-12700H | 16G | 2080Ti 及以上 |

### 2.2 软件配置要求

|  |  |  |
| --- | --- | --- |
| 项目 | 名称 | 版本 |
| 操作系统 | Windows | Windows10 及以上 |

## 软件使用说明

操作流程可分为四步：添加模型、创建知识库、创建应用、应用集成。



### 3.1添加模型

登录AISERVERS系统后，在供应商列表中选择xxxx模型(例如deepseekR1:1.5b)，然后点击【添加模型】，进入模型配置表单配置参数如下：



模型名称：爱微思服中自定义的模型名称。

权限：分为私有和公用两种权限，私有模型仅当前用户可用，公用模型即系统内所有用户均可使用，但其它用户不能编辑。

模型类型：大语言模型

基础模型：通义千问支持的 LLM 模型名称，下拉选项是常用的一些大语言模型名称，支持自定义输入。

### 3.2 创建知识库

打开【知识库】页面，点击【创建知识库】，输入知识库名称、知识库描述、选择内置的向量模型 AISERVERS-Embedding，并设置知识库类型为通用型，然后将离线文档通过拖拽方式或选择文件上传方式进行上传。

上传文档AISERVERS支持智能分段和高级分段两种分段方式，产生分段预览，确认无误点击【创建并导入】，系统后台会对文档进行自动分段 -> 存储 -> 向量化处理操作，执行完成后在知识库文档列表中各个文件状态显示为成功。

### 3.3创建应用

#### 3.3.1简单配置

点击【创建应用】，输入应用名称，选择【简易配置应用】，点击【创建】

应用创建完成，进入简易配置应用的设置页面，左侧为应用信息，右侧为调试预览界面。

应用名称：用户提问时对话框的标题和名字。

应用描述：对应用场景及用途的描述。

AI模型：在【系统设置】-【模型管理】中添加的大语言模型。

提示词：系统默认有智能知识库的提示词，用户可以自定义通过调整提示词内容，引导大模型聊天方向。可对无引用知识库和引用知识库两种情况分别设置不同的提示词。

历史聊天记录：大模型提交当前会话中最后 N 条对话内容，否则仅向大模型提交当前问题。

关联知识库：用户提问后将在关联的知识库中检索分段。

开场白：用户打开对话时，系统弹出的问候语。支持 Markdown 格式；[-]后的内容为快捷问题，一行一个。

#### 3.3.2高级配置

点击【创建应用】，输入应用名称，选择【高级编排】，点击【创建】，进入工作流编排页面。新创建的高级编排应用会默认生成简易工作流，可以根据自己的需求进行自定义编排，点击发布后生效。

每个工作流都有基本信息与开始两个基础节点：

基本信息：应用的基本信息设置节点，如应用名称、描述、开场白等设置，每个应用只有一个基本信息节点，不能删除和复制。

开始：工作流程的开始，每个应用只能有一个开始节点，不能删除和复制。

点击右上角的【添加组件】，可以点击或拖拽到画布进行工作流编排。以下是每个组件的用途说明：

AI对话：与AI大模型进行对话节点。

图片理解：识别并理解图片所包含的信息。

图片生成：根据提供的文本内容生成图片。

知识库检索：关联知识库，检索与问题相关分段的节点。

多路召回：使用重排模型队多个知识库的检索结果进行二次召回。

判断器：根据不同条件执行不同的节点。

指定回复：直接指定回复内容。

表单收集：通过表单的方式收集问答所需要的必要信息。

问题优化：AI对话的一种，设定了默认的角色和提示词，根据上下文优化问题。

文档内容提取：提取文档中的内容。

语音转文本：将音频转换为文本。

文本转语音：将文本转换为语音。

注：应用信息设置完成后，可以在右侧调试预览中进行提问预览，调试预览中提问内容不计入对话日志。经过调试后，保存设置并发布。在应用列表页面，在概览页面点击演示或在浏览器中复制公开访问链接进入问答页面进行提问。

### 3.4 AISERVERS应用支持零编码快速嵌入第三方 Web系统

在应用概览页面，点击【嵌入第三方】，复制对应的嵌入代码，以全屏模式或浮窗模式嵌入到第三方系统中，嵌入完成后便可在第三方系统中进行问答

## 功能介绍

## 

### 4.1知识库

企业私有的专业知识库，包含各种类型的数据，是问答对话中回答用户问题的知识来源。AISERVERS 中知识库分为通用型知识库和 Web 站点知识库两种类型。

通用型知识库：对离线文档上传管理，支持的文本文件、表格以及 QA 问答对。

Web 站点知识库：用于获取在线静态文本数据管理，输入 Web 根地址后自动同步根地址及子级地址的文本数据。

AISERVERS 支持知识库创建、重新向量化、设置、同步、导出、删除等功能

#### 4.1.1文档

对于通用型知识库，点击【上传文档】进入上传文档页面，可以拖拽文件或选择文件/文件夹进行上传，支持的文件格式包括：TXT、Markdown、PDF、DOCX、HTML、XLS、XLSX、CSV、ZIP。如果选择文件夹，将通过文件后缀进行自动过滤，每次最多上传 50 个文件，且单个文件不超过 100 MB。

#### 4.1.2问题

AISERVERS 知识库管理员可以收集用户可能提出的相关问题，创建问题并关联知识库文档中的内容，维护知识库常见问题列表，以提高问答的准确度。

#### 4.1.3命中测试

知识库命中测试是对提问进行提前验证，点击【命中测试】，输入测试问题，系统在当前知识库中向量检索与问题相关的分段，然后按照相似度和返回Top 分段数等参数返回命中结果，可以根据命中测试返回分段内容，对分段进行再次编辑或添加关联问题，以进一步提升和优化回答效果。

检索模式：当前支持向量检索、全文检索和混合检索 3 种检索模式。  
（1）向量检索：使用向量模型通过向量距离计算与用户问题最相似的文本分段。  
（2）全文检索：通过关键词检索，返回包含关键词最多的文本分段。  
（3）混合检索：同时执行全文检索和向量检索，再进行重排序，从两类查询结果中选择匹配用户问题的最佳结果。

相似度：相似度越高代表问题和分段的相关性越强。

### 4.2函数库

AISERVERS 知识库问答系统支持了函数库功能。用户可以根据自身的业务需求，在函数库中创建特定的函数脚本，例如数据处理、逻辑判断、信息提取或其它任何满足实际工作场景的操作。

函数创建完成后，在应用编排时以添加组件的方式调用这些函数，从而更好地满足各种复杂的业务需求。这种组件化的设计不仅提高了工作效率，还使得整个系统更加模块化，便于维护和扩展

### 4.3应用

#### 4.3.1应用导出与导入

应用支持导出操作，导出后文件名称为：应用名称.mk。 对于高级编排类型的应用，导出内容为包括工作流中所有节点的参数设置以及函数内容，流程节点中所选择的知识库和模型信息不导出。

在应用页面，点击【导入应用】，即可导入应用文件（文件后缀为.mk）。

#### 4.3.2应用概览

在应用列表页面，点击应用的面板，进入到应用概览页面。

在概览页面可以进行启用/禁用/重新生成公开访问链接、演示、嵌入第三方、访问限制以及 API Key 管理。

#### 4.3.3命中测试

应用命中测试功能是简易配置类型应用针对用户提问进行一个模拟测试，用来调整分段内容和关联问题从而提高分段内容的匹配度，从而提升回答效果。输入测试问题，系统根据检索模式在当前应用关联的知识库中进行检索，然后按照相似度和返回Top 分段数筛选返回满足条件的分段。如有必要，可对指定返回分段内容进行再次编辑。

#### **4.3.4对话日志**

在对话日志记录了所有用户会话中的问答详情，包括用户对AI 回答的反馈信息，维护人员可以通过查看对话日志详情并参考用户反馈进一步修正答案。

在对话日志支持查询过去 7 天、30 天、90 天、过去半年的对内容，以及按摘要和用户反馈查询进行查询。

### **4.4模型管理**

模型管理用于对接供应商的大语言模型，支持对接主流的大模型，包括本地私有大模型（Llama 3 / Qwen 2 等）、国内公共大模型（ DeepSeek / SILICONFLOW / Kimi / 通义千问 / 智谱 AI / 百度千帆 等）和国外公共大模型（OpenAI / Azure OpenAI / Anthropic / Gemini 等），支持的类型包括文本生成、向量模型、重排、语音识别、语音合成、视觉模型、图片生成等。

**4.5团队管理**

在 AISERVERS 中，支持团队协同维护知识库和应用。内置所有者和成员两种角色，知识库/应用创建者即为所有者角色，添加的成员用户都是成员角色。所有者可以添加和移除成员。

## 联系客服

遇到任何问题可扫码联系客服

