

混合云一体机推理基础版用户手册

1、前提条件

1、已获取管理员账号和密码

2、浏览器适配情况：

- 支持谷歌 Chrome 135+，微软 Edge 127+，火狐 Firefox 135+，360 个人安全浏览器最新版 15+、奇安信浏览器 V1.1.45746.52、360 企业定制版本号:13.1.2001.321、13.1.2002.425、13.3.3120.116
- 不支持红莲花国密浏览器等其他浏览器

2、用户登录

1. 打开浏览器，在地址栏中，输入登录地址，按“Enter”。进入云平台。



欢迎登录iStack2.0

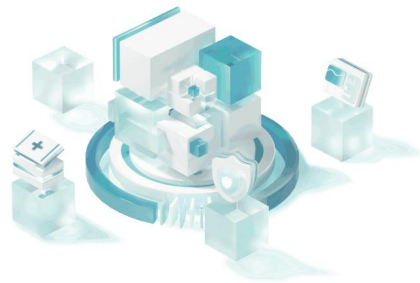
从 请输入您的电子邮箱账号

到 请输入密码

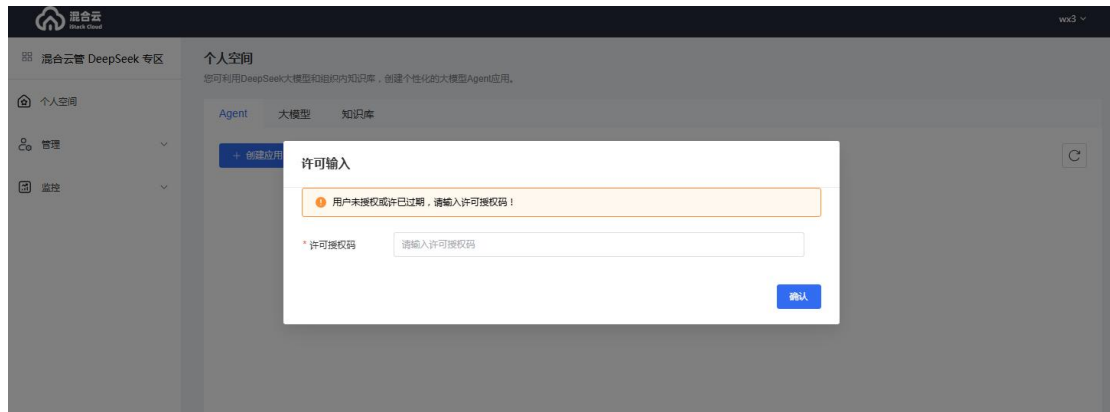
登录

浏览器支持

360浏览器 火狐浏览器 Chrome Safari



2. 输入用户名、密码，登录云平台。
3. 选择融合模型，进入混合云管 DS 模块。
4. 若为新创建的用户，则需要输入许可授权，授权码输入正确后才可进入融合模型

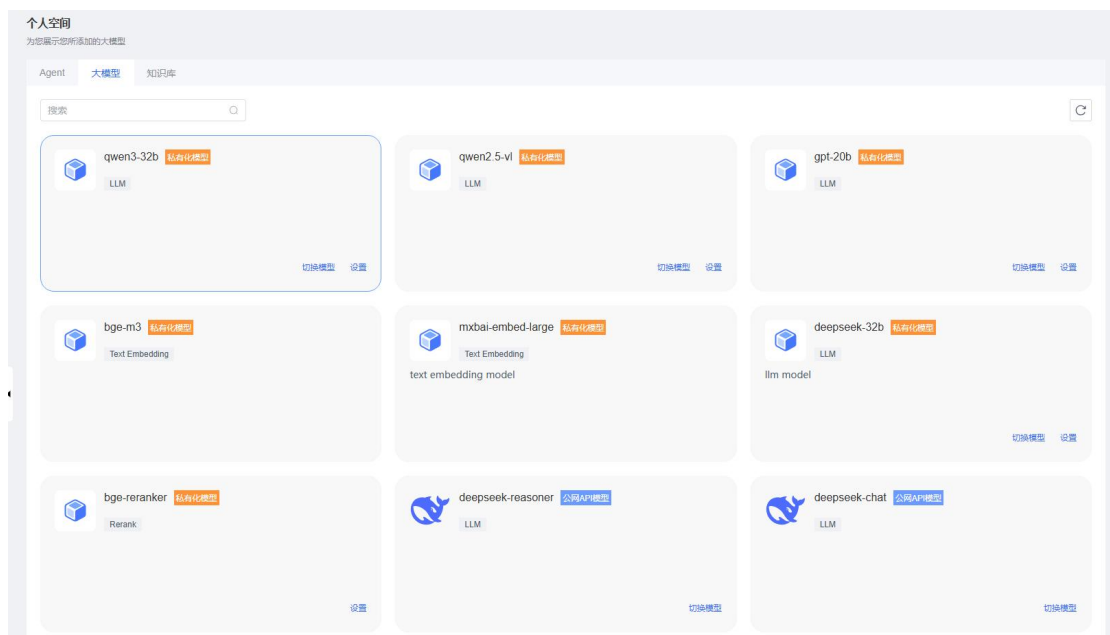


注意：

- admin 用户默认提供 3 年有效期
- 如下操作均为进入混合云管 DS 模块的操作

3、个人空间-大模型

1. 点击“个人空间”→“大模型”，可以查看到当前平台纳管的模型，包括启用的私有化模型和公网 API 模型。



2. 其中私有化模型类型为 LLM 和 Rerank 支持设置模型参数，可根据模型处理文本的能力和限制，配置模型上下文长度和最大 token 上限。text-embedding 模型无法设置相关参数。

说明：

1. 为保障模型处理长文本和多轮对话的能力，请根据实际需求调整 Token 上限及上下文长度。配置过小可能导致信息截断、回答不完整或对话连贯性丢失。
2. 此处默认的参数为管理员配置，若修改后按照个人配置的参数为准。



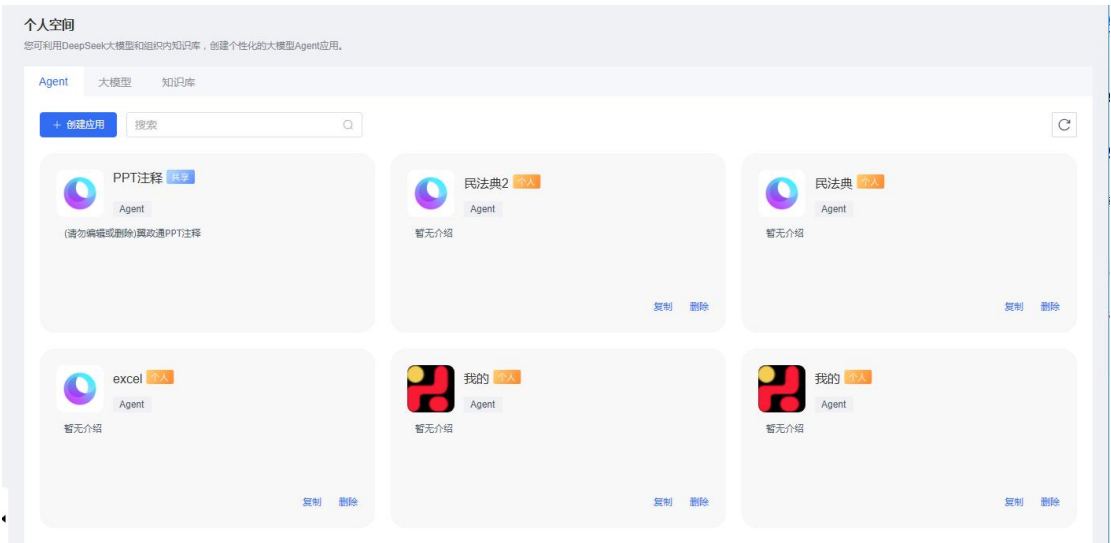
3. 其中不论私有化模型还是公网 API 模型，只要类型为 LLM，均支持切换模型，选择某一对话类型的模型点击“切换”，即可将所有应用使用的模型切换为该模型

注意：切换模型会影响现有 Agent 的运行

4、个人空间-Agent

4.1 创建智能体应用

1. 点击“个人空间-Agent”，进入个人智能体应用列表页。搜索框可根据智能体名称模糊搜索。



2. 单击**创建应用**。进入应用创建页面。

创建空白应用

选择应用类型

Agent ?

基础编排, 适合初步尝试

修改头像

给应用起个名字

请输入应用名称0 / 50

添加应用描述

请输入应用描述0 / 100

取消

创建

3. 参照表进行参数配置，配置完成后，单击**确定**。

应用类型	选择智能体应用类型。Agent 类型允许使用简单的设置编排聊天机器人应用程序。
应用头像	选填，支持选择图片或表情符号更换头像。
应用名称	必填，自定义应用名称，便于识别。
描述	选填，应用的描述。

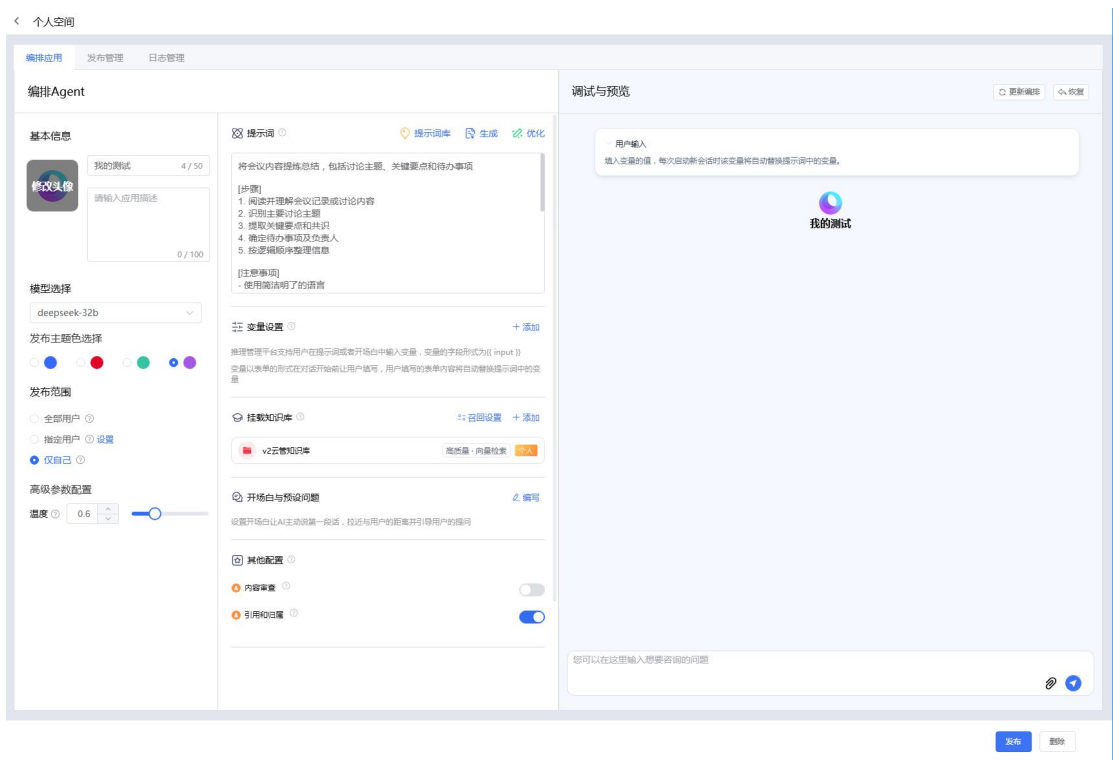
4. 创建完成后，可以在列表中查看应用信息。

4.2 编辑智能体应用

个人类型的智能体支持编排操作。

操作步骤：

1. 点击创建好的智能体应用，进入编辑智能体页面。
2. 左侧为智能体应用的基础信息，可修改名称及描述。
3. 支持选择平台纳管的模型。



4.2.1 发布主题色设置

编排智能体应用时，支持设置应用发布后的主题色。

操作步骤：

1. 进入编排应用页面
2. 选择发布主题色，内置 4 种主题色，包括：蓝色、红色、绿色、紫色
3. 点击“发布”按钮，发布后的应用问答页面将按照设置的主题配色进行展示。

4.2.2 高级参数配置

编排智能体应用时，提供温度的高级参数配置

操作步骤：

1. 进入编排应用页面
2. 选择设置温度参数：取值范围：0.0-2.0，控制模型输出的随机性。数值越高（接近 1 时），输出结果越随机；数值越低，输出结果越具有确定性
3. 点击“发布”按钮，发布后的模型问答将按照设置温度参数进行回答

4.2.3 发布范围设置

编排智能体应用时，支持设置应用发布的范围。即此应用发布后，指定哪些用户可以使用。

- 当选择“全部用户”后，则此应用发布后，所有平台和访客用户均可使用此应用（访问用户的详细介绍可参考 7.3 章节）
- 当选择“指定用户”后，则需要选择设置哪些用户支持访问此应用，当应用发布后，仅有选择的的用户和自己才能访问此应用
- 当选择“仅自己”后，则此应用发布后仅有自己才能访问

操作步骤：

1. 进入编排应用页面
2. 选择发布范围，支持选择全部用户、指定用户、仅自己选项。默认为“仅自己”
3. 点击“发布”按钮，发布后即可按照设置的用户范围访问此应用
4. 当需要变更发布范围，打开需要编辑的应用，修改发布范围后，点击“发布”按钮即可

4.2.4 提示词设置

1. 提示词的关键在于清晰、准确地传达需求，引导模型生成符合期望的内容。可以从明确目标、设定背景、细化要求等角度入手。
2. 点击添加新变量，即可插入变量



4. 编写提示词时可插入变量，变量用{{a}}来表示。变量名称必须包含字母，添加完成之后，括号内容变成蓝色，即代表添加成功。

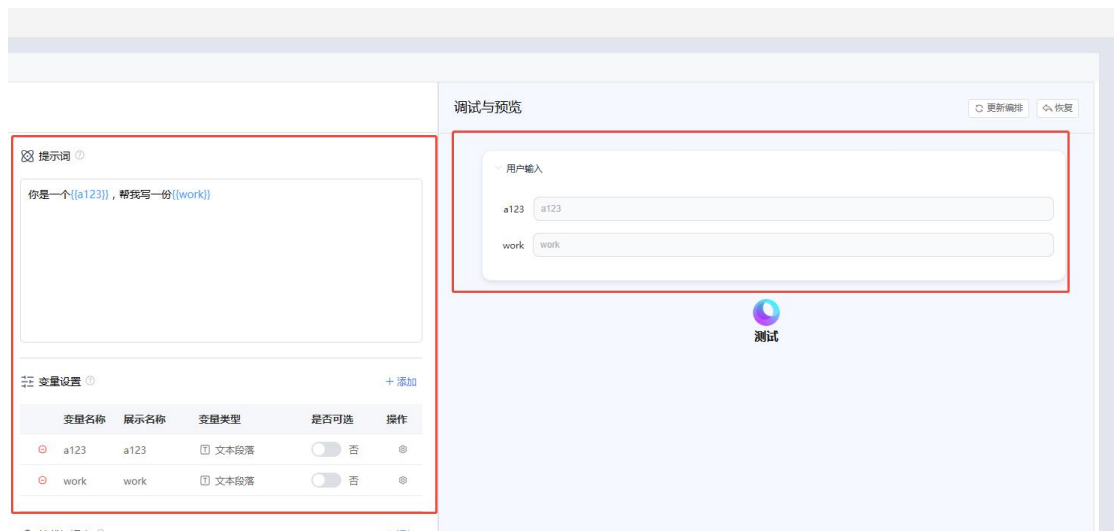
提示词 ?

你是一个{{a123}},帮我写一份{{work}}

5. 写完提示词大纲后，将变量添加到用户输入表单中。



6. 添加变量后效果，右侧为智能体调试界面。可看到用户输入好的变量。



7. 点击变量设置下，变量最右侧的设置按钮，对变量进行编辑。变量类型可选文本段落、下拉选项。可编辑变量名称、展示名称、文本长度（变量类型为文本段落时）。

变量设置

+ 添加

	变量名称	展示名称	变量类型	是否必选	操作
⊖	name	小助手	下拉选项	☐ 否	
⊖	work	周报	下拉选项	☐ 否	

编辑变量

变量类型

文本段落

变量名称

name4 / 30

展示名称

name4 / 30

文本长度

10

取消

保存

8. 修改提示词的变量的展示名称

☰ 变量设置 ⓘ

+ 添加

	变量名称	展示名称	变量类型	是否必选	操作
⊖	name	周报助手	📄 文本段落	☐ 否	⚙️
⊖	work	周报	📄 文本段落	☐ 否	⚙️

9. 点击右下角**发布**按钮

10. 发布模型后，你的智能体将获得一个新的身份——周报助手！

 **新建对话**

a123

a123

work

work

 开始对话

11. 变量设置为下拉选项时，可同时为一个智能体，定义多种身份和工作。

编辑变量

×

变量类型

下拉选项

变量名称

name4 / 30

展示名称

小助手3 / 30

选项

公文助手

周报助手

财务分析

+ 添加选项

取消

保存

12. 点击右上角更新编排，可将刚才重新配置的变量应用到模型中。

调试与预览

更新编排

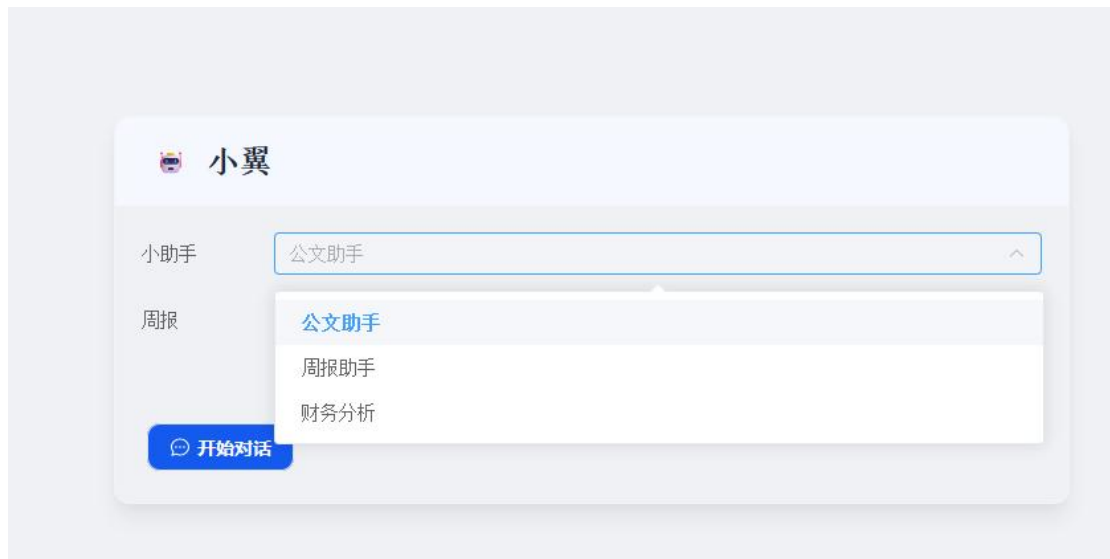
恢复

用户输入

填入变量的值，每次启动新会话时该变量将自动替换提示词中的变量。

test

13. 发布模型后，为你的智能体选择角色吧！



4.2.5 提示词库

系统内置了四种提示词库，供用户选择使用。支持用户按照特定的模板填写对应的信息，帮助客户编写规范的提示词。

四种提示词包括：通用结构、任务执行、角色扮演和基于知识库回答

操作步骤：

1. 点击提示词上方的**提示词库**按钮
2. 弹框展示内置的 4 种提示词库



3. 用户根据应用需求，选项对应的提示词模板，将提示词模板中紫色字体部分编辑为此智能体应用需要达到的需求。
4. 点击**插入提示词**按钮，即可将编辑的内容插入到提示词框光标放置的位置。

5. 点击**复制提示词**按钮，即可将编辑的内容复制到剪贴板

注意：模板中紫色字体的提示信息不会被插入和复制；提示词库中的内置内容无法增加或删除，可插入或复制到提示词框内后再进行编辑

6. 插入或者复制到提示词框内后，仍然可以针对内容进行编辑

4.2.6 提示词优化

针对用户已编写的提示词，可进行优化。

操作步骤：

1. 在提示词框内编写相应的提示词内容
2. 点击提示词上方的**优化**按钮
3. 等待一段时间后，即可生成优化后的提示词
4. 若针对优化后的提示词不满意，可点击**重新生成**按钮进行重新生成
5. 若需要使用优化后的提示词，则可点击**应用**按钮，即可将生成的提示词替换之前的提示词

4.2.7 提示词生成

帮助用户一键生成更高的质量的提示词。按照用户要求，点击生成按钮，可通过模型生成相应的提示词，另外根据生成的结果，可再次输入要求进行优化。

操作步骤：

1. 点击提示词上方的**生成**按钮
2. 在输入框中输入具体要求，点击确定
3. 模型根据要求进行生成，需要等待一段时间
4. 生成后，可继续输入优化点的要求，系统将再次生成新的提示词
5. 针对生成的多条提示词，最后选择一个满意的提示词，点击**应用**按钮，即可将此提示词填入提示词框中

4.2.8 知识库挂载

1. 点击添加，展示已有知识库列表。单击目标知识库，点击添加。一次可点击多个目标知识库进行添加。

注意：

- 若同时选择高质量和经济型知识库，则建议配置 Rerank 模型
- 支持在弹框中创建新知识库



2. 召回设置

说明：默认情况下使用多路召回，从多个知识库中检索知识，然后重新排序。若同时选择高质量和经济型知识库，建议配置 Rerank 模型，保证内容检索的准确性！若同时选择不同 Embedding 模型的高质量知识库，建议配置 Rerank 模型，保证内容检索的准确性。

支持为应用配置召回测试，支持选择 Rerank 模型，配置 TopK，设置 Score 阈值。

召回设置



① 默认情况下使用多路召回，从多个知识库中检索知识，然后重新排序。若同时选择高质量和经济型知识库，建议配置 Rerank 模型，保证内容检索的准确性！若同时选择不同 Embedding 模型的高质量知识库，建议配置 Rerank 模型，保证内容检索的准确性。

Rerank 模型

Top K

Score 阈值 ☒

3. 添加后的知识库显示在调试区域的列表中。

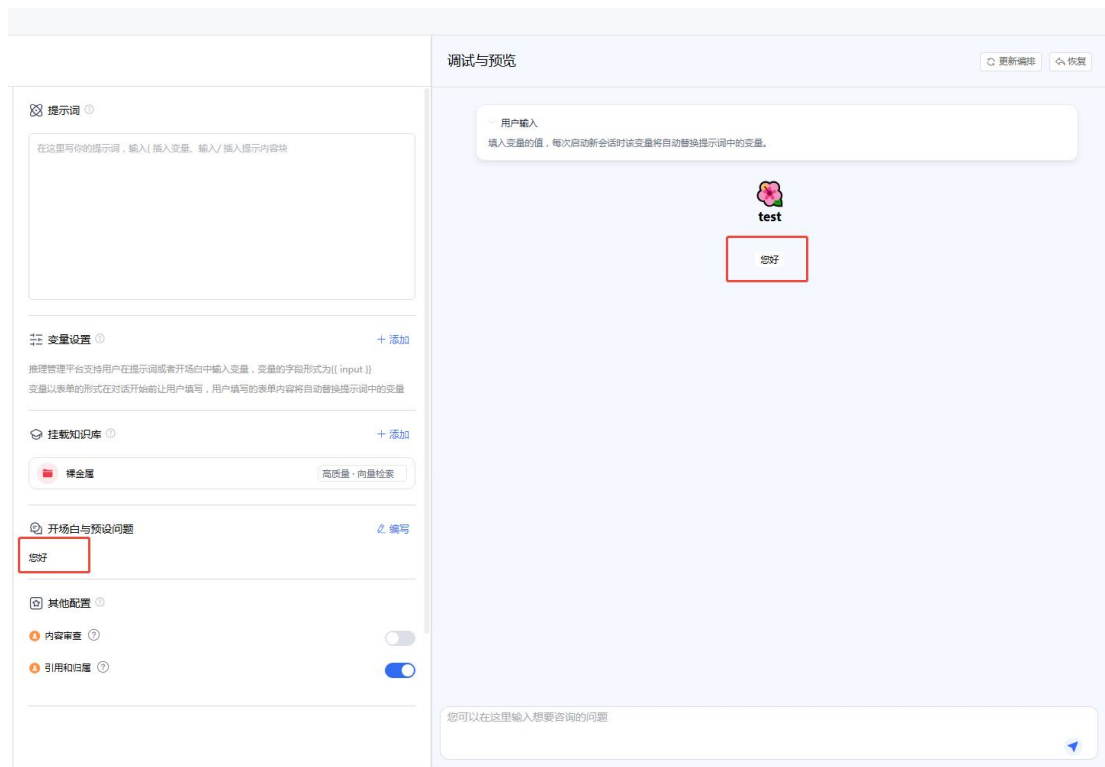
挂载知识库

召回设置 + 添加

test2	高质量·混合检索	企业
揭惠	高质量·混合检索	个人
变压器	高质量·混合检索	个人

4.2.9 开场白和预设问题

1. 为聊天型应用添加预对话设置，可以提升用户体验。
2. 在对话开场白下面的输入框内输入开场白内容
3. 在预设问题下面点击添加选项，在出现的输入框内输入问题。最多添加 5 个问题
4. 点击右上角按钮保存，即可将内容同步至右侧调试面板。



4.2.10 引用和归属

支持开启引用和归属，开启后，挂载知识库文档时，可列出被引用的知识库文件。

1. 点击开启引用和归属按钮



2. 效果展示

知识库问答可展示引用的文件



点击引用文档，可查看引用段落



4.2.11 内容审查

1. 支持开启内容审查，可设置敏感词。当用户输入敏感词时，输出用户预置好的文本内容

内容审查设置

关键词

每行一个，用换行符分隔。每行最多 100 个字符。

公安

审查输入内容

预设回复

支持markdown

内容敏感

审查输出内容

取消

保存

2. 效果展示



3. 同时，可对输出内容进行审查。包含敏感词时，输出用户预置好的文本内容

内容审查设置

关键词

每行一个，用换行符分隔。每行最多 100 个字符。

公安

审查输入内容

审查输出内容

预设回复

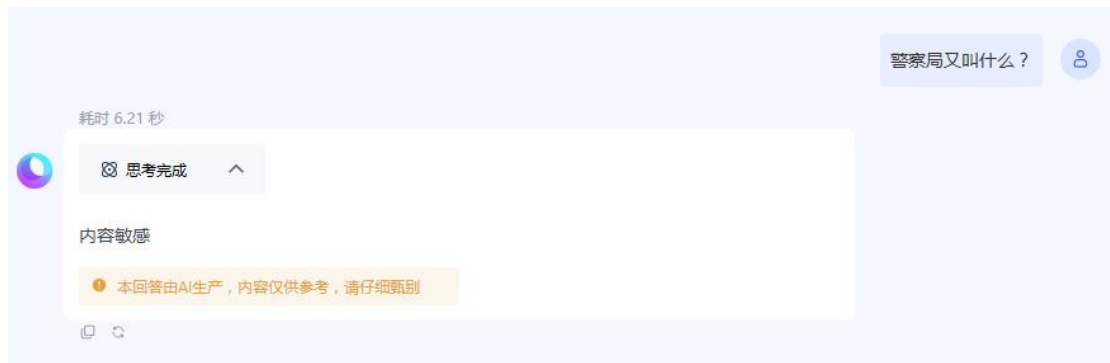
支持markdown

内容敏感

取消

保存

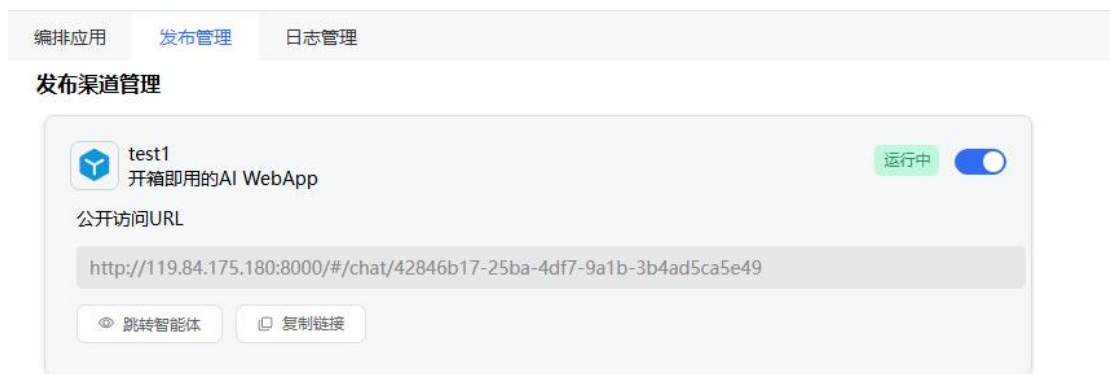
4. 此例中，模型想输出敏感词，但被规则禁用了



4.3 发布管理

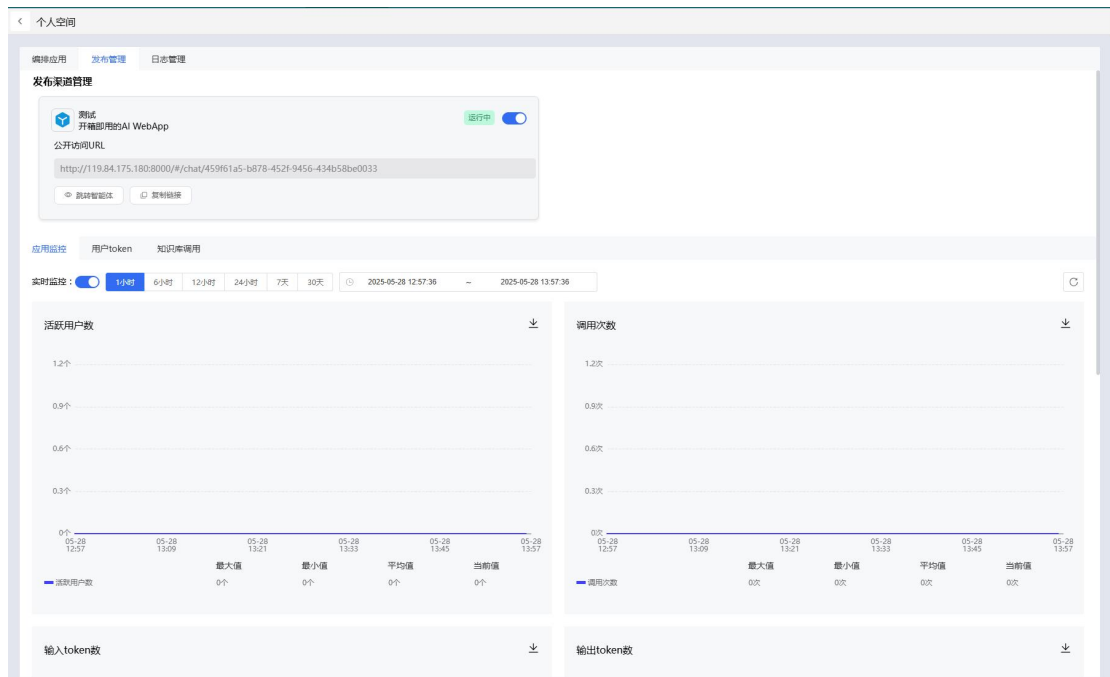
个人类型的智能体支持发布管理。

点击“发布管理”，支持启用和禁用应用。禁用后发布的应用无法使用



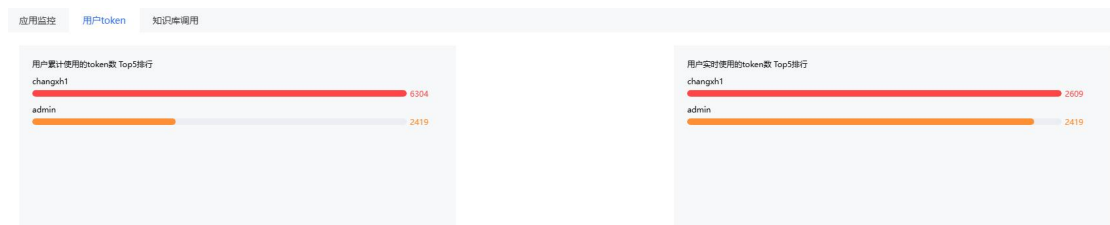
4.3.1 应用监控

支持查看发布应用的监控信息，可查看实时和历史的应用监控信息，监控指标包括：应用的活跃用户数、应用的调用次数、应用输入的 token 数、应用输出 token 数、token 输出速度、调用总 token 数。



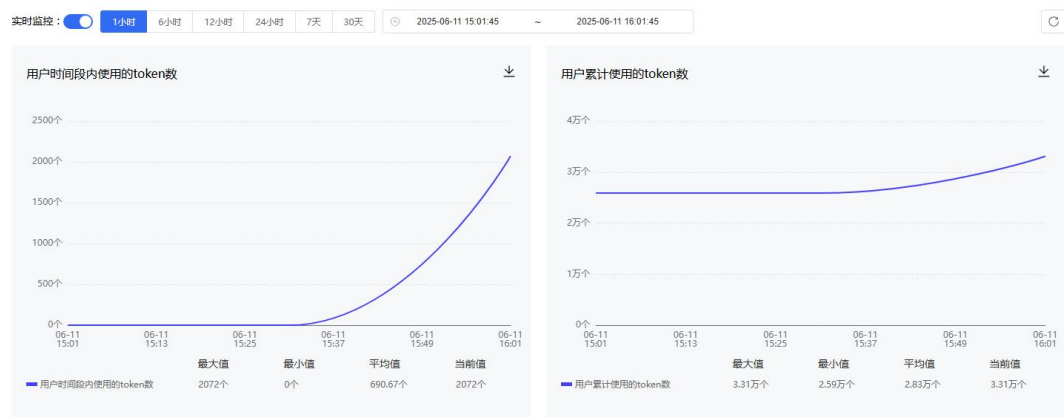
4.3.2 用户 token 监控

1. 支持查看用户累计使用 Token 数量的 Top5 排名情况、用户实时使用 Token 数量的 Top5 排名情况



2. 支持查看所有使用此应用的用户列表，且支持选择某一用户，查看用户时间段内使用的 token 数，以及该用户的累计使用的 token 数的趋势图

说明：用户时间段内使用的 token 数指标，每个点代表该时间点前半小时使用的 token 数

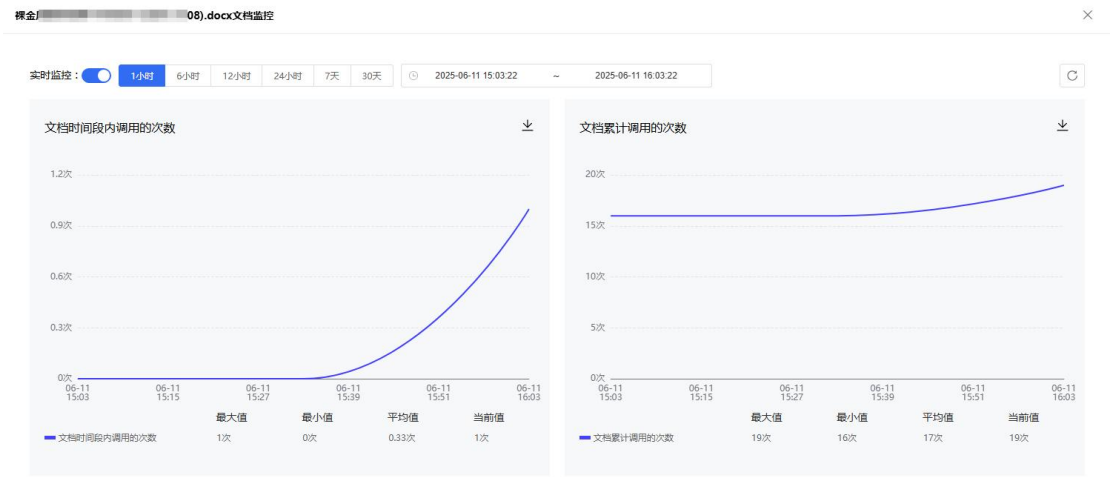


4.3.3 知识库调用监控

1. 支持查看文档累计调用次数的 Top5 排名情况、文档实时调用次数的 Top5 排名情况

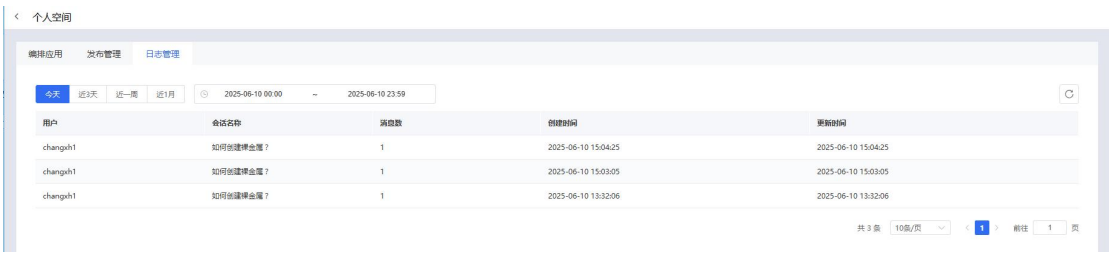


2. 支持查看该应用当前和历史绑定的知识库所调用过的文档列表，且支持选择某一文档，查看文档时间段内调用的次数，以及该文档累计调用次数的趋势图
说明：文档时间段内调用的次数指标，每个点代表该时间点前半小时使用的 token 数。



4.4 日志管理

个人类型的智能体支持日志管理。
日志管理中支持查看使用此应用的用户名、会话名称和消息数
同时支持按照时间维度进行筛选查看，默认展示当天的日志。



4.5 共享智能体问答

当选择共享类型的智能体时，点击后，则直接进入智能体问答界面。

操作步骤：

1. 进入个人空间-Agent 界面；
2. 点击某一共享的应用；
3. 自动跳转到此应用的问答界面，可直接进行问答

说明：访客用户登录后，仅支持查看到共享给自己的智能体应用，点击后直接使用

4.6 上传文档问答

无论是个人智能体还是共享智能体，在问答界面均支持上传文档问答。

注意：公网 API 模型暂不支持上传文档问答

操作步骤：

1. 进入个人空间-Agent 界面；
2. 入口 1：点击个人应用进入编排页码，在调试与预览界面，可选择“文档”类别进行上传
入口 2：个人应用发布后，在问答界面，可选择“文档”类别进行上传
入口 3：点击共享应用，在问答界面，可选择“文档”类别进行上传



说明：如果平台中部署了 VL 模型，则会自动解析文档中图片内容

3. 选择需要问答的文档，支持.md,.txt,.doc,.docx,.pdf,.html,.htm 格式
4. 文档选择完成后，在问答框中输入需要咨询的问题，点击发送。

限制条件：

- 由于模型处理能力，默认（GPU 2 张卡）支持大约 2w 字的文本处理，若上传文档的内容过多可能无法进行问答

- 文档类型的会话在问答过程中，不支持继续上传文件进行问答。若需要对新文件进行问答，请新建会话
- 文档类型的历史会话仍可继续对上传的文档进行问答

4.7 上传表格问答

无论是个人智能体还是共享智能体，在问答界面均支持上传表格问答。

注意：公网 API 模型暂不支持表格问答

限制条件：

- 模型需要部署为 qwen3-32b 模型
- 目前仅支持针对表格的第一个 sheet 进行问答
- 表格需要支持 UTF-8；若不支持 UTF-8，则问答时会报错
- 表格类型的会话在问答过程中，不支持继续上传表格进行问答。若需要对新表格进行问答，请新建会话
- 表格类型的历史会话仍可继续对上传的表格进行问答

操作步骤：

1. 进入个人空间-Agent 界面；
2. 入口 1：点击个人应用进入编排页码，在调试与预览界面，可选择“表格”类别进行上传
入口 2：个人应用发布后，在问答界面，可选择“表格”类别进行上传
入口 3：点击共享应用，在问答界面，可选择“表格”类别进行上传



3. 选择需要问答的表格，支持.csv,.xls,.xlsx 格式，最多 3 个文件，每个不超过 5MB
4. 表格选择完成后，在问答框中输入需要咨询的问题，点击发送。

4.8 上传图片问答

无论是个人智能体还是共享智能体，在问答界面均支持上传图片问答。

注意：公网 API 模型暂不支持图片问答

限制条件：

- 模型必须支持开启 visio，原 deepseek 模型不支持，至少需要单独一张卡新部署 VL 模型
- 图片类型的会话在问答过程中，不支持继续上传图片进行问答。若需要对新图片进行问答，请新建会话
- 图片类型的历史会话仍可继续对上传的图片进行问答

操作步骤：

1. 进入个人空间-Agent 界面；
2. 入口 1：点击个人应用进入编排页码，在调试与预览界面，可选择“图片”类别进行上传
入口 2：个人应用发布后，在问答界面，可选择“图片”类别进行上传
入口 3：点击共享应用，在问答界面，可选择“图片”类别进行上传

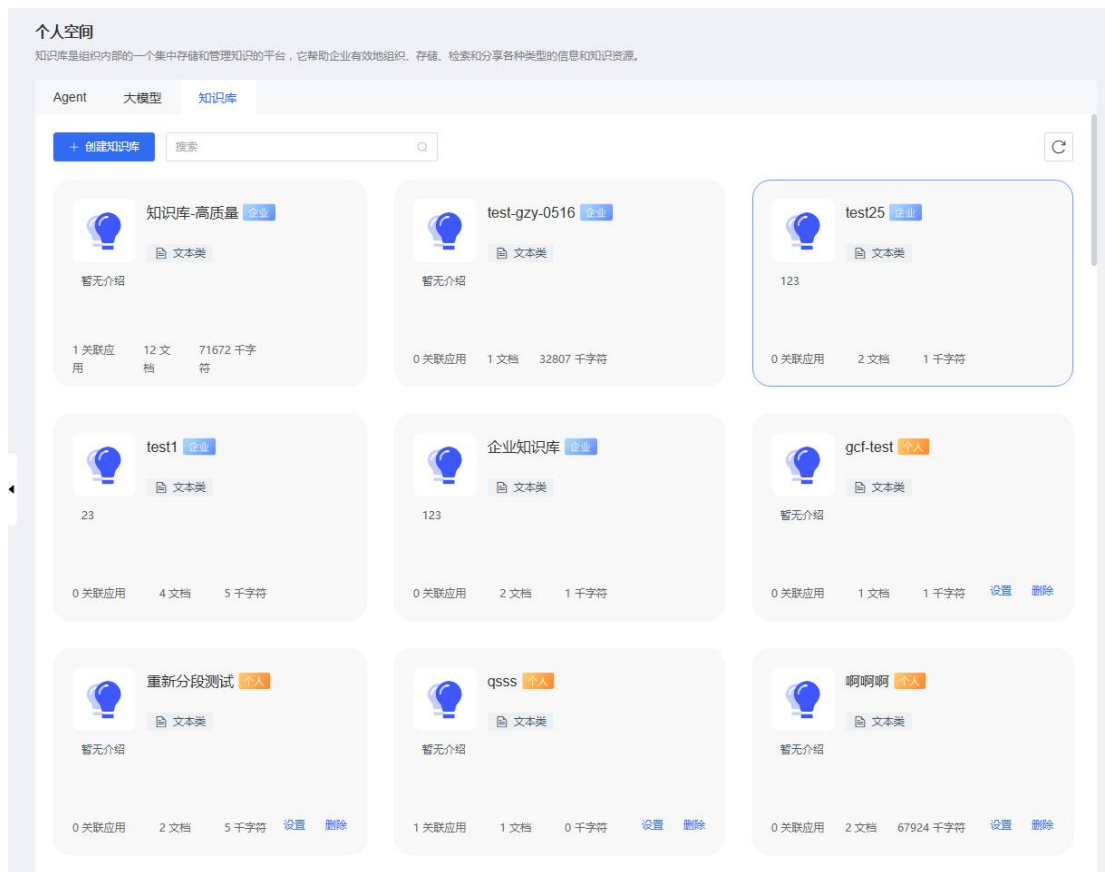


3. 选择需要问答的图片，支持.png、.jpg、.jpeg 格式，最多 1 个文件，每个不超过 2M
4. 图片选择完成后，在问答框中输入需要咨询的问题，点击发送。

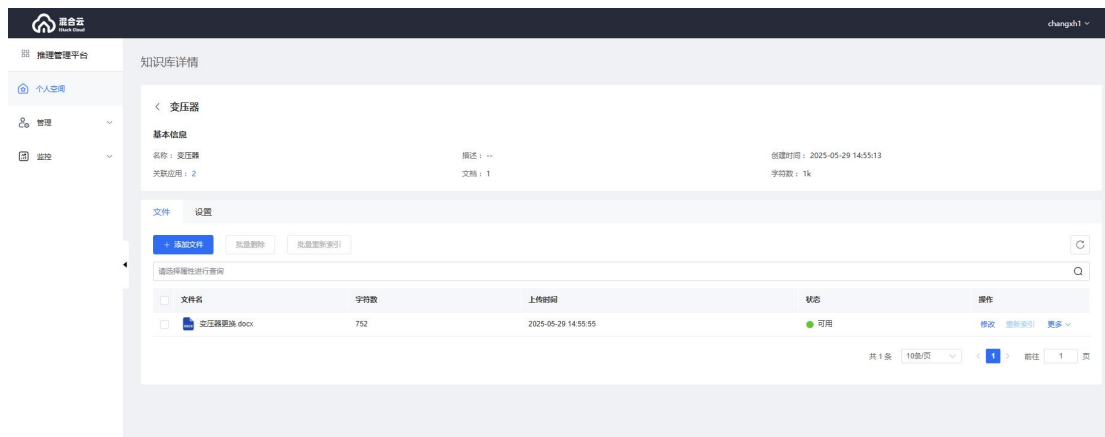
5 个人空间-知识库

5.1 文件管理

1. 点击知识库，进入知识库管理列表



2. 点击进入知识库，可为知识库添加文件



3. 文件右侧开关可修改文件内容、重新索引、禁用文件、重新分段和删除文件。其中，删除文件支持单个文件删除和批量删除；重新索引页支持单个操作和批量操作。

说明：重新索引仅针对错误和索引中状态，且单个文档十分钟内仅支持一次重新索引操作；仅正常、错误状态的文件支持重新分段

4. 添加文件后，可对文档进行预处理。为了使输出结果更精准，文档需要做分段处理，文档分段处理可保证文档的逻辑结构及可读性，同时提高大模型回答的效率和精准度。分段重合字符数（即重叠长度）是优化信息连贯性与检索效果的关键参数。如果遇到知识库理解不明确时，可尝试降低分段长度，增加分段重合字符数进行文档预处理。



5. 自动分段与清洗：本地非结构化文档的自动分段设置采用递归字符文本分割器，分段最大长度 500，分段重合字符数 10。
6. 自定义分段：可设置分段长度和分段重叠字符数。

自定义
自定义分段规则、分段长度以及预处理规则等参数

分段标识符

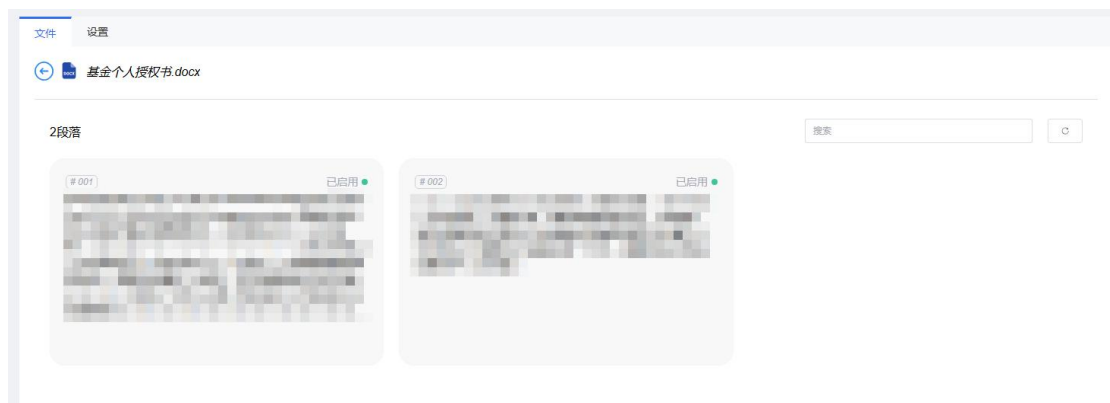
分段最大长度

分段重叠长度

文本预处理规则 ? ☒ 替换掉连续的空格、换行符和制表符
☐ 删除所有 URL 和电子邮件地址

确认并预览 **重置**

7. 上传到知识库的文件支持预览。



8. 点击文件分段，可查看所有分段内容，并可编辑分段内容。

001 分段内容

编辑 已启用

揭惠高速土建养护基础信息 揭惠高速有101座桥梁，无大跨钢结构桥梁，桥梁分为主线桥54座（特大桥9座）、匝道桥38座、天桥5座、线外桥4座。揭惠高速共6座隧道，分别是：小北山1号隧道、小北山2号隧道、坝峰山隧道、仙田明洞、排金山隧道、雷岭峰隧道，隧道单洞累计长14362m。揭惠高速有1座特长隧道，是小北山1号隧道。揭惠高速10处互通立交，互通立交中有枢纽互通立交3处：湖心互通立交、司马浦北互通立交、美园互通立交。揭惠高速主线路基长34.851km，软基段217处（主线126处，匝道91处），连续软基段落长6.8km 揭惠高速有170座涵洞：主线涵洞110座、匝道涵洞60座。揭惠高速有487处边坡：331处填方边坡，156处挖方边坡。揭惠高速有39处高边坡：6处填方、33处挖方。揭惠高速有68处半填半挖段。揭惠高速有14处高陡路堤。揭惠高速有3处临水路堤。揭惠高速有0处山洪易发段。揭惠高速填平区有1处。揭惠高速有328处堆坡、171处护坡。揭惠高速有1处斜坡高墩：林招林场大桥2#墩。揭惠高速有14处隧道口仰坡。

9. 支持禁用和启用分段，分段被禁用后，则无法被检索到

5.2 设置管理

- 1. 选择某一知识库，进入知识库详情
- 2. 点击设置 tab，可编辑知识库名称、知识库描述，选择索引模式。

知识库详情

您可以在这里查看并维护您的企业知识

基本信息

名称：我的 描述： 创建时间： 2025-04-02 18:51

0 关联应用 1 文档 3 千字符

文件 设置

知识库名称

我的 2 / 40

知识库描述

请输入知识库描述 0 / 255

索引模式

高质量

调用 Embedding 模型进行处理，以在用户查询时提供更高的准确度。

经济

使用离线的向量引擎、关键词索引等方式，降低了准确度但无需花费 Token

检索设置

倒排索引

倒排索引是一种用于高效检索的结构。按术语组织，每个术语指向包含它的文档或网页

Top K

2

保存

索引模式介绍：

索引模式	说明	可设置参数
------	----	-------

高质量	通过调用 Embedding 模型进行处理，以在用户查询时提供更高的准确度 注意： ● 前提需要在部署 Embedding 模型才能使用 ● 如需使用 Rerank 模型，则需要提前部署 Rerank 模型	1、选择 Embedding 模型 2、配置检索设置，包括：向量检索、全文检索、混合检索。 ● 向量检索：可设置开启和关闭 Rerank 模型，支持配置 TopK 和 Score 阈值 ● 全文检索：可设置开启和关闭 Rerank 模型，当关闭 Rerank 模型时，可配置 TopK，当开启 Rerank 模型，可配置 TopK 和 Score 阈值 ● 混合检索：支持选择权重设置和 Rerank 模型。支持配置 TopK 和 Score 阈值
经济	使用离线的向量引擎、关键词索引等方式，降低了准确度但无需花费 Token	选择倒排索引，配置 top K

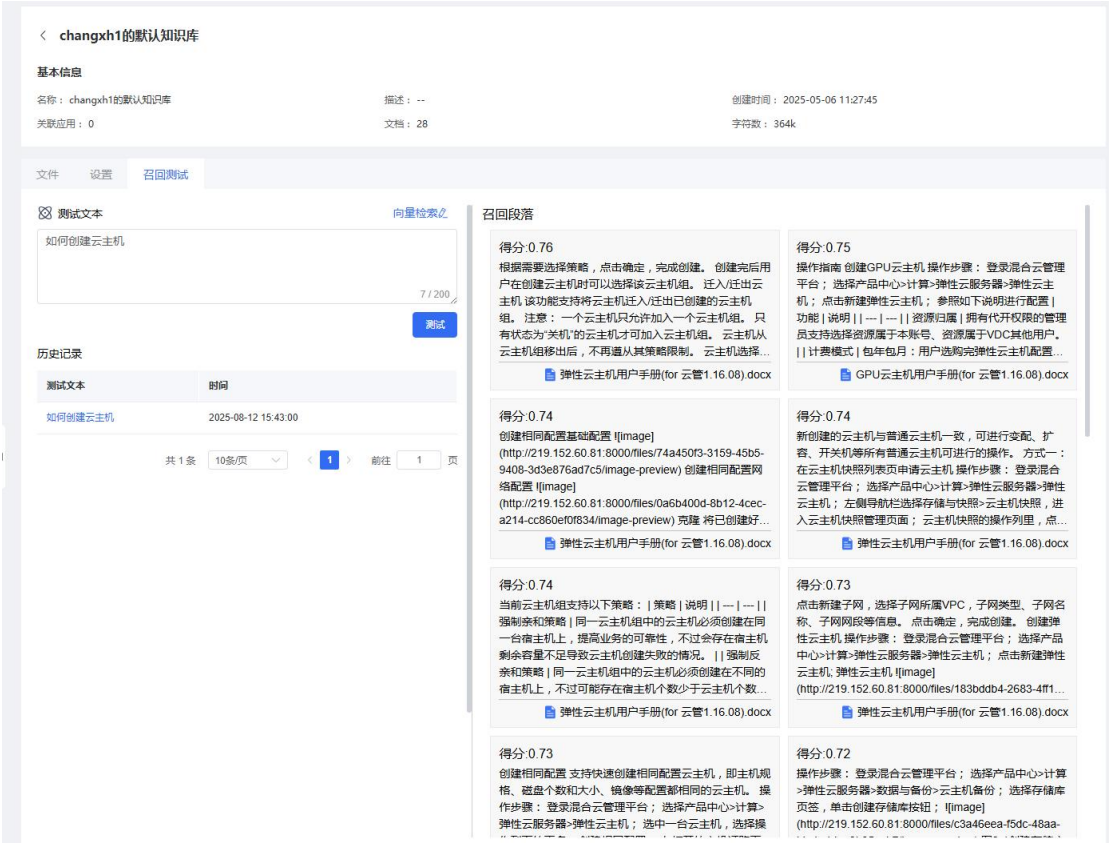
参数说明：

参数	说明
倒排索引	倒排索引是一种用于高效检索的结构。按术语组织，每个术语指向包含它的文档或网页
向量检索	通过生成查询嵌入并查询与其向量表示最相似的文本分段
全文检索	索引文档中的所有词汇，从而允许用户查询任意词汇，并返回包含这些词汇的文本片段
混合检索	同时执行全文检索和向量检索，并应用重排序步骤，从两类查询结果中选择匹配用户问题的最佳结果，用户可以选择设置权重或配置重新排序模型。
Rerank 模型	重排序模型将根据候选文档列表与用户问题语义匹配度进行重新排序，从而改进语义排序的结果
top K	用于筛选与用户问题相似度最高的文本片段。系统同时会根据选用模型上下文窗口大小动态调整分段数量。
Score 阈值	用于设置文本片段筛选的相似度阈值。
权重设置	通过调整分配的权重，重新排序策略确定是优先进行语义匹配还是关键字匹配。

5.3 召回测试

用户可以基于当前知识库进行召回测试，输入测试文本后可测试知识库的召回效果，展示召回的每个分段，并可查看每个分段的得分情况、字符数和具体的分段序号。

- 1. 选择某一个知识库，进入知识库详情
- 2. 点击召回测试 tab 页面
- 3. 输入需要测试的测试文本，点击测试按钮。即可查看到此问题召回的段落和每个段落的得分情况



- 4. 同时支持调整检索方式，调整后，可继续点击测试按钮，查看不同检索方式召回的段落得分
- 5. 支持查看测试文本的历史记录，可点击某一历史记录测试文本，可快速将测试文本填入到测试文本框

6 用户管理

管理员可在此界面查看平台用户、管理平台用户的用户组。

创建、查看和管理翼政通注册的用户。

创建、查看和管理访客用户

几种不同用户的区别：

用户类型	定位	管理员针对不同类型用户支持的操作
平台用户	1、在混合云管平台创建而来 2、支持登录混合云管平台，支持访问融合模型；针对角色不同，融合模型可查看的菜单不同 3、支持登录翼政通	查看、用户组、续期
翼政通用 户	1、通过翼政通客户端注册或者平台管理员创建而来 2、仅支持登录访问翼政通 3、不支持登录混合云管平台	查看、创建、导入、用户组、重置密码、修改、续期、删除、禁用、启用
访客用户	1、通过平台管理员创建而来 2、仅支持登录访问发布的应用，仅可查看个人空间--Agent 界面 3、不支持登录混合云管平台，不支持登录翼政通	查看、创建、导入、重置密码、修改、删除、禁用、启用

6.1 平台用户

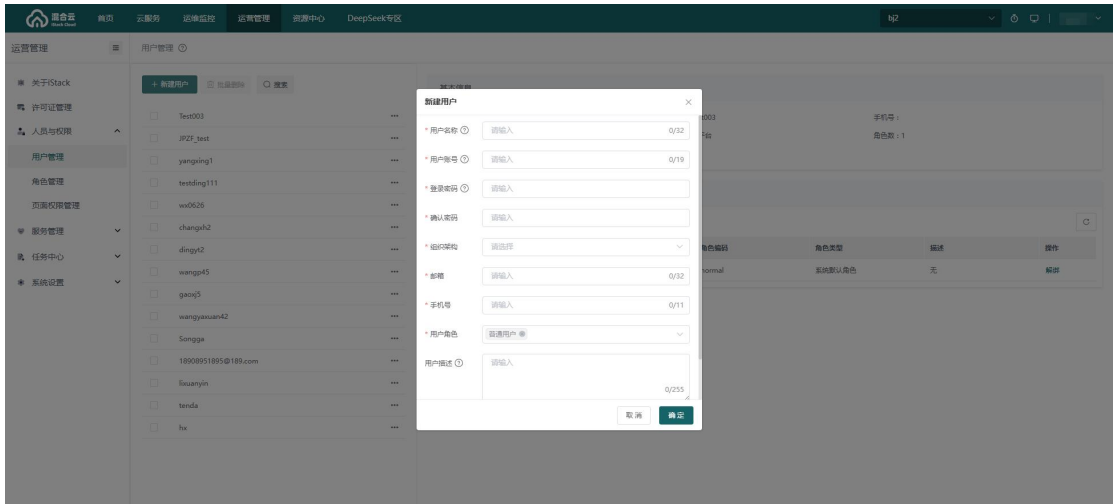
说明：平台用户需要在**运营管理>人员与权限>用户管理**中创建。可参考 istack 云平台使用手册。用户角色说明：

角色名称	融合模型可查看和管理的菜单
普通管理员、超级管理员	全部菜单
普通用户	个人空间

6.1.1 创建用户

操作步骤：

1. 管理员登录平台；
2. 选择**运营管理>人员与权限>用户管理**菜单；
3. 点击**新建用户**按钮



4. 参考如下参数进行配置

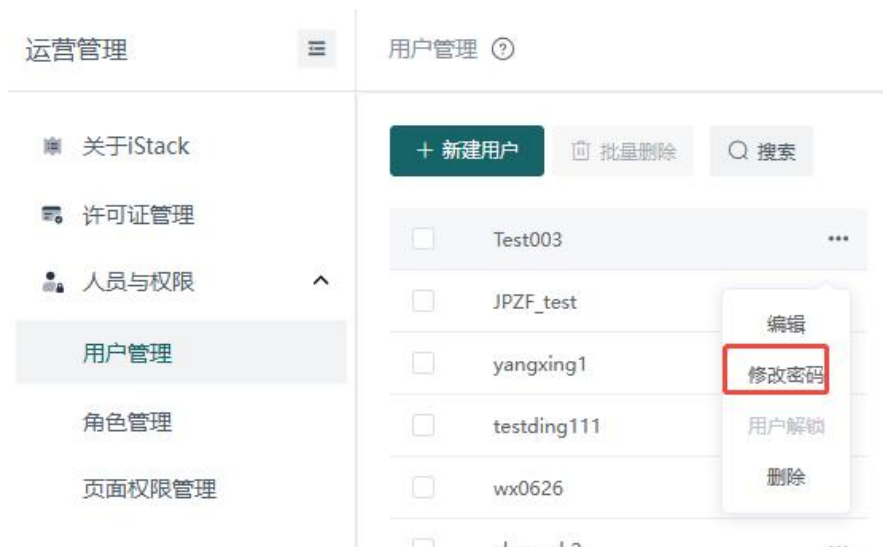
参数名称	说明
用户名称	长度不能超过 32 个字符
用户账号	登录平台使用的账号，只能用英文、数字、下划线、中划线，长度在 5 到 19 个字符
登录密码	按照要求输入用户登录的密码
确认密码	按照要求输入用户登录的密码
组织架构	选择用户所属的组织架构
邮箱	输入邮箱信息
手机号	输入手机号信息
用户角色	1、如果创建的用户是融合模型的管理员，可选择角色：普通管理员、超级管理员 2、如果创建的用户是融合模型的普通用户，可选择角色：普通用户
描述	支持填写用户描述信息
状态	1、有效状态：可正常登录平台 2、无效状态：不可登录平台

5. 点击**确定**，将跳转到用户管理列表

6.1.2 重置用户密码

操作步骤：

1. 管理员登录平台；
2. 选择**运营管理>人员与权限>用户管理**菜单；
3. 选择需要修改的用户，点击 ，选择**修改密码**按钮



4. 在修改密码的弹框中输入新密码和确认密码，即可重置平台用户密码

6.1.3 修改用户

操作步骤：

1. 管理员登录平台；
2. 选择**运营管理>人员与权限>用户管理**菜单；
3. 选择需要修改的用户，点击 ，选择**编辑**按钮
4. 可修改用户名称、组织架构、邮箱、手机号、用户描述和状态

6.1.4 删除用户

操作步骤：

1. 管理员登录平台；
2. 选择**运营管理>人员与权限>用户管理**菜单；
3. 选择需要修改的用户，点击 ，选择**删除**按钮
4. 弹框中点击确定，即可删除用户

6.1.5 绑定角色

当需要为用户绑定角色时，可对用户进行绑定角色操作。

操作步骤：

- 1. 管理员登录平台;
- 2. 选择**运营管理>人员与权限>用户管理**菜单;
- 3. 选择需要绑定角色的用户, 点击右侧的**绑定角色**按钮, 弹框中选择需要绑定的角色, 点击**确定**即可绑定角色

角色说明:

角色名称	DeepDeek 专区可查看和管理的菜单
普通管理员、超级管理员	全部菜单
普通用户	个人空间

6.1.6 解绑角色

操作步骤:

- 1. 管理员登录平台;
- 2. 选择**运营管理>人员与权限>用户管理**菜单;
- 3. 选择需要解绑角色的用户, 点击右侧列出的角色列表, 选择需要解绑的角色, 点击**解绑**按钮, 在确认弹框中点击**确定**, 即可解绑角色。

6.1.7 查看用户

支持查看平台用户。

操作步骤:

- 4. 管理员登录平台;
- 5. 选择**融合模型>管理>用户管理**菜单;
- 6. 选择**平台用户 tab** 页面支持查看: 用户名称、类型、角色、状态、邮箱、手机号、单位名称、所属用户组个数、创建时间、到期时间

说明: 邮箱和手机号默认脱敏展示

6.1.8 加入用户组

管理员可将平台用户加入用户组

操作步骤:

入口 1:

- 1. 管理员登录平台;
- 2. 选择**融合模型>管理>用户管理**菜单;
- 3. 选择**平台用户 tab** 页面;
- 4. 选择某一用户, 点击**用户组**按钮;

5. 弹框中在“可选用户组”区域选择需要加入的用户组，点击移动按钮，移入“已选用户组”区域，点击**确定**，即可将所选用户添加到用户组。

入口 2:

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择平台用户 tab 页面；
4. 点击进入某一用户详情，选择用户组 tab 页面，点击**加入用户组**按钮；
5. 弹框中在“可选用户组”区域选择需要加入的用户组，点击移动按钮，移入“已选用户组”区域，点击**确定**，即可将所选用户添加到用户组。

6.1.9 移除用户组

管理员可将平台用户移出用户组

操作步骤:

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**平台用户 tab** 页面；
4. 点击进入某一用户详情，选择用户组 tab 页面；
5. 选择需要移出的用户组，点击**移除**按钮，在确认弹框中点击确定，即可将用户移除用户组。

6.1.10 续期

管理员可为平台用户进行续期操作。

操作步骤:

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**平台用户 tab** 页面，选择某一用户，点击**续期**按钮；
4. 弹框中输入许可授权码，点击确定，即可为用户进行续期。

6.2 翼政通用户

6.2.1 查看用户

支持查看翼政通用户。

操作步骤：

1. 管理员登录平台；
2. 选择融合模型>管理>用户管理菜单；
3. 选择翼政通用户 tab 页面支持查看：用户名称、类型、状态、邮箱、手机号、单位名称、所属用户组个数、创建时间、到期时间

说明：邮箱和手机号默认脱敏展示

6.2.2 创建翼政通用户

支持创建翼政通用户

操作步骤：

1. 管理员登录平台；
2. 选择融合模型>管理>用户管理菜单；
3. 选择翼政通用户 tab 页面，点击添加用户按钮；
4. 在弹框中填写如下参数后，点击确定，即可创建用户。

管理员支持创建用户。参数说明：

参数名称	说明
用户名称	必填，名称长度在 1-30 个字符之间
密码	密码必须包含字母和数字，且长度不能小于 8 位
授权码	必填
邮箱	非必填
手机号	非必填
单位名称	非必填
用户组	非必填，选择可加入的用户组

6.2.3 批量导入

支持批量导入翼政通用户，导入成功后，用户可直接使用导入的密码登录翼政通应用。

操作步骤：

1. 管理员登录平台；

2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面，点击列表上方的**批量导入**按钮；
4. 在弹框中下载导入模板，按照模板提示信息填写相关信息后，再弹框中导入已填写的表格。

注意：

- 必须按照模板格式及字段填写用户信息后进行上传，一次仅可上传一个文件
 - 单次上传最多同时添加 500 条用户信息
 - 文件格式为 xlsx，请勿修改为其他格式
 - 需要将模板中的红字提示信息删除后再导入
5. 导入后，列表上方将展示导入详情，若导入失败，可点击查看详情。

注意：若导入的用户名称已存在会导入失败，若需要更新用户信息，需要将原有用户删除后再导入！



6. 将导入失败的条目按照要求修改正确后，单独重新上传。

6.2.4 修改用户

支持修改翼政通用户的邮箱和手机号。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面，选择某一用户，点击**更多>修改**按钮；
4. 在弹框中填写需要修改的参数后，点击**确定**，即可修改用户邮箱、手机号、单位名称。

6.2.5 重置密码

支持管理员为翼政通用户重置密码。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面，选择某一用户，点击**重置密码**按钮；
4. 在弹框中填写新密码后，点击**确定**，即可重置用户密码。

重置密码

×

*

密码

请输入密码

ⓘ

*

确认密码

请确认密码

ⓘ

取消

确定

6.2.6 启用/禁用用户

支持管理员启用或禁用翼政通用户，用户被禁用后无法登录翼政通。

禁用操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面，选择某一“正常”状态的用户，点击**更多>禁用**按钮；
4. 在确认弹框中，点击**确定**，即可将用户禁用，用户被禁用后无法登录翼政通。

启用操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面，选择某一“禁用”状态的用户，点击**更多>启用**按钮；
4. 在确认弹框中，点击**确定**，即可将用户启用，用户被启用后可正常登录翼政通。

6.2.7 批量启用/禁用用户

支持批量启用、禁用用户

批量启用操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面，选择多个翼政通用户，点击列表上方的**更多>启用**按钮；
4. 弹框中仅列出“禁用”状态的用户，点击确定，即可针对禁用状态的用户进行批量启用。

批量禁用操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面，选择多个翼政通用户，点击列表上方的**更多>禁用**按钮；
4. 弹框中仅列出“正常”状态的用户，点击确定，即可针对禁用状态的用户进行批量禁用。

6.2.8 删除用户

禁用状态的用户，支持删除。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面，选择某一“禁用”状态的用户，点击**更多>删除**按钮；
4. 在确认弹框中，点击**确定**，即可将用户删除，删除后无法恢复，请谨慎操作！。

6.2.9 批量删除用户

禁用状态的用户，支持批量删除。

操作步骤：

5. 管理员登录平台；
6. 选择**融合模型>管理>用户管理**菜单；
7. 选择**翼政通用户 tab** 页面，选择多个翼政通用户，点击列表上方的**更多>删除**按钮；
8. 弹框中仅列出“禁用”状态的用户，点击**确定**，即可针对禁用状态的用户进行批量删除。

6.2.10 加入用户组

管理员可将翼政通用户加入用户组

操作步骤：

入口 1：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面；
4. 选择某一用户，点击**用户组**按钮；
5. 弹框中在“可选用户组”区域选择需要加入的用户组，点击移动按钮，移入“已选用户组”区域，点击**确定**，即可将所选用户添加到用户组。

入口 2：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面；
4. 点击进入某一用户详情，选择用户组 tab 页面，点击**加入用户组**按钮；
5. 弹框中在“可选用户组”区域选择需要加入的用户组，点击移动按钮，移入“已选用户组”区域，点击**确定**，即可将所选用户添加到用户组。

组”区域，点击**确定**，即可将所选用户添加到用户组。

6.2.11 移除用户组

管理员可将翼政通用户移出用户组

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面；
4. 点击进入某一用户详情，选择用户组 tab 页面；
5. 选择需要移出的用户组，点击**移除**按钮，在确认弹框中点击确定，即可将用户移除用户组。

6.2.12 续期

管理员可为翼政通用户进行续期操作。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**翼政通用户 tab** 页面，选择某一用户，点击**更多>续期**按钮；
4. 弹框中输入许可授权码，点击确定，即可为用户进行续期。

6.3 访客用户

6.3.1 查看用户

支持查看访客用户。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**访客用户 tab** 页面支持查看：用户名称、类型、状态、邮箱、手机号、单位名称、创建时间

说明：邮箱和手机号默认脱敏展示

6.3.2 创建访客用户

支持创建访客用户

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**访客用户 tab** 页面，点击**添加用户**按钮；
4. 在弹框中填写如下参数后，点击**确定**，即可创建用户。

管理员支持创建用户。参数说明：

参数名称	说明
用户名称	必填，名称长度在 1-30 个字符之间
密码	密码必须包含字母和数字，且长度不能小于 8 位
邮箱	非必填
手机号	非必填
单位名称	非必填

6.3.3 批量导入

支持批量导入访客用户，导入成功后，平台用户发布应用时，选择发布范围，可支持选择到对应的访客用户。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**访客用户 tab** 页面，点击列表上方的**批量导入**按钮；
4. 在弹框中下载导入模板，按照模板提示信息填写相关信息后，再弹框中导入已填写的表格。

注意：

- 必须按照模板格式及字段填写用户信息后进行上传，一次仅可上传一个文件
 - 单次上传最多同时添加 500 条用户信息
 - 文件格式为 xlsx，请勿修改为其他格式
 - 需要将模板中的红字提示信息删除后再导入
5. 导入后，列表上方将展示导入详情，若导入失败，可点击查看详情。

注意：若导入的用户名称已存在会导入失败，若需要更新用户信息，需要将原有用户删除后再导入！

6. 将导入失败的条目按照要求修改正确后，单独重新上传。

6.3.4 修改用户

支持修改访客用户的邮箱和手机号。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**访客户户 tab** 页面，选择某一用户，点击**修改**按钮；
4. 在弹框中填写需要修改的参数后，点击**确定**，即可修改用户邮箱、手机号、单位名称。

6.3.5 重置密码

支持管理员为访客用户重置密码。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**访客户户 tab** 页面，选择某一用户，点击**重置密码**按钮；
4. 在弹框中填写新密码后，点击**确定**，即可重置用户密码。

重置密码

* 密码

请输入密码

* 确认密码

请确认密码

取消

确定

6.3.6 启用/禁用用户

支持管理员启用或禁用访客用户，用户被禁用后无法登录应用问答平台。

禁用操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**访客户户 tab** 页面，选择某一“正常”状态的用户，点击**更多>禁用**按钮；
4. 在确认弹框中，点击**确定**，即可将用户禁用，用户被禁用后无法访问应用问答平台。

启用操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**访客用户 tab** 页面，选择某一“禁用”状态的用户，点击**更多>启用**按钮；
4. 在确认弹框中，点击**确定**，即可将用户启用，用户被启用后可正常访问应用问答平台。

6.3.7 批量启用/禁用用户

支持批量启用、禁用用户

批量启用操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**访客用户 tab** 页面，选择多个访客用户，点击列表上方的**更多>启用**按钮；
4. 弹框中仅列出“禁用”状态的用户，点击**确定**，即可针对禁用状态的用户进行批量启用。

批量禁用操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**访客用户 tab** 页面，选择多个访客用户，点击列表上方的**更多>禁用**按钮；
4. 弹框中仅列出“正常”状态的用户，点击**确定**，即可针对禁用状态的用户进行批量禁用。

6.3.8 删除用户

禁用状态的用户，支持删除。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**访客用户 tab** 页面，选择某一“禁用”状态的用户，点击**更多>删除**按钮；
4. 在确认弹框中，点击**确定**，即可将用户删除，删除后无法恢复，请谨慎操作！。

6.3.9 批量删除用户

禁用状态的用户，支持批量删除。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单；
3. 选择**访客用户 tab** 页面，选择多个访客用户，点击列表上方的**更多>删除**按钮；
4. 弹框中仅列出“禁用”状态的用户，点击**确定**，即可针对禁用状态的用户进行批量删除。

7 企业知识库

支持管理员创建企业知识库，并可将企业知识库共享给用户组使用。共享成功后，用户组内的用户即可查看对应的企业知识库，并可选择相应的企业知识库进行问答。

企业知识库相比个人知识库而言，不同之处在于完全由企业管理员进行管理，包括上传文件、删除文件等，设置对应的共享范围。个人用户不可针对企业知识库中的文件进行上传和修改。

7.1 创建企业知识库

管理员可创建企业知识库

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 点击**创建知识库**按钮；
4. 在弹框中输入如下参数，填写完成后，即可创建知识库。

参数	说明
知识库名称	1-32 字符，支持使用中文、字母、数字、下划线（_）、中划线（-）。不支持用数字、连字符（-）、下划线（_）作为命名开头，不能包含续连字符（--）
知识库描述	不超过 255 字符
索引模式	高质量、经济
Embedding 模型	仅高质量索引模式支持选择
检索设置	● 经济：选择 TopK ● 高质量：支持选择向量检索、全文检索、混合检索

7.2 设置企业知识库

管理员可针对已创建的企业知识库进行设置

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择设置 tab 页面，进入设置页面；
4. 可编辑知识库名称、知识库描述，选择索引模式。

注意：当知识库中的文件存在“索引中”状态，则不可切换 Embedding 模型、也不可切换

索引模式。

索引模式介绍：

索引模式	说明	可设置参数
高质量	通过调用 Embedding 模型进行处理，以在用户查询时提供更高的准确度 注意： ● 前提需要在部署 Embedding 模型才能使用 ● 如需使用 Rerank 模型，则需要提前部署 Rerank 模型	3、选择 Embedding 模型 4、配置检索设置，包括：向量检索、全文检索、混合检索。 ● 向量检索：可设置开启和关闭 Rerank 模型，支持配置 TopK 和 Score 阈值 ● 全文检索：可设置开启和关闭 Rerank 模型，当关闭 Rerank 模型时，可配置 TopK，当开启 Rerank 模型，可配置 TopK 和 Score 阈值 ● 混合检索：支持选择权重设置和 Rerank 模型。支持配置 TopK 和 Score 阈值
经济	使用离线的向量引擎、关键词索引等方式，降低了准确度但无需花费 Token	选择倒排索引，配置 top K

参数说明：

参数	说明
倒排索引	倒排索引是一种用于高效检索的结构。按术语组织，每个术语指向包含它的文档或网页
向量检索	通过生成查询嵌入并查询与其向量表示最相似的文本分段
全文检索	索引文档中的所有词汇，从而允许用户查询任意词汇，并返回包含这些词汇的文本片段
混合检索	同时执行全文检索和向量检索，并应用重排序步骤，从两类查询结果中选择匹配用户问题的最佳结果，用户可以选择设置权重或配置重新排序模型。
Rerank 模型	重排序模型将根据候选文档列表与用户问题语义匹配度进行重新排序，从而改进语义排序的结果
top K	用于筛选与用户问题相似度最高的文本片段。系统同时会根据选用模型上下文窗口大小动态调整分段数量。
Score 阈值	用于设置文本片段筛选的相似度阈值。
权重设置	通过调整分配的权重，重新排序策略确定是优先进行语义匹配还是

	关键字匹配。
--	--------

7.3 共享企业知识库

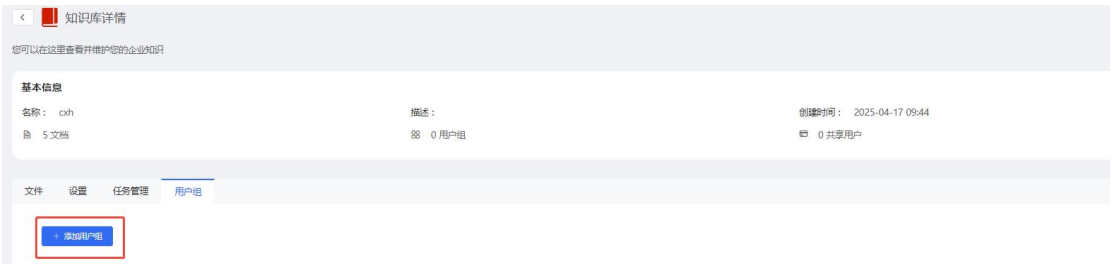
管理员可选择已创建的企业知识库进行共享。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 入口 1：选择某一知识库，点击右下角区域按钮选择**共享**；



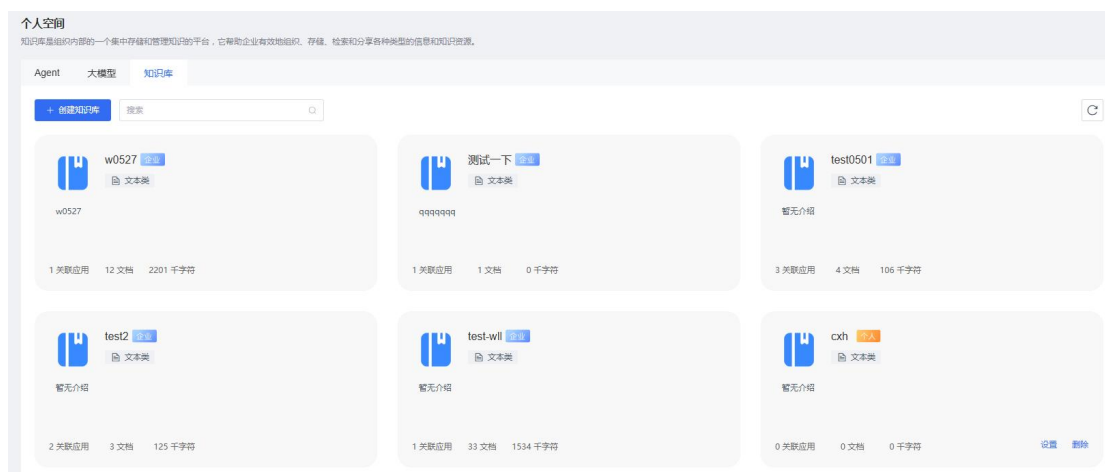
入口 2：点击进入知识库详情，选择用户组 tab 页面，点击**添加用户组**按钮；



4. 弹框中在“可选用户组”区域选择需要共享的用户组，点击移动按钮，移入“已选用户组”区域，点击**确定**，即可将此企业知识库共享给所选用户组。



5. 被共享的用户登录平台后，即可在个人空间>知识库中看到已共享的企业知识库。



若为翼政通用户，登录翼政通，在知识库选择时也可看到此知识库



7.4 删除企业知识库

管理员可删除企业知识库，删除知识库是不可以逆的。用户将无法再访问您的知识库，所有的配置和日志将被永久删除。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一知识库，点击右下角区域按钮选择**删除**；



4. 在删除确认弹框中点击确定，即可删除企业知识库。

7.5 文件管理

7.5.1 添加文件

企业知识库创建完成后可添加文件。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 点击**添加文件**按钮；
5. 在添加文件页面，拖拽文件至此或点击上传按钮，选择如下文件格式：TXT、MARKDOWN、MD、PDF、HTML、HTM、DOCX、CSV。且每个文件不能超过 15M。单次最多选择 100 个文件；文件选择完成后，点击**下一步**，进入文本分段清洗页面；
6. 在文本分段清洗页面可选择分段设置和检索设置。其中索引方式为知识库已经设置好的索引方式，如需修改，请在知识库的设置中修改。其他参数配置完成后，可点击**保存并**

处理按钮。

7. 待文档处理完成后自动返回文件 tab 页面，可查看文件名称、字符数上传时间和状态。

7.5.2 URL 上传

除了手动添加文件，还提供了 URL 批量上传的功能。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 点击 **URL 上传**按钮；
5. 在 URL 导入文件弹框中填写下载 URL 链接，每行一个 URL。当前支持的文档格式包括：TXT、MARKDOWN MD、PDF、HTML、HTM、DOCX、CSV。
6. 输入 URL 地址后，点击**确定**按钮，可在任务管理 tab 页面中查看文件的上传状态。

7.5.3 禁用/启用文件

针对已上传的文件，可对其进行禁用和启用操作，禁用后此文件将不会被问答时引用。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 选择某一文件，点击**更多>禁用或启用**按钮，可对此文件进行禁用和启用操作。

7.5.4 修改文件

针对已上传的文件，可对其进行重命名操作。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 选择某一文件，点击**修改**按钮，可对此文件进行重命名操作。

7.5.5 删除文件

针对已上传的文件，可对其进行删除操作。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 选择某一文件，点击**更多>删除**按钮，在确认弹框中点击确定，可对此文件进行删除操作。

7.5.6 批量删除文件

针对已上传的文件，可对其进行删除操作。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 选择多个文件，点击列表上方的**批量删除**按钮，在确认弹框中点击确定，可对多个文件进行删除操作。

7.5.7 预览文件

针对已上传的文件，可对其进行预览操作。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 点击某一文件名称，即可预览该文件分段后的内容

7.5.8 重新索引文件

针对错误状态和长时间索引中的文档，支持重新索引操作。帮助用户再次对文件进行分段和索引操作。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 选择某一“错误”状态或长时间处于“索引中”状态的文件，点击“**重新索引**”按钮，弹框中点击确定，即可重新开始索引。

注意：单个文档十分钟内仅支持一次重新索引操作。

7.5.9 批量重新索引文件

针对错误状态和长时间索引中的文档，支持批量重新索引操作。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 选择多个文件，点击列表上方的**批量重新索引**按钮，仅列出针对"错误"和"索引中"状态的文件，点击确定，即可重新开始索引。

注意：单个文档十分钟内仅支持一次重新索引操作。

7.5.10 重新分段

针对正常和错误状态的文档支持重新分段操作。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 选择某一“错误”状态或“可用”状态的文件，点击“**更多>重新分段**”按钮；
5. 选择分段设置，点击确定，即可针对此文档进行重新分段。

7.5.11 编辑分段

支持用户编辑启用状态的分段内容，禁用状态的分段不可编辑

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 选择某一文件，进入文件分段详情；
5. 选择某一段，在弹框中点击**编辑**按钮；
6. 在文本框中进行修改，点击确定，即可编辑此分段内容。

7.5.12 禁用/启用分段

支持用户启用和禁用分段

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 选择某一文件，进入文件分段详情；
5. 选择某一分段，可点击启用或禁用开关，即可对此分段进行启用或禁用操作，分段被禁用后则无法被检索到。

7.5.13 删除分段

支持用户删除分段。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择文件 tab 页面；
4. 选择某一文件，进入文件分段详情；
5. 选择某一分段，可点击删除按钮，即可删除此分段。删除后若需要恢复，可针对此文档进行重新分段操作。

7.6 任务管理

针对 URL 上传的文件，可在任务管理模块中查看文件上传的进度。每次点击 URL 上传弹框确定后，记录为一次任务，任务列表可查看每次任务中上传的文件总数和完成情况。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择任务管理 tab 页面；
4. 可查看任务 ID、任务类型、任务状态、上传文件总数、已完成、处理中、失败、开始时间、更新时间

任务状态说明：

- 全部完成：全部文件上传成功

- 部分成功：部分文件上传成功
 - 失败：全部文件上传失败
 - 上传中：上传进行中
5. 若发生失败任务，可查看上传失败的文件个数，点击后可展示失败的 URL，并且可点击**失败重传**按钮，进行重新上传。

7.7 用户组管理

支持查看和管理企业知识库共享的用户组。

7.7.1 添加用户组

同共享企业知识库功能，操作步骤可参考 7.3 章节

7.7.2 移除用户组

针对已共享的用户组，可进行移除操作，移除后该用户组内的用户将无法再看到此企业知识库。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**菜单；
3. 选择某一企业知识库，进入知识库详情，选择用户组 tab 页面；
4. 选择某一用户组，点击**移除**按钮，在确认弹框中点击**确定**，即可将此用户组移除该企业知识库。

8 用户组管理

用户组用于企业知识库共享使用，可以将用户加入到不同用户组，不同用户组加入不同的企业知识库，做到不同企业知识库之间的用户隔离。

8.1 创建用户组

管理员可创建用户组

操作步骤：

1. 管理员登录平台；

- 2. 选择**融合模型>管理>用户组管理**菜单；
- 3. 点击**创建用户组**按钮；
- 4. 在弹框中输入用户组名称，选择企业知识库，描述，点击**确定**，即可创建用户组。

参数名称	说明
用户组名称	必填，长度为 2-32 字符 支持使用中文、字母、数字、下划线（_）、中划线（-） 不支持用数字、连字符（-）、下划线（_）作为命名开头，不能包含续连字符（--）
企业知识库	非必填项，可多选
描述	非必填，0-128 字符

8.2 添加用户

管理员可为用户组添加用户

操作步骤：

入口 1：

- 1. 管理员登录平台；
- 2. 选择**融合模型>管理>用户组管理**菜单；
- 3. 选择某一用户组，点击**添加用户**按钮；
- 4. 弹框中在“可选用户”区域选择需要添加的用户，点击移动按钮，移入“已选用户”区域，点击**确定**，即可将所选用户添加到用户组。

入口 2：

- 1. 管理员登录平台；
- 2. 选择**融合模型>管理>用户组管理**菜单；
- 3. 点击进入某一用户组详情，选择用户 tab 页面，点击**添加用户**按钮；
- 4. 弹框中在“可选用户”区域选择需要添加的用户，点击移动按钮，移入“已选用户”区域，点击**确定**，即可将所选用户添加到用户组。

8.3 移除用户

针对已加入用户组的用户，可进行移除操作，移除后，此用户原来所属用户组共享的知识库，将无法看到。

操作步骤：

- 1. 管理员登录平台；
- 2. 选择**融合模型>管理>用户组管理**菜单；

3. 选择某一用户组，进入用户组详情，选择用户 tab 页面；
4. 选择某一用户，点击**移除**按钮，在确认弹框中点击**确定**，即可将此用户移除该用户组。

8.4 加入企业知识库

管理员可将某一用户组加入到多个企业知识库。

操作步骤：

入口 1：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户组管理**菜单；
3. 选择某一用户组，点击**企业知识库**按钮；
4. 弹框中在“可选企业知识库”区域选择需要添加的企业知识库，点击移动按钮，移入“已选企业知识库”区域，点击**确定**，即可将用户组加入到所选企业知识库。

入口 2：

1. 管理员登录平台；
2. 选择**管理>用户组管理**菜单；
3. 点击进入某一用户组详情，选择企业知识库 tab 页面，点击**添加企业知识库**按钮；
4. 弹框中在“可选企业知识库”区域选择需要添加的企业知识库，点击移动按钮，移入“已选企业知识库”区域，点击**确定**，即可将用户组加入到所选企业知识库。

8.5 移除企业知识库

针对已加入的企业知识库，可进行移除操作，移除后，此用户组内的所有用户将无法查看此企业知识库。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户组管理**菜单；
3. 选择某一用户组，进入用户组详情，选择企业知识库 tab 页面；
4. 选择某一企业知识库，点击**移除**按钮，在确认弹框中点击**确定**，即可将此企业知识库移除该用户组。

8.6 修改用户组

管理员支持修改用户组名称和描述。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户组管理**菜单；
3. 选择某一用户组，点击**修改**按钮；
4. 在弹框中可修改用户组名称和描述。

8.7 删除用户组

管理员支持删除用户组，删除后自动解除用户组、企业知识库、用户的关联。用户和企业知识库不会被删除！

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>用户组管理**菜单；
3. 选择某一用户组，点击**删除**按钮；
4. 在弹框中点击**确定**，即可删除用户组。

9 模型管理

支持管理员添加和管理私有化和公网 API 模型模型，添加后，普通用户创建应用时可选择此模型。

9.1 私有化模型

9.1.1 添加模型

管理员支持添加 LLM、Text Embedding、Rerank 模型。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>模型管理**菜单，选择私有化模型 Tab 页面；
3. 点击**添加模型**按钮；
4. 在弹框中填写相应参数，点击**确定**，即可添加模型。

LLM 参数说明：

参数名称	说明
模型类型	单选：LLM、Text Embedding、Rerank

模型名称	任意字符，1-50 字符
API Key	任意字符，1-100 字符
模型地址	http 或 https://服务器 IP 或域名 + 端口
模型生成模式	仅支持对话
模型上下文长度	默认 4096，取值范围：100-10240，整数
最大 token 上限	默认 4096，取值范围：100-10240，整数
Vision 支持	默认不支持，可选：支持、不支持
部署方式	仅支持 VLLM
MCP 支持	可选择不支持和不支持

Test Embedding 模型参数说明：

参数名称	说明
模型类型	单选：LLM、Text Embedding、Rerank
模型名称	任意字符，1-50 字符
API Key	任意字符，1-100 字符
模型地址	http 或 https://服务器 IP 或域名 + 端口
部署方式	支持 VLLM 和 Ollama

Rerank 模型参数说明：

参数名称	说明
模型类型	单选：LLM、Text Embedding、Rerank
模型名称	任意字符，1-50 字符
API Key	任意字符，1-100 字符
模型地址	http 或 https://服务器 IP 或域名 + 端口
部署方式	支持 Text-Embedding-Inference

9.1.2 设置模型

已创建完成的模型支持修改相关参数。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>模型管理**菜单，选择私有化模型 Tab 页面；
3. 选择某一模型，点击**设置**按钮；
4. 在弹框中修改模型的相关参数，点击**确定**，即可修改模型参数。

参数说明：

- LLM 模型支持修改如下参数：API Key、模型地址、模型上下文长度、最大 token 上

限、Vision 支持

- Text Embedding 模型支持修改如下参数：API Key、模型地址
- Rerank 模型支持修改如下参数：API Key、模型地址

说明：

管理员设置的模型参数为模型全局默认的参数，若用户在个人空间修改参数后按照个人设置的为准！

9.1.3 禁用模型

被应用正在使用的模型无法禁用！模型禁用后，用户将无法使用此模型。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型**>**管理**>**模型管理**菜单，选择私有化模型 Tab 页面；
3. 选择某一启用的模型，点击**禁用**按钮；
4. 在弹框中点击**确定**，即可禁用模型。

注意：被应用正在使用的模型无法禁用！

9.1.4 启用模型

模型启用后，用户可选择此模型进行问答。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型**>**管理**>**模型管理**菜单，选择私有化模型 Tab 页面；
3. 选择某一禁用的模型，点击**启用**按钮；
4. 在弹框中点击**确定**，即可启用模型。

9.1.5 删除模型

禁用的模型支持删除。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型**>**管理**>**模型管理**菜单，选择私有化模型 Tab 页面；
3. 选择某一禁用的模型，点击**删除**按钮；
4. 在弹框中点击**确定**，即可删除模型。

9.2 公网 API 模型

9.2.1 添加模型

管理员支持添加 DeepSeek 公网 API 模型, 目前支持 deepseek-chat、deepseek-reasoner 两个模型

操作步骤:

1. 管理员登录平台;
2. 选择**融合模型>管理>模型管理**菜单, 选择公网 API 模型 Tab 页面;
3. 点击**添加模型**按钮;
4. 在弹框中填写相应参数, 点击**确定**, 即可添加模型。

参数名称	说明
模型供应商	单选: 目前仅支持 deepseek
API Key	任意字符

9.2.2 禁用模型

被应用正在使用的模型无法禁用! 模型禁用后, 用户将无法使用此模型。

操作步骤:

1. 管理员登录平台;
2. 选择**融合模型>管理>模型管理**菜单, 选择公网 API 模型 Tab 页面;
3. 选择某一启用的模型, 点击**禁用**按钮;
4. 在弹框中点击**确定**, 即可禁用模型。

注意: 被应用正在使用的模型无法禁用!

9.2.3 启用模型

模型启用后, 用户可选择此模型进行问答。

操作步骤:

1. 管理员登录平台;
2. 选择**融合模型>管理>模型管理**菜单, 选择公网 API 模型 Tab 页面;
3. 选择某一禁用的模型, 点击**启用**按钮;
4. 在弹框中点击**确定**, 即可启用模型。

9.2.4 删除模型

禁用的模型支持删除。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>模型管理**菜单，选择公网 API 模型 Tab 页面；
3. 选择某一禁用的模型，点击**删除**按钮；
4. 在弹框中点击**确定**，即可删除模型。

注意：当模型被删除后，再次添加模型可加回；若未删除，重复添加同一模型供应商时，模型不会新增。

10 企业模板

此功能需要和翼政通关联使用！当用户使用翼政通软件时，支持管理员从平台上传企业模板后，翼政通用户使用翼政通时，即可在 **公文模板- 企业定制** 菜单中查看管理员上传的企业模板

10.1 添加企业模板

支持管理员上传企业模板

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业模板**菜单；
3. 点击**添加模板**按钮；
4. 输入模板名称、上传模板图片、上传模板文件

模板名称要求：1-32 位字符，不支持用数字、连字符（-）、下划线（_）作为命名开头，不能包含连续字符（--）

模板图片要求：仅支持 jpg、jpeg、png 格式

模板文件要求：仅支持 .wps、.wpt、.doc、.dot、.docx 格式

5. 选择完成后，点击确认，即可添加模板

10.2 启用企业模板

新添加后的模板默认为禁用状态，需手动启用。启用后，翼政通用户才能正常使用。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业模板**菜单；
3. 选择某一个禁用状态的模板，点击**启用**按钮；
4. 弹框中点击确认按钮，即可启用此模板。

10.3 禁用企业模板

启用的模板支撑禁用。禁用后，翼政通用户将无法查看到此模板。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业模板**菜单；
3. 选择某一个启用状态的模板，点击**禁用**按钮；
4. 弹框中点击确认按钮，即可禁用此模板。

10.4 下载企业模板

支持下载企业模板。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业模板**菜单；
3. 选择某一个模板，点击**下载**按钮；
4. 根据浏览器提示信息，确认下载，即可将模板下载至本地查看。

10.5 删除企业模板

禁用状态的模板支持删除

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>管理>企业模板**菜单；
3. 选择某一个禁用状态的模板，点击**删除**按钮；
4. 弹框中点击确认按钮，即可删除此模板。

11 服务器监控

11.1 服务器性能监控

支持查看服务器的监控信息，包括：

- CPU 使用率、CPU 的空闲时间占比、CPU 的中断时间占比、CPU 的 IO 等待时间占比、CPU 的软中断时间占比、CPU 的系统态运行时间占比、CPU 的用户态运行时间占比、CPU 每分钟的负载值
- 内存使用率、
- 磁盘读速率、磁盘读请求速率、磁盘写速率、磁盘写请求速率
- 网络流入速率、网络流出速率、网络上行包数、网络下行包数

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>监控>服务器**菜单；
3. 选择服务器 tab 页面，可查看服务器 CPU 使用率、内存使用率和磁盘使用率；
4. 选择某一服务器，点击查看监控详情，可查看每个服务器具体的监控指标。

11.2 服务器 GPU 卡监控

支持查看服务器 GPU 卡监控信息，包括：

- 服务器 GPU 整机监控：GPU 总功耗、GPU 平均温度、GPU 平均使用率、GPU 显存平均使用率
- 服务器单张 GPU 卡监控：GPU 功耗、GPU 温度、GPU 使用率、GPU 显存使用率

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>监控>服务器**菜单；
3. 选择 GPU 卡 tab 页面，可查看服务器 GPU 总功耗、GPU 平均温度、GPU 平均使用率、GPU 显存平均使用率；
4. 选择某一服务器，点击查看监控详情，可查看每个服务器单卡和整机的 GPU 监控信息。

12 中间件监控

支持查看容器、数据库、redis 缓存的监控信息

12.1 容器监控

支持查看容器监控信息，包括查看每个容器的 CPU 使用率、内存使用率。详细监控中，可查看实时和历史的监控信息，监控指标包括：CPU 使用率、内存使用率、网络流入速率、网络流出速率、磁盘读速率、磁盘写速率、磁盘读 IOPS、磁盘写 IOPS。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>监控>中间件**菜单；
3. 选择**容器监控** tab，可查看所有容器的状态、CPU 使用率和内存使用率；
4. 选择某一容器，点击查看监控详情，可查看每个容器具体的监控指标。

12.2 数据库监控

支持查看实时和历史的数据库监控信息，监控指标包括：连接数、缓存命中率、死锁数、冲突数、数据库读写时间统计、生成临时文件统计、数据查询统计、数据更新统计、事务统计。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>监控>中间件**菜单；
3. 选择**数据库监控** tab，可查看数据库的具体监控信息。

12.3 redis 缓存监控

支持查看实时和历史的 redis 缓存监控信息，可查看运行时间、客户端数量。监控指标包括：执行最多的前五个指令、已/未设置过期时间、每个 DB 的 KEY 数量、命令执行个数、网络 IO 性能、内存使用量

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>监控>中间件**菜单；
3. 选择 **redis 监控** tab，可查看 redis 缓存的具体监控信息。

13 助手管理

助手管理功能提供管理员对翼政通 APP 的助手进行管理，可开启和关闭是否开启对应助手功能。目前提供政策问答助手功能。政策问答助手主要功能为：管理员可上传本企业的相关政策或公共文档，上传完成后，用户在可翼政通 APP 上进行问答。

13.1 政策问答助手

13.1.1 文件管理

管理政策问答助手中的文档。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>助手管理**菜单；
3. 选择**政策问答助手**，点击**文件管理**
4. 支持添加文件、配置索引方式、文件名称修改、重新索引、启用/禁用、重新分段、删除功能。
 - 添加文件：可参考 7.5.1 章节
 - 配置索引模式：支持修改检索设置
 - 修改：可参考 7.5.4 章节
 - 重新索引：可参考 7.5.8 章节
 - 启用/禁用文件：可参考 7.5.12 章节
 - 重新分段：可参考 7.5.10 章节
 - 删除文件：可参考 7.5.5 章节

13.1.2 切换模型

支持切换政策问答助手问答时使用的模型。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>助手管理**菜单；
3. 选择**政策问答助手**，点击**切换模型**；
4. 可选择私有化模型和公网 API 模型
5. 选择后，点击**确定**，立即生效

13.1.3 设置

支持设置此助手的名称和头像。

操作步骤：

1. 管理员登录平台；

2. 选择**融合模型>助手管理**菜单；
3. 选择**政策问答助手**，点击**设置**；
4. 可重新修改名称和头像，修改后翼政通 APP 展示的名称和头像也会一起修改

13.1.4 禁用

禁用后翼政通 APP 将无法查看此助手入口。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>助手管理**菜单；
3. 选择**政策问答助手**，若应用为启用状态，可点击**禁用**；
4. 弹框中点击**确定**，即可禁用此助手

13.1.4 启用

启用后翼政通 APP 刷新后，可查看此助手入口。

操作步骤：

1. 管理员登录平台；
2. 选择**融合模型>助手管理**菜单；
3. 选择**政策问答助手**，若应用为禁用状态，可点击**启用**；
4. 弹框中点击**确定**，即可启用此助手

14 最佳实践

14.1 管理员构建企业知识库共享用户使用

场景描述：

企业用户在使用翼政通或 DS 模块的的知识库功能时，企业管理员可以发布一些企业自己的知识库，例如公司规定、某些系统的使用规范或某些系统的常见问题等。同时，针对不同部分的知识库不同，也可设置不同用户的可见范围。

操作步骤：

(1) 创建企业知识库

1. 管理员登录平台；
2. 选择**融合模型>管理>企业知识库**管理菜单；
3. 点击**创建企业知识库**按钮，填写相应参数创建企业知识库，具体参数填写可参考 8.1 章节；
4. 点击某一个企业知识库上传企业文件，具体操作可参考 8.5 文件管理章节

(2) 创建用户组

1. 管理员登录平台；
2. 选择**融合模型>管理>用户组**管理菜单；
3. 点击**创建用户组**按钮，若需要针对不同的企业知识库设置不同用户的可见权限，则按照需要创建多个用户组；
4. 选择某一用户组，点击**添加用户**按钮；
5. 将对于企业知识库可见范围一致的用户加入到同一用户组

(3) 企业知识库关联已创建的用户组

1. 管理员登录平台；
2. 选择**融合模型>管理>用户组**管理菜单；
3. 选择某一创建的用户组，点击 **企业知识库**，选择需要共享的企业知识库；或选择某一知识库详情，点击**添加用户组**，选择需要共享的用户组。
4. 共享完成后，对应用户组内的用户登录后即可看到此企业知识库

14.2 构建企业应用并发布到指定用户使用

场景描述：

企业构建企业应用后，可选择发布给普通用户直接使用。可支持发布多个企业应用发布给某些指定用户。

操作步骤：

(1) 创建访客用户

访客用户仅支持使用共享给其的应用，不可创建和编排应用。

1. 管理员登录平台；
2. 选择**融合模型>管理>用户管理**菜单，选择访客页签；

3. 支持导入和创建访客用户，具体操作参数可参考 7.3

(2) 创建企业应用

1. 管理员登录平台；
2. 选择**融合模型>个人空间**菜单，
3. 点击创建应用并编排，具体操作可参考章节 4

(3) 选择应用发布范围

1. 管理员登录平台；
2. 选择**融合模型>个人空间**菜单，
3. 选择需要发布的应用，点击进入后，选择发布范围，根据需要可选择全部用户后指定用户。
4. 被选择共享的用户登录平台后，即可查看到被共享的应用，支持直接点击使用问答。