

SAG 企业版

产品帮助文档

生成日期：2026 年 09 月 14 日

本文档由帮助中心自动导出，供离线查阅

目录

产品介绍 1

快速开始 6

普通成员使用

知识库

进入知识库与上传文档 9

同步外部文档 14

邀请成员 22

图谱 25

搜索 27

问答 30

SAG 开放集成

开放连接 33

消息渠道 37

MCP 与 HTTP API 43

CLI 接入 45

超级管理员使用

管理后台概览 56

人员管理与创建账号 58

配置模型与 API Key 60

配置外部数据源 65

配置检索方式与处理参数 71

计划与套餐审批 73

任务中心 75

任务、审计与回收站 77

审计日志 80

回收站 82

企业部署与升级 84

更新日志 88

常见问题 93

产品介绍

SAG 企业知识增强引擎——面向复杂业务问题的关系型知识库，让企业已有的知识找得到、找得准、敢拿去用。

SAG 是智跃 Zleap 推出的企业知识增强引擎，一套面向复杂业务问题的关系型知识库。它把制度、工艺、合同、项目与沟通记录等企业资料组织成可检索、可溯源的知识底座，让员工用一句话提问，就能得到带原文依据的业务答案——而不是一堆还要自己翻的文件列表。

> **一句话理解 SAG**
>
> 找到相关文件，不等于得到可执行答案。SAG 的目标是把「资料找到了，答案仍然没有」变成「审批路径、材料清单与原文依据一次给全」。

为什么需要 SAG

企业不是没有资料，而是系统没把正确的那份找出来。普通知识库通常卡在三件事上：

瓶颈	具体表现
信息被切碎	制度、工艺、项目分散在不同文件与系统里，同一件事要翻三个系统才凑得齐
关系被丢掉	普通检索只比「像不像」，读不懂金额区间、区域条款、项目类型这些条件之间的关联
答案无法核验	回答给不出明确出处，员工不敢直接拿去审批、开工或答复客户，最后还是回去问人

知识库效果的上限，首先由召回是否正确决定——模型可以写得很流畅，但检索错了，答案只会更像真的。

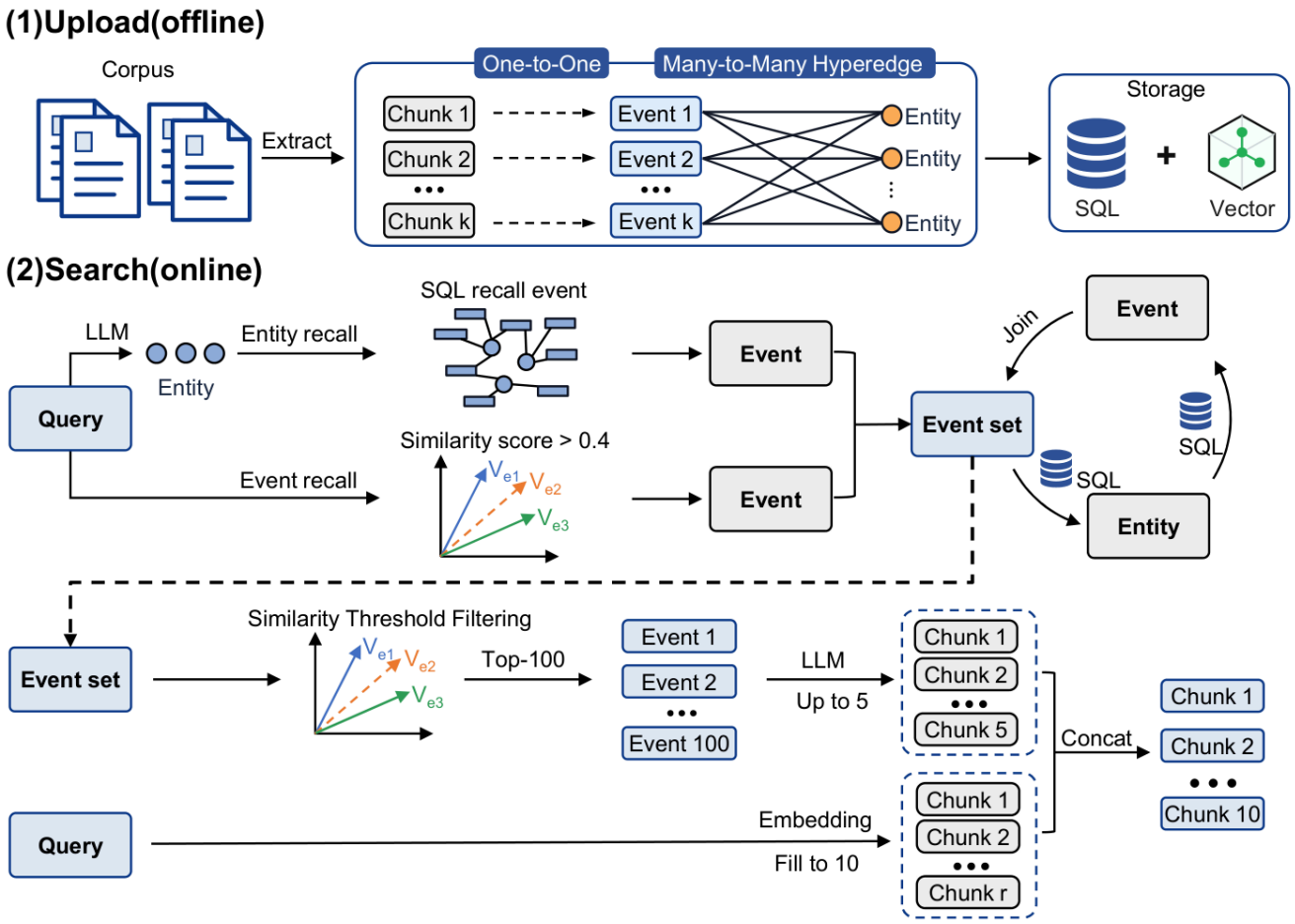
SAG 的做法：组织完整业务事项

SAG 不只切分文档，而是把资料整理成完整的业务事项：

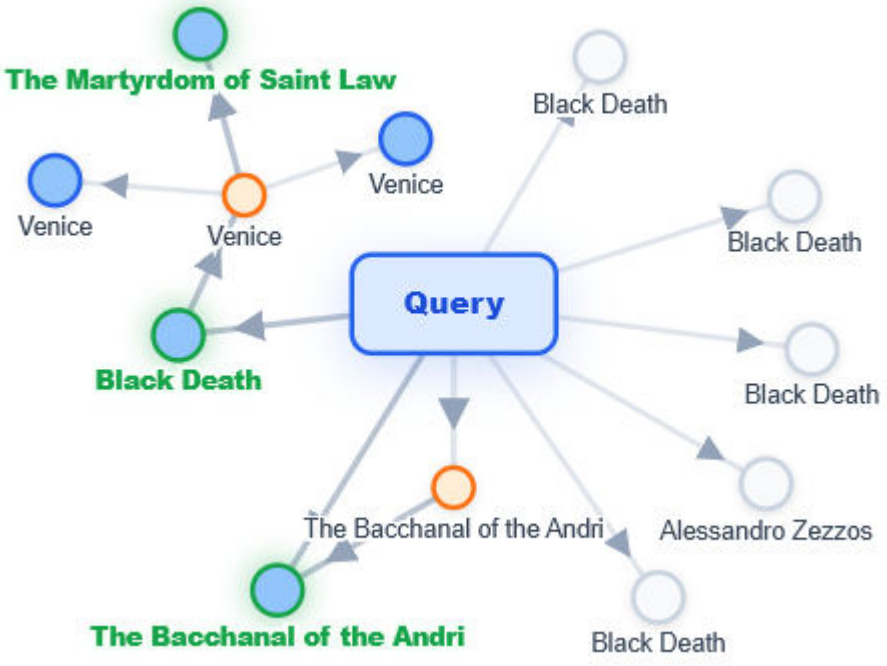
1. 文档原样导入——制度、工艺、合同、会议纪要与网页不做破坏性切分，原文位置全程保留；

2. 整理成完整事项——每段内容成为一件能独立读懂的事：谁、什么时间、什么条件、按哪条规则、出自哪份文件；
3. 抽取关键实体——人员、设备、金额、项目与制度条款等实体让分散的文件之间有了连接点；
4. 提问时才展开关系——只把与当前问题真正相关的事项拉进来，不需要预先维护全局图谱；
5. 结论附原文依据——回答连同原文片段一起返回，点开即可核对出自哪份文件、哪一条。

资料进来一次、整理一次，之后每一次提问都在复用这份红利。



上图来自 SAG 技术论文（arXiv:2608.12129）：离线阶段（Upload）把每个文本块整理成语义完整的事件，并与抽取出的实体构成多对多的「动态超边」索引；在线阶段（Search）从问题出发做实体召回与事件召回，通过共享实体像数据库 Join 一样把相关事件连起来，最后筛选出证据交给大模型。



这张论文中的检索示例展示了 SAG 的关键机制：初始检索命中两个相关事件后，二者共享的实体（图中的 Venice）会激活一条「潜在超边」，把仅靠向量相似度找不到的第三份关键文档也带回来——这正是多跳问题上 SAG 领先普通检索的原因。

检索效果（公开可查）

- 三大公开多跳检索基准 9 项 Recall 指标全部第一：在 HotpotQA、2WikiMultiHopQA、MuSiQue 三个数据集的 Recall@1 / @2 / @5 共 9 项指标上均取得最佳，结果来自公开论文 arXiv:2608.12129，配置与数据集可复现；
- 推理链越长优势越大：在最难的 MuSiQue 数据集上 Recall@5 达到 80.36%，领先最强基线 11.52 个百分点；
- 综合数据集召回率 90–99%，显著高于常见 RAG 方案的 70–85%；
- 5 亿级生产数据验证：在真实生产数据规模下稳定运行、秒级检索。

权限与数据边界

知识可以被调用，也必须守住边界：

- 权限：同一套知识，不同人看到的范围不同；搜索和问答只返回当前账号有权限且已发布的资料，Agent 继承调用者的相同权限；
- 部署：部署在客户服务器或私有云，支持内网离线环境；知识数据保留在客户环境，使用本地模型时可全程内网运行；
- 接入：UI 开箱即用，同时提供 REST API、OpenAI 兼容接口和 MCP，可接入既有业务系统。

典型应用场景

一套 SAG 覆盖企业最常见的四类知识需求：

知识类型	覆盖内容	典型提问
管理知识	制度、规范、合同与审批流程	这件事按现行制度，应该谁批、交什么材料？
生产研发知识	工艺参数、设备手册、故障记录	这个异常以前发生过吗，应该先查什么？
业务知识	客户背景、历史方案、报价与服务记录	类似客户过去用过什么方案，效果如何？
组织知识	项目经验、岗位做法、培训材料与复盘总结	这项工作以前谁做过，有哪些坑要避开？

快速导航

登录
登录后认识搜索、对话、知识库、集成四个主入口。

知识库
进入知识库，上传文档，等待发布，检查解析内容。

搜索
快速/精确检索，AI 总结，引用与原文核对。

问答
基于证据回答，引用可核对，会话可管理。

SAG 开放集成

创建开放连接，交给 Agent（MCP）或应用（HTTP API）。

管理后台

人员、模型与 API Key、检索、计划、任务、审计、回收站。

角色说明

角色	说明
普通成员	使用企业版进行资料检索、文档问答和知识库协作
知识库负责人	在具体知识库内管理文档与成员
超级管理员	创建成员账号、管理企业配置与开放连接

⚠ 关于账号注册

企业版默认由超级管理员创建账号并交付一次性凭证；开放注册时激活也需要管理员审批。

快速开始

三分钟上手：上传文档、搜索资料、向知识库提问。

SAG 企业版的核心工作流分三步：把资料放进知识库，用搜索找到它，用问答直接问它。

上传文档

进入左侧【知识库】，打开目标知识库，点击【上传文档】或直接拖入文件。文档经过解析、建索引等阶段后状态变为【已发布】——只有已发布的文档才会出现在搜索和问答结果中。



搜索企业资料

进入左侧【搜索】，输入关键词（用 `@` 可限定知识库范围）。页面顶部显示 AI 总结，下方列出相关证据条目，点击任意条目打开来源面板核对原文。



向知识库提问

进入左侧【对话】，选择知识范围后输入问题。助手会主动检索企业文档作为依据；回答完成后点击引用来源可展开引用侧栏，按层次核对原文出处。



下一步

三个入口的完整操作说明分别在知识库、搜索和问答章节。

进入知识库与上传文档

找到有权限访问的知识库，上传文档并处理重复文件，确认资料进入可检索状态。

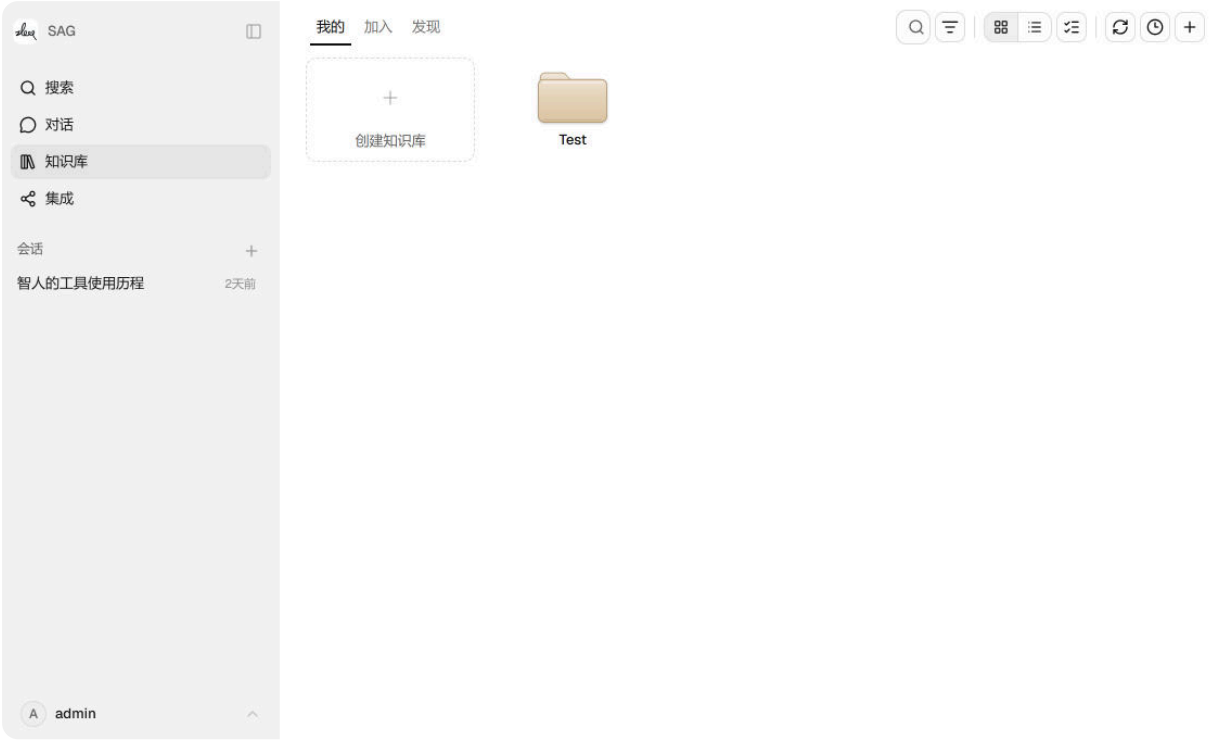
场景：查看本人有权限访问的企业资料，并将新资料放入知识库供搜索和问答使用。

角色：已获授权的普通成员、知识库负责人。

路径：左侧【知识库】 → 打开目标知识库。

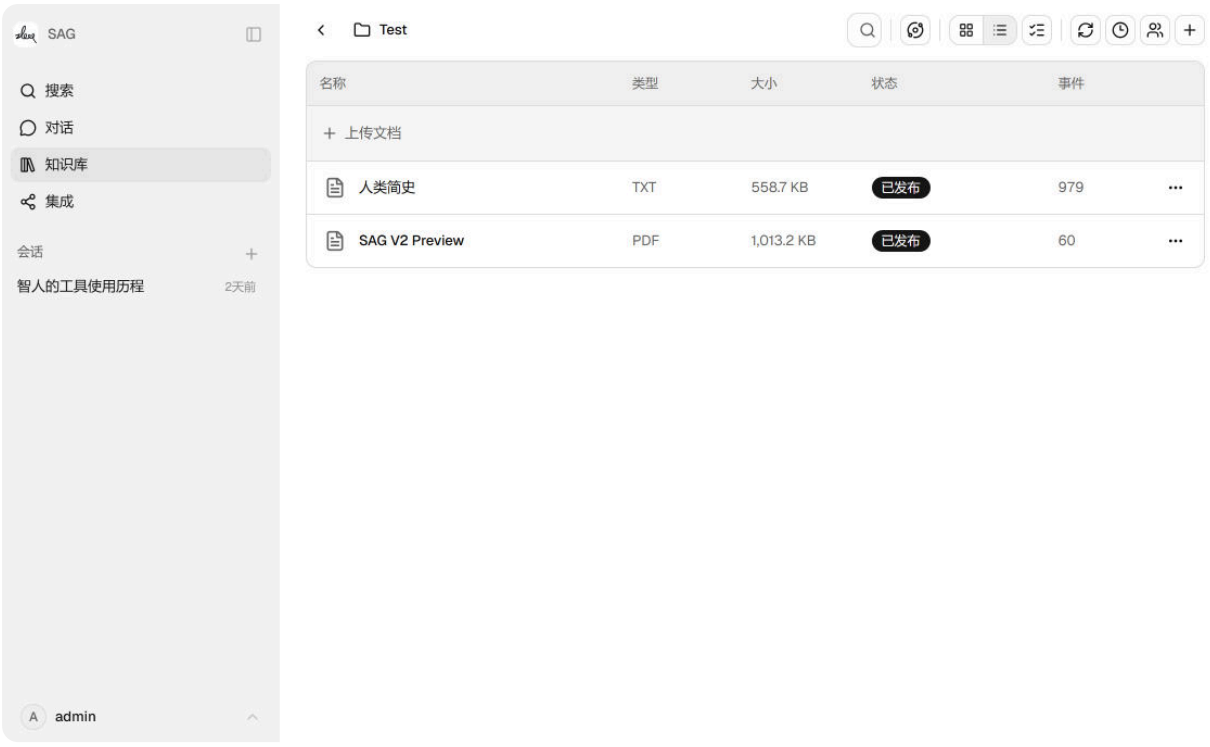
1 找到目标知识库

在知识库页面找到目标知识库并打开。



2 认识文档列表

进入知识库后默认显示文档列表。顶部工具栏提供文档筛选、视图切换、批量操作、刷新、知识库动态、成员管理和【上传文档】等入口。



列表中每份文档显示类型、大小、状态、事件数和修改日期；只有状态为【已发布】的文档才会出现在搜索和问答结果中。

3 上传文件

点击【上传文档】选择文件并提交，也可以直接将文件拖入文档列表。



4 处理重复文档

如果上传文件与知识库中的文档文件名相同，系统会先提示确认：

- 如果文件内容完全相同，页面会提示检测到重复文档。选择【取消上传】可保留原文档；选择【覆盖原文档】后还需要确认，系统才会继续替换原文档。
- 如果文件名相同但内容已经变化，页面会提示检测到同名文档。选择【更新为新版本】后，系统会处理新文件，成功后替换当前可搜索内容。
- 如果希望把文件作为新文档保留，请先取消本次上传，修改文件名后重新上传。

已有文档仍在处理时，暂时不能用新文件覆盖；请等待当前处理完成后再上传。

5 等待发布完成

提交后查看文档列表中的处理状态，等待状态变为【已发布】。只有发布完成的文档才会进入搜索和问答。

6 检查解析内容

点击文档名打开预览，在【解析内容】中检查正文是否完整；需要核对文件时切换【原始文件】。

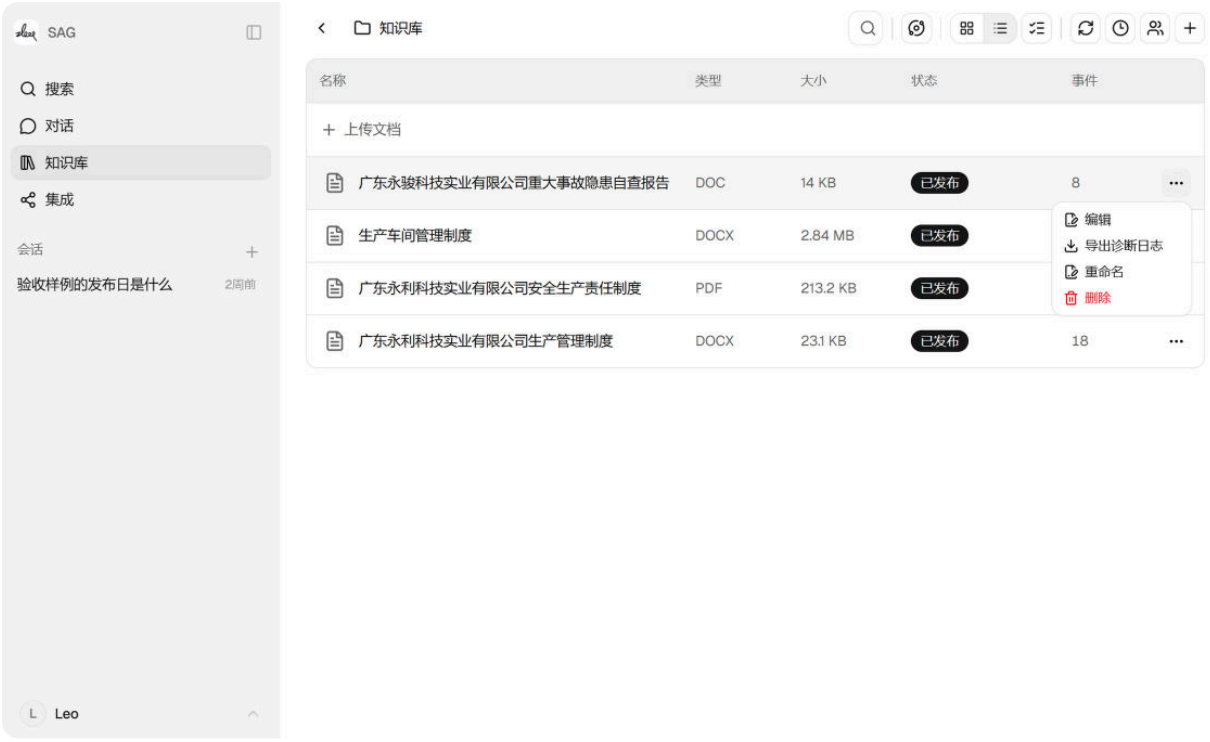




7 处理失败文档

处理失败时先查看页面提示，再通过文档行尾的【...】菜单执行【重试】、【重命名】或【删除】等操作。重新尝试解析时不需要再次上传原文件。

如果更新已有文档后的新版本处理失败，原来已发布的版本仍然可以继续搜索。



> ⚠ 注意
>
> 1. 上传成功不等于文档已经**可用**：只有【已发布】状态的文档才会出现在搜索和问答结果中。
> 2. 覆盖已有文档前请确认文件内容和目标文档；如果只是想保留两份资料，请先修改文件名再上传。
> 3. 单份文档内容额度由站点管理员配置，超出额度的上传会被明确拒绝。

同步外部文档

先确认数据源可用，再将飞书或钉钉中有权访问的在线文档添加到知识库并按计划更新。

场景：先确认飞书或钉钉数据源可用，再把本人有权访问的在线文档添加到 SAG 知识库，供搜索、问答和引用使用。

角色：拥有目标知识库集成管理权限的成员、知识库负责人。

路径：先进入左侧【集成】→【数据源】，确认平台可用；再进入左侧【知识库】→打开目标知识库→【添加文档】。

使用前准备

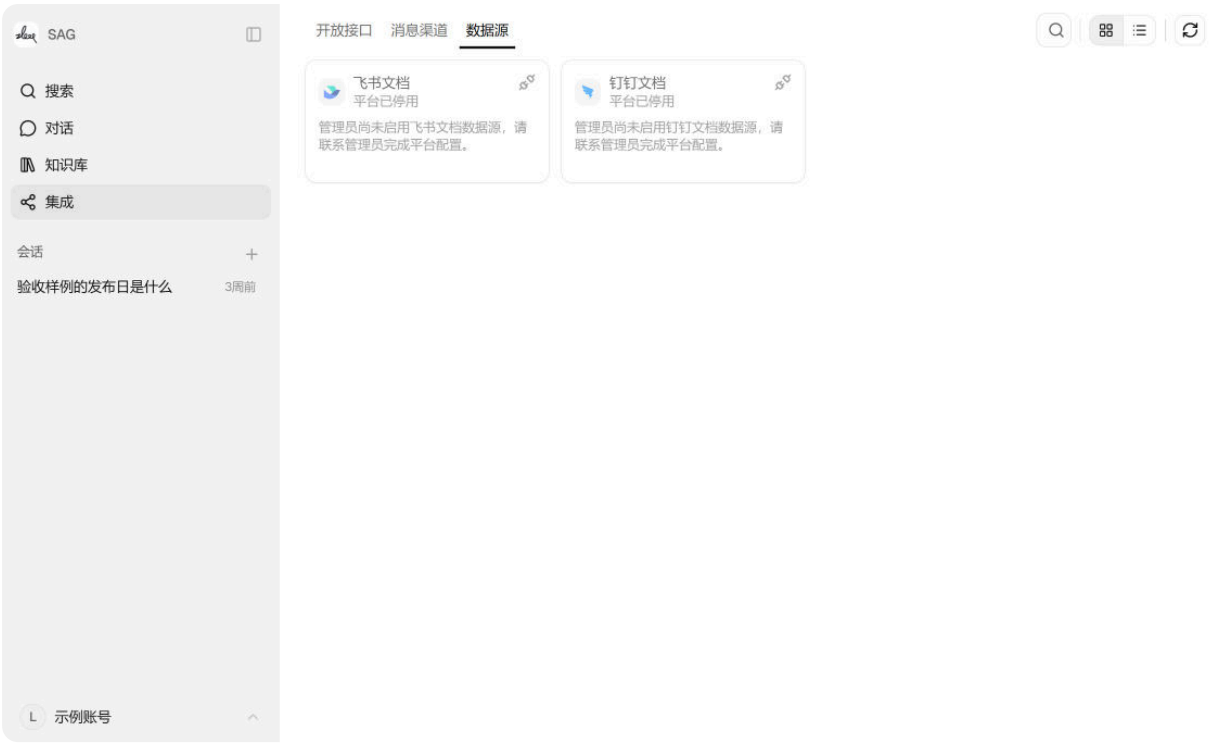
开始前请确认：

1. 站点管理员已经配置并启用对应的外部数据源；
2. 你在飞书或钉钉中拥有来源文档的访问权限；
3. 你拥有目标知识库的集成管理权限。

先确认数据源可用

1 查看数据源状态

进入【集成】→【数据源】，找到【飞书文档】或【钉钉文档】。只有平台状态为可用时，才能继续添加对应平台的文档；下图展示的是【平台已停用】的不可用状态，出现这种状态时先联系站点管理员。



2 连接平台账号

点击可用数据源卡片上的【连接账号】，使用自己的飞书或钉钉账号完成授权。授权只用于读取本人有权访问的文档，不会增加原账号的文档权限。

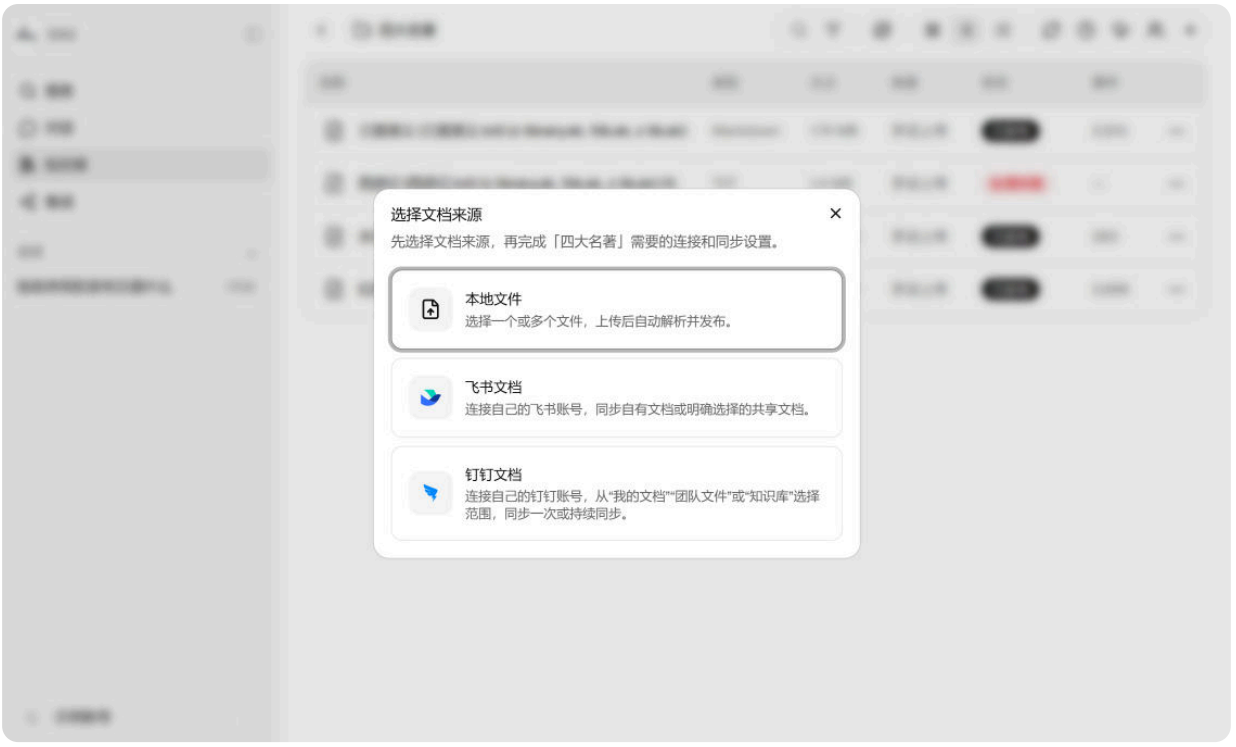
如果数据源显示【平台已停用】或管理员尚未启用，请联系站点管理员，不要继续创建同步计划。

在知识库添加文档

1 打开添加文档

进入【知识库】，打开目标知识库，点击【添加文档】。在【选择文档来源】窗口中选择【飞书文档】或【钉钉文档】。

如果还没有连接对应账号，可以先关闭窗口，回到【集成】→【数据源】完成连接，再重新打开【添加文档】。



2 选择来源范围

按平台选择需要同步的文档或文件夹：

- 飞书：从本人有权访问的文档或共享文档中选择；
- 钉钉：从页面提供的【我的文档】、【团队文件】或【知识库】等区域选择。

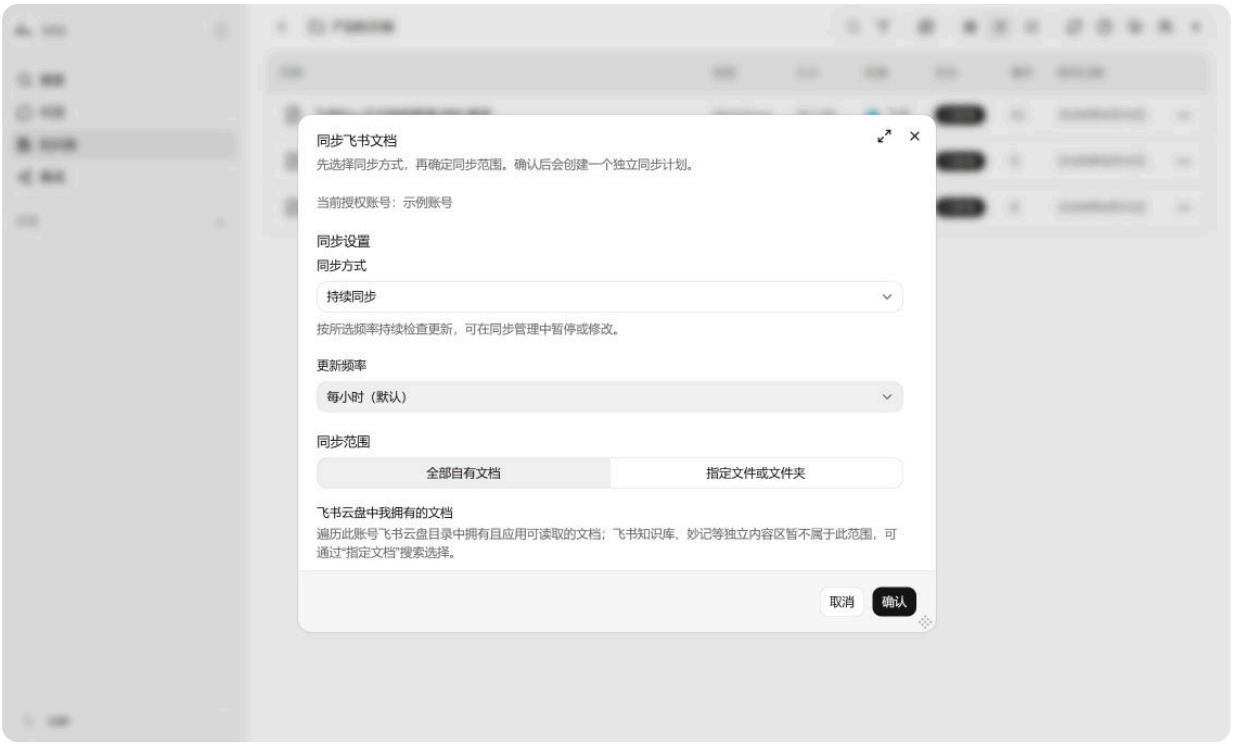
只选择确认需要进入当前知识库的范围。选择器展示的目录和文档类型取决于当前账号权限及管理员配置。



3 填写同步设置

在同步表单中核对并填写：

- 目标知识库：确认文档将进入哪个知识库；
- 来源账号：确认使用的是已授权的飞书或钉钉账号；
- 文档或文件夹：确认同步范围；
- 同步方式：选择【只同步一次】或【持续同步】；
- 更新频率：持续同步时，选择页面提供的频率。



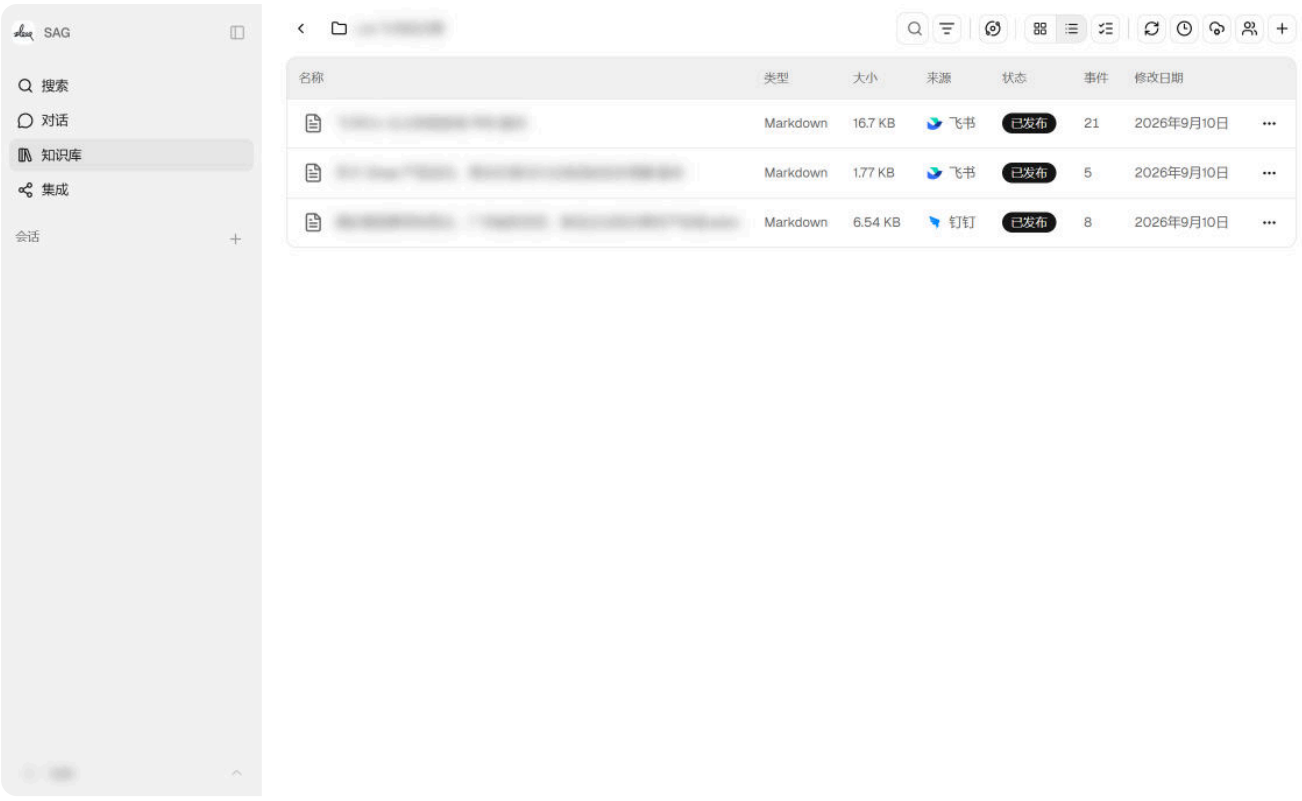
4 保存同步计划

确认目标知识库、来源范围和同步方式后，点击【保存】或【创建计划】。系统会扫描来源并分别处理文档。

文档状态变为【可用】并完成发布后，才会进入搜索和问答；【已发现】或【已安排处理】不代表已经可以搜索。

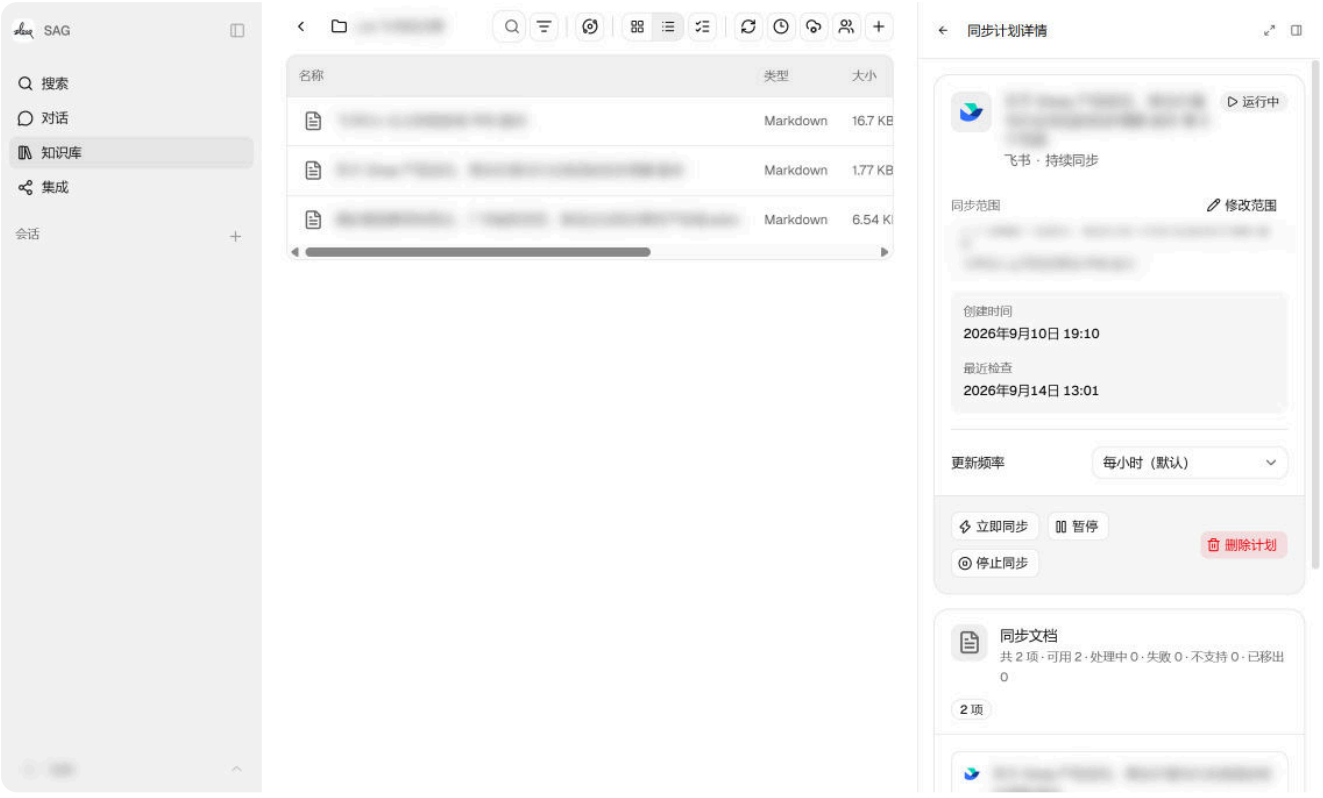
查看同步结果

回到知识库文档列表，外部文档会标记为外部文档。打开文档后，可以查看【同步副本】，也可以点击【打开飞书原文】或【打开钉钉原文】返回来源平台。

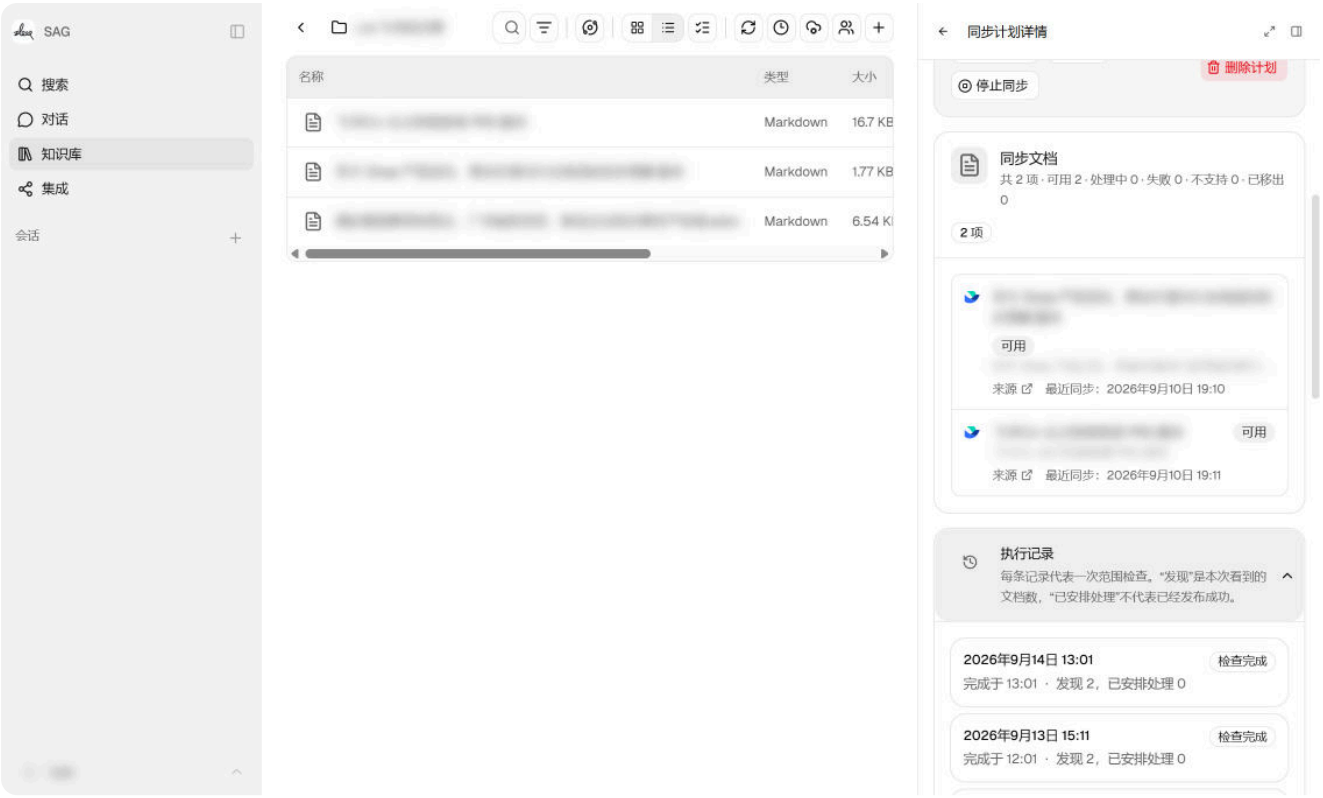


如果需要查看计划状态，在知识库工具栏中打开【同步管理】，可查看同步计划、执行记录和每篇文档的【等待同步】【同步中】【可用】【失败】等状态，并对运行中的计划执行【立即同步】【暂停】【继续】或【修改范围】。

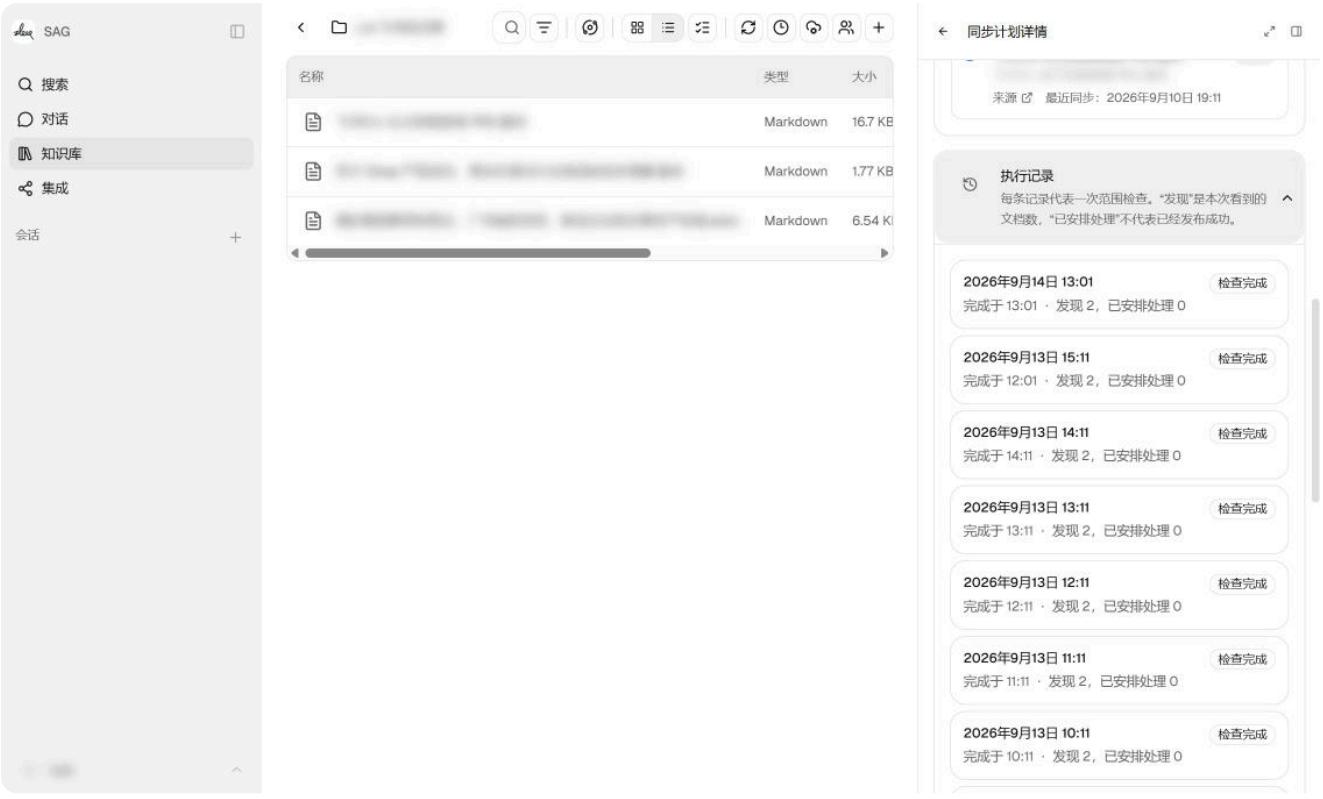
同步计划上半部分展示当前运行状态、同步范围和计划操作：



向下滚动后，可以查看每篇文档的同步状态和最近同步时间：



继续向下滚动，可以查看每次范围检查的执行记录：



状态判断

1. 数据源显示【平台已停用】或未配置时，先联系站点管理员完成企业应用配置。
2. 数据源可用但未连接账号时，先完成个人账号授权，再进入知识库添加文档。
3. 文档显示【可用】并完成发布后，才能在搜索和问答中使用。

⚠ 注意

1. 断开平台账号会停止依赖该账号的同步计划，但已经导入的文档不会删除；如需恢复更新，请重新授权。
2. 来源文档被删除、移动或取消访问权限后，计划会显示对应失败原因；请调整来源权限或同步范围后再执行【立即同步】。
3. 暂停或删除同步计划只影响后续更新，不会自动清理已经发布到知识库的内容。

邀请成员

查看知识库成员，邀请已有企业账号的同事加入。

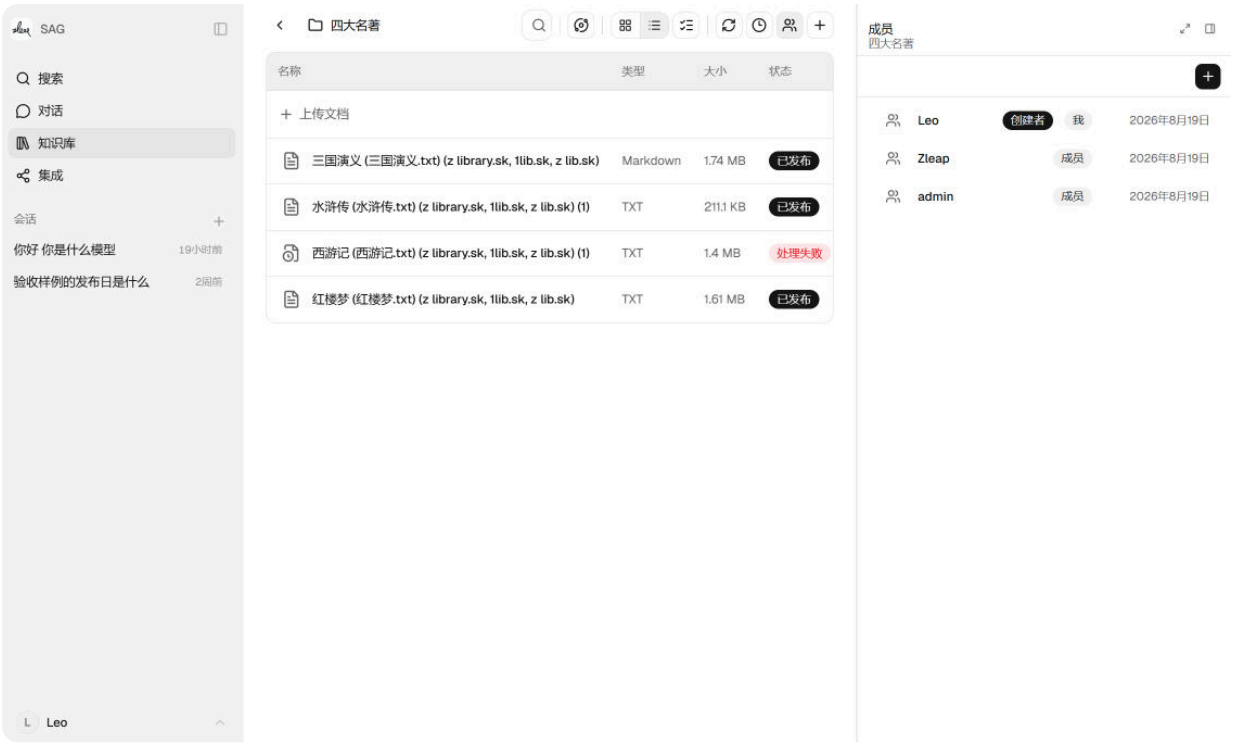
场景：确认当前知识库成员，或邀请一名已有企业账号的同事加入。

角色：知识库创建者或拥有成员管理权限的成员。

路径：知识库详情 → 工具栏【成员】。

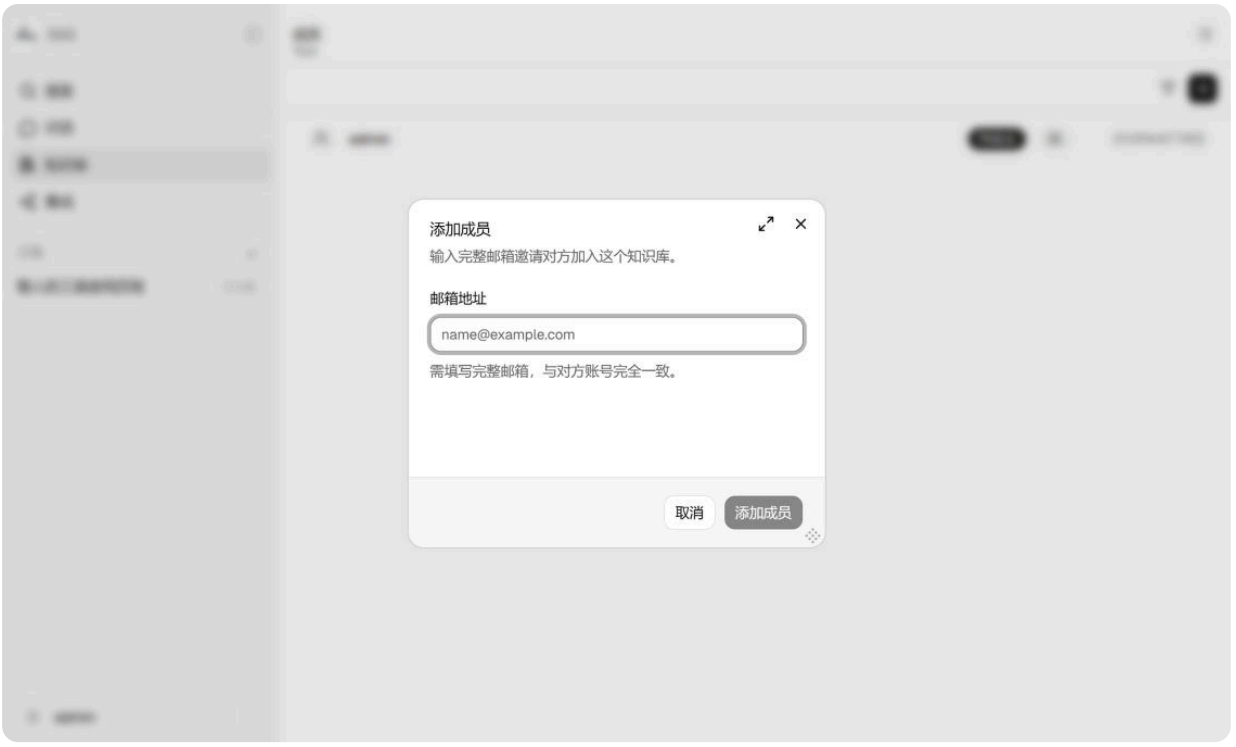
1 查看成员

点击工具栏的【成员】图标打开成员侧栏，可查看成员的姓名、角色和加入时间。



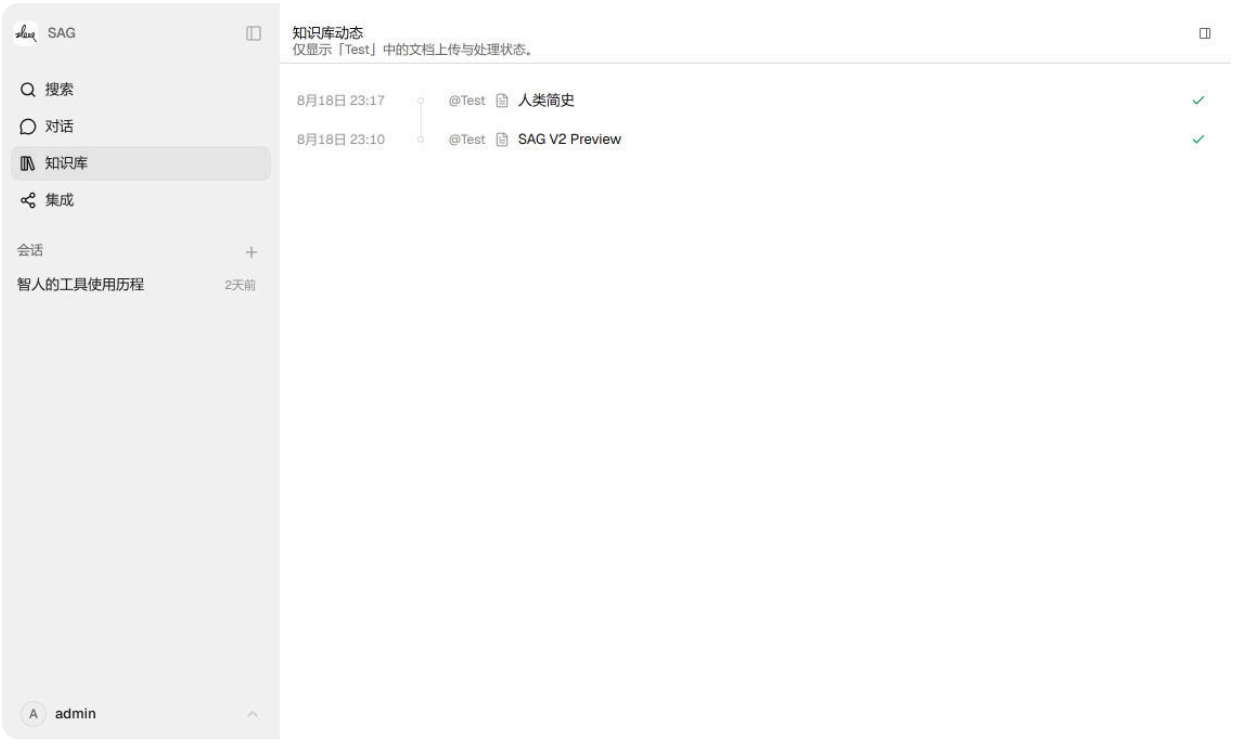
2 邀请成员

点击侧栏中的【+】，填写与对方账号完全一致的邮箱并提交。



3 查看知识库动态

工具栏的【知识库动态】记录文档上传、发布、成员变更等操作历史，协作时可用来了解最近发生了什么。



注意

成员权限只对当前知识库生效，不等于超级管理员权限。邮箱不完整或与账号不一致时，邀请可能无法完成。

图谱

在 2D/3D 图谱中浏览文档的实体与事件关系。

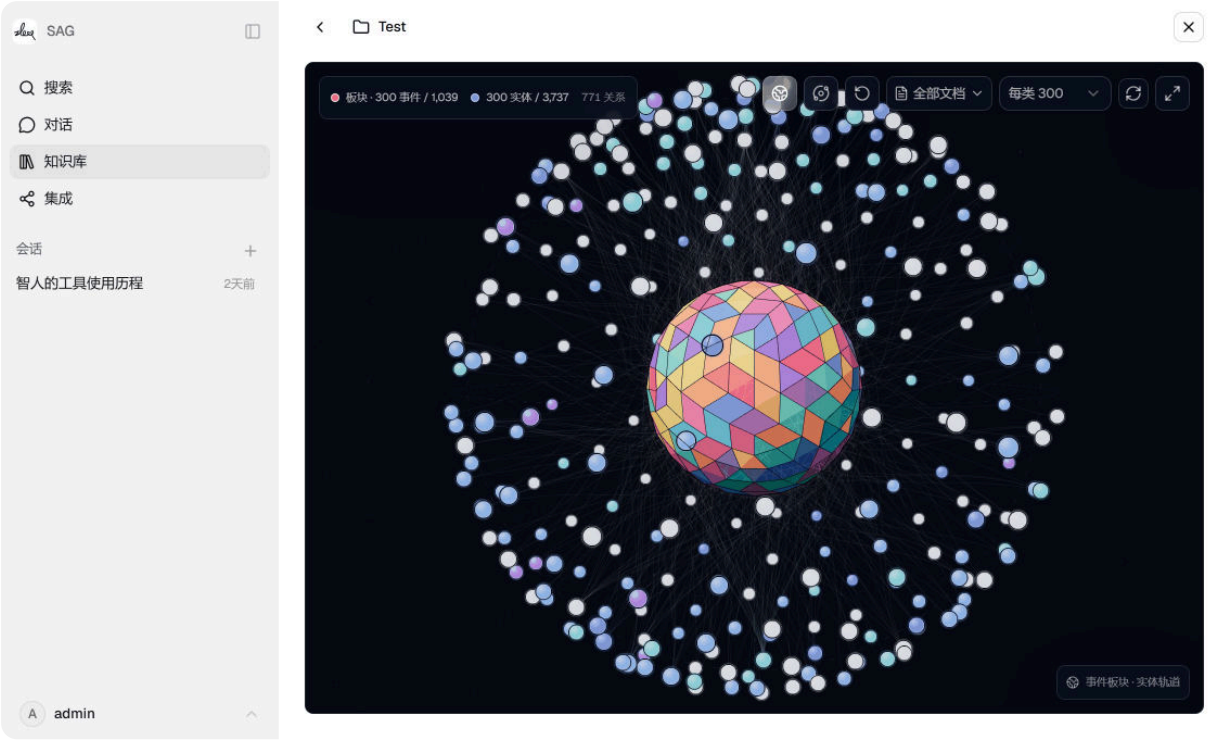
场景：以图谱方式浏览知识库中的实体、事件及其关联，辅助发现资料间的联系。

角色：已获授权的成员。

路径：知识库详情 → 工具栏【进入图谱】。

1 进入图谱

在知识库工具栏点击【进入图谱】。图谱内部可在 2D 和 3D 之间切换，右下角提供布局、复位、刷新与镜头控制。



2 查看节点来源

点击节点打开来源面板，查看该实体或事件对应的文档证据；面板与搜索、问答使用同一套来源展示。

> **注意**
>
> 文档发布、重命名或删除后图谱会同步刷新。搜索结果页也提供【图谱】视图，展示当次检索命中的关系网络。

普通成员使用

搜索

关键词搜索、快速/精确模式、AI 总结与引用核对。

场景：通过关键词定位企业资料，并查看结果依据。

角色：所有已获授权的成员。

路径：左侧【搜索】 → 输入关键词 → 选择检索方式。

1 选择检索方式

在搜索框输入业务关键词或问题；输入 `@` 可以把范围限定到某个知识库。点击搜索按钮左侧的检索方式图标，在【快速】和【精确】之间选择。



快速适合先看结果；精确在多路召回和图谱扩展后用大模型排序，结果更完整但耗时更长。

2 查看结果与 AI 总结

结果页顶部是「基于相关证据的总结」，下方按相关度列出证据条目：每条显示标题、摘要、命中的实体标签、所属知识库和日期。右上角可在【列表】和【图谱】两种视图间切换。



3 核对引用与原文

点击任意一条结果打开右侧来源面板，按「引用事件 / 关联事件 / 来源片段 / 原始文件」的层次核对依据；点击实体名称会立即以该实体发起新搜索。



注意

搜索只返回当前账号有权限且已发布的资料；未发布或无权限的文档不会出现在结果中，即使已上传也不例外。

普通成员使用

问答

围绕企业资料提问，核对引用，管理会话。

场景：围绕企业资料提问，并回到引用内容确认回答依据。

角色：所有已获授权的成员。

路径：左侧【对话】 → 选择知识库范围 → 输入问题。

1 提问

进入【对话】，选择需要参考的知识库范围，输入问题并发送。



2 查看回答

助手会先主动检索证据：第一轮没有结果会自动调整检索方式，必要时升级到精确模式，最多执行 4 轮证据查找；仍无证据时会明确说明无法可靠回答，不会按模型常识编造库内结论。

回答完成后，正文末尾会显示引用来源的数量。



3 核对引用

点击回答下方的引用来源，打开右侧引用侧栏，按「引用 / 来源 / 原始文件」核对回答是否来自企业文档。引用侧栏与搜索共用同一套证据展示；在引用中点击实体，会在当前会话追加一轮带上下文的追问。



4 管理会话

左侧「会话」列表支持重命名、归档和删除；删除需要二次确认，删除后进入管理员回收站。

> **注意**
>
> 没有引用时，请先检查**知识库范围**、**文档状态**和**当前账号权限**。企业业务判断应以可核对的引用内容为准。

开放连接

通过开放连接，把知识库检索和问答能力提供给 AI Agent（MCP）或企业应用（HTTP API）。

场景：将知识库能力连接到外部 AI Agent 或自建应用，实现自动化检索和问答调用。

角色：查看与使用：已获授权的成员；创建与管理：超级管理员或被授权的连接负责人。

路径：左侧【集成】。

知识库连接与助手连接的区别

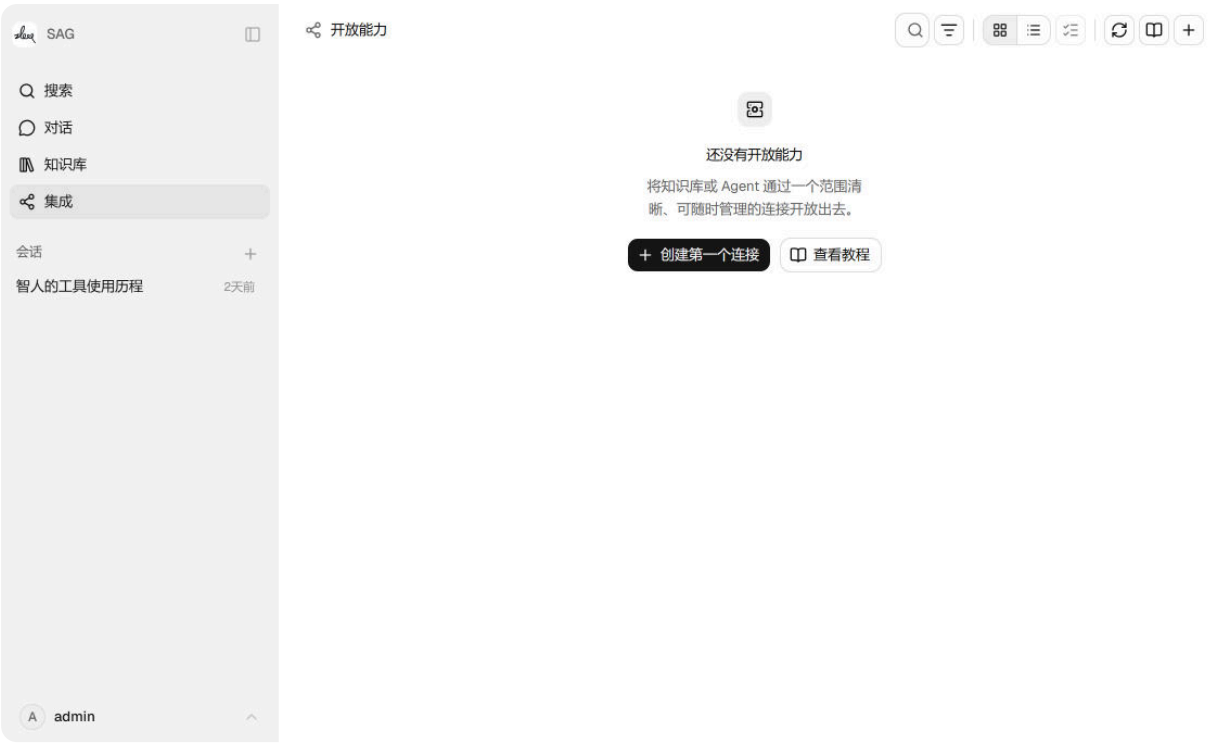
SAG 企业版提供两种开放连接，适用场景不同：

	知识库连接	助手连接
开放内容	知识库中的检索、阅读和管理能力	Agent 的直接问答能力
适合场景	自己编排检索逻辑的 Agent 或应用	直接问答，无需关心检索细节
调用方负责	根据开放能力组织检索请求	只传入对话历史，服务端完成检索

创建开放连接

1 认识集成页

集成页列出当前可见的开放连接，支持搜索、筛选和卡片/列表视图切换。没有任何连接时，页面提供【创建第一个连接】和【查看教程】入口。



2 查看接入教程

点击工具栏【教程】，查看完整接入路径：创建开放连接 → 配置开放能力 → 为 Agent 安装 MCP 或通过 HTTP API 接入 → 管理 API Key。



3 创建开放连接

点击【创建开放连接】，填写名称和说明，选择开放对象（知识库或 Agent），设置开放范围和开放能力。创建完成后立即保存只显示一次的 API Key。



表单较长，向下滚动后继续配置开放能力：



配置知识库开放能力

创建或编辑知识库连接时，在【开放能力】中选择希望提供给外部 Agent 或应用的业务能力：

能力	作用
查看知识库	查看当前连接开放范围内的知识库及简介。
查看文档目录	浏览知识库中已经发布的文档。
搜索知识库	检索知识库中最相关的内容，并读取搜索证据。
阅读文档	在线读取已发布文档的正文和结构。
管理文档	上传、查看导入状态、重试、重命名或删除文档。
管理知识库	创建、修改或删除工作区中的知识库。

配置时只需要选择业务能力，页面会将相关工具显示为【关联工具（自动包含）】；这些工具会随业务能力自动包含，不能单独关闭。

如果连接只开放指定知识库，管理知识库能力会受到范围限制；需要管理整个工作区时，必须开放整个工作区。

消息渠道

接入飞书、钉钉和企业微信，让成员通过日常沟通工具使用 Agent。

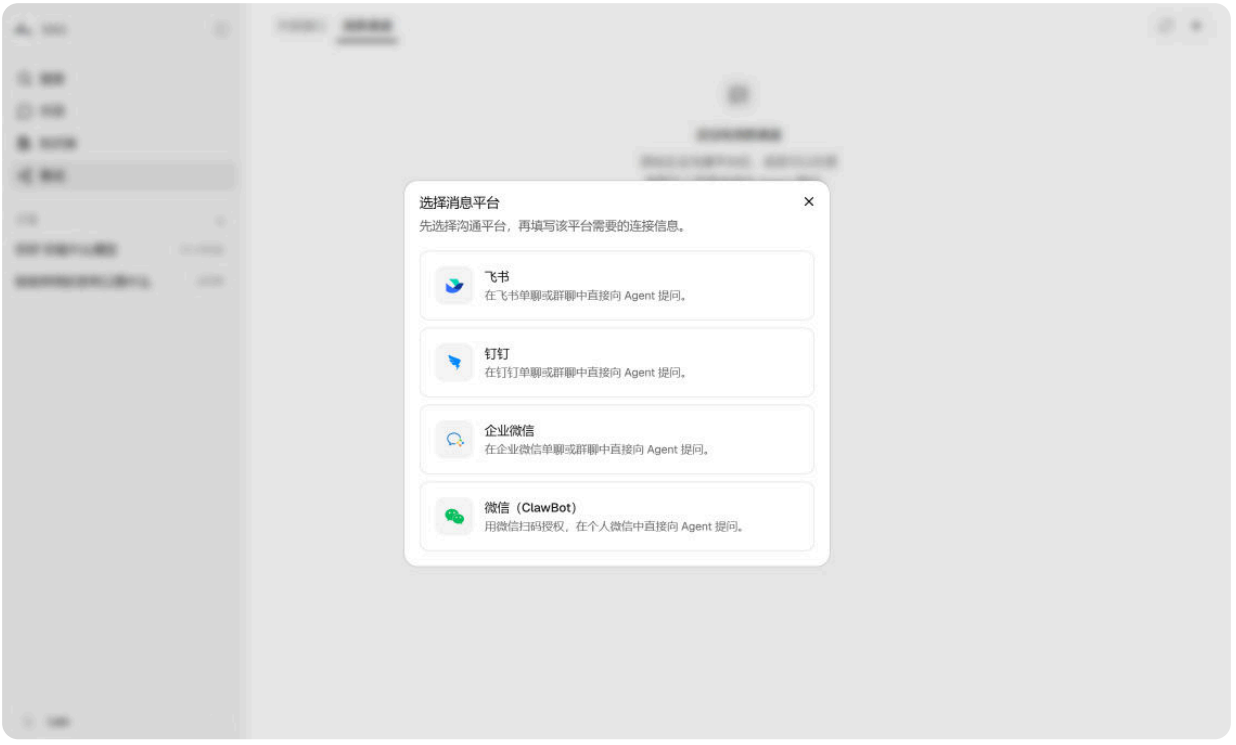
配置消息渠道

如果希望成员直接在日常沟通工具中向 Agent 提问，可以在【消息渠道】中接入飞书、钉钉或企业微信。消息渠道会把收到的消息交给已选择的【Agent 连接】处理，并沿用该连接绑定的知识库。

前提：超级管理员需要先在【开放接口】中创建并启用一条 Agent 连接；该连接只绑定一个知识库，并关闭个人记忆。

1 选择消息平台

进入【集成】，在【集成类型】中切换到【消息渠道】，点击【添加】，再选择【飞书】、【钉钉】或【企业微信】。



2 填写渠道信息

填写【渠道名称】，选择要使用的【Agent 连接】，按平台准备并填写对应凭据，再按需开启【允许群聊】、填写【新对话欢迎语（可选）】。

- 飞书：在飞书创建企业自建应用并启用机器人能力，填写【App ID】和【App Secret】。在飞书侧选择长连接接收事件，订阅 `im.message.receive_v1`，并开通机器人收发消息和消息表情回应权限。
  - 查看飞书机器人创建教程

下图依次展示飞书渠道表单的上半部分和向下滚动后的下半部分。





- 钉钉：创建企业内部应用并添加机器人能力，填写【Client ID】和【Client Secret】。为机器人启用 Stream 模式和接收消息能力，不要使用普通 Webhook 模式。
  - 查看钉钉机器人创建教程（Stream 模式）
  - 查看钉钉机器人使用教程（加群）

下图依次展示钉钉渠道表单的上半部分和向下滚动后的下半部分。





- 企业微信：创建企业微信智能机器人，填写【Bot ID】和【Secret】。为机器人选择 API 模式，通过 WebSocket 长连接收发消息。
  - 查看企业微信智能机器人教程

下图依次展示企业微信渠道表单的上半部分和向下滚动后的下半部分。





3 验证并保存

点击【验证并添加】。平台标识和密钥验证成功后，消息渠道才会启用长连接并出现在列表中；钉钉需要先发布应用版本，企业微信需要配置机器人的可见范围，之后再把机器人添加到需要使用的单聊或群聊中。

编辑消息渠道

在消息渠道列表打开对应渠道的【...】菜单：

- 点击【编辑】，可修改【渠道名称】、【Agent 连接】、【允许群聊】和欢迎语；保存后立即生效。
- 点击【更新凭据】，填写新的平台标识和密钥，再点击【验证并更新】。新凭据验证成功后才会替换旧凭据。
- 需要暂时停止服务时，使用【停用】；恢复时使用【启用】。

> **注意**
>
> 1. 飞书、钉钉和企业微信使用的是企业应用或智能机器人凭据，不需要使用成员个人账号登录。
> 2. 消息渠道只能使用已选择的【Agent 连接】及其绑定的知识库；如列表中没有可用 Agent，请先创建并启用 Agent 连接。

3. 已保存的渠道凭据不会再次明文显示；更换凭据时请使用【更新凭据】，不要把密钥写入文档、代码或日志。

MCP 与 HTTP API

为 AI Agent 或企业应用配置 MCP 和 HTTP API 接入。

接入方式一：为 AI Agent 安装 MCP

MCP（Model Context Protocol）是让 AI Agent 调用外部能力的标准协议。SAG 企业版知识库通过 MCP 接入后，Agent 可以按照连接授权的范围检索和阅读企业文档。

1 获取 MCP 安装配置

打开连接详情，进入【给 Agent 使用 · MCP】标签，展开【完整客户端配置】，复制 JSON 配置。配置示例如下：

```
{
  "mcpServers": {
    "sag-knowledge": {
      "type": "streamable-http",
      "url": "https://your-sag-domain/api/v1/integrations/open-connections/{
      "headers": {
        "Authorization": "Bearer YOUR_API_KEY"
      }
    }
  }
}
```

2 在 Agent 中安装 MCP（以 CodeBuddy 为例）

以 CodeBuddy 为例：

1. 打开 CodeBuddy 侧边栏，点击右上角【设置】图标。
2. 切换到【MCP】标签。
3. 点击【添加 MCP】，将上一步复制的 JSON 配置粘贴到配置文件中。
4. 保存后，MCP 服务器状态变为绿色即表示连接成功。

其他支持 MCP 的 Agent（如 Cursor、Claude Desktop、Cline 等）操作类似，均在其 MCP 配置文件中添加相同格式的服务器配置。

3 验证开放能力

安装完成后，Agent 的能力列表中会出现该连接已经开放的知识库能力。向 Agent 提问或发出检索请求时，它只能使用连接授权范围内的文档。

接入方式二：应用通过 HTTP API 接入

1 获取接口地址和示例

打开连接详情，进入【给应用使用 · HTTP API】标签，复制接口地址和调用示例。所有请求使用同一个 API Key，通过 `Authorization: Bearer <API_KEY>` 认证发送。

使用 CLI 接入

如果希望通过命令行让 Agent 直接使用 SAG 知识库能力，而不是在 Agent 中手动配置 MCP，请查看 CLI 接入 页面，了解安装和使用方式。

> ⚠ 注意
>
> 1. 普通成员只能使用已有的开放连接，无法自行创建或管理；如需新建连接，请联系超级管理员。
> 2. API Key 是敏感凭证，不要在代码或日志中明文记录，也不要分享给他人；创建后请立即保存。
> 3. 开放连接只能访问创建时授权的知识库范围，不会自动扩展到其他知识库。

CLI 接入

通过 CLI 让 AI Agent 使用 SAG 知识库能力，完成检索、阅读和文档管理。

场景：通过 CLI 将 SAG 知识库能力交给 AI Agent，由 Agent 按自然语言完成资料检索、文档阅读或知识库管理。

角色：已登录 SAG 且拥有目标知识库权限的成员；创建开放连接需要超级管理员或被授权的连接负责人。

路径：左侧【集成】→【开放接口】→打开知识库类型的开放连接。

CLI 接入是什么

CLI 接入让 AI Agent 可以通过命令行安装并使用 SAG 知识库能力。安装完成后，你可以在 ChatGPT、Codex 或其他支持执行命令的 Agent 中，直接用自然语言让 Agent 查询、阅读或管理 SAG 中的知识库内容。

它适合以下场景：

- 让 Agent 根据企业资料回答问题，并保留搜索结果中的引用信息；
- 让 Agent 阅读指定文档，整理或比较文档内容；
- 让 Agent 浏览多级文件夹，并把新文档上传到指定文件夹；
- 让 Agent 协助上传、重命名或移入回收站文档。

使用前准备

开始前请确认：

1. 你已经登录 SAG，并且拥有目标知识库的访问权限；
2. 你可以使用一个已配置好范围的知识库开放连接；

3. 你使用的 Agent 支持执行 `npx` 命令，并且能够读取安装结果。

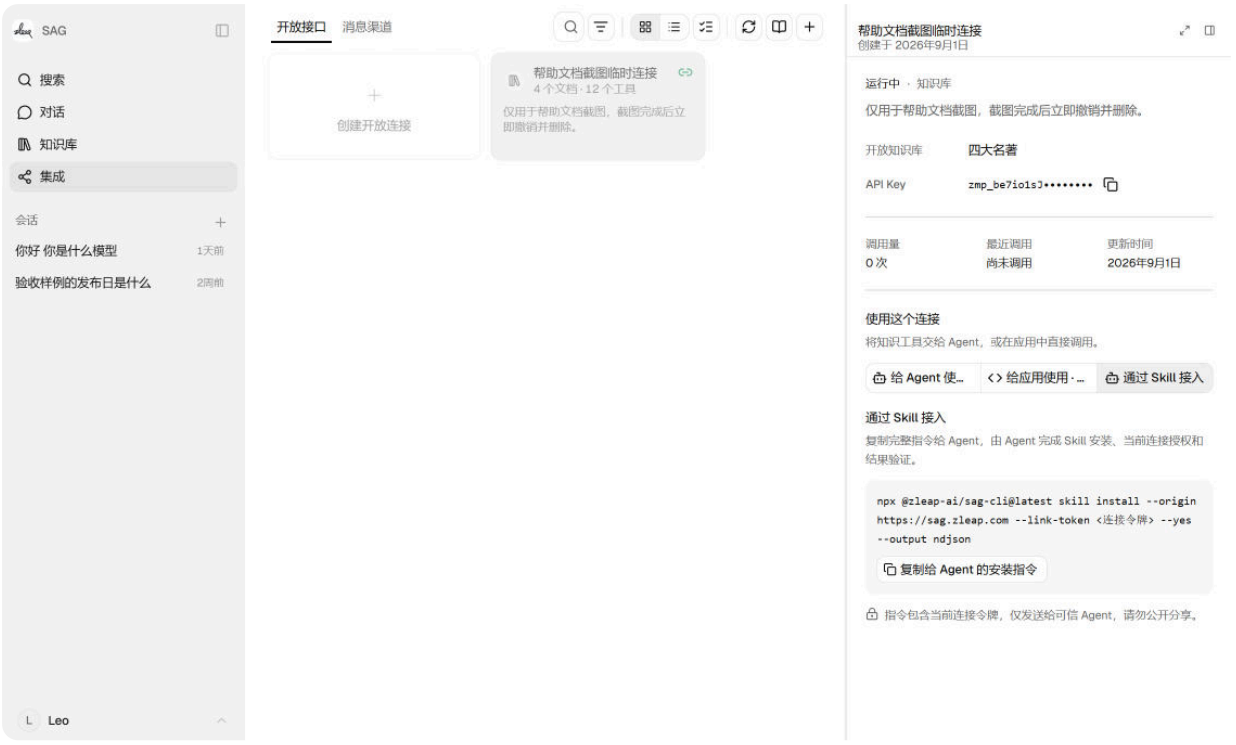
通过开放连接安装

开放连接适合需要限定知识库范围的 Agent。连接只能访问创建时授权的知识库，不会自动扩大范围。

1 打开 CLI 接入入口

进入【集成】→【开放接口】，打开知识库类型的开放连接，向下找到【使用这个连接】区域。

选择【通过 Skill 接入】，页面会显示完整安装指令。



2 复制安装指令

点击【复制给 Agent 的安装指令】，将完整指令交给可信的 Agent 执行。不要手动修改指令中的连接信息，也不要把指令公开发布。

Agent 会自动完成 CLI 安装、当前连接授权和结果验证。安装过程中请检查每条安装结果，确认没有失败项。

整个过程不需要另行打开浏览器授权；以连接页面生成的安装指令为准。

3 验证 CLI 是否可用

安装完成后，可以让 Agent 执行以下验证命令，查看当前连接可访问的知识库：

```
npx @zleap-ai/sag-cli@latest skill knowledge list --output ndjson
```

如果当前会话没有识别新安装的 CLI 能力，请刷新 Agent 或开启新会话后再试。

场景示例

以下为模拟对话，展示不同部门如何在自己的 Agent 中调用 SAG 知识库能力；实际可访问的知识库和执行结果取决于授权范围。

财务部丨制度更新与资料检索



查找差旅报销标准，
并附上原文引用。



我整理了交通、住宿和餐补三类标准；
原文引用来自《差旅管理办法》第三章
“费用标准”。



把 D:\\部门资料\\财务制度\
2026 年报销管理办法.docx
上传到财务知识库。



文件已接收，正在上传到【财务知识库】；
当前状态为“处理中”，完成后我会
继续查询结果。



导入完成后告诉我处理结果。



导入已完成，文档状态为“已发布”；
现在可以在财务知识库中搜索并
查看引用。

人事部 | 员工政策问答

查找年假申请流程，并整理成员工可以执行的步骤。





申请流程是：先确认可用假期，再提交申请，等待直属负责人审批；所需材料和原文依据已一并列出。

新员工入职需要准备哪些材料？





我按办理顺序整理了身份证明、学历信息、银行卡资料和入职表单，报到前准备好即可。

这项员工政策的原文依据是什么？



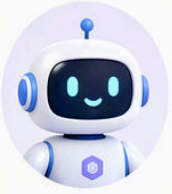


依据是《员工假期管理办法》相关条款；我已标出文档名称和引用位置，方便你逐条核对。

产研部｜技术资料与故障处理

查找这个故障的历史处理记录。





我找到三条相关记录，并按故障现象、处理动作和最终结果整理成了时间线。

比较不同版本的解决方案有什么差异。





旧版侧重临时恢复，当前版本增加了根因修复和回滚步骤；差异已按适用条件标注。

找出最适合当前版本的处理建议，并附引用。





当前版本建议先执行配置校验，再按文档中的恢复步骤处理；引用位置已附在每条建议后。

销售部｜产品资料与客户准备



查找适合这个客户行业的
产品能力。



结合客户行业和当前需求，
我筛出三项最匹配的产品能力，
并附上对应资料。



找几份相关行业案例，
并附原文引用。



我找到两份相关案例，
分别标注了客户背景、落地做法
和原文引用位置。



把这些资料整理成销售拜访提纲。



拜访提纲已按客户现状、
产品价值、案例证明和下一步行动
四部分整理完成。

安装完成后如何使用

安装完成后，直接用自然语言告诉 Agent 目标和范围即可。例如：

- “列出我可以访问的知识库，并说明每个知识库的用途。”
- “在指定知识库中搜索关于报销流程的资料，并给出引用。”
- “阅读这份文档，整理出适合新员工的操作步骤。”
- “列出财务制度知识库根目录中的文件夹，再进入报销制度文件夹查看它的子文件夹。”
- “在年度制度文件夹下新建一个 2026 年文件夹。”
- “把 `D:\部门资料\财务制度\2026 年报销管理办法.docx` 上传到指定知识库；导入完成后告诉我结果。”
- “把这份文档上传到财务制度知识库的年度制度 / 2026 年文件夹。”

涉及上传、重命名或移入回收站等管理操作时，应明确指定目标知识库和文档，并在 Agent 执行前确认操作意图。

在子文件夹中管理资料

CLI 支持浏览多级文件夹、创建子文件夹、重命名或删除文件夹，以及把新文档上传到指定文件夹。安装完成后不需要自己查找文件夹 ID，也不需要输入 CLI 命令，直接把目标知识库、文件夹路径和操作告诉 Agent 即可。

直接这样告诉 Agent

想完成的操作	自然语言示例
查看根目录	“列出「财务制度」知识库根目录下的文件夹。”
进入子文件夹	“进入「财务制度」知识库的「年度制度」文件夹，继续列出里面的子文件夹。”
核对同名文件夹	“显示所有名为「2026 年」的文件夹及其完整路径，让我选择。”
创建子文件夹	“在「财务制度」知识库的「年度制度」文件夹下创建「2026 年」文件夹。”
上传到文件夹	“把 `D:\资料\2026 年报销管理办法.docx` 上传到「财务制度」知识库的「年度制度 / 2026 年」文件夹，完成后告诉我导入结果。”
上传到根目录	“把这份文档上传到「财务制度」知识库根目录。”
重命名文件夹	“把「财务制度 / 年度制度 / 2026 年」重命名为「2026 年制度」。”
删除文件夹	“删除「财务制度 / 历史制度 / 2023 年」文件夹；先告诉我会影响哪些子文件夹、文档和同步计划，我确认后再执行。”

Agent 会根据完整路径逐层定位目标文件夹。如果存在同名文件夹或目标不明确，Agent 会先展示完整路径，请你选择后再继续；涉及删除时，会先展示影响范围并等待确认。

> ⚠ 文件夹操作须知
>
> 1. 删除文件夹会递归删除其子文件夹，并将其中的文档移入回收站；如涉及持续同步计划，预览中会同时显示影响范围。
> 2. 文件夹只用于整理知识库内容，不会改变文档的检索范围或访问权限。
> 3. CLI 暂不支持移动已有文件夹或文档；需要调整已有内容的位置时，请在 SAG 知识库页面中操作。

当前支持的知识库能力

当前 CLI 接入支持以下用户可见能力：

能力	说明
查看知识库	查看当前授权范围内的知识库及简介。
查看文档目录	浏览知识库中已经发布的文档。
浏览多级文件夹	从根目录逐层查看子文件夹，并核对文件夹完整路径。
管理文件夹	创建、重命名或删除知识库文件夹。
搜索知识库并进行问答	检索知识库内容，回答问题并读取搜索结果中的引用信息。
阅读文档	查看已发布文档的正文和结构。
上传单个文档	向指定知识库的根目录或指定文件夹上传一个文档。
查看文档导入进度	查看上传文档的处理进度和结果。
查看上传大小限制	查看当前允许的文档上传大小。
重命名文档	修改知识库中文档的名称。
将文档移入回收站	将指定文档移入回收站。

权限范围与限制

- 通过开放连接安装时，只能访问连接创建时授权的知识库范围；
- 文档上传、重命名和移入回收站属于管理操作，请在执行前确认目标和结果；
- 文档仍在导入时，搜索结果可能尚未包含最新内容。

> ⚠ 安全提示
>
> 1. 安装指令包含连接令牌，只发送给可信 Agent，请勿公开分享。
> 2. 如果原令牌无法再次读取或怀疑泄露，请先轮换 API Key，使旧令牌立即失效。

3. 如果安装结果出现失败项，请先处理失败项，再开始使用该 CLI 能力。

管理后台概览

管理后台的入口与各模块速览。

场景：超级管理员进入管理后台，掌握站点整体状态。

角色：超级管理员。

路径：左下角账户入口 → 【管理后台】。

1 概览页

管理后台首页汇总人员、计划、任务记录、模型服务和检索策略的关键数字，并列出需要关注的事项（如待审批、清理失败）。左侧导航依次是：概览、人员、计划、模型、数据源、检索、任务、审计、回收站。



各模块的具体操作见后续章节：创建成员账号、配置模型与 API Key、配置外部数据源、配置检索、计划与套餐审批、任务/审计/回收站。

人员管理与创建账号

查看人员列表，创建账号并交付首次登录凭证，处理注册审批。

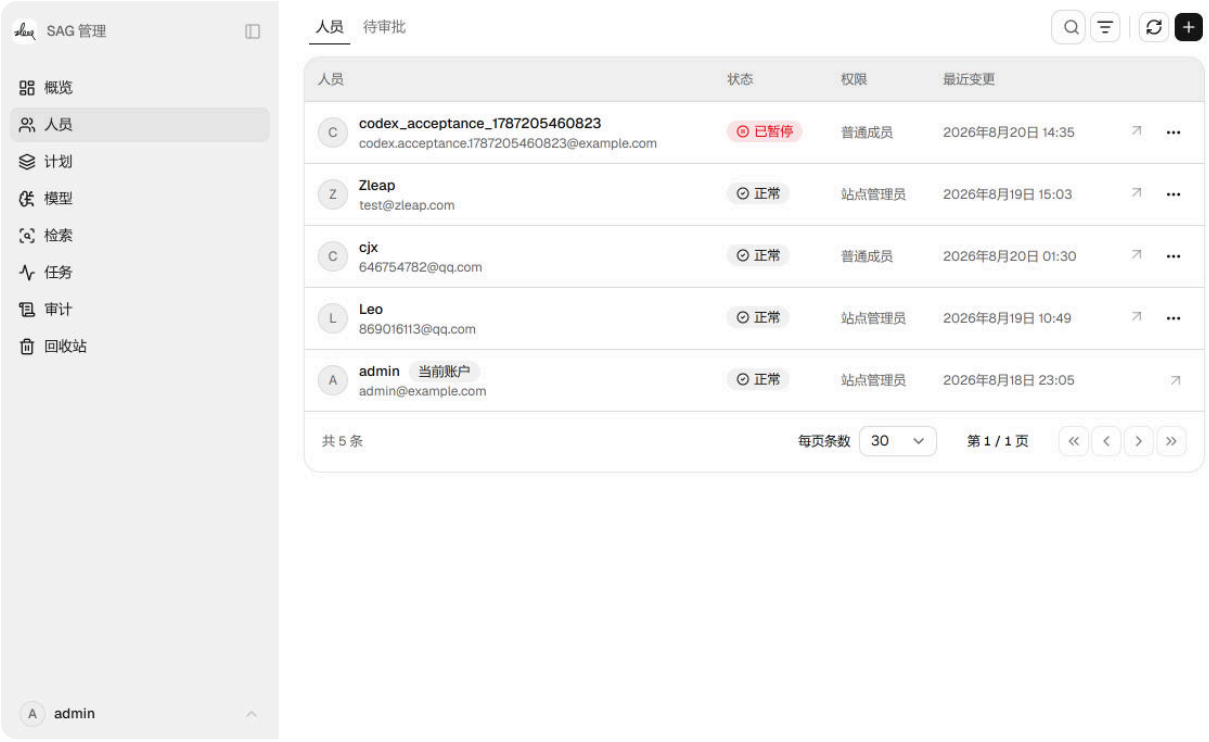
场景：为新同事创建企业版账号，管理已有人员，处理注册审批。

角色：超级管理员。

路径：管理后台 → 【人员】。

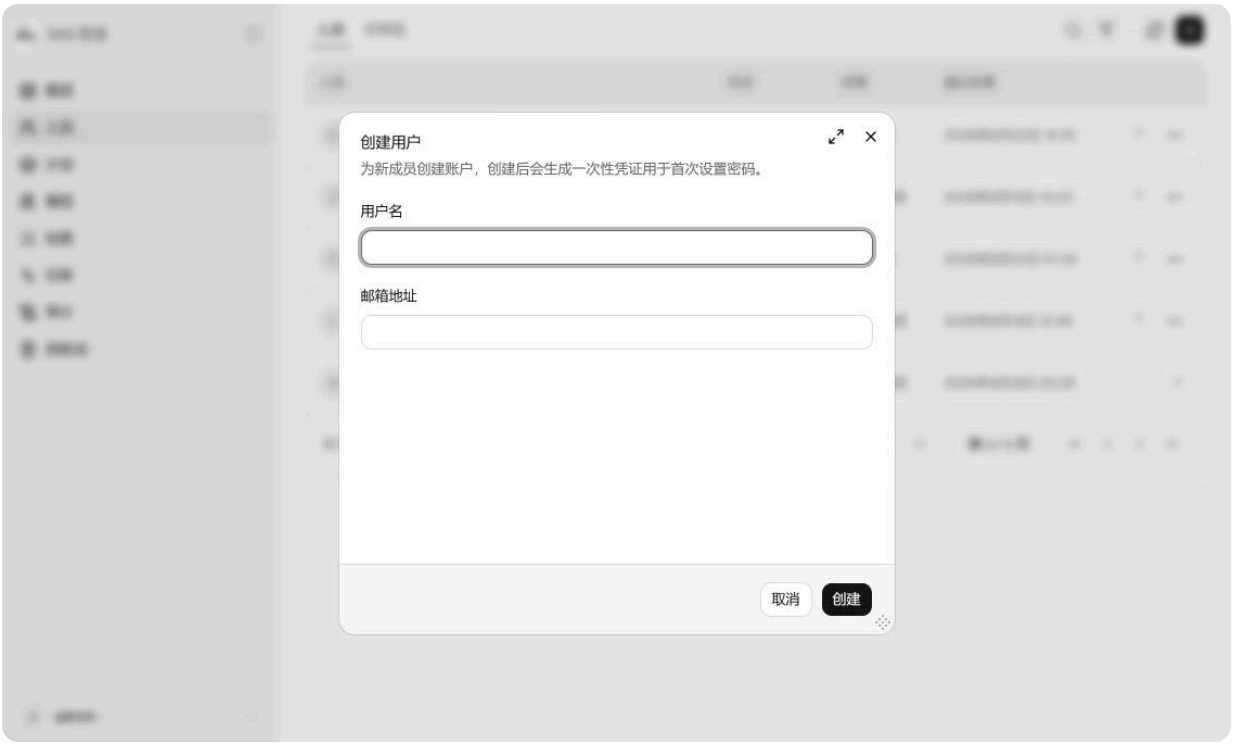
1 查看人员列表

【人员】页列出全部账号，可搜索、按状态和角色筛选；点击成员可查看详情或执行操作。顶部的【待审批】标签处理开放注册时的激活申请。



2 创建用户

点击【创建用户】，填写用户名和邮箱后提交。



3 交付一次性凭证

将系统生成的一次性凭证通过安全方式交给新成员；新成员使用自己的账号完成首次登录并设置密码。

> **注意**
>
> 一次性凭证只用于首次设置密码，不要在公开聊天或文档中传播。成员忘记密码或凭证失效时，由超级管理员重新发放。

超级管理员使用

配置模型与 API Key

供应商连接、API Key、模型资源、知识库应用位与 Agent 模型池。

场景：为企业版配置模型供应商的 API Key，并管理模型资源在知识库和 Agent 中的使用。

角色：超级管理员。

路径：管理后台 → 【模型】。

模型页顶部的卡片显示供应商连接数、可用能力资源数、知识库绑定数和 Agent 模型数，下方分为三个标签页：资源目录、知识库应用和 Agent 模型池。

推荐供应商：302.AI

SAG 企业版的文档解析、语义检索和问答推理都依赖 LLM 和 Embedding 模型。如果企业尚未自建模型服务，推荐使用 302AI 作为模型供应商——国内可直连，支持 OpenAI、Claude、DeepSeek 等主流模型，按量付费无需服务器。

注册和获取 API Key 的步骤：

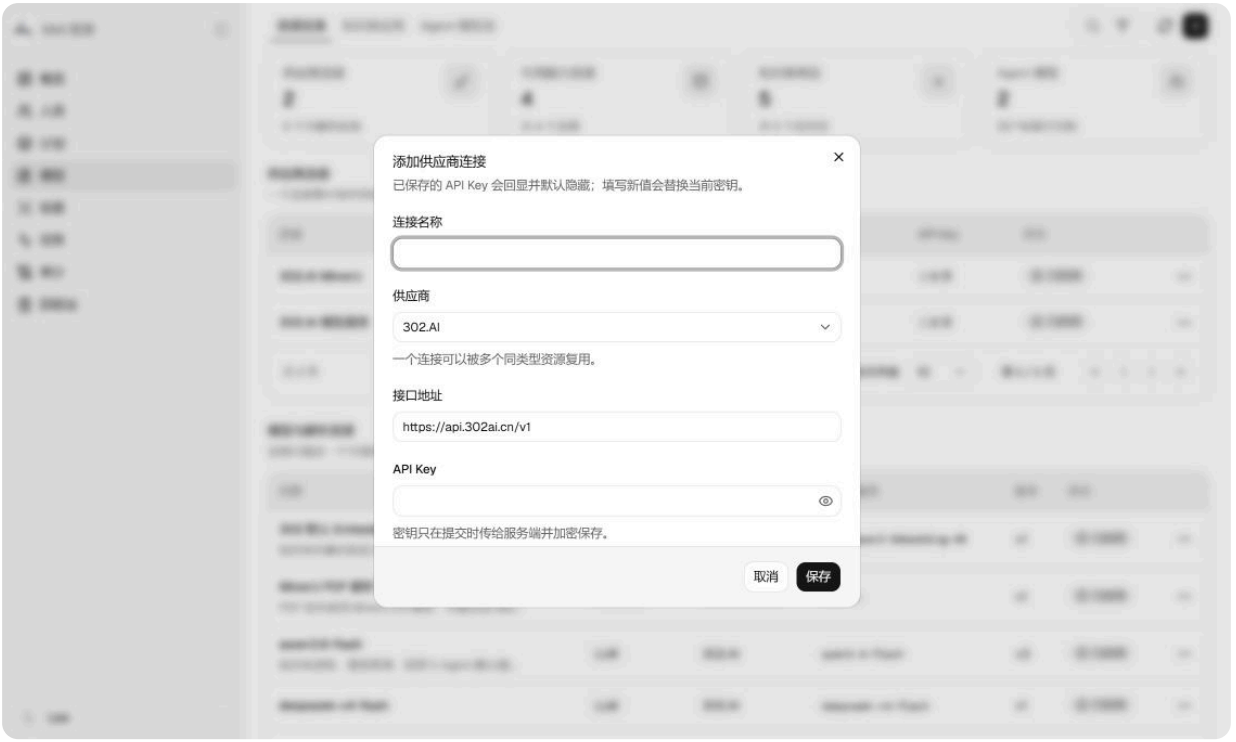
1. 访问 302ai.cn 注册账号并完成充值
2. 进入管理面板 dash.302ai.cn
3. 在「API KEY 管理」页面创建 API Key
4. 复制 Key 备用，回到 SAG 管理后台填入

> **302AI 使用说明**
>
> 302AI 提供完整的帮助文档，包含 API 使用教程、定价说明和在线调试器。如需使用 MinerU 解析复杂 PDF，也在 302AI 的工具市场中开通对应服务后填入相同的 API Key 即可。

配置步骤

1 添加供应商连接并填写 API Key

在【资源目录】点击右上角【+】，选择添加供应商连接：填写连接名称，选择供应商（如 302.AI），确认接口地址，填入 API Key 后【保存】。



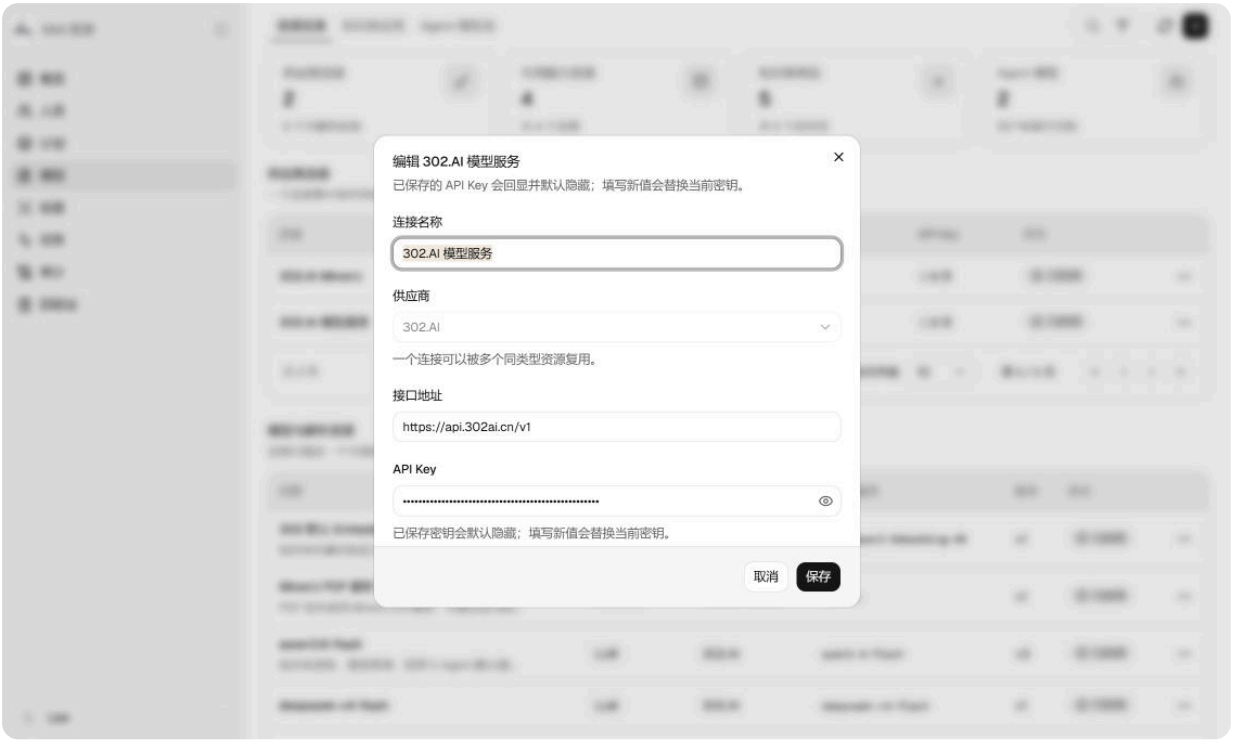
一个连接集中保存供应商、接口地址和 API Key，可被多个同类型资源复用。新建连接时 API Key 为必填项；点击 API Key 右侧的眼睛图标，可以在显示和隐藏之间切换。

2 查看模型与解析资源

「供应商连接」列表显示各连接的 API Key 配置状态；「模型与解析资源」列出每个可调用能力：类型（Embedding / 文档解析 / LLM）、供应商、模型/服务、版本和状态。可以对已保存的 LLM 和 Embedding 资源发起真实连通性测试。

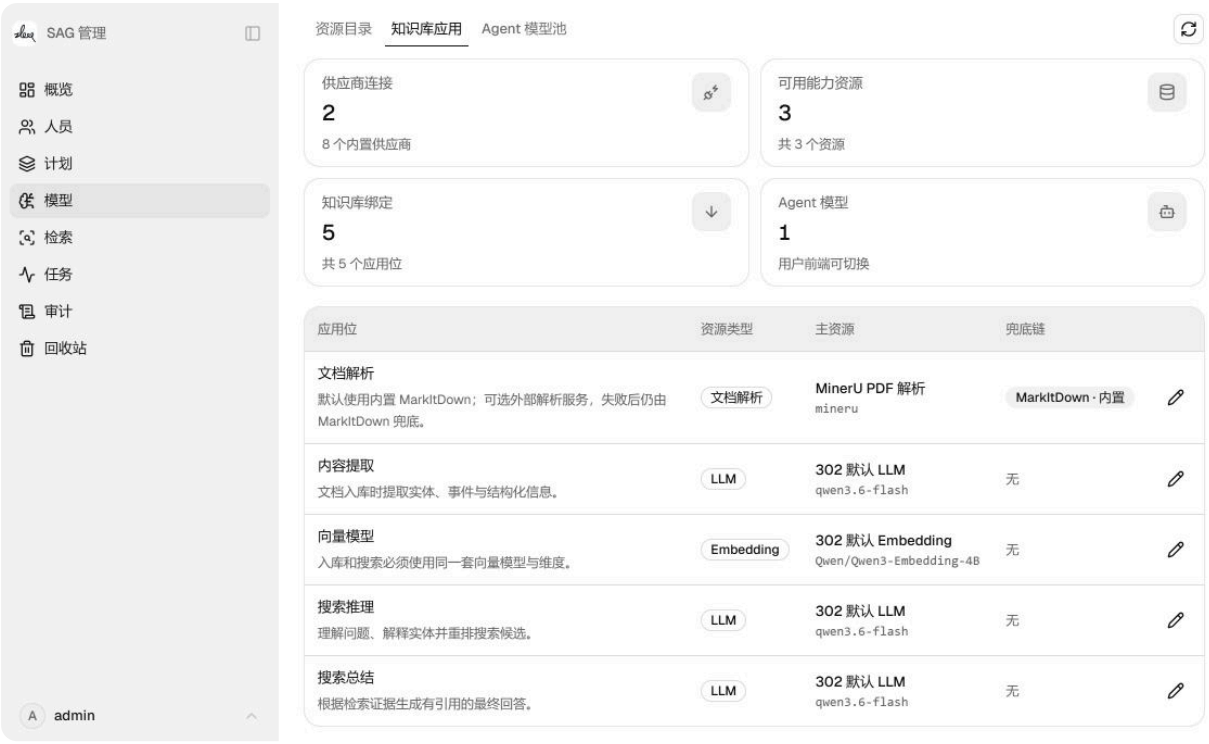


修改模型参数会创建新版本；修改名称和说明只更新展示信息。编辑供应商连接时不能更换供应商；可以修改连接名称和接口地址。已保存的 API Key 默认隐藏，保持原值时直接保存即可；只有更换凭证时才输入新的 API Key，保存后会替换当前密钥。



3 配置知识库应用位

【知识库应用】为五个应用位指定主资源和兜底链：文档解析、内容提取、向量模型、搜索推理、搜索总结。文档解析默认使用内置 MarkItDown，可选外部解析服务（如 302.AI MinerU），失败后仍由 MarkItDown 兜底；清空绑定即表示只用内置解析。



4 管理 Agent 模型池

【Agent 模型池】控制对话前端可以看到哪些 LLM、调整开放顺序，并设置默认推荐模型；知识库后台任务不使用这个列表。



> ⚠ 注意
>
> 1. 已保存的 API Key 默认隐藏；编辑时不输入新值会保留原值，只有输入新的 API Key 才会替换当前密钥。
> 2. 如 API Key 泄露，请立即在对应的【供应商连接】中输入新密钥并【保存】。
> 3. 入库和搜索必须使用同一套向量模型与维度，更换向量模型前先评估对已有索引的影响。

超级管理员使用

配置外部数据源

配置飞书和钉钉企业应用，让成员授权自己的账号并同步有权访问的在线文档。

场景：站点接入飞书或钉钉文档，统一配置企业应用，同时保留每名成员原有的文档权限边界。

角色：站点管理员；飞书或钉钉企业应用管理员。

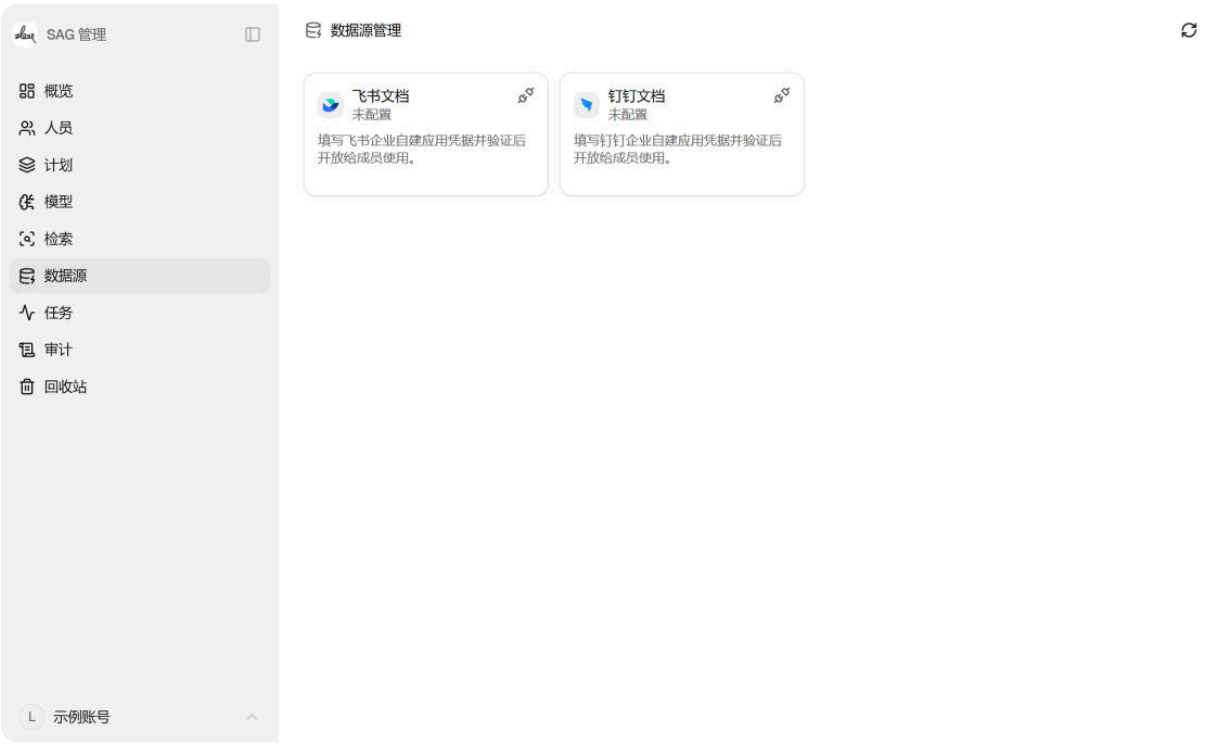
路径：左下角账户入口 → 【管理后台】 → 【数据源】。

配置方式

飞书和钉钉都采用相同的两层授权方式：站点管理员配置一个企业应用，成员再使用自己的平台账号完成授权。成员只能同步本人原本有权访问的文档，企业应用不会扩大成员权限。

1 打开数据源

进入【管理后台】 → 【数据源】，选择【飞书文档】或【钉钉文档】，点击【配置】。配置弹窗会显示当前 SAG 站点专用的 OAuth 回调地址，后面需要把它原样复制到对应开放平台。



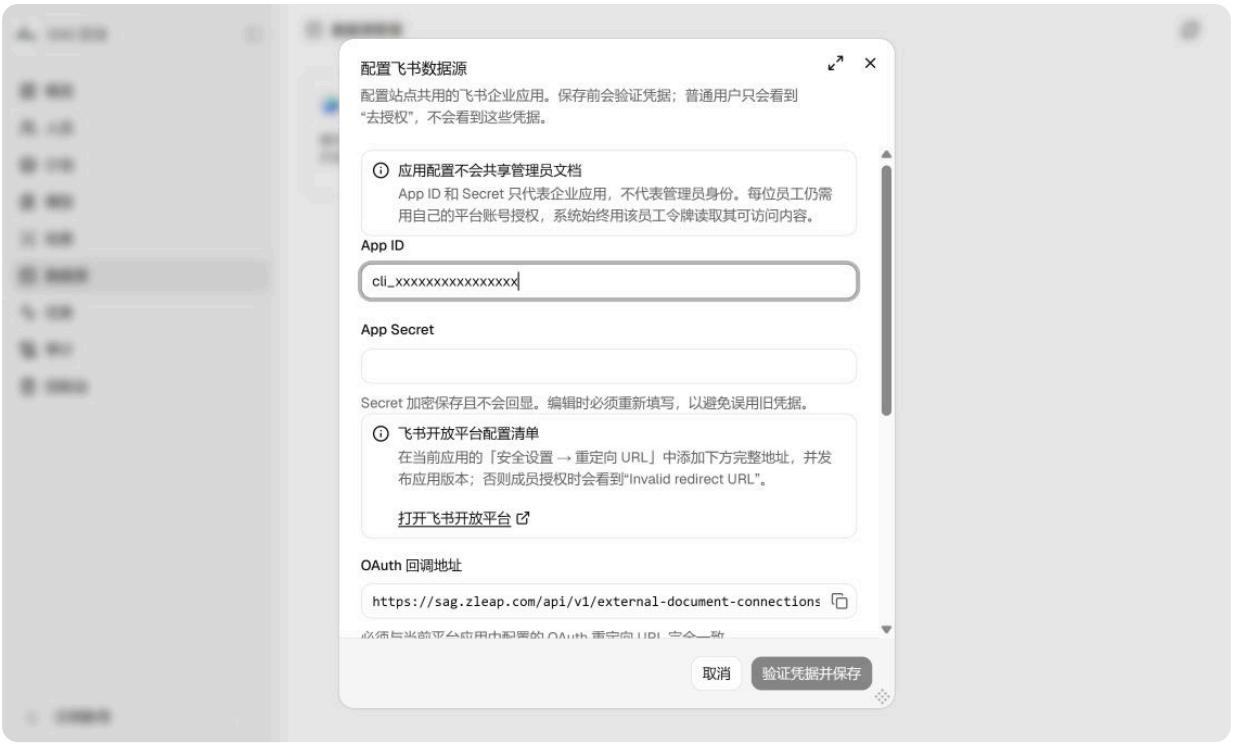
2 配置飞书应用

打开飞书开放平台，使用企业管理员账号完成以下配置：

1. 创建或打开一个企业自建应用，在应用凭证页面复制【App ID】和【App Secret】。
2. 在应用的权限管理中，按 SAG 配置弹窗列出的清单开通文档读取权限，至少核对 `offline_access`、`drive:drive:readonly`、`docx:document:readonly` 和 `search:docs:read`。
3. 在安全设置的重定向 URL / OAuth 设置中，新增 SAG 弹窗显示的【OAuth 回调地址】；必须完整复制，协议、域名、路径和末尾斜杠都不能自行修改。
4. 在应用的可用范围中加入需要使用同步功能的成员，并按飞书要求发布或启用应用版本。

回到 SAG，在飞书配置弹窗中填写【App ID】和【App Secret】，先点击【验证凭证】，验证成功后保存并启用。

下图依次展示配置弹窗的上半部分和向下滚动后的下半部分。





3 配置钉钉应用

打开钉钉开放平台，使用企业管理员账号完成以下配置：

1. 创建或打开一个企业内部应用，在应用凭证页面复制【Client ID / AppKey】和【Client Secret / AppSecret】。

2. 在应用权限管理中，按 SAG 配置弹窗列出的清单开通钉钉文档、知识库或云盘的只读权限；具体权限名称以当前钉钉控制台展示为准。
3. 在应用的 OAuth / 安全设置中，新增 SAG 弹窗显示的【OAuth 回调地址】；不要填写钉钉控制台自动生成的其他地址，也不要把 `localhost` 或服务器内网地址填入生产配置。
4. 设置应用可见范围，确保需要授权同步的成员能够使用该应用，并按钉钉要求发布或启用应用版本。

回到 SAG，在钉钉配置弹窗中填写【Client ID / AppKey】和【Client Secret / AppSecret】，先点击【验证凭证】，验证成功后保存并启用。

下图依次展示配置弹窗的上半部分和向下滚动后的下半部分。





4 验证并启用

确认数据源卡片状态变为【已启用】后，成员即可在【集成】→【数据源】中连接自己的平台账号。成员首次连接时，仍需要在飞书或钉钉的授权页确认本人权限。

> **配置字段怎么对应**
>
> 1. 飞书：SAG 的【App ID】对应飞书应用凭证中的 App ID，【App Secret】对应 App Secret。
> 2. 钉钉：SAG 的【Client ID / AppKey】对应钉钉应用凭证中的 Client ID 或 AppKey，【Client Secret / AppSecret】对应 Client Secret 或 AppSecret。
> 3. 两个平台的回调地址都以 SAG 配置弹窗当前显示的地址为准，不要手工拼接。

管理数据源

- 【编辑配置】：更新应用密钥或调整平台配置；
- 【停用】：停止新的账号授权和定时扫描，已有连接、同步计划和已同步文档会保留；
- 【启用】：恢复成员授权和同步计划运行；
- 【删除配置】：删除 SAG 保存的应用密钥和成员授权令牌，已有同步范围和已同步文档仍会保留。

成员连接账号并创建同步计划的操作，请查看同步外部文档。

> ⚠ 注意
>
> 1. 企业应用密钥只由站点管理员保管，不要发送给普通成员，也不要放入截图、群聊或工单。
> 2. 更换 App ID 或 Client ID 会清除现有成员授权并暂停相关同步计划；成员重新授权后可以恢复，已同步文档不会删除。
> 3. 平台后台的菜单和权限名称可能调整，请以配置弹窗列出的权限要求和开放平台实际页面为准。

配置检索方式与处理参数

检索方式（快速/精确）、单文档内容额度、批次与并发。

场景：调整企业资料的检索方式、单文档内容额度和处理批次参数。

角色：超级管理员。

路径：管理后台 → 【检索】。

1 查看当前生效配置

页面顶部的「当前生效配置」显示站点已保存的显式配置：索引模式、检索方式、并发、文档额度、向量模型批次和最近更新时间。新发布任务会把这些值固定为任务快照。



2 选择检索方式

在「配置新发布任务」中选择检索方式：快速优先速度，适合先看有没有相关内容；精确在多路召回和图谱扩展后使用大模型排序，结果更完整但耗时更长。索引范围固定为

完整模式，两种方式基于同一份完整索引，用户搜索时仍可临时切换。

3 调整文档额度

「单文档内容额度」限制新上传解析、Markdown 编辑和新发布的单份规范化内容，默认 8 MiB、最低 1 MiB；页面同时显示管理员配置、运维硬上限和实际生效值。超过运维上限的配置会被明确拒绝。调低额度不影响历史内容，但之后重新编辑或发布必须满足新额度。

4 调整批次与并发

三个参数各有用途：文档 LLM 分块并发（单份文档同时进行的抽取数）、向量模型批次（一次向量请求包含的文本数）、向量索引写入批次（一次写入的记录数）。修改后点击【保存配置】。

> **注意**
>
> 检索方式和执行参数只影响之后创建的发布任务，已发布内容不受影响。调整部署硬上限需由运维修改环境变量并重启服务。

计划与套餐审批

查看站点计划配置，审批成员的套餐申请。

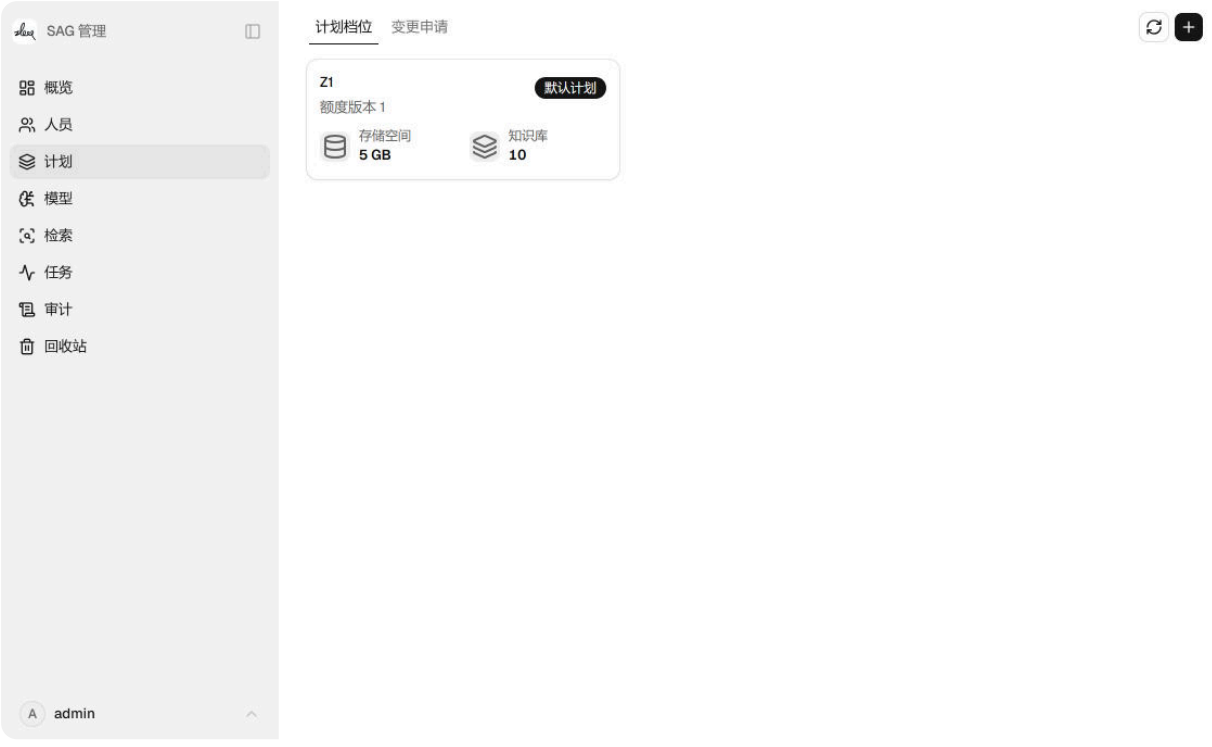
场景：管理站点计划（知识库数量、存储空间等额度），审批成员提交的套餐申请。

角色：超级管理员。

路径：管理后台 → 【计划】。

1 查看计划

【计划】页列出站点的计划配置和各成员的额度使用情况；顶部【申请】标签处理成员提交的套餐调整申请，批准后立即生效。



注意

套餐审批不涉及支付。成员在自己的【用量】里提交申请，管理员在这里审批；额度不足或连接异常时，先确认当前计划和资源状态，再联系交付团队处理。

任务中心

查看全站后台任务进度和失败详情。

场景：监控全站文档处理进度，排查处理失败的原因。

角色：超级管理员。

路径：管理后台 → 【任务】。

1 查看全站任务

【任务】是站点级只读视图，可查看所有用户发起的后台任务和失败详情。文档导入任务按解析、准备发布、准备、抽取、建索引、核对六个阶段展示进度；只有抽取阶段显示真实分块百分比。



2 处理失败任务

失败的任务会显示具体失败分类（解析失败、模型凭证被拒等）和建议操作。解析失败可直接从知识库文档列表重试，系统复用已保存原文件重新处理，不需要重新上传。

> **注意**
>
> 任务中心只有**只读查看权限**，不支持跨用户取消他人的任务。

任务、审计与回收站

站点级任务视图、操作审计与删除内容恢复。

场景：查看全站处理进度、追踪管理操作，或恢复误删内容。

角色：超级管理员。

路径：管理后台 → 【任务】【审计】或【回收站】。

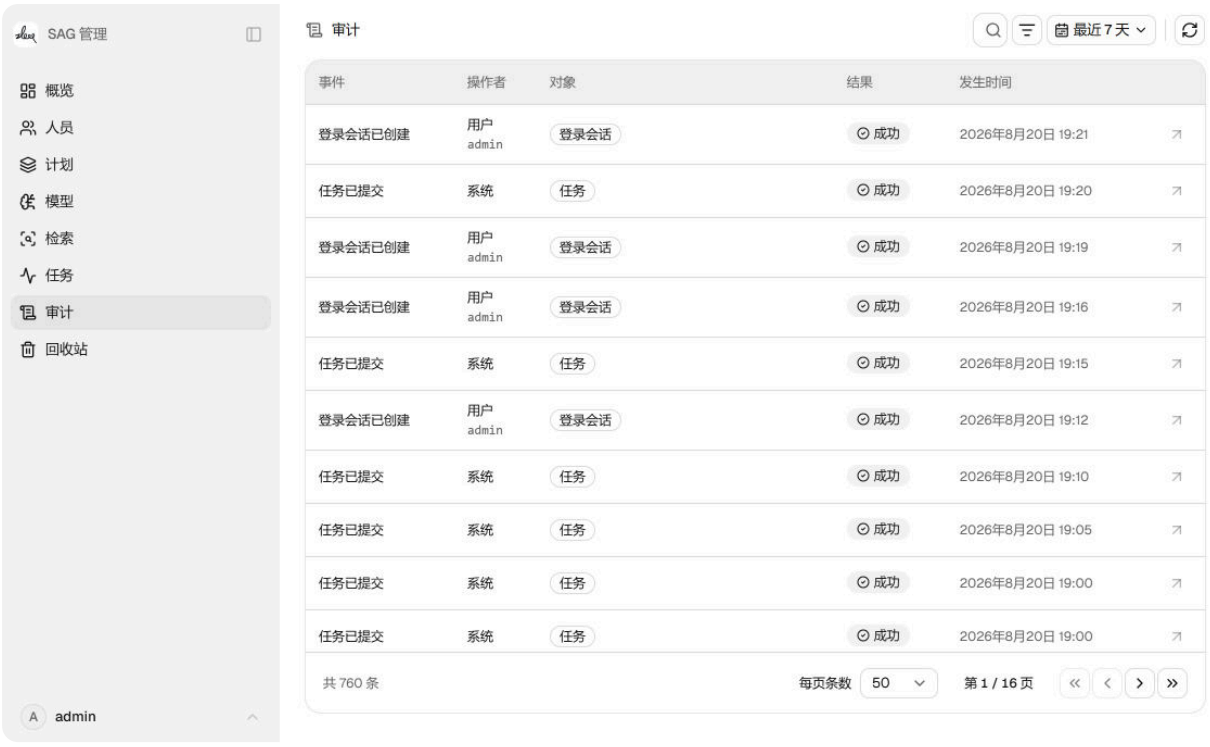
1 查看任务

【任务】是站点级只读视图，可查看全部用户的后台任务和失败详情；文档导入任务按解析、准备发布、准备、抽取、建索引、核对六个阶段展示进度。



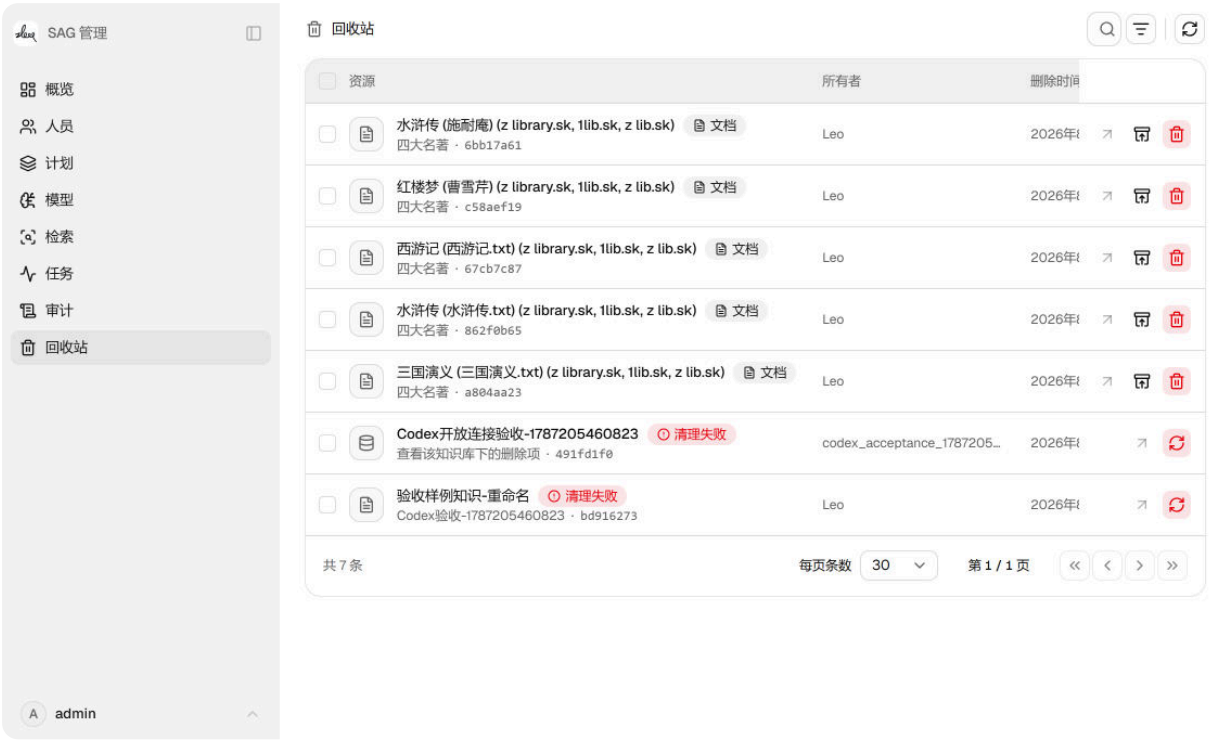
2 查看审计记录

【审计】按时间或操作类型查看变更记录，覆盖人员、知识库、文档、连接和配置等管理操作。



3 使用回收站

【回收站】可恢复被删除的知识库、文档和会话；确认不再使用后再执行永久删除。永久清理失败时会显示对应任务编号和失败代码，可跳转任务详情继续排查。



注意

永久删除后无法恢复。执行前请确认对象、权限和业务影响。

审计日志

追踪全站管理操作记录。

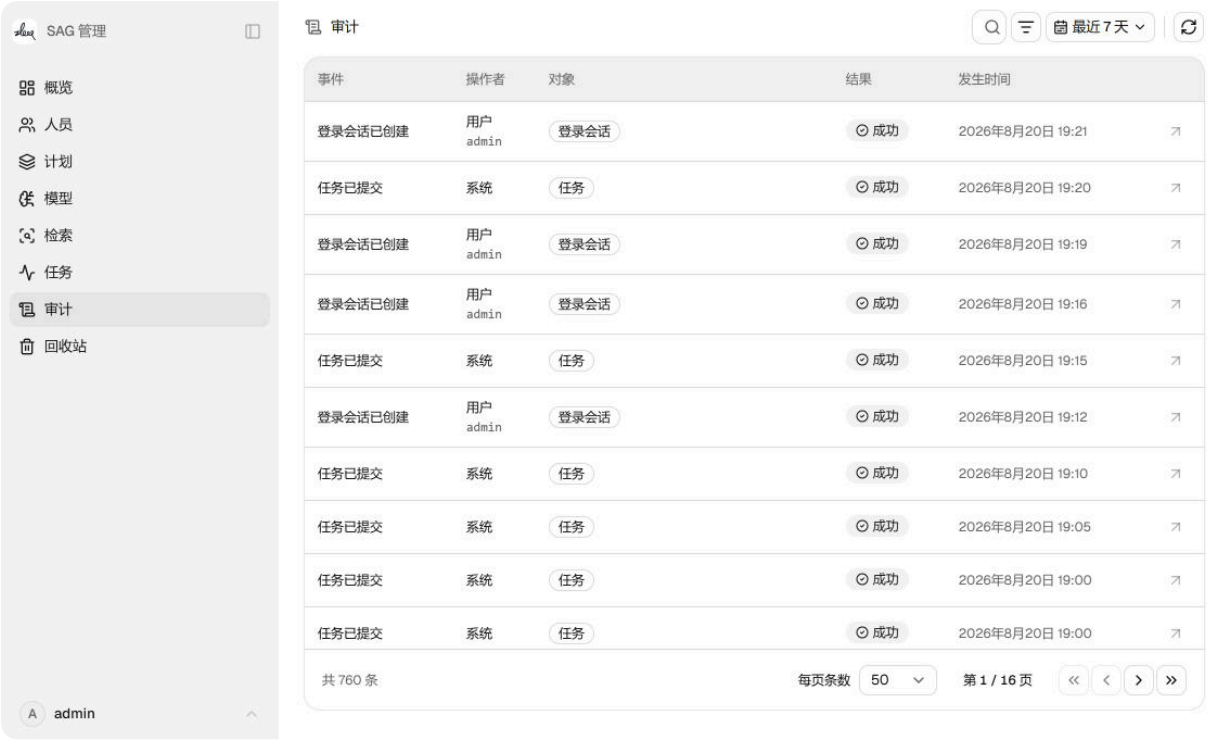
场景：查看人员、知识库、文档、连接和配置的变更历史，追踪操作来源。

角色：超级管理员。

路径：管理后台 → 【审计】。

1 查看审计记录

【审计】按时间或操作类型过滤变更记录，覆盖人员管理、知识库操作、文档变更、开放连接和系统配置等管理操作。每条记录显示操作者名称、操作类型、目标资源和时间。



回收站

恢复被删除的知识库、文档和会话，或执行永久删除。

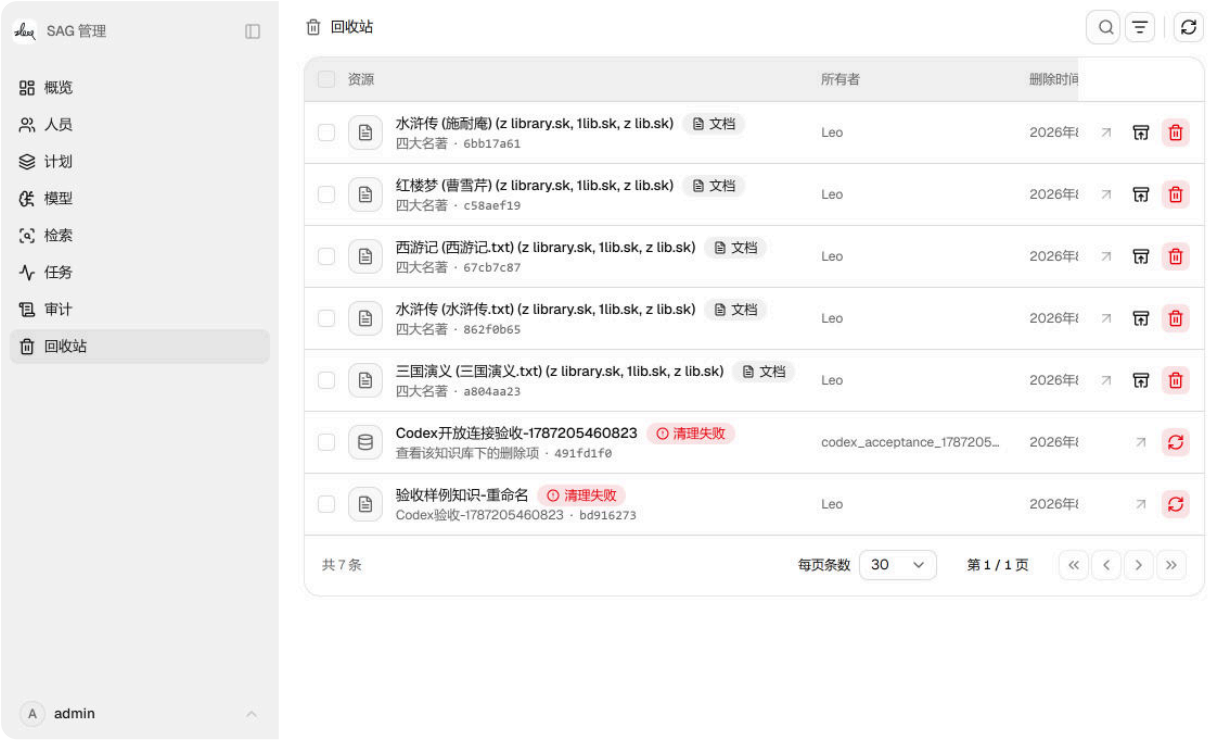
场景：成员误删内容后需要恢复，或确认不再使用时彻底清理。

角色：超级管理员。

路径：管理后台 → 【回收站】。

1 查看回收站

回收站汇总会话、知识库和文档的删除记录，显示删除时间、所有者和所属知识库。支持按资源类型、所有者、知识库和关键词筛选。



2 恢复内容

找到需要恢复的条目，点击【恢复】；如果父知识库也已删除，系统会提示并询问是否同时恢复父级。恢复成功后内容立即重新出现在成员工作区。

3 永久删除

确认内容不再使用后，点击【永久删除】并确认。永久删除后数据无法恢复，对象存储和索引会在后台异步清理。永久清理失败时会显示任务编号，可跳转任务中心继续排查。

> **注意**
>
> 永久删除后无法恢复。执行前请确认对象、权限和业务影响。

企业部署与升级

使用企业版交付包完成首次部署，并由站点管理员安全检查和执行版本升级。

场景：在客户服务器安装 SAG 企业版，并在新版本发布后完成备份、升级和验证。

角色：客户侧部署运维人员、站点管理员。

路径：首次部署使用企业版交付包；版本升级进入左下角账户入口 → 【设置】 → 【关于与更新】。

部署前准备

企业版交付包按服务器架构提供，例如：

```
zleap-enterprise-<版本>-linux-amd64.tar.gz
zleap-enterprise-<版本>-linux-arm64.tar.gz
```

部署服务器需要满足以下条件：

- Linux x86_64 或 Linux ARM64，且交付包架构与服务器一致；
- 已安装 Docker、Docker Compose v2、systemd 和 curl；
- 已准备 SAG 对外访问地址、反向代理、DNS 和 HTTPS；
- 可以访问交付方为当前客户授权的镜像仓库。

交付包中的客户配置不包含密码。镜像仓库密码和初始管理员密码应分别保存在服务器上的独立文件中，并设置为 root 所有、权限 `0600`。

完成首次部署

1 校验交付包

将压缩包和对应的 `.sha256` 文件上传到服务器，先校验文件完整性：

```
sha256sum -c zleap-enterprise-<版本>-linux-amd64.tar.gz.sha256
```

校验失败时不要继续解压或安装，应重新取得正确的交付包。

2 解压安装包

```
tar -xzf zleap-enterprise-<版本>-linux-amd64.tar.gz
cd zleap-enterprise-<版本>-linux-amd64
```

ARM64 服务器请使用对应的 `linux-arm64` 包和目录名。

3 执行安装

```
sudo ./install.sh \
  --registry-password-file /secure/acr-password \
  --admin-password-file /secure/admin-password
```

安装程序会进行环境预检、解析当前客户可用版本、部署平台和更新服务，并在最后执行安装验证。再次运行同一交付包不会覆盖已有产品数据、站点密钥或当前版本。

4 验证站点

安装完成后确认：

1. 命令输出显示 `企业版安装完成`；
2. SAG 对外地址可以打开；
3. 初始站点管理员可以登录；
4. API 就绪检查正常，核心容器处于运行状态；
5. 【设置】→【关于与更新】能够显示当前产品版本。

> **部署范围**
>
> 标准企业交付包不会自动安装 Docker 等系统依赖，也不自动配置外部 DNS、HTTPS、代理或自签 CA。完全离线或有特殊网络要求的环境，请使用交付方确认的离线部署方案，不要直接套用在线镜像安装步骤。

检查并执行升级

只有站点管理员可以发起升级。其他成员可以查看当前版本，但不会显示升级操作。

1 检查新版本

进入【设置】→【关于与更新】，查看【当前产品版本】，点击【检查更新】。



如果页面显示【更新服务暂不可用】，请联系部署运维人员检查更新服务配置；当前应用可以继续运行。

2 核对升级信息

发现新版本后，核对版本号、更新摘要、预计维护时间和最低起始版本。页面明确提示需要维护时，应提前通知成员并选择合适的维护窗口。

3 确认开始升级

点击【立即升级】，重新输入当前登录密码，再点击【开始升级】。

系统会先生成并验证升级前备份，然后更新服务。升级进行期间不要重复提交操作，也不要在服务器上手动替换容器或删除数据卷。

4 等待服务恢复

页面会显示当前升级阶段；服务重启期间浏览器会自动尝试重新连接。恢复后重新登录，确认【关于与更新】显示目标版本，并查看最近更新记录是否为【成功】。

5 完成业务验证

升级完成后至少检查：

1. 站点管理员和普通成员可以正常登录；
2. 原有知识库可以打开、搜索和问答；
3. 新文档可以上传并完成发布；
4. 本次版本涉及的新增功能可以正常进入和使用。

> ⚠ 注意
>
> 1. 升级前不要手动清理历史数据、数据库或数据卷；升级程序需要使用现有状态完成备份和迁移。
> 2. 升级失败时先保留页面中的阶段、错误信息和更新记录，再联系部署运维人员处理，不要连续重复发起升级。
> 3. 生产环境只应安装交付方正式发布并完成验收的版本，不要把内部候选版本用于正式业务。

更新日志

SAG 企业版各版本的新功能、改进与修复记录。

> **关于本日志**
>
> 仅记录成员和管理员可感知的产品变化。内部架构调整、性能优化和安全加固不在此列。

2026-08-18

新增

- 文档额度可由管理员配置 — 单份文档的内容大小上限不再固定。站点管理员可在「检索配置」中调整额度（默认 8 MiB），超出上限的上传会被明确拒绝，不再静默截断。
- 外部文档解析服务可选 — 管理员现在可以在「模型」页选择 302.AI MinerU 作为文档解析方式，用于处理含图片的 PDF 等复杂格式。内置解析器仍作为默认方案；清空绑定即回到内置解析。

改进

- 知识库文档模式与图谱模式分离 — 上传、筛选、批量操作只在文档模式下显示；切换到图谱后界面更简洁，减少图谱阅读时的干扰。
- 会话管理更清晰 — 会话菜单新增明确的「归档」和「删除」两个选项；删除需二次确认，操作后会话立即从目录移除，不会残留。
- 套餐配额提示更准确 — 知识库数量或存储空间到达上限时，页面现在明确提示联系管理员调整套餐，不再显示「稍后重试」类误导提示。

2026-08-17

新增

- 搜索结果与问答引用展示统一 — 搜索结果和问答引用侧栏现在使用同一套证据展示：标题、摘要、命中词和实体标签保持一致，来源更清晰可溯。
- 点击实体追问 — 搜索页点击实体标签会立即以该实体发起新搜索；问答引用侧栏中点击实体则在当前会话追加一轮带上下文的追问，正在输入的草稿不会丢失。
- 引用侧栏层次更清晰 — 引用侧栏改为「引用事件 / 关联事件 / 来源片段 / 原始文件」四层结构，方便从摘要一路追溯到原文。

改进

- 图谱视图操控更统一 — 2D 和 3D 图谱的布局、复位、刷新和镜头控制统一为带说明标签的紧凑按钮组，不再分散在不同位置。
- 账户菜单精简 — 去掉了和设置页重复的语言、主题快捷入口，账户菜单只保留账户信息、设置、套餐、管理后台入口和退出登录。

2026-08-16

新增

- 开放能力接入教程 — 「集成」页顶部新增可随时打开的接入教程，说明创建开放连接、为 Agent 安装 MCP、应用通过 HTTP API 接入和 API Key 管理的完整路径。没有任何连接时，空状态同时提供「创建第一个连接」和「查看教程」入口。

改进

- 管理员任务中心 — 改为站点级只读视图，可查看全部用户的后台任务和失败详情，方便管理员排查全站文档处理问题。

- 文档导入进度更透明 — 导入流程现在按解析、准备发布、准备、抽取、建索引、核对六个阶段展示真实进度；只有抽取阶段显示分块百分比，其他阶段不再用虚假进度条。
- 失败文档可直接重试 — 解析失败的文档可以原地重试，系统复用已保存的原文件重新处理，不需要重新上传；发布失败同理，复用现有草稿进入原发布流程。
- 管理员可测试模型连接 — 管理员在「模型」页可以对已保存的 LLM 和 Embedding 资源发起真实连通性测试，快速验证 API Key 是否有效。

2026-08-15

新增

- 知识库开放 MCP 和 HTTP API 接入 — 管理员在「集成」中创建开放连接后，可将知识库的搜索和问答能力通过 MCP 协议接入 AI Agent，或通过 HTTP API 接入企业应用，详见集成说明。

改进

- 搜索结果点击更流畅 — 去掉了独立的「详情」按钮，点击整条结果即可展开右侧来源面板，减少操作步骤。
- 模型接入更灵活 — 支持接入 OpenAI、DeepSeek、百炼、方舟、硅基流动及所有 OpenAI 兼容端点，不再只限于预设供应商。

2026-08-14

新增

- 问答主动寻找证据 — 问答助手现在会主动调用知识工具检索证据；第一轮无结果会自动调整检索策略，必要时升级精确模式，最多执行多轮查找。仍无证据时明确告知无法可靠回

答，不会根据通用知识编造库内结论。

- 失败问答给出补充线索 — 实在找不到证据时，助手会给出补充建议（如换个关键词、确认文档是否已发布），而不是只说「没有找到相关信息」。

改进

- 快速 / 精确两种模式统一 — 搜索和问答均使用同一套「快速 / 精确」模式名称，已有设置无需手动调整，平台自动迁移。

2026-08-13

改进

- 删除后立即从目录消失 — 知识库、文档和会话删除后立即从工作区目录、搜索结果和问答引用中消失，不再需要刷新页面。
- 管理后台回收站 — 站点管理员可在回收站查看并恢复被删除的知识库、文档和会话，或执行永久删除。永久删除后无法恢复。
- 管理员可删除未使用套餐 — 管理员现在可删除没有任何成员使用的套餐，长套餐列表支持滚动选择。
- 用户创建冲突提示清晰 — 用户名或邮箱已被占用时，系统现在明确说明冲突原因，不再只显示通用错误。

2026-08-12

改进

- PDF 在线预览 — 知识库文档的原始文件（PDF）现在在浏览器内直接预览，不再触发下载。

- 解析内容与原始文件标签分离 — 文档详情区分「解析内容」和「原始文件」两个标签，方便核对平台解析结果是否完整。

2026-08-11

新增

- 管理后台统一回收站 — 汇总会话、知识库和文档的删除记录，支持按资源类型、所有者、知识库筛选，可恢复或永久删除。

改进

- 删除与归档明确区分 — 归档只影响会话目录排序，删除才会进入回收站；两者不再混用。
- 卡片 / 列表视图切换样式优化 — 切换按钮改为紧凑尺寸，当前视图有明确选中样式，鼠标 hover 时不会覆盖选中状态。

常见问题

使用过程中常见问题的排查方法。

刚上传的文档搜不到

先回到知识库确认状态是否为「已发布」。只有已发布文档才会稳定出现在搜索和问答结果中；导入过程按解析、准备发布、准备、抽取、建索引、核对六个阶段推进，可在文档列表或管理后台【任务】中查看进度。

文档显示「处理失败」

打开文档查看具体失败分类：解析失败可直接重试（复用已保存的原文件，不需要重新上传）；模型凭证被拒请联系超级管理员检查模型配置；发布失败会复用现有草稿重新进入发布流程。

上传被拒绝，提示超出额度

单份文档内容额度由站点管理员配置（默认 8 MiB）。请拆分文件后重试，或联系管理员在【检索】配置中调整额度。

回答没有引用

检查知识库范围、文档状态和当前账号权限。助手最多执行 4 轮证据查找，仍无证据时会明确说明无法可靠回答——企业业务判断应以可核对的引用内容为准。

搜索提示结果不完整

查询扩展模型暂时不可用时，系统会用原问题继续执行受限检索并明确标记结果不完整。稍后重试，或联系超级管理员检查模型连接状态。

普通成员看不到管理后台

这是正常的权限表现，管理后台仅面向超级管理员。页面中没有某个入口，通常表示当前账号未被授予对应权限。

找不到注册入口 / 无法登录

新增成员由超级管理员创建；无法登录时请联系站点管理员重新发放初始凭证。

知识库或存储空间达到上限

这是套餐配额限制。在左下角账户菜单的【用量】中提交套餐申请，由站点管理员审批，批准后立即生效。

MCP 无法调用

请联系超级管理员确认连接已创建、已授权且 API Key 或访问凭证仍有效；连接可能被暂停或撤销。工具列表读取失败时，先在 MCP 客户端刷新工具列表。