Zleap Router

产品帮助文档

生成日期：2026 年 11 月 13 日

本文档由帮助中心自动导出，供离线查阅

目录

认识Zleap Router 1

快速开始 3

模型接入

连接供应商 15

配置模型 18

发布与采用企业配置 23

管理 API Key 29

成员完成方案接入 32

组织与权限

管理成员与控制台权限 36

审批客户端接入申请 38

维护部门 42

管理项目 44

分配角色与权限 47

管理登录连接与通讯录来源 50

工作洞察与审计

选择用量查看入口 53

查看概览与报表 55

分析团队使用情况 57

查询请求与上游尝试 60

查询审计日志 61

费用管理

查看费用总览 63

设置员工与项目预算 65

核对内部成本 66

导入供应商账单 67

成本对账 69

固定成本 71

汇率 73

运行监控

检查运行状态 75
检查核心组件与依赖 76
处理告警通知 77
通知与投递 79
管理日志容量与清理 81

系统设置

设置隐私保护 83
维护企业信息与接入地址 85
连接与管理宿主网络 86
查看在线客户端设备 87
设置请求限流 88
管理留存与数据清理 90
备份并恢复配置与报表 91
核对许可证状态 93

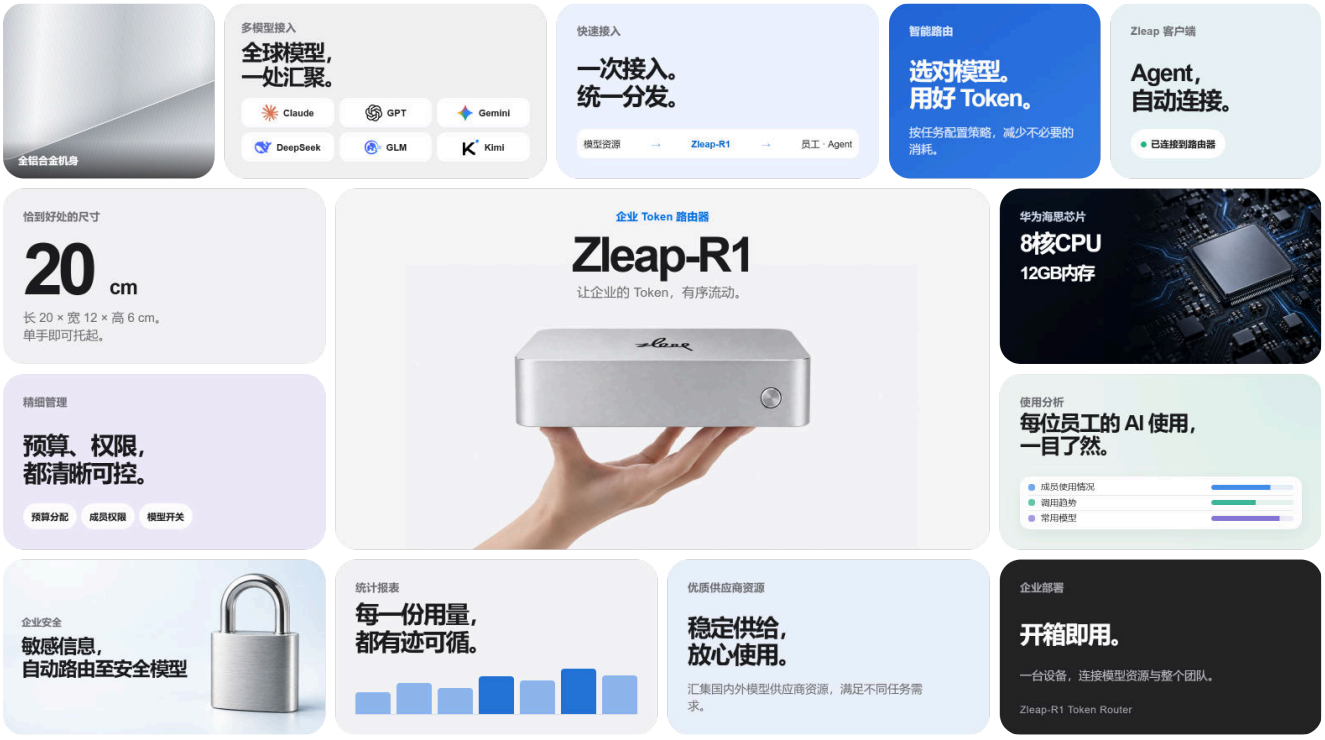
部署与更新

升级并确认运行版本 94
排查常见问题 96
更新日志 99

认识Zleap Router

连接企业模型，管理成员权限、预算与用量。

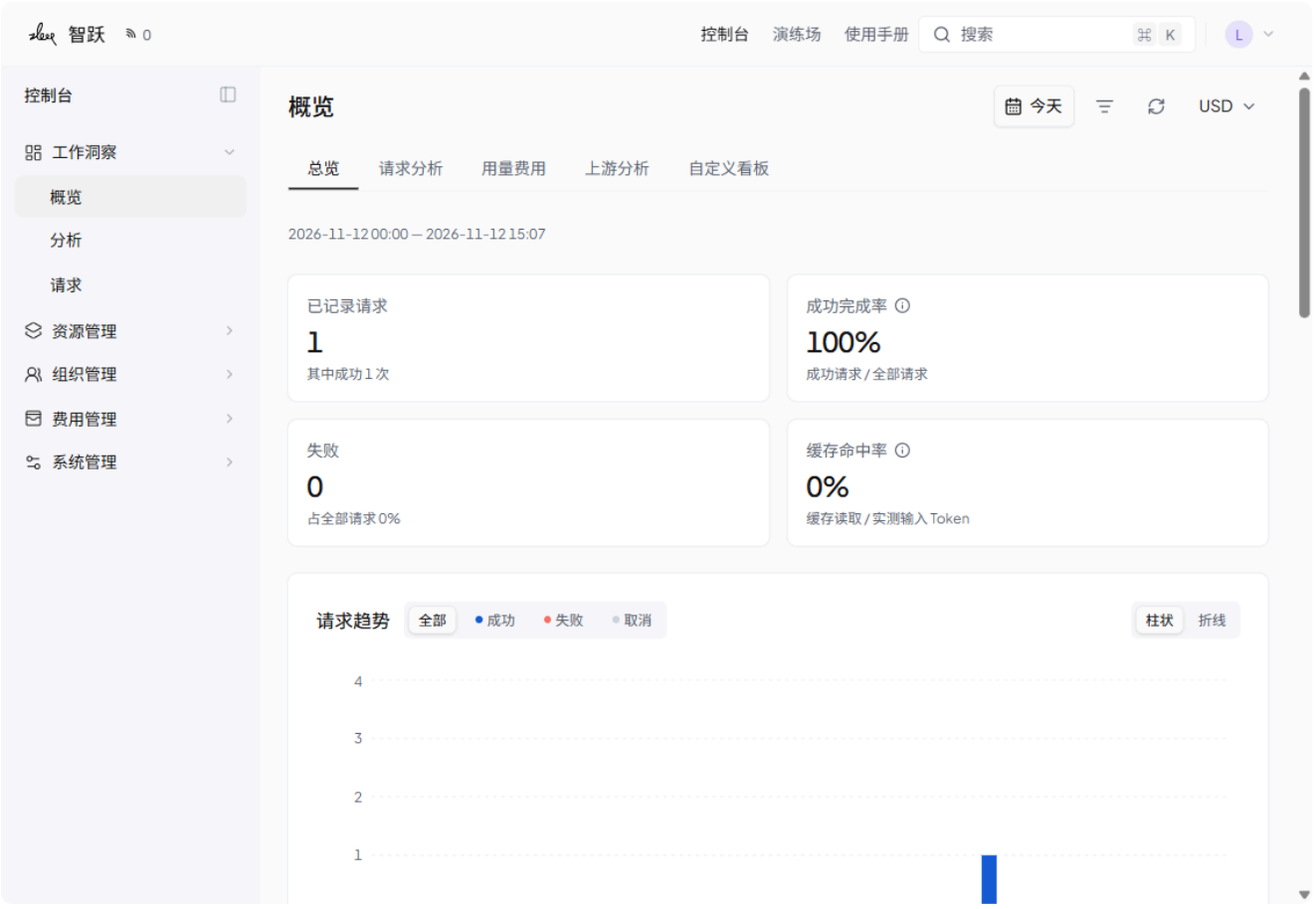
Zleap Router 为团队提供统一的模型入口。管理员维护供应商、模型、成员授权与预算；员工用 Token路由客户端连接企业，并将模型配置应用到自己的工具。



找到对应工作区

入口	用途	接着阅读
【工作洞察】→【概览】	看汇总、请求分析、用量费用及上游分析	查看概览与报表
【工作洞察】→【分析】	所有者查看团队使用情况、设置自动分析	分析团队使用情况
【工作洞察】→【请求】	查一条请求的结果、上游尝试与费用	查询请求
【资源管理】	连接供应商、维护模型、发布企业配置	连接供应商
【组织管理】→【成员】	管理成员、接入申请、部门与来源	审批接入申请
【系统管理】	设置安全、数据、备份与系统维护	设置隐私保护

菜单与数据范围取决于当前账号权限。 看不到入口时，请企业管理员检查权限；【分析】为所有者入口。



完成第一次接入

按快速开始依次准备资源、处理员工申请、发布配置，再发起真实请求。控制台账号开通与 Token路由客户端接入分别管理，员工不必先登录控制台才能申请接入。



了解本次更新

本轮手册按管理后台部署版本 1.0.56、Token路由客户端 0.1.19 的功能边界修订；版本变化与文档维护记录见更新日志。实例实际版本、权限及部署能力以当前界面为准。

快速开始

从部门与成员、供应商与模型，到客户端连接、组合配置和首次调用。

管理员先准备部门、成员、供应商、模型及工具组合；员工再通过 Token路由客户端连接企业管理后台，把配置应用到工具，最后发起请求核对结果。

完整流程：准备部门与成员 → 连接供应商 → 更新并导入模型 → 检查路由、授权与预算 → 配置并发布组合 → 客户端扫描路由 → 自主申请或使用连接码接入 → 一键应用企业组合 → 发起请求。

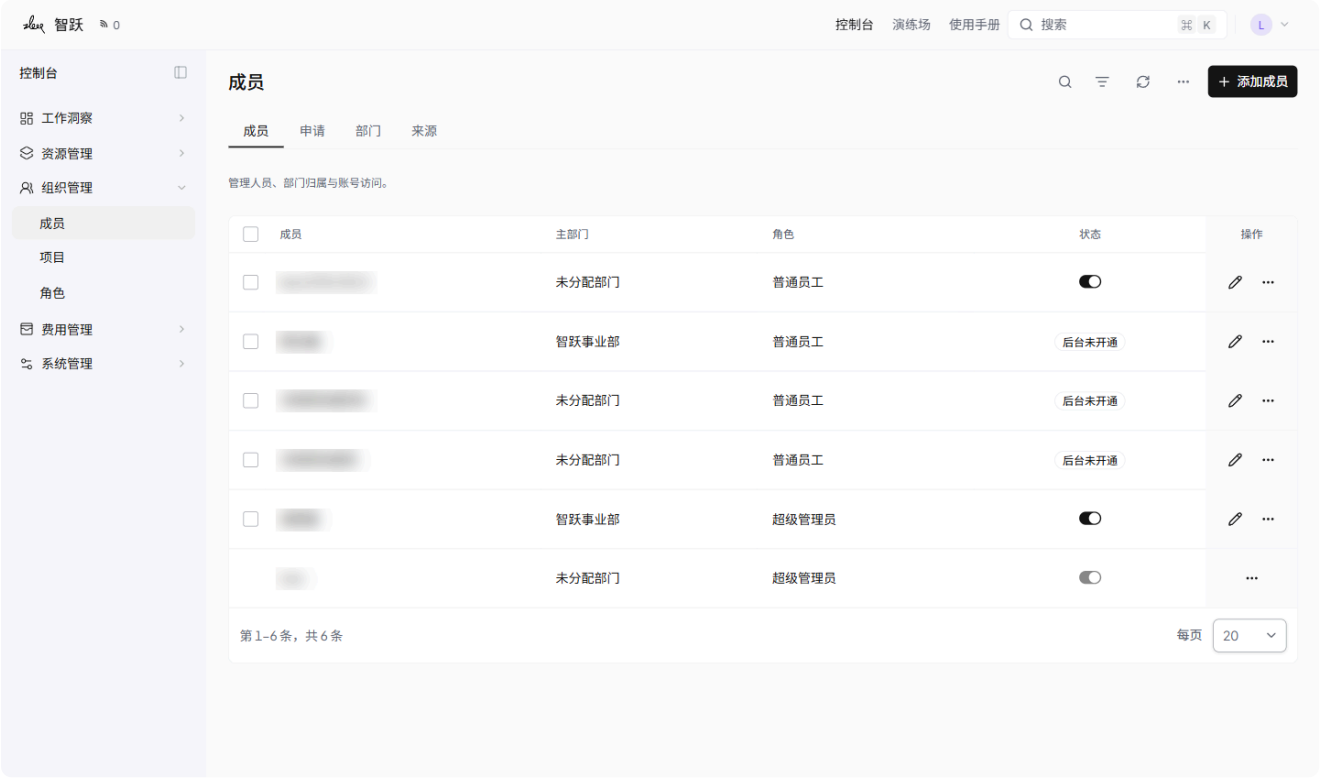
管理员需要资源管理、成员管理及相应预算权限。员工准备企业提供的 Token路由客户端安装包及要使用的工具。首次先用一名成员、一个模型、一个工具验证整条流程；员工电脑必须能访问管理后台，管理员电脑的 `localhost` 不能作为其他电脑的接入地址。

1. 准备部门与成员

打开【组织管理】→【成员】→【部门】，按需添加部门、选择上级部门。已有合适部门时直接使用，具体操作见维护部门。

成员可由管理员提前添加，也可在员工申请接入时建立：

- 提前添加：切回【成员】，点击【添加成员】，填写姓名及联系方式，选择部门和角色，再检查模型授权与预算。已有成员直接核对其设置。
- 申请时建立：到【申请】→【去设置】开放申请，设置新成员的默认部门和额度；审批时再核实身份与模型授权。若申请人已有成员记录，应关联原成员。





后台角色、模型授权和调用预算分别生效。 仅分配角色还不能保证成员能调用模型；员工使用 Token路由客户端接入，也不要求先开通控制台登录。详见管理成员与设置预算。

2. 连接供应商

1. 打开【资源管理】→【供应商】，点击【添加供应商】，填写便于辨认的连接名称。
2. 接入 302.AI 时选择【302.AI】；其他服务选择【自定义供应商】，按真实服务填写基础地址。
3. 向下填写供应商的【API Key】。按需在【高级配置】中设置代理等连接参数。
4. 要在创建后选模型，勾选【创建后选择要导入的模型】，点击【添加】；不勾选时只建立连接，按下一节手动更新模型。

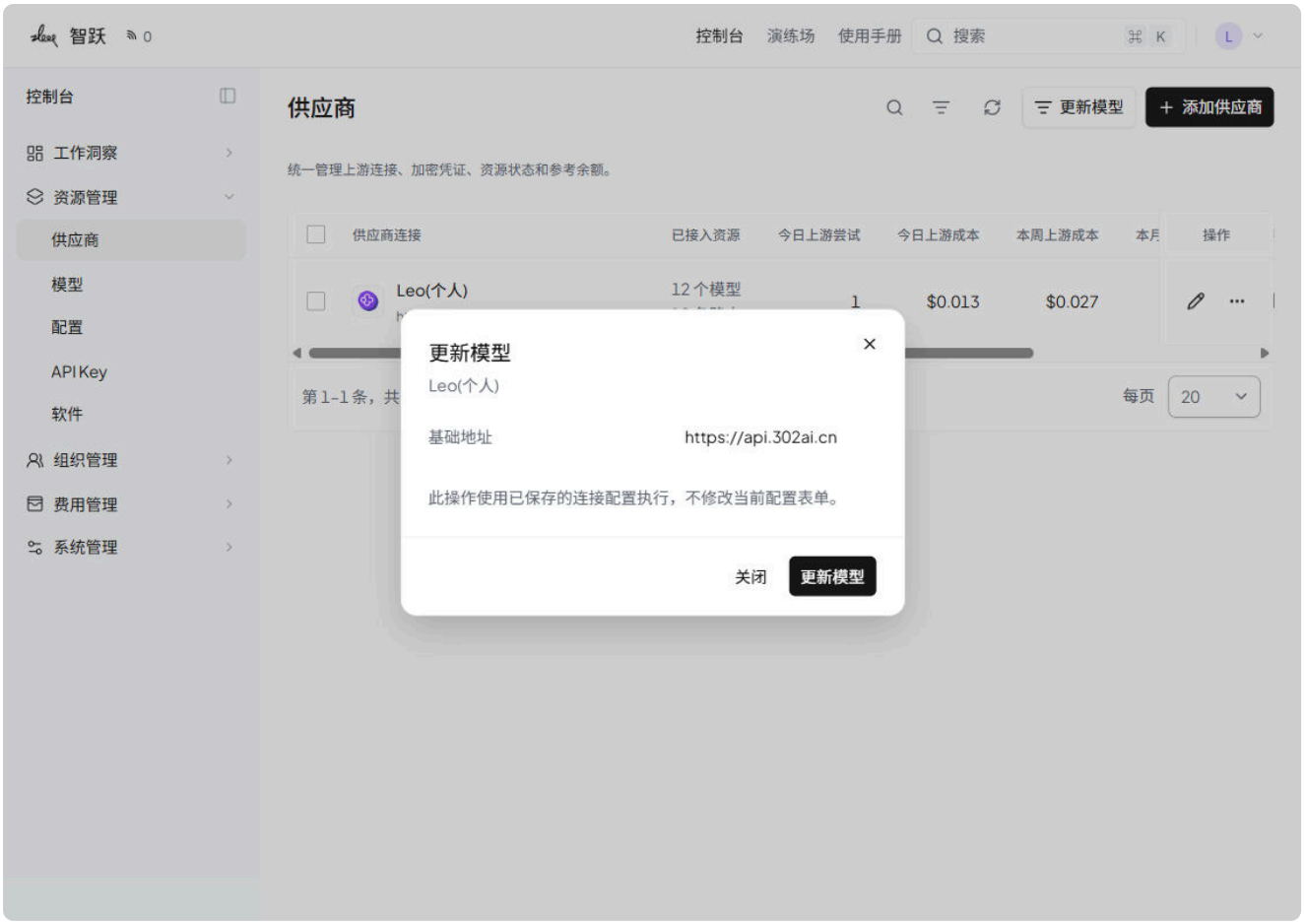
保存供应商连接后，还要选择并导入模型。 该选项会打开上游模型列表，不代表全部模型已自动导入。若修改已有供应商的地址或凭证，先保存，再执行更新模型；更新使用已保存的连接配置。

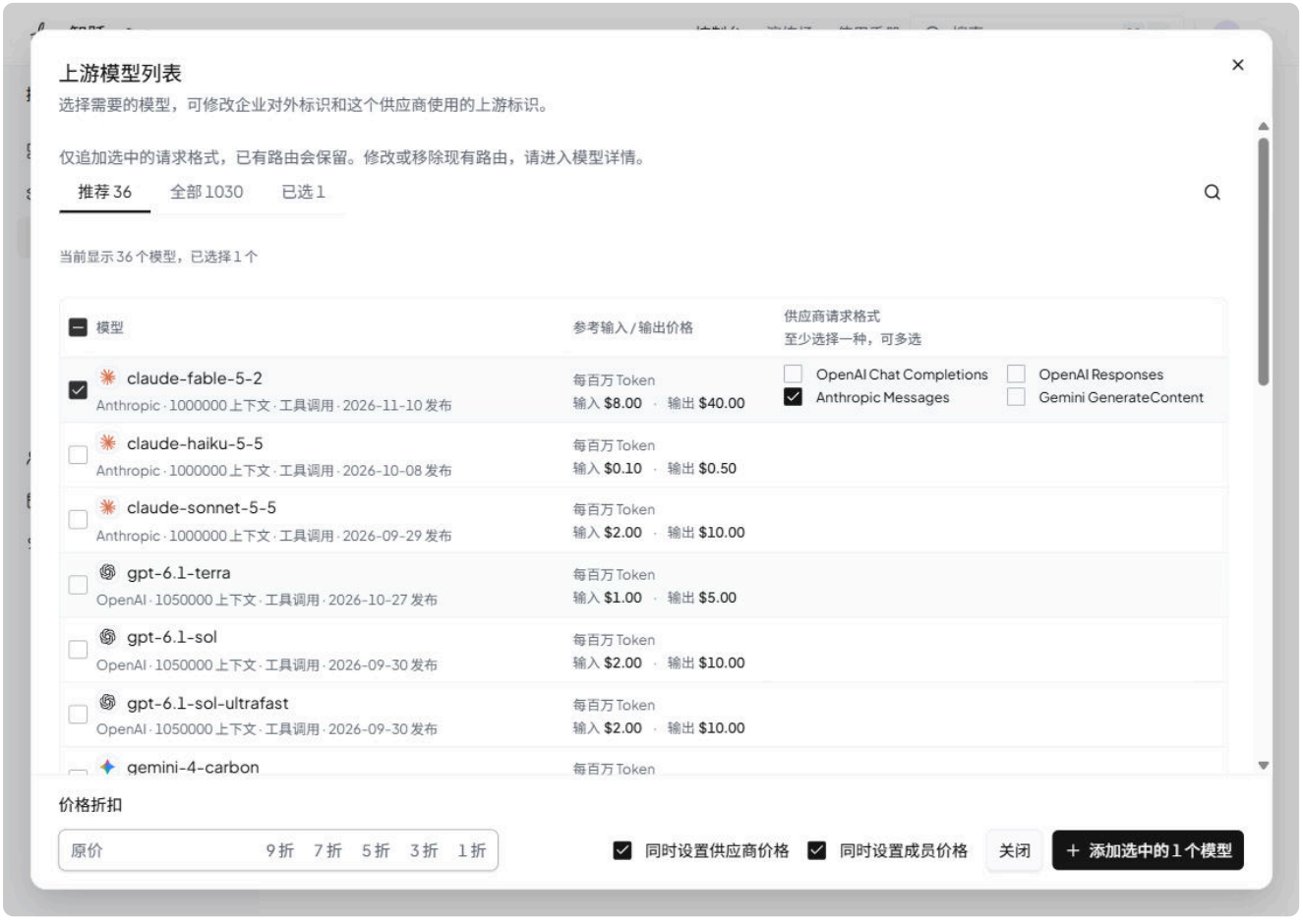


字段及连接设置见连接供应商。

3. 更新并导入模型

1. 在供应商页点击【更新模型】；有多个供应商时，选择刚建立的连接。按窗口提示执行，等待取得【上游模型列表】。创建后已打开列表的，可直接从下一步开始。
2. 从【推荐】或【全部】查找要使用的模型，勾选所需项；首次只选择准备验证的模型。
3. 核对【对外模型标识】与【上游模型标识】：前者供企业内工具调用，后者用于向供应商发送请求。核对【供应商请求格式】，至少选择一种与上游实际支持情况一致的格式。
4. 核对【价格设置】中的【同时设置供应商价格】和【同时设置成员价格】，按本次需求选择；参考价格未收录时，不把空缺理解为免费。
5. 点击【添加选中的 … 个模型】，查看结果中的新建、追加路由、已存在及失败数量。失败项按提示修正后再处理。

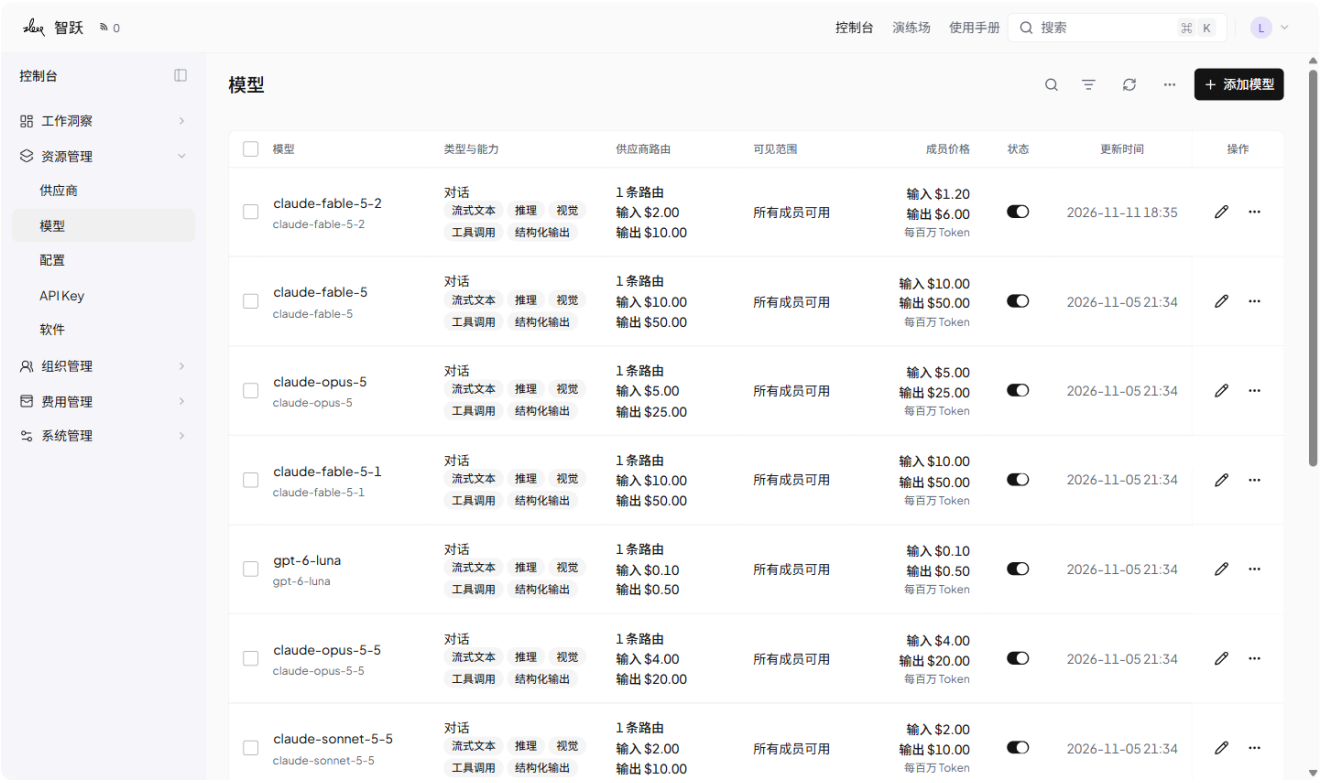

智跃
控制台
演练场
使用手册
搜索
供应商
更新模型
添加供应商
统一管理上游连接、加密凭证、资源状态和参考余额。
工作洞察
资源管理
供应商
模型
配置
API Key
软件
组织管理
费用管理
系统管理
供应商连接
已接入资源
今日上游尝试
今日上游成本
本周上游成本
操作
Leo(个人)
12 个模型
1
$0.013
$0.027
更新模型
Leo(个人)
基础地址
https://api.302ai.cn
此操作使用已保存的连接配置执行，不修改当前配置表单。
关闭
更新模型
每页
20


上游模型列表
选择需要的模型，可修改企业对外标识和这个供应商使用的上游标识。
仅追加选中的请求格式，已有路由会保留。修改或移除现有路由，请进入模型详情。
推荐 36
全部 1030
已选 1
当前显示 36 个模型，已选择 1 个
模型
参考输入 / 输出价格
供应商请求格式
至少选择一种，可多选
claude-fable-5-2
Anthropic · 1000000 上下文 · 工具调用 · 2026-11-10 发布
每百万 Token
输入 $8.00 · 输出 $40.00
OpenAI Chat Completions
OpenAI Responses
Anthropic Messages
Gemini GenerateContent
claude-haiku-5-5
Anthropic · 1000000 上下文 · 工具调用 · 2026-10-08 发布
每百万 Token
输入 $0.10 · 输出 $0.50
claude-sonnet-5-5
Anthropic · 1000000 上下文 · 工具调用 · 2026-09-29 发布
每百万 Token
输入 $2.00 · 输出 $10.00
gpt-6.1-terra
OpenAI · 1050000 上下文 · 工具调用 · 2026-10-27 发布
每百万 Token
输入 $1.00 · 输出 $5.00
gpt-6.1-sol
OpenAI · 1050000 上下文 · 工具调用 · 2026-09-30 发布
每百万 Token
输入 $2.00 · 输出 $10.00
gpt-6.1-sol-ultrafast
OpenAI · 1050000 上下文 · 工具调用 · 2026-09-30 发布
每百万 Token
输入 $2.00 · 输出 $10.00
gemini-4-carbon
每百万 Token
价格折扣
原价
9折
7折
5折
3折
1折
同时设置供应商价格
同时设置成员价格
关闭
添加选中的 1 个模型

这也是以后供应商上线新模型时的操作入口。更新列表不会替你删除旧路由；要修改或移除现有路由，进入模型详情处理。

4. 检查模型、路由与成员可用范围

打开【资源管理】→【模型】，找到刚导入的模型，检查启用状态、上游成本及成员价格。



进入模型编辑页，核对名称、模型标识与模型能力；在【供应商路由】检查供应商连接、上游模型标识和请求格式。需要手动新建模型时，点击【添加模型】填写这些信息，再保存。



确认上游账户有可用额度，目标成员有模型授权及可用预算，再继续配置工具。能读取模型列表还不能证明实际调用成功。路由、价格及访问设置见配置模型与供应商路由。

5. 由管理员准备企业方案

由管理员操作。已有合适且已发布的方案时可直接使用；尚未准备时，按下面步骤新建并发布。

1. 打开【资源管理】→【配置】，点击右上角【新建组合】，填写名称，例如“方案”。



2. 在【可用模型】勾选要提供给员工的模型；向下找到【组合中的工具】，打开要使用的工具开关。
3. 点击工具的【设置】，选择默认模型；提供角色模型的工具还需设置各角色的模型。处理工具行的【待完善】提示后，点击右上角【保存】。
4. 在组合详情点击【发布给员工】，检查确认窗口中的工具与模型后，点击【发布】。也可从配置列表的【更多】→【发布】进入。
5. 返回列表，确认方案状态为【已发布】，且员工有对应模型的授权与可用预算。只保存还不能供员工获取，必须完成发布。 后续修改方案后，点击【更新发布】交付新版本。



各工具的详细设置与发布操作见发布与采用企业配置。员工继续下面的接入步骤，在客户端一键应用管理员已发布的方案即可。

6. 在客户端扫描管理后台

安装并打开 Token路由客户端，进入【路由】→【组织】。首次尚无组织路由时会自动发现；已有路由时，点击【组织路由】标题右侧的刷新图标，打开【可连接路由器】。

等待扫描结果，选择公司的管理后台，点击对应行右侧的连接箭头。未发现时可点击【再次扫描】，或通过【手动添加】填写管理员提供的地址。



7. 自主申请并接通管理后台

推荐由员工自主申请，管理员审批，无须逐个提前创建账号。 管理员先在【组织管理】→【成员】→【申请】中开放申请，并设置新成员的默认部门和额度；模型授权在成员设置中核对。

1. 员工选择管理后台后，等待接入能力加载，切换到【申请接入】。
2. 填写姓名及邮箱或手机号，按需补充部门和说明，点击【申请接入】。
3. 管理员在后台【申请】中核实身份并批准；已有成员记录时关联原成员，避免重复建立。
4. 员工保持客户端联网，批准后由客户端领取并保存连接，再进入对应路由。



提交后客户端显示【等待管理员审批】；可以关闭窗口，保持客户端联网，审批通过后由客户端自动接入。

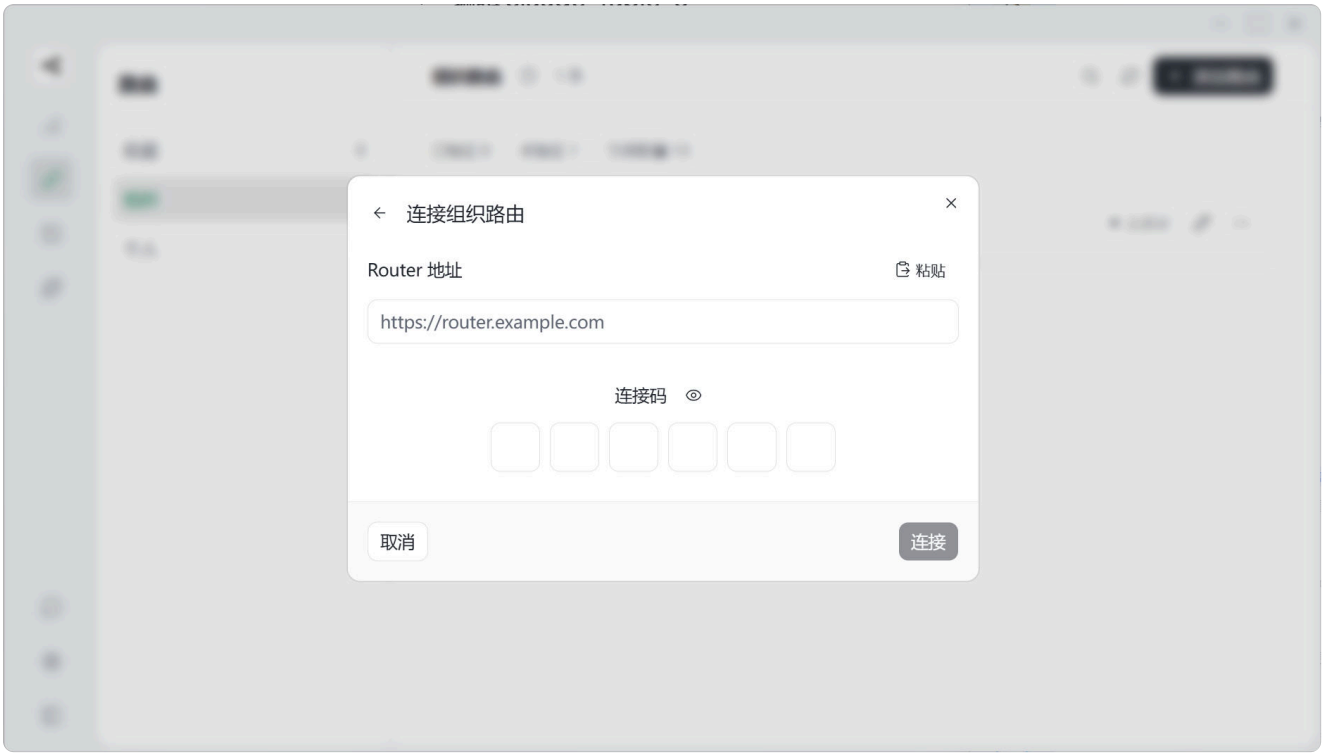


管理员在【组织管理】→【成员】→【申请】筛选【待审批】，核实申请详情，再点击【通过】。新成员检查本次授权并确认；已有成员改选【关联已有成员】，核对关联对象。审批操作见审批接入申请。





若未显示申请页签，先请管理员确认当前管理后台是否开放申请。已取得连接码的员工，也可选择【使用连接码】，填写六位连接码后点击【连接】。

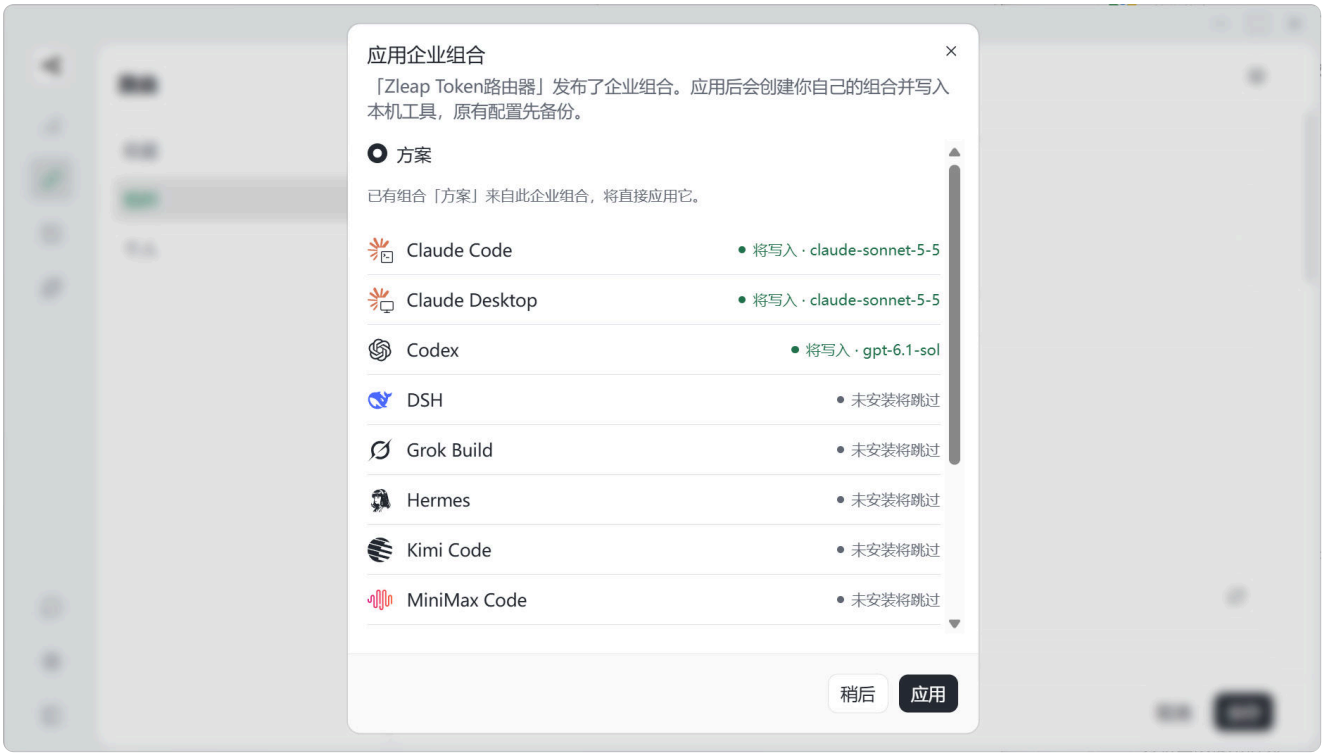


申请状态与异常处理见成员完成方案接入。

8. 一键应用管理员的方案

1. 安装要使用的工具。接通管理后台后，在【应用企业组合】窗口选择管理员发布的【方案】。若先前选择了【稍后】，可进入对应组织路由，在编辑页右上角点击【应用企业组合】重新打开。
2. 查看工具预览，点击【应用】。客户端自动准备本地组合并把配置写入工具，原有配置先备份；员工无需手动新建组合。未安装的工具会跳过。
3. 按窗口提示查看应用结果。应用完成后，务必先关闭并重新打开目标工具，再新建会话发起请求。

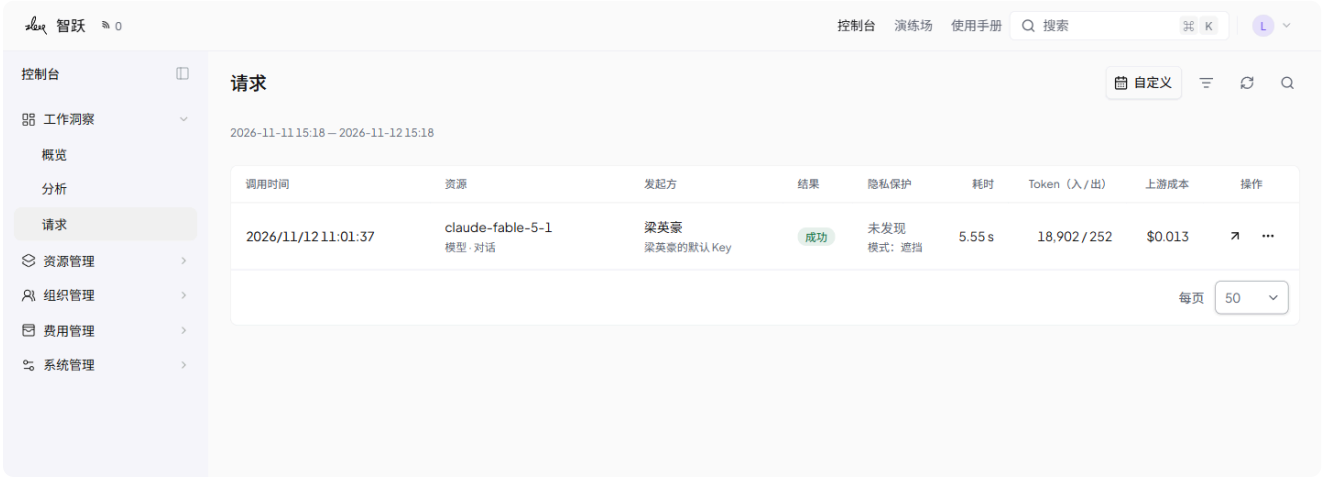
配置应用成功后，已打开的工具可能仍在使用旧配置。请先重启目标工具，让它读取新配置，再验证请求是否经过公司管理后台。



如果没有可用企业组合，请管理员先发布方案。应用后继续完成下面的首次请求，确认工具实际可用。

9. 发起并核对首次请求

1. 确认目标工具已关闭并重新打开，再新建会话并发送不含业务资料的测试请求，例如“请只回复：连接成功”。
2. 管理员到【工作洞察】→【请求】，按时间和成员找到本次记录，核对成员、模型、最终状态、上游尝试和费用。
3. 工具确实收到回答，且请求归属正确，才确认首次接入完成。再到【概览】查看汇总；所有者需要 AI 洞察时，到【分析】配置自动分析。



若失败，先核对连接地址、成员授权、预算、模型状态及上游响应，按常见问题排查。

模型接入

连接供应商

添加上游连接，选择导入模型，并更新模型与价格。

打开【资源管理】→【供应商】。

添加供应商连接

1. 点击【添加供应商】，填写连接名称。接入 302.AI 时选择【302.AI】；其他服务选择【自定义供应商】，可从【常用服务】填入地址，也可手动输入。
2. 向下找到【API Key】，填写供应商提供的凭证。代理和固定请求头等设置在【高级配置】中。
3. 需要立即选择模型时，勾选【创建后选择要导入的模型】；只建立连接时取消勾选，稍后再用【更新模型】获取列表。
4. 点击【添加】保存。API Key 保存后只显示隐藏值。

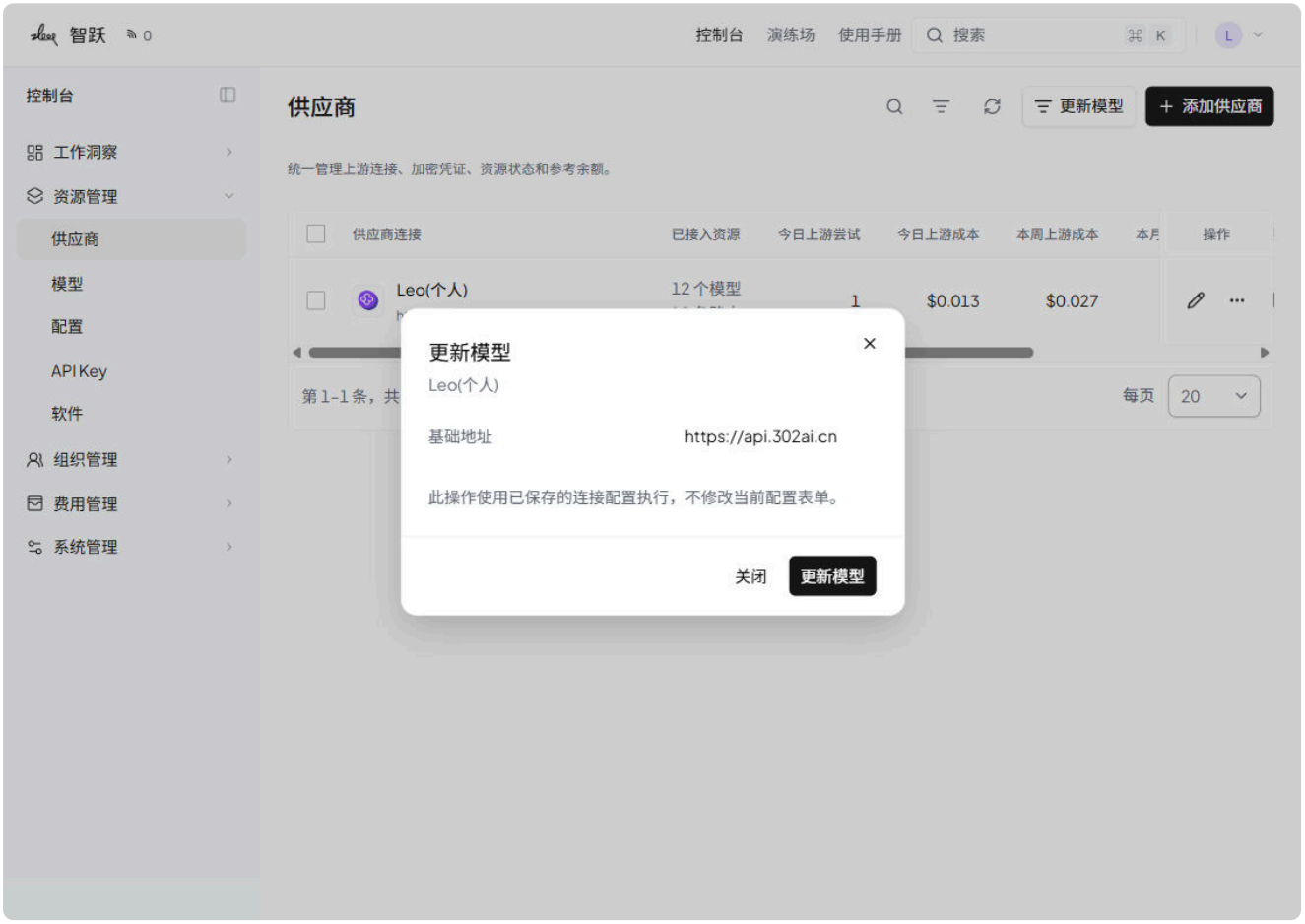


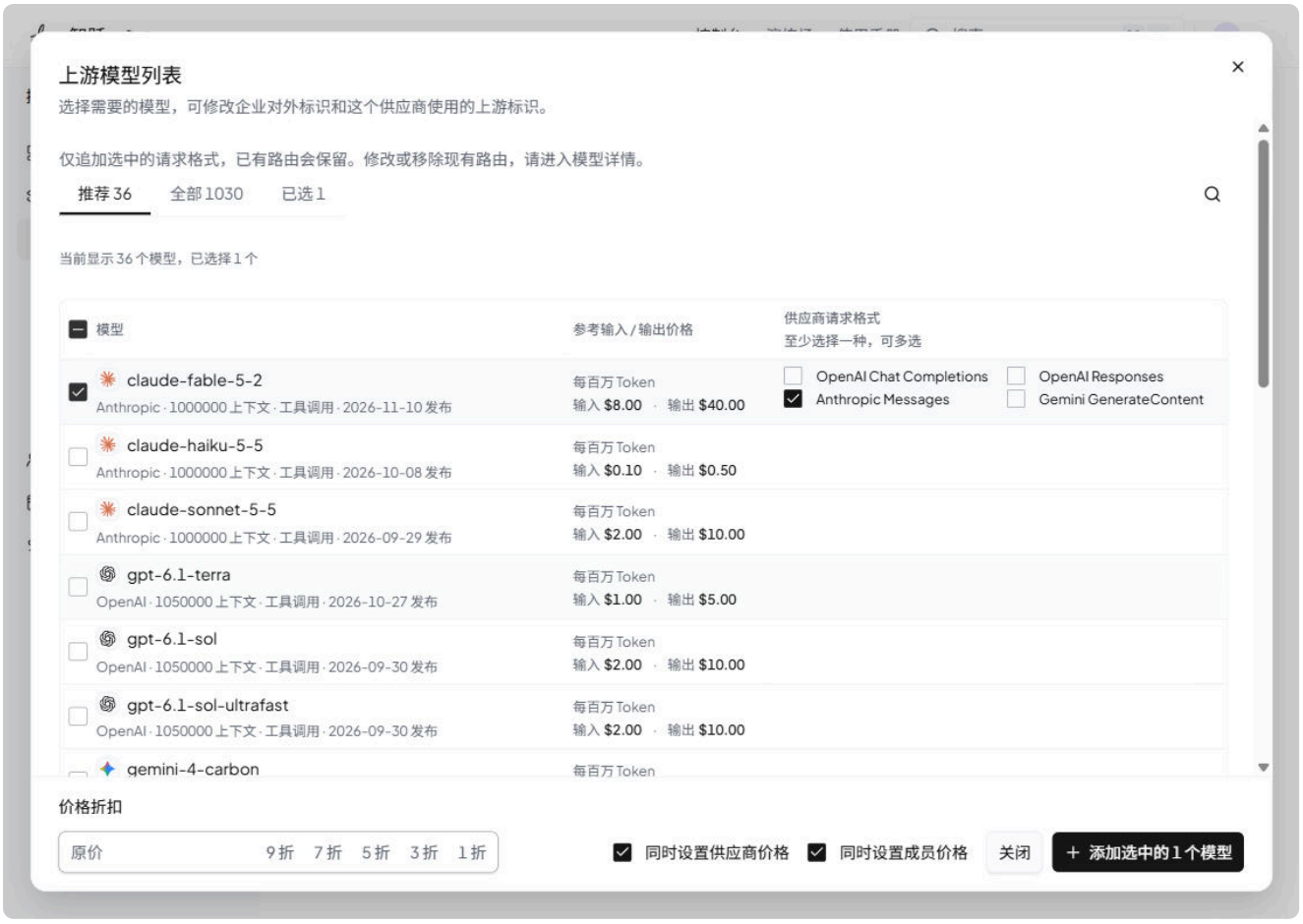
该选项会在创建后打开上游模型列表，还需选择模型并执行添加。供应商连接已保存，不代表模型已经全部导入。

更新并选择模型

1. 在供应商页点击【更新模型】；有多个连接时选择本次供应商。更新使用已保存的连接配置，修改地址或凭证后需先保存。
2. 按窗口提示执行并等待【上游模型列表】。从【推荐】或【全部】查找需要的模型，勾选要导入的项。
3. 核对【对外模型标识】、【上游模型标识】和【供应商请求格式】。请求格式至少选择一种，按上游真实能力设置。
4. 通过【价格设置】决定是否【同时设置供应商价格】、【同时设置成员价格】，核对参考价格、折扣及最终单价，避免重复折扣。价格未收录不表示免费。

5. 点击【添加选中的 … 个模型】，查看新建、追加路由、已存在和失败数量。失败项按结果说明修正后再处理。





已有路由会保留，本次仅追加选中的请求格式。要修改或移除现有路由，进入模型详情处理。供应商推出新模型后也按上述流程更新；只打开窗口不等于完成导入。

核对导入结果

到【资源管理】→【模型】，检查模型启用状态、供应商路由、上游成本和成员价格，再为目标成员确认授权与预算。读取到模型列表后，仍需通过真实请求验证上游可用。

本页截图展示真实的创建表单、更新确认和上游模型选择；尚未提交图示选择，导入结果截图仍待补拍。

模型接入

配置模型

在统一模型页面维护能力、路由、成本、成员价格与可见范围。

打开【资源管理】→【模型】。

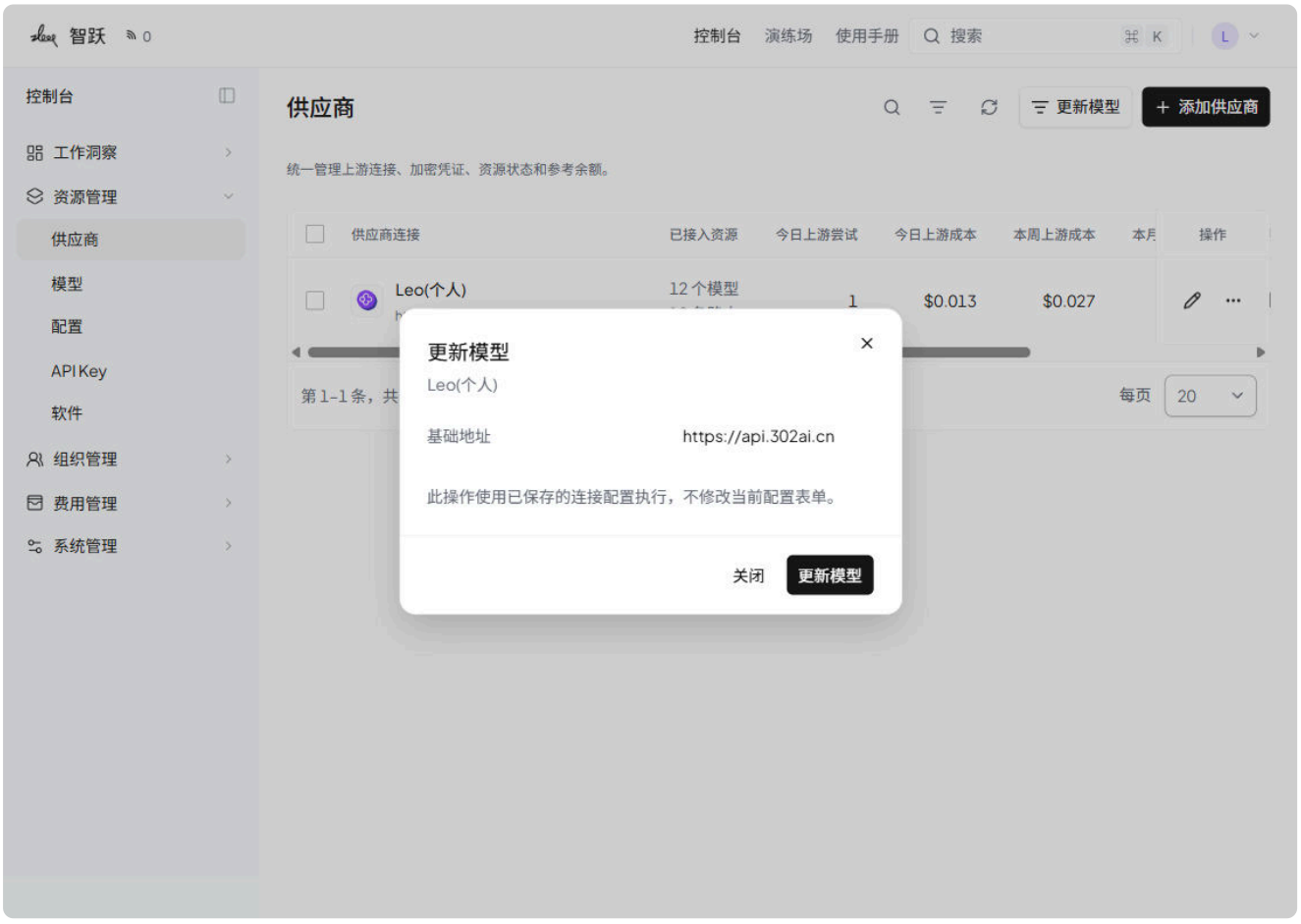
选择添加模型的方式

模型可以从供应商获取，也可以自行新建。两种方式最终都在【资源管理】→【模型】统一管理。

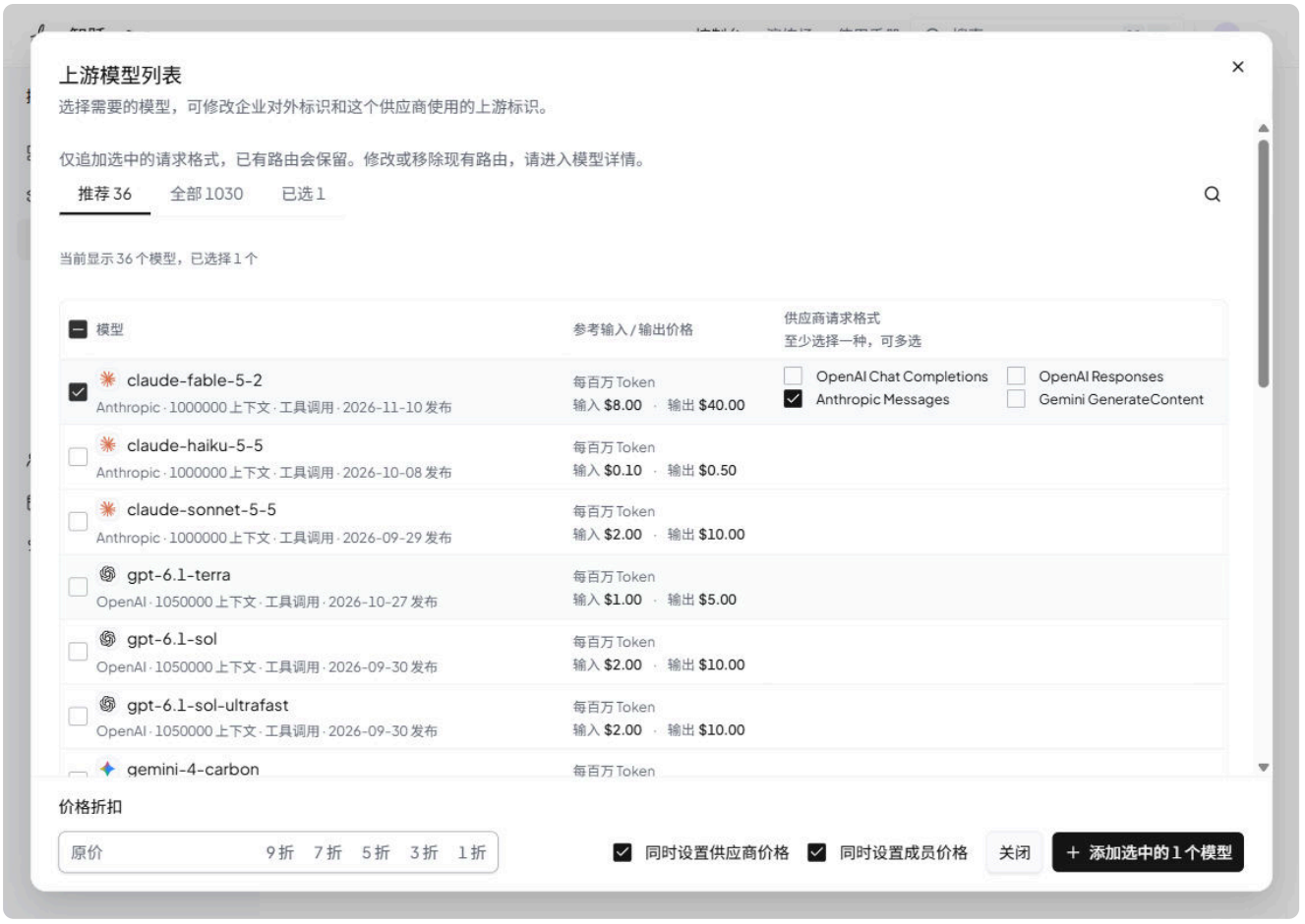
方式	入口	适用情况
从供应商拉取并导入	【供应商】→【更新模型】	上游支持获取模型列表，需要批量选择模型或追加路由。
自行新建	【模型】→【添加模型】	上游未提供模型列表，或需要自行指定对外模型标识与供应商路由。

从供应商拉取并导入模型

1. 打开【资源管理】→【供应商】，先添加供应商连接并保存地址与 API Key。已有连接时直接使用。
2. 点击【更新模型】，选择本次供应商，按窗口提示获取上游模型列表。



3. 在【上游模型列表】从【推荐】或【全部】查找并勾选需要的模型，核对【对外模型标识】【上游模型标识】与【供应商请求格式】。
4. 展开【价格设置】，按需设置供应商价格、成员价格及折扣，再点击【添加选中的 … 个模型】。

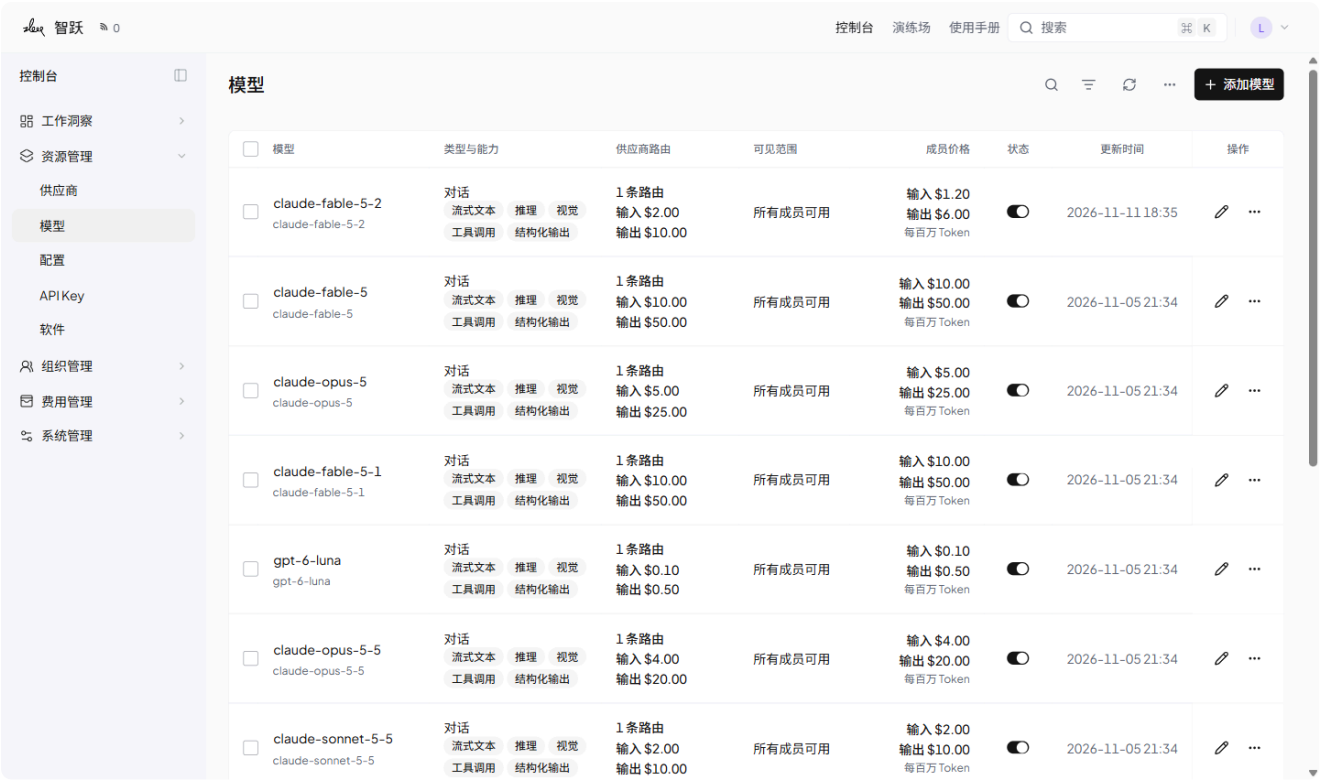


5. 查看添加结果，再打开【资源管理】→【模型】，找到导入的模型，检查路由、价格与可见范围。获取列表后还需勾选并添加，才算完成导入。

连接创建与价格同步的详细说明见连接供应商。

在模型列表自行新建模型

1. 打开【资源管理】→【模型】，点击右上角【添加模型】。



2. 填写名称和【模型标识】，选择模型类型及上游实际支持的能力。模型标识是员工调用时使用的名称，供应商展示名不能直接替代上游模型标识。
3. 配置【供应商路由】：选择已保存的供应商连接，填写上游模型标识，并选择该上游支持的请求格式。
4. 设置上游成本、成员计费价格与可见范围，点击【保存】。保存后返回列表，确认模型状态，再由有授权和预算的成员发起测试请求。

新建时的供应商路由与价格填写说明见下文。已有模型需要调整时，点击行内编辑按钮进入编辑页。



使用模型列表管理已有模型

在列表查看模型类型与能力、供应商路由、可见范围、成员价格和状态。需要修改时使用行内编辑按钮；需要临时停止提供模型时，使用该行状态开关。停用前确认是否有员工或企业方案仍在使用。

模型行右侧【…】提供的操作受权限和当前状态限制，具体以实际菜单为准。列表入口见上图；展开菜单的操作截图尚待补拍。

维护供应商路由

进入【供应商路由】，为路由选择供应商连接、上游模型标识和请求格式。上游为 Anthropic 接口时使用对应 Messages 格式；不要只凭模型名称判断协议。

多条路由用于提供可用资源。逐条检查启用状态、尝试次数和健康状态；具体尝试设置以当前表单为准。需要检查真实连通性时使用检测入口，检测可能产生上游调用和费用。

区分上游成本与成员价格

【上游成本】用于企业核算，【成员计费价格】用于内部计量。检查币种、计费方式、输入、输出、缓存读取与写入等适用价格。音频、图像、视频等类型按各自支持的计量项填写，不能全部套用文本 Token 单价。

供应商同步价格时应核对折扣及手工价格保留方式，避免在已经折扣的价格上再次手工叠加同一折扣。

检查成员可用范围

在【可见范围】设置可调用成员，保存后检查模型状态和供应商路由。让目标成员刷新 Token路由客户端模型目录，再发起真实请求；失败时到请求查看授权、预算及上游响应。

模型接入

发布与采用企业配置

设置组合中的模型和工具，发布给员工，再在客户端采用和应用。

管理员在【资源管理】→【配置】组织各工具的模型设置，发布给员工。员工在客户端通过【应用企业组合】一键应用到本地工具。企业组合中可包含多个工具，每个工具分别使用自己的配置。



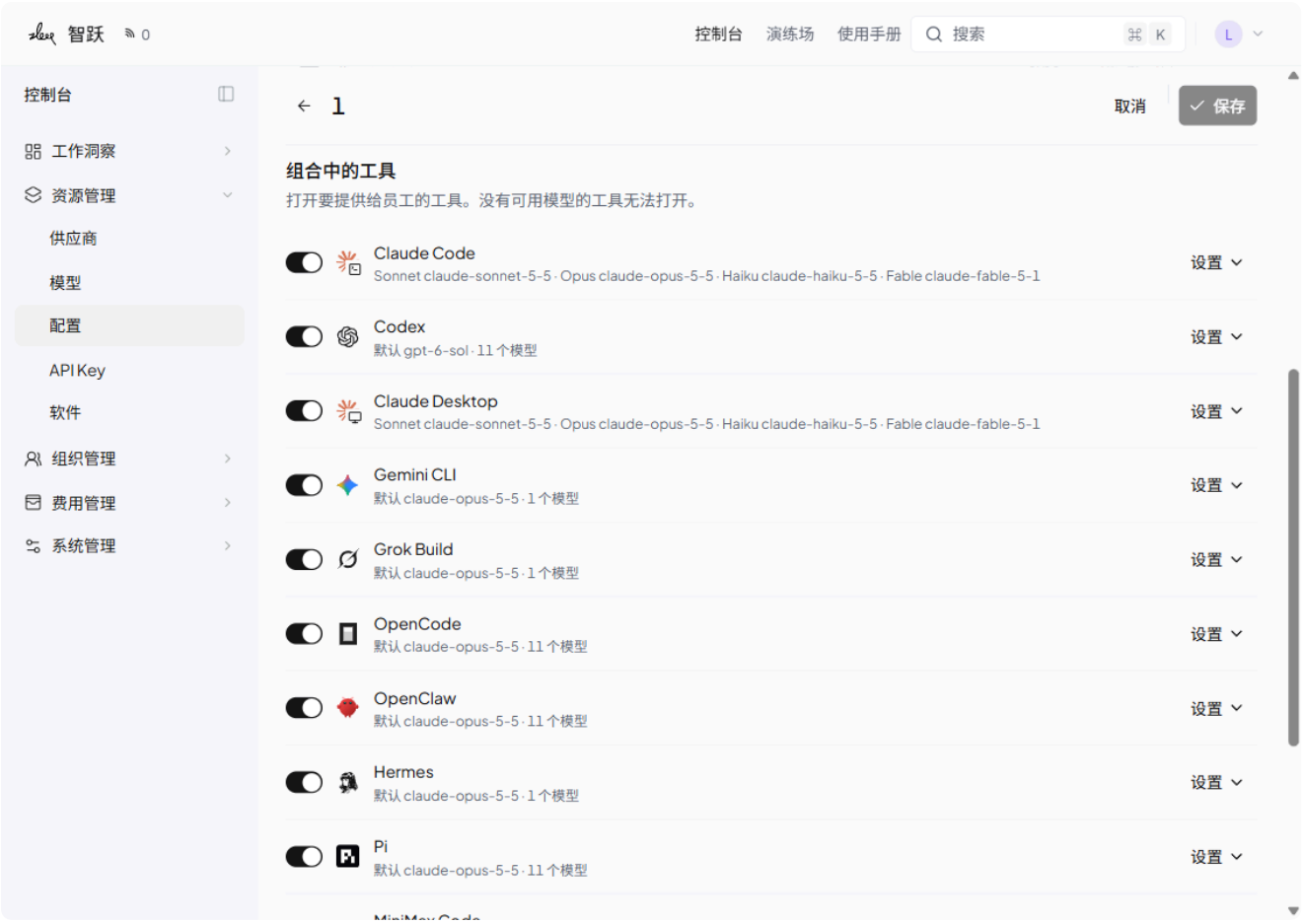
选择组合中的模型

1. 点击【新建组合】，填写便于员工辨认的组合名称。
2. 在【可用模型】勾选要提供的模型，可通过搜索定位。这里选的是供各工具使用的模型范围，各工具只使用其中自己支持的项。
3. 若列表没有可用模型，先到【模型】检查模型是否导入并启用。成员是否能使用这些模型，还取决于其授权与预算。



设置每个工具

向下找到【组合中的工具】，打开要提供的工具开关，再点击该工具的【设置】展开配置。没有兼容模型的工具无法打开。



设置项	如何填写
默认模型	选择该工具平时使用的模型，核对是否来自本次可用模型范围。
角色模型	工具提供角色设置时，逐个检查角色与模型的对应关系；自动匹配后也应复核。
协议	仅在工具提供该选项时设置，选择模型与路由实际支持的协议。
推理强度、上下文等	按工具实际字段和模型能力设置；没有调整需求时保留默认，不把声明的上下文大小当作上游能力保证。

例如同时提供 Claude Code 和 Codex 时，先勾选它们各自支持的模型，再分别打开两个工具。核对 Claude Code 的角色模型，以及 Codex 的默认模型和所需选项；两个工具的字段并不相同。

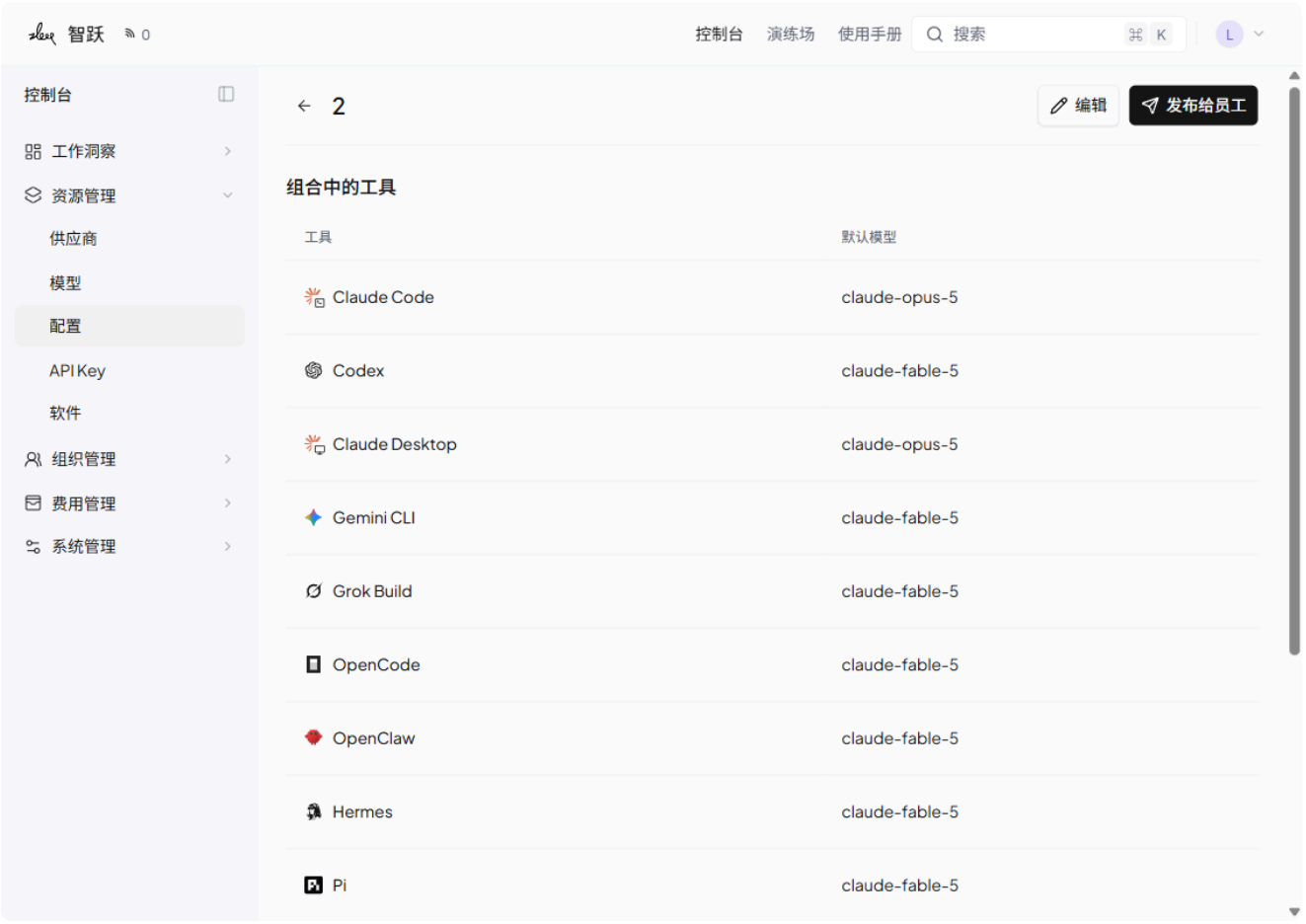


展开 Codex 的【高级】，可按需配置子代理模型、推理强度、推理摘要、响应详细程度、服务等级、上下文窗口及自动压缩阈值。保留【跟随工具默认】或【使用工具默认设置】表示沿用对应默认行为，不要求首次接入时逐项填写。

工具行显示【待完善】时，进入设置修正。对每个启用的工具检查完毕后点击右上角【保存】。没有修改时，【保存】保持灰色不可用。

发布给员工

保存后在组合详情点击【发布给员工】，也可以从列表的【更多】→【发布】进入。检查确认窗口中的版本、工具、默认模型和模型列表，再点击【发布】。返回列表，确认状态为【已发布】；仅保存的组合还不能作为已发布方案供员工获取。

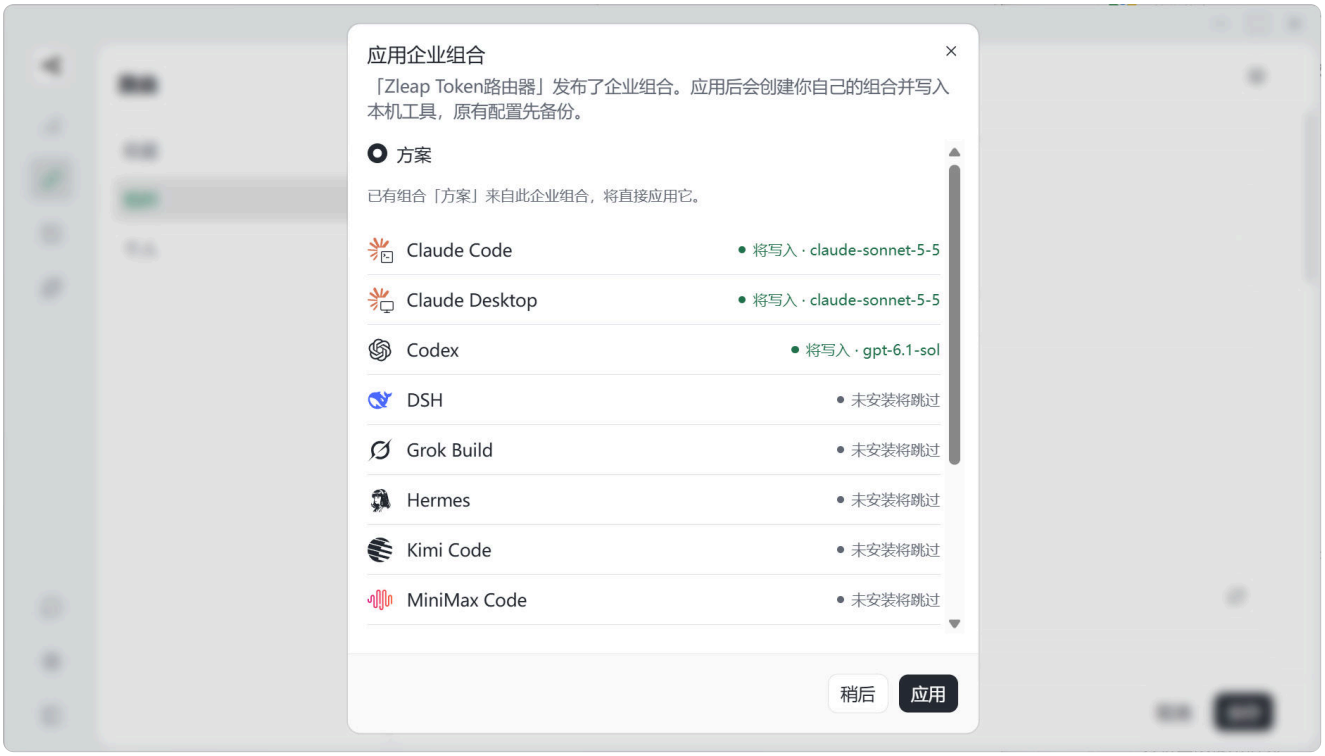


后续修改后，检查当前草稿并点击【更新发布】。要停止继续提供时使用撤回发布。发布、撤回都不应理解为立即覆盖或删除员工已应用到工具的本地文件。

一键应用企业组合

员工先连接正确的企业管理后台并安装目标工具。接通后在【应用企业组合】选择管理员发布的【方案】，查看工具预览并点击【应用】。客户端会自动准备本地组合、备份原有配置并写入工具；未安装的工具会跳过。

若先前选择【稍后】，进入对应组织路由，在编辑页右上角点击【应用企业组合】重新打开。已有来自该企业方案的本地组合时，会直接应用已有组合。



按提示查看逐项结果。应用完成后，务必先关闭并重新打开目标工具，再新建会话验证真实请求。 已打开的工具可能仍在使用旧配置，重启后才能确认它已读取新配置。该入口无需手动创建组合。

按需手动创建本地组合

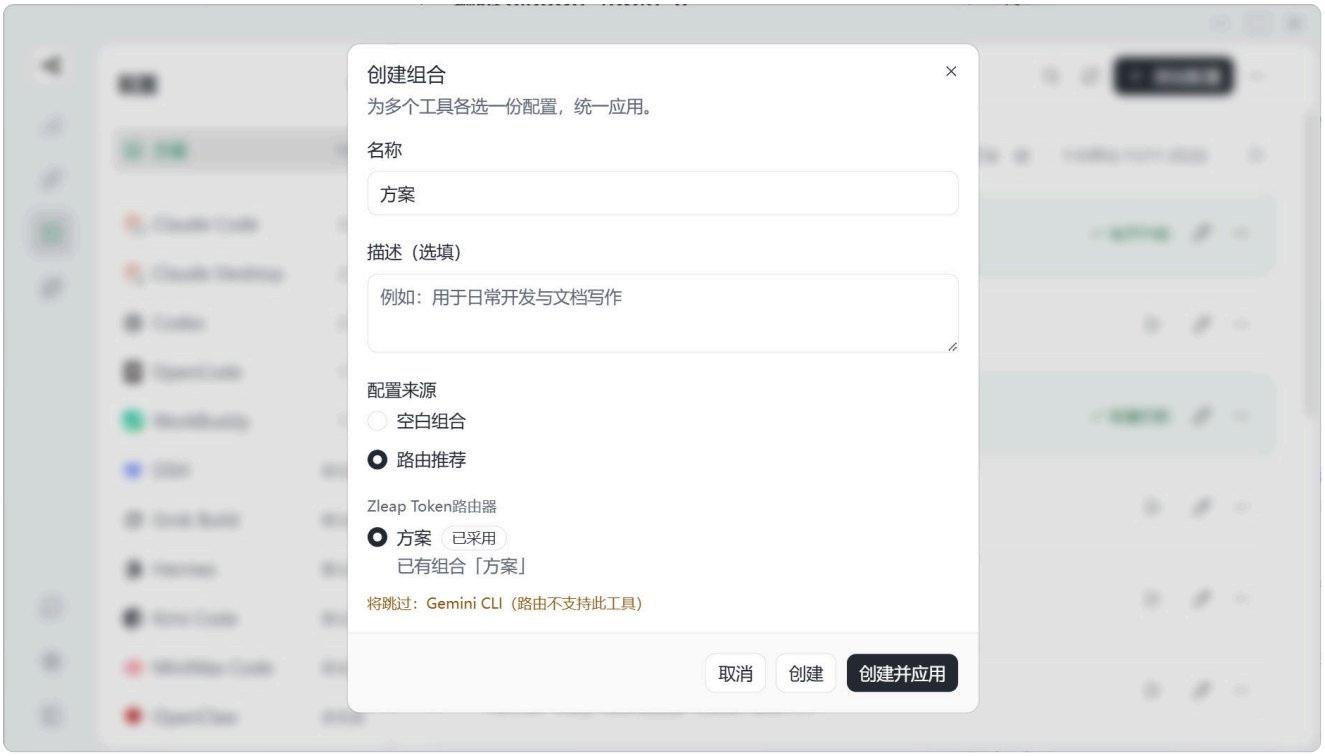
需要自行组织配置时，在【配置】点击标题旁的【＋】→【添加组合】，选择【路由推荐】或【空白组合】。采用路由推荐时选择企业方案并填写本地名称，查看工具预览后【创建】；进入详情再【全部应用】，或使用【创建并应用】。

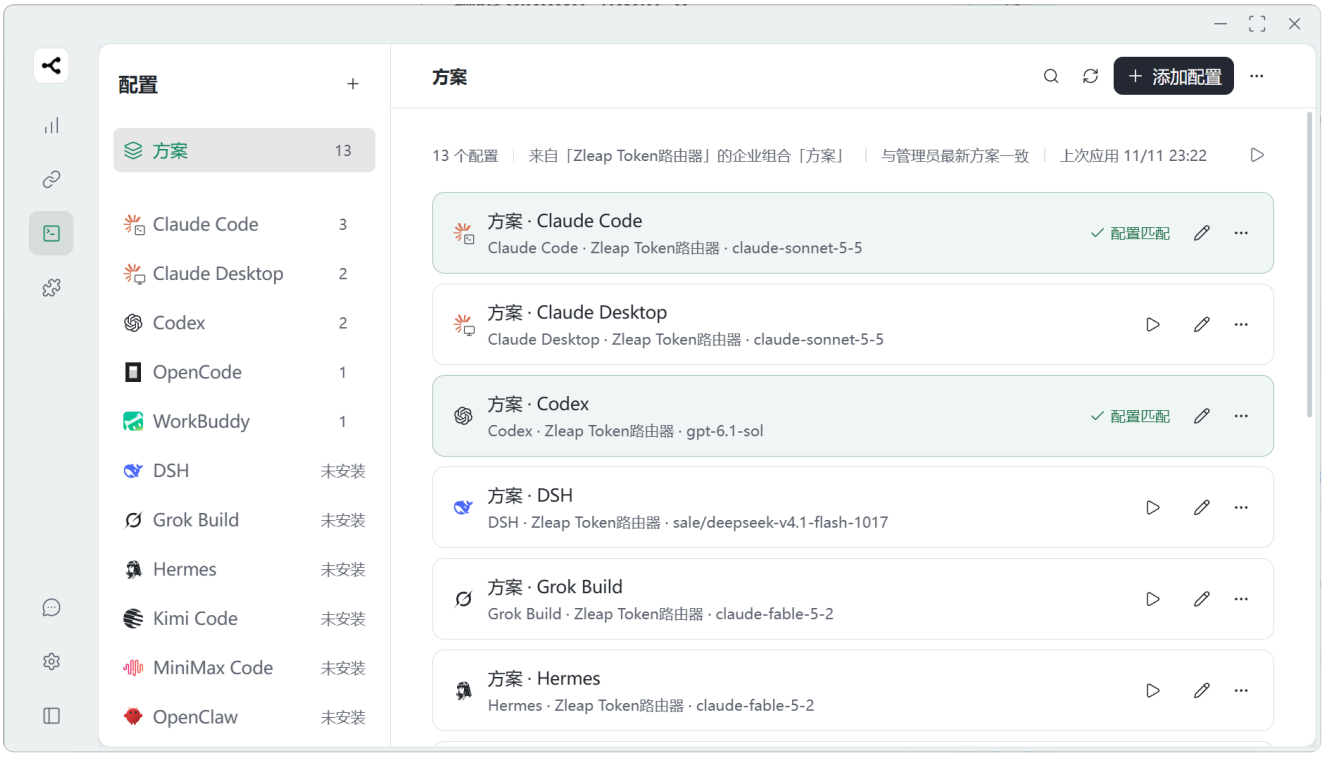
只创建或保存配置尚未写入工具；结果待核对时，先查看【应用记录】。

确认管理员发布的更新

员工在组合详情看到管理员更新提示后，点击【查看差异】，检查本地个人修改及管理员新设置。确认后按提示从企业组合重新填入，再应用到工具。

重新填入会先更新 Token路由客户端中的配置，不会立即修改工具文件。没有内容变化时按界面提示处理，无需反复应用。连接和授权异常时，先恢复管理后台连接或联系管理员。





区分组合配置与软件下载

【配置】组织工具的模型及选项；【软件】提供企业允许的软件下载。下载或安装工具后，员工仍需连接管理后台、取得模型授权并应用配置。

模型接入

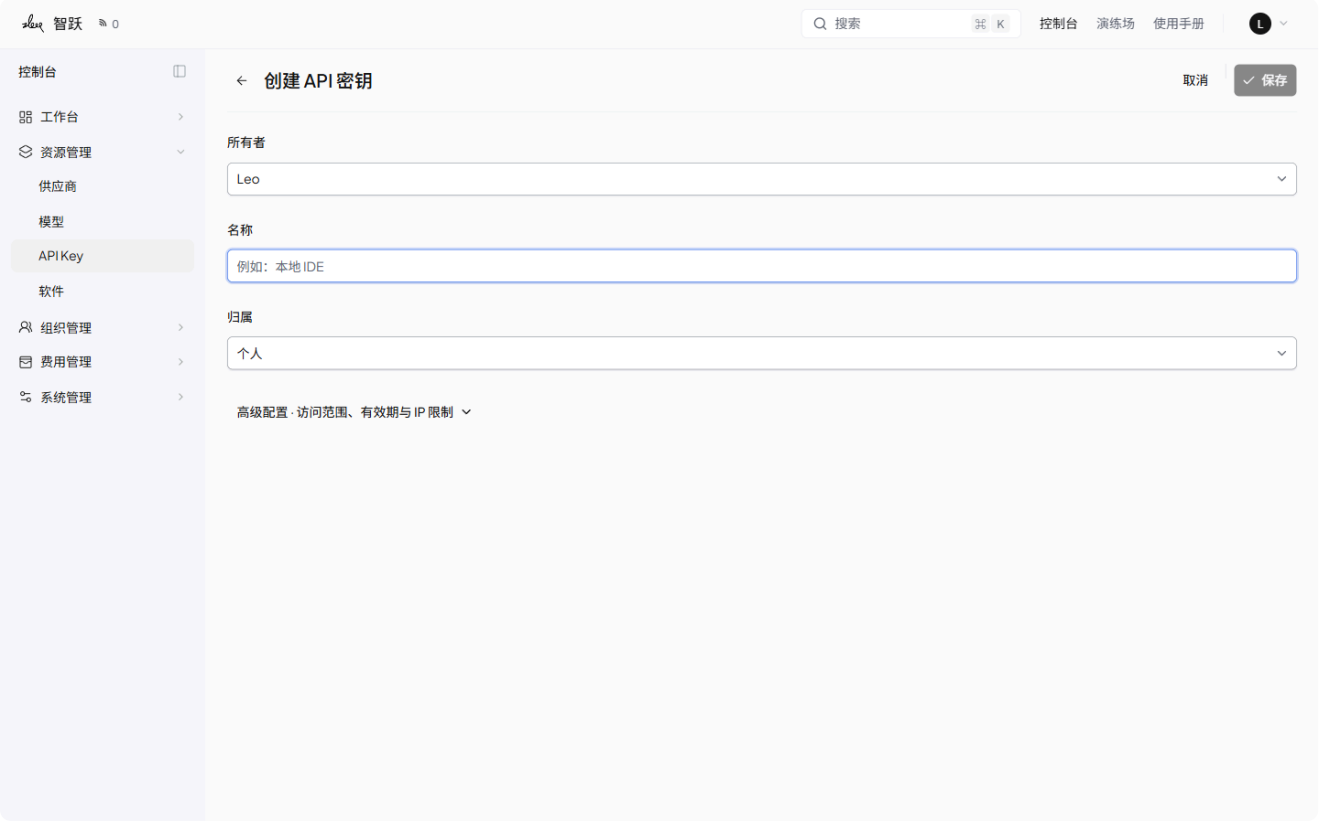
管理 API Key

创建个人或项目密钥并管理访问范围。

路径：【控制台】→【资源管理】→【API Key】

创建密钥

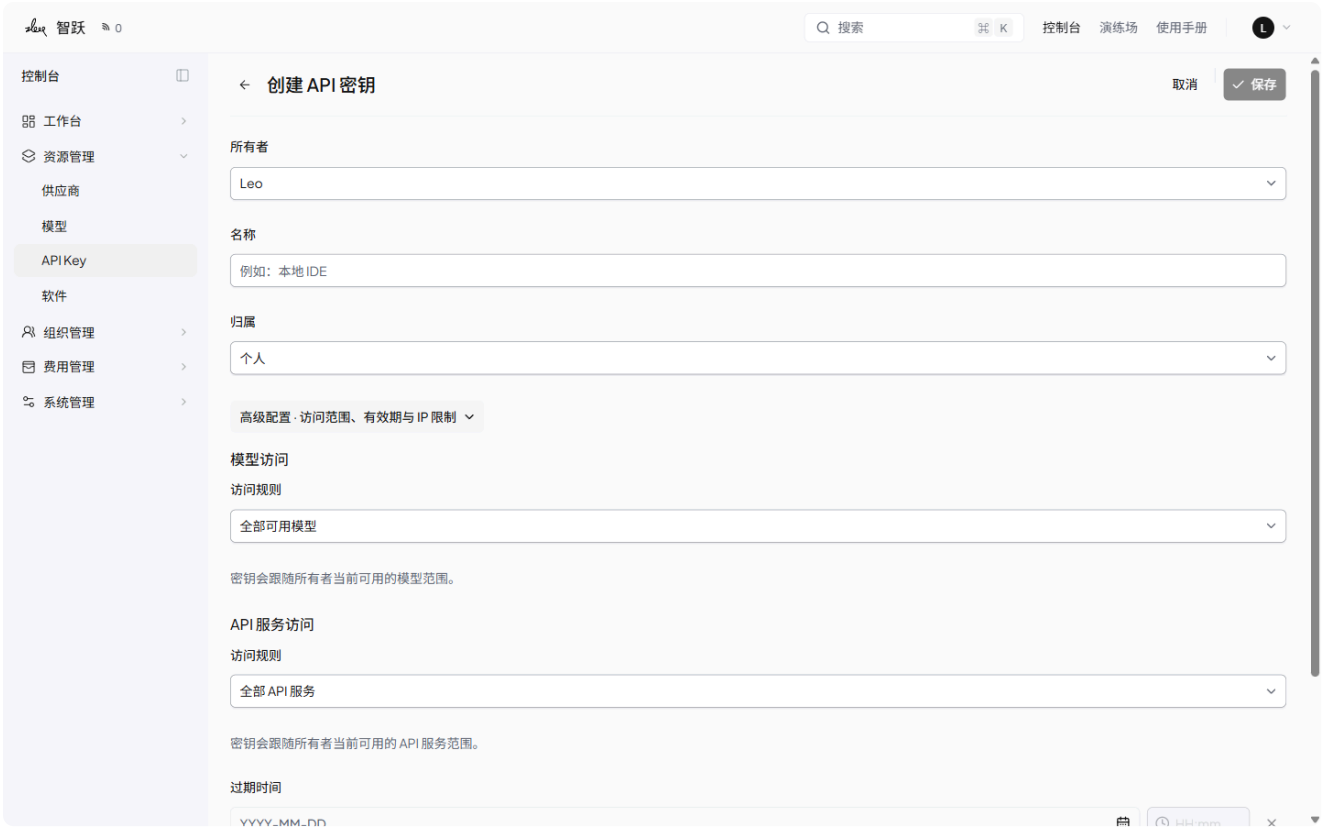
点击【添加凭证】，选择【所有者】，填写名称，再选择【个人】或项目归属。

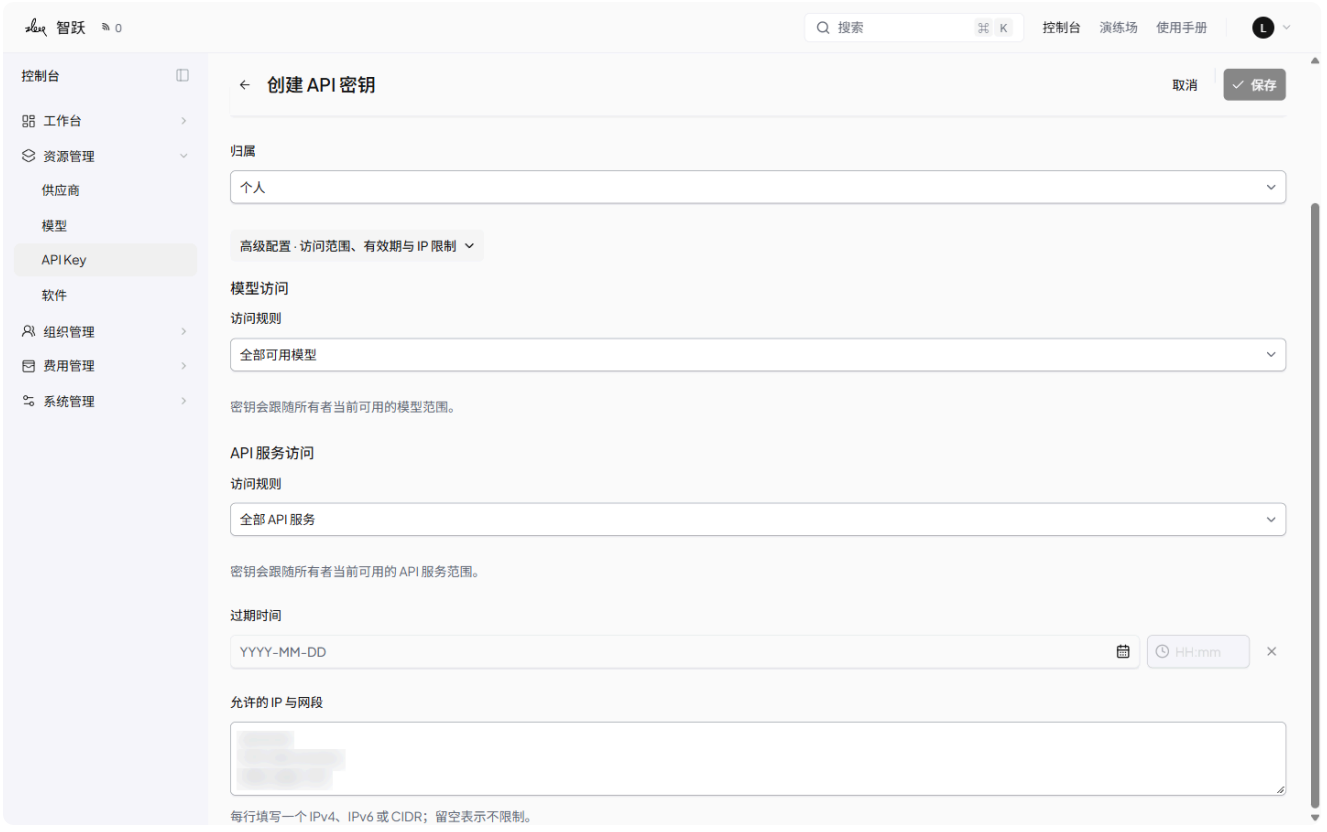


设置访问范围

展开【高级配置】，设置可用模型、API 服务、过期时间和允许的 IP。IP 每行一个，支持 IPv4、IPv6 或 CIDR；留空表示不限制。

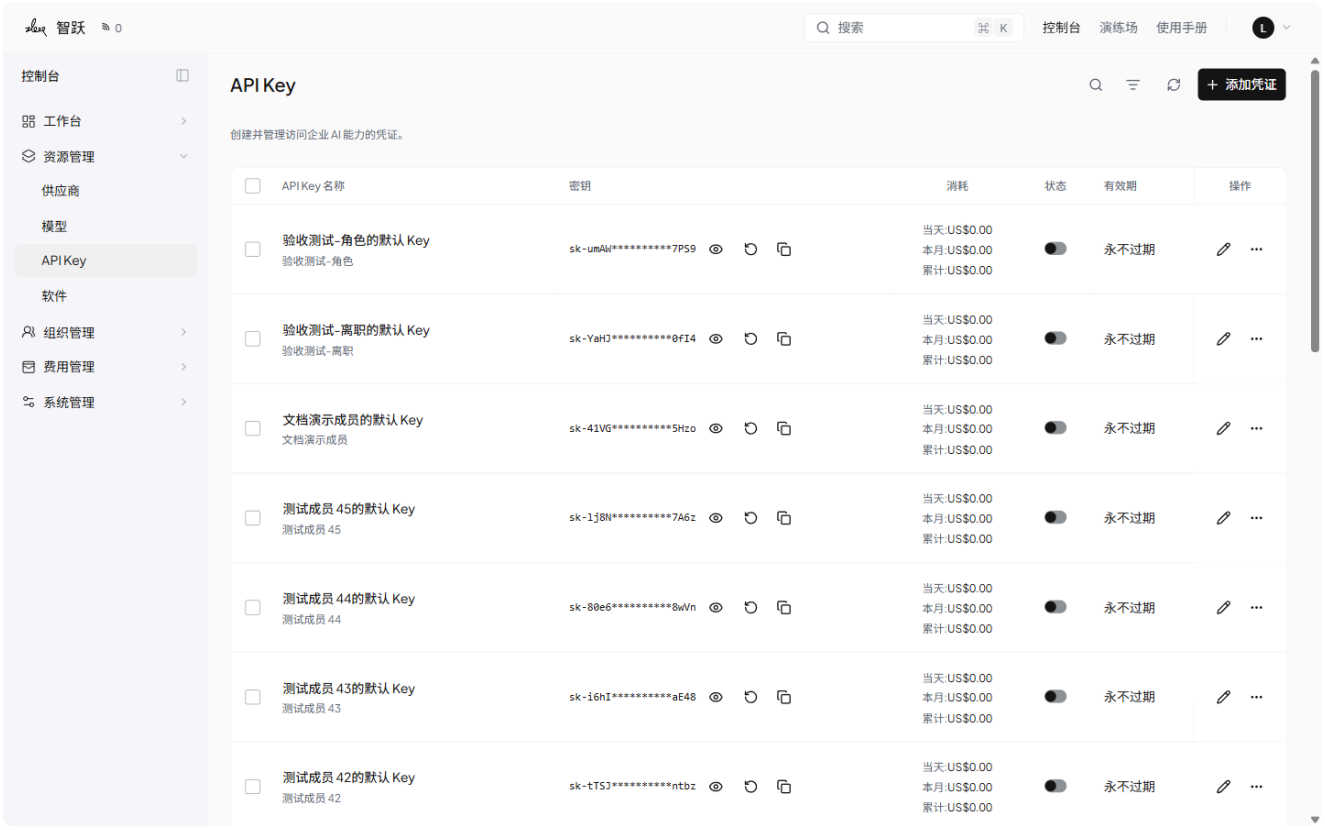
选择【全部可用模型】时，密钥会跟随所有者的模型权限变化。





保存与使用

点【保存】，按创建结果复制密钥。用它调用一次模型，再到请求查看调用归属和结果。



成员通过客户端接入时，使用成员连接授权。密钥泄露后应停用并更换。

模型接入

成员完成方案接入

扫描路由、自主申请或使用连接码接入，并一键应用企业方案。

员工电脑需要能访问企业管理后台；管理员先开放申请或提供连接码，并准备模型授权、预算和已发布方案。

扫描企业管理后台

进入【路由】→【组织】。首次尚无组织路由时会自动发现；已有路由时，点击【组织路由】标题右侧的刷新图标，打开【可连接路由器】。等待扫描结果，点击公司管理后台右侧的连接箭头。

未发现时点击【再次扫描】，或通过【手动添加】填写管理员提供的地址。发现结果不代表网络一定可达。



自主申请并等待批准

企业开放申请后，选择管理后台，等待能力加载，再切换到【申请接入】。填写姓名及邮箱或手机号，按需补充部门和说明，点击【申请接入】。管理员审批时可以建立新成员或关联已有成员，无须逐个提前创建控制台账号。



提交后会显示【等待管理员审批】。保持联网并查看申请状态。需要纠正信息时按界面提示处理，不重复提交来催促审批。若没有申请页签，请管理员检查当前管理后台是否开放申请。



管理员批准后，客户端在后台领取并保存连接。关闭申请窗口不会自动取消申请；网络恢复或重新启动后先查看原申请和连接状态。

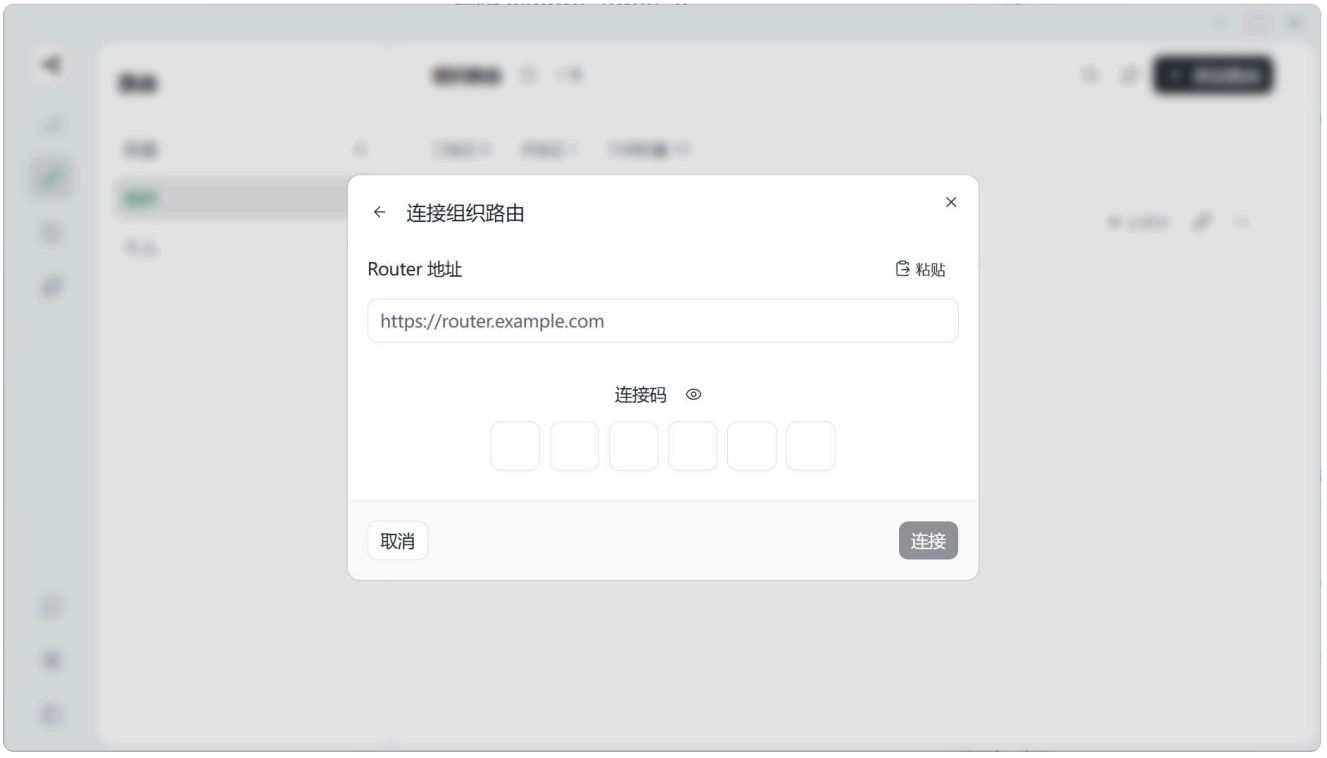
管理员处理步骤见审批接入申请。

管理员从【组织管理】→【成员】→【申请】进入列表，选择【待审批】并打开申请详情核实，再通过或拒绝。



使用连接码接入

管理员已提供连接信息时，填写管理后台地址，在【使用连接码】中输入六位连接码，点击【连接】。手动添加入口也可填写连接信息。

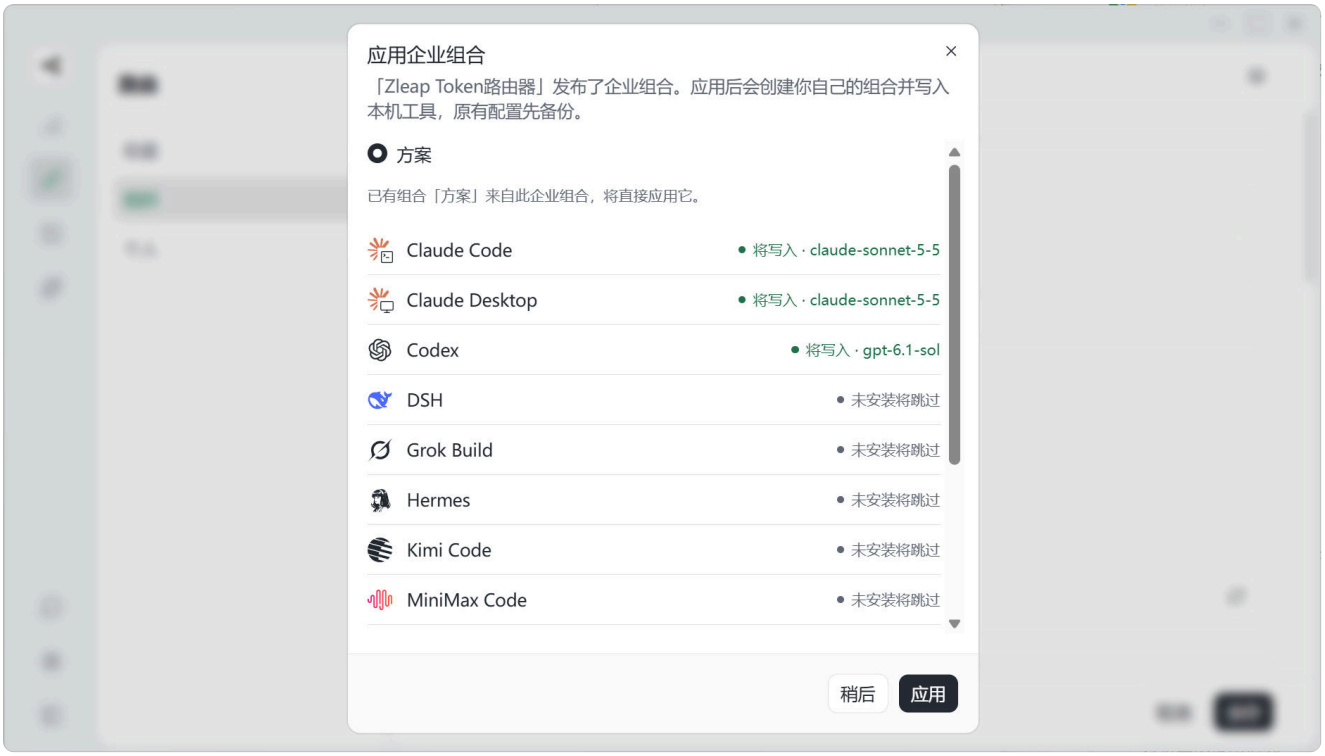


不要将连接码发给无关人员。连接失败时先检查地址可达性、连接码及成员授权。

一键应用企业方案

接通管理后台后，在【应用企业组合】中选择管理员发布的【方案】，查看工具预览，点击【应用】。客户端自动准备本地组合并写入本机工具，原有配置先备份；未安装的工具会跳过。

若之前选择【稍后】，进入对应组织路由，在编辑页右上角点击【应用企业组合】重新打开。无需先手动新建组合。



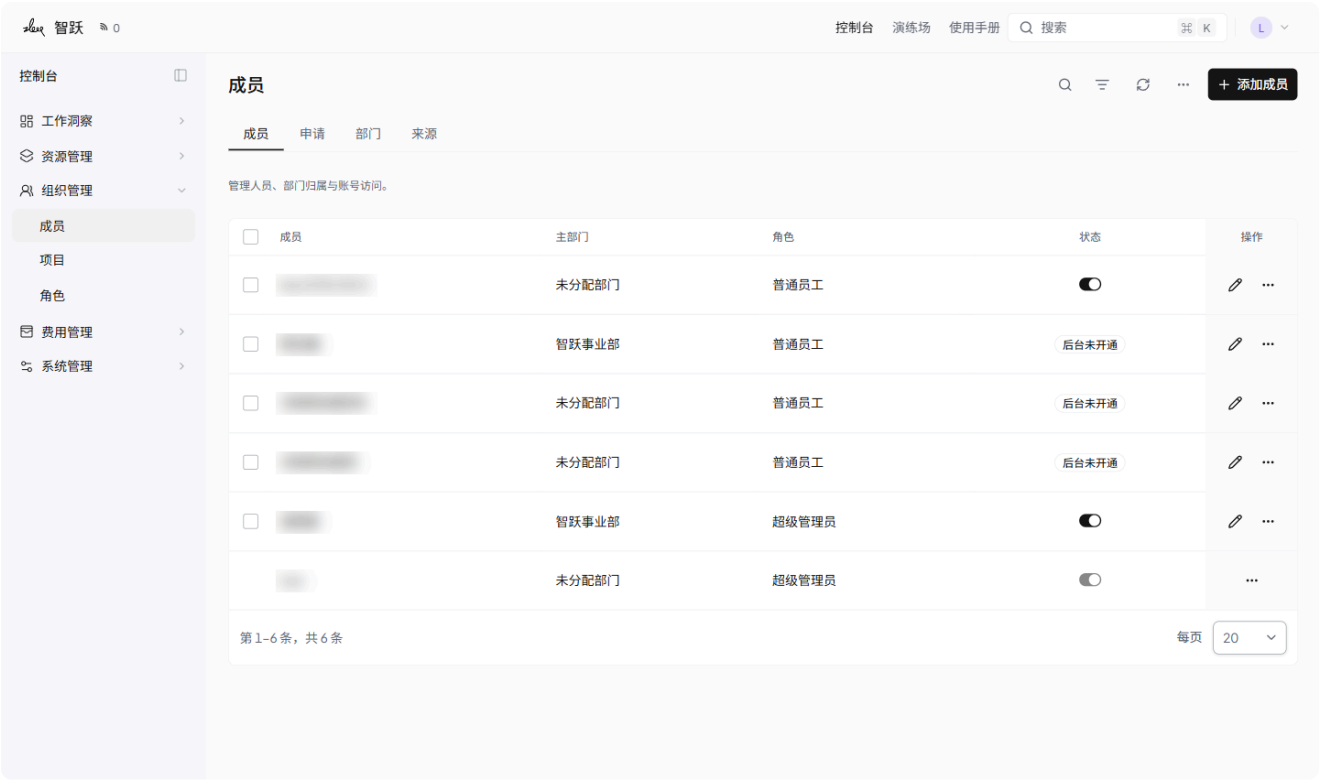
按提示查看逐项应用结果。应用完成后，务必先关闭并重新打开目标工具，再新建会话发起真实请求。 已打开的工具可能仍在使用旧配置。【配置匹配】仅表示文件匹配，不能代替调用验证。管理员更新与个性化设置见发布与采用企业配置。

组织与权限

管理成员与控制台权限

区分成员身份、模型授权、预算和后台账号。

打开【组织管理】→【成员】。工作区包含【成员】【申请】【部门】【来源】。



添加或关联成员

管理员可以【添加成员】，也可以通过申请审批建立成员。审批时发现已有身份，应关联原成员，避免重复身份导致授权和费用归属分散。

按职责设置成员所属部门、可用资源和预算。修改后检查成员详情；成员能否使用模型，同时取决于模型授权、模型状态及适用预算。



分开管理后台与客户端接入

【后台未开通】表示该成员尚未开通控制台账号，不等于不能通过 Token路由客户端申请并使用获准的模型。需要管理后台的成员，再按界面开通账号并分配角色，登录或激活操作仅用于该控制台账号。

Token路由客户端接入使用申请批准后领取的连接，或管理员提供的连接码；不应要求员工先完成控制台登录。详见成员完成方案接入。

调整部门、授权和预算

部门与来源在成员工作区内切换管理。调整模型授权后让成员刷新模型目录，再检查真实调用；预算调整见设置预算。角色用于决定后台管理能力，不能用后台角色代替模型授权。

组织与权限

审批客户端接入申请

开放申请、核实身份，并为新成员或已有成员完成接入。

打开【组织管理】→【成员】→【申请】。

开放申请并设置默认规则

点击【去设置】，开启【允许成员申请接入】，按需选择新成员的默认部门并设置默认额度，然后保存。已有成员保持原部门与额度；模型授权另在成员设置中核对。开放申请不等于自动批准。



员工在 Token路由客户端找到企业管理后台后，填写姓名及邮箱或手机号申请。申请开放、网络可达及版本支持均是必要条件；未开放的企业仍可按其接入要求提供连接码。

核实并处理申请

在【申请】页右上角选择【待审批】，从申请行右侧打开【申请详情】，核对申请人、联系方式、设备和申请编号。

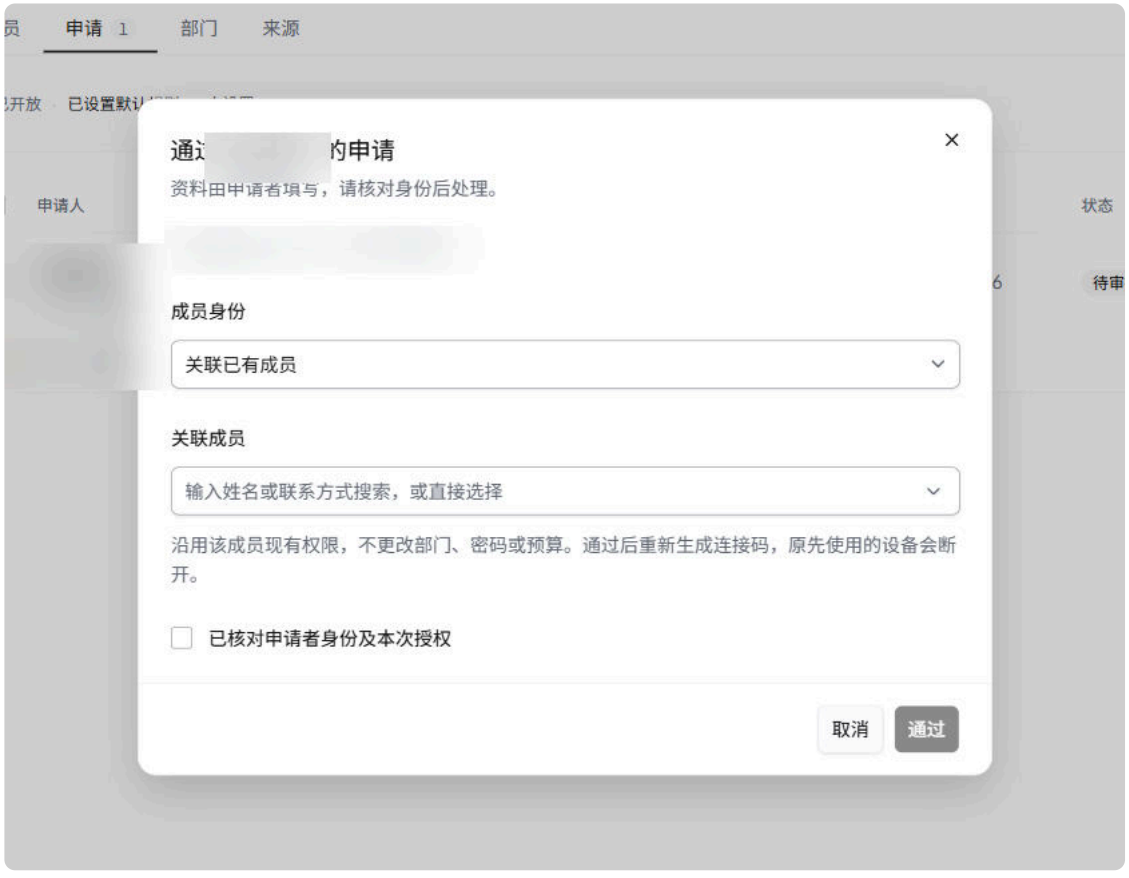




1. 筛选待审批申请，核对身份、联系方式及接入对象。
2. 点击【通过】。新成员保持【创建新成员】，核对本次授权中的部门和额度，需要变更时点击【调整】；勾选身份与授权确认后再点击弹窗中的【通过】。模型及其他资源沿用成员权限规则，创建后到成员设置核对模型授权。



3. 若已经有成员记录，在【成员身份】选择【关联已有成员】，搜索并选择正确的成员，再核对确认。此路径沿用原成员权限，不更改其部门、密码或预算；通过后重新生成连接码，原先使用的设备会断开。



4. 无法确认或不应接入的申请予以拒绝；批量处理前逐项核对所选申请及统一授权范围。

等待员工领取连接

批准后由 Token路由客户端领取并保存连接；员工关闭申请窗口不会自动取消已提交的申请。暂时断网时恢复连接，查看原申请进度，避免重复申请。

审批完成仍需在客户端检查模型目录，并通过真实请求确认模型可调用。申请状态、连接成功、目录可读与模型调用成功是不同结果。

组织与权限

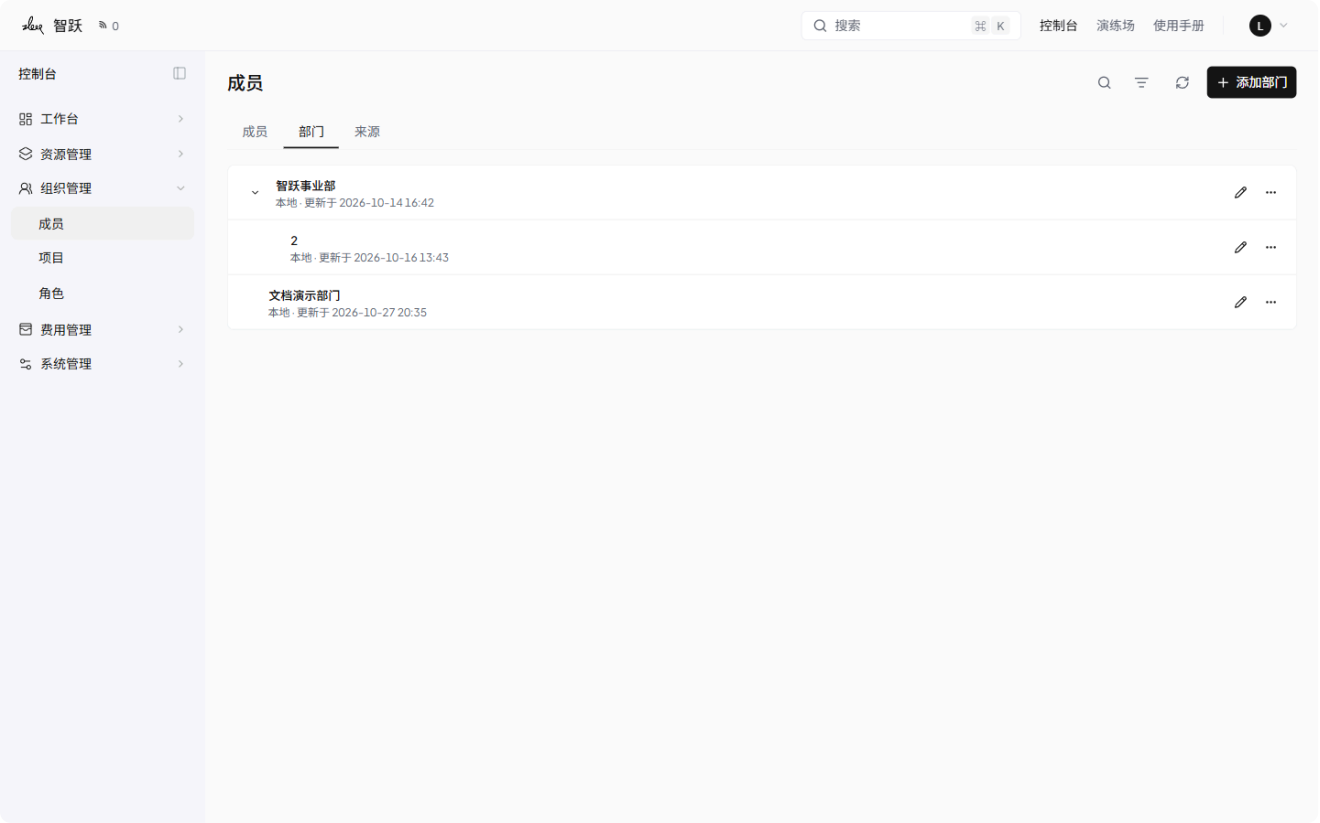
维护部门

维护部门层级与成员归属。

路径：【控制台】→【组织管理】→【成员】→【部门】

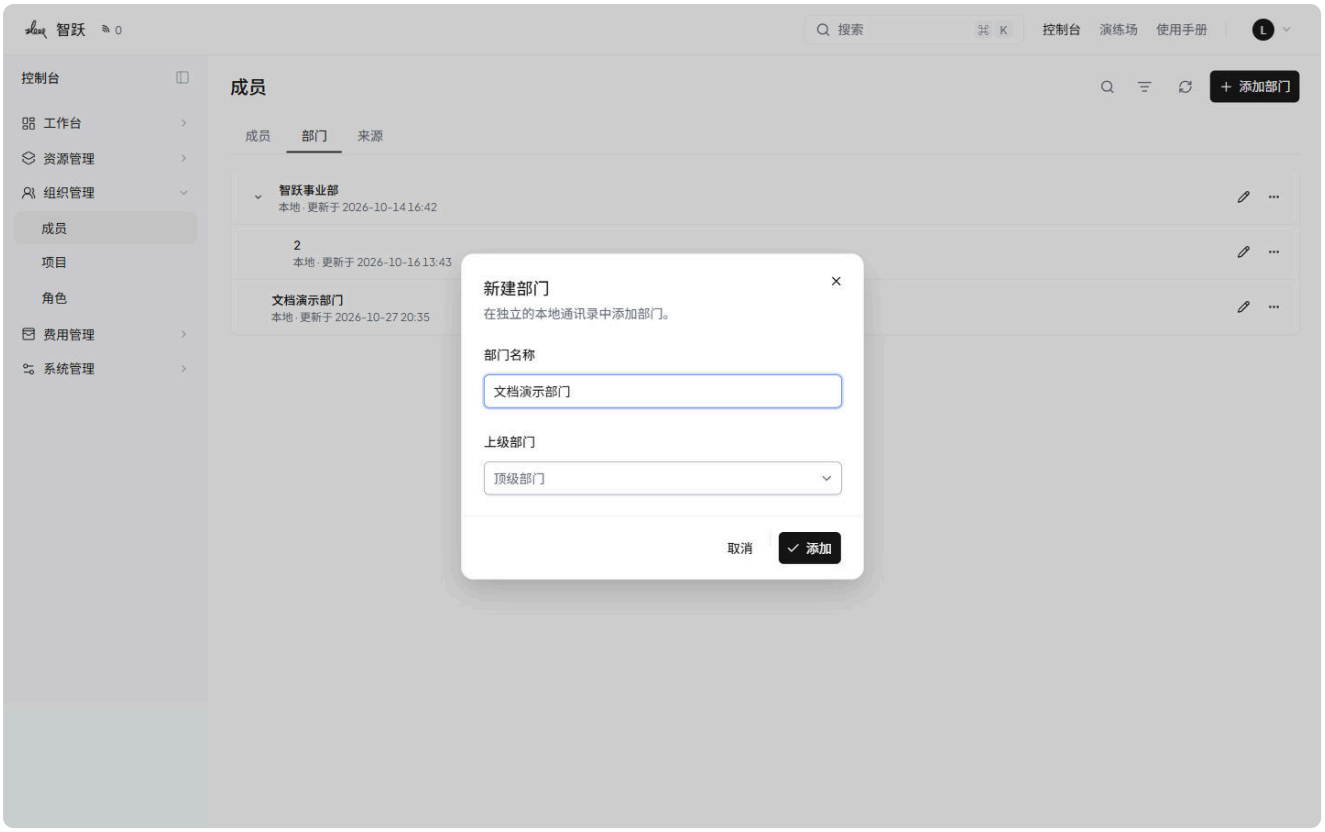
查看部门结构

在【成员】页切换到【部门】，查看现有部门及层级。



添加部门

点击【添加部门】，填写名称，选择上级部门，再点【添加】。独立部门选择【顶级部门】。



将成员归入部门

切回【成员】，添加或编辑成员，在【部门】中选择目标部门并保存。新建成员未选部门时，会归入【未分配部门】。



组织与权限

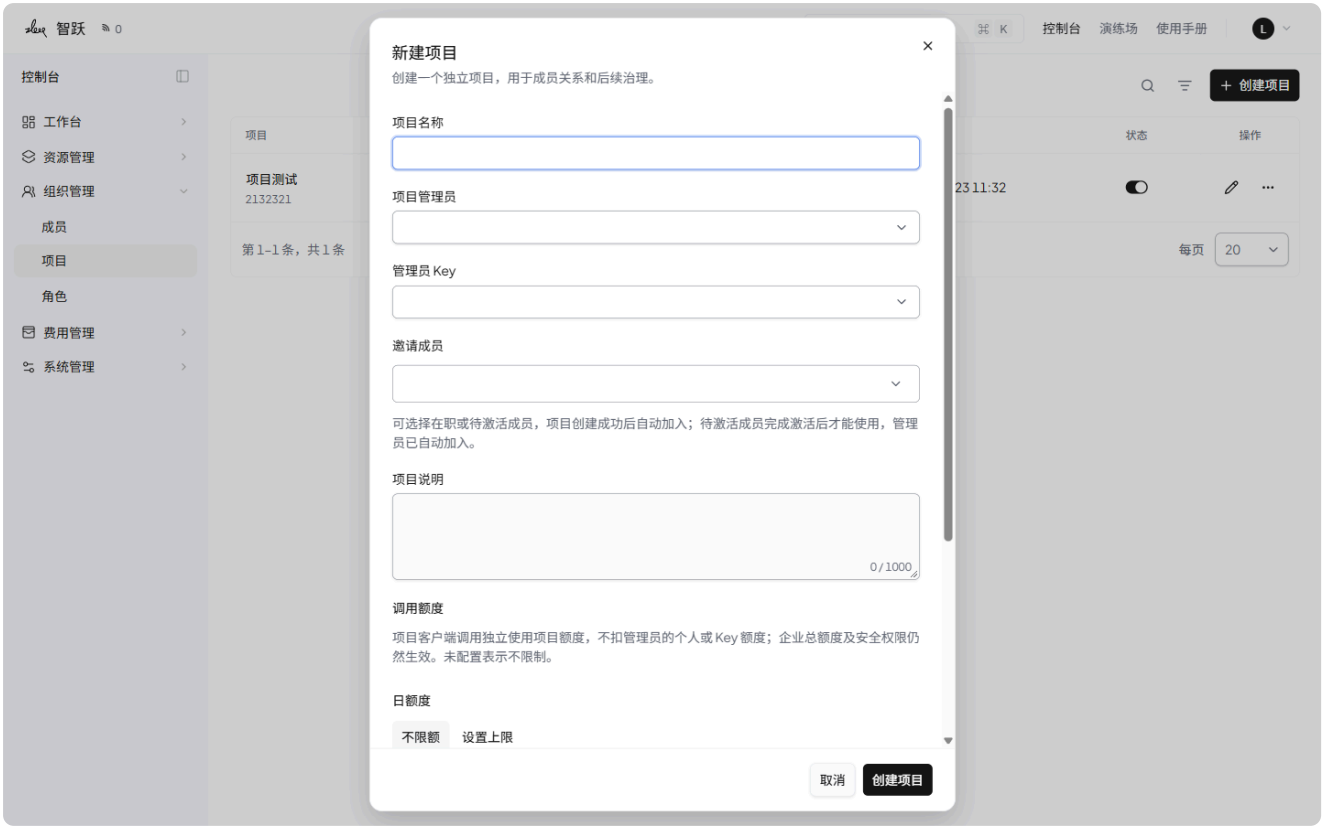
管理项目

组织项目成员、管理员、Key 与预算。

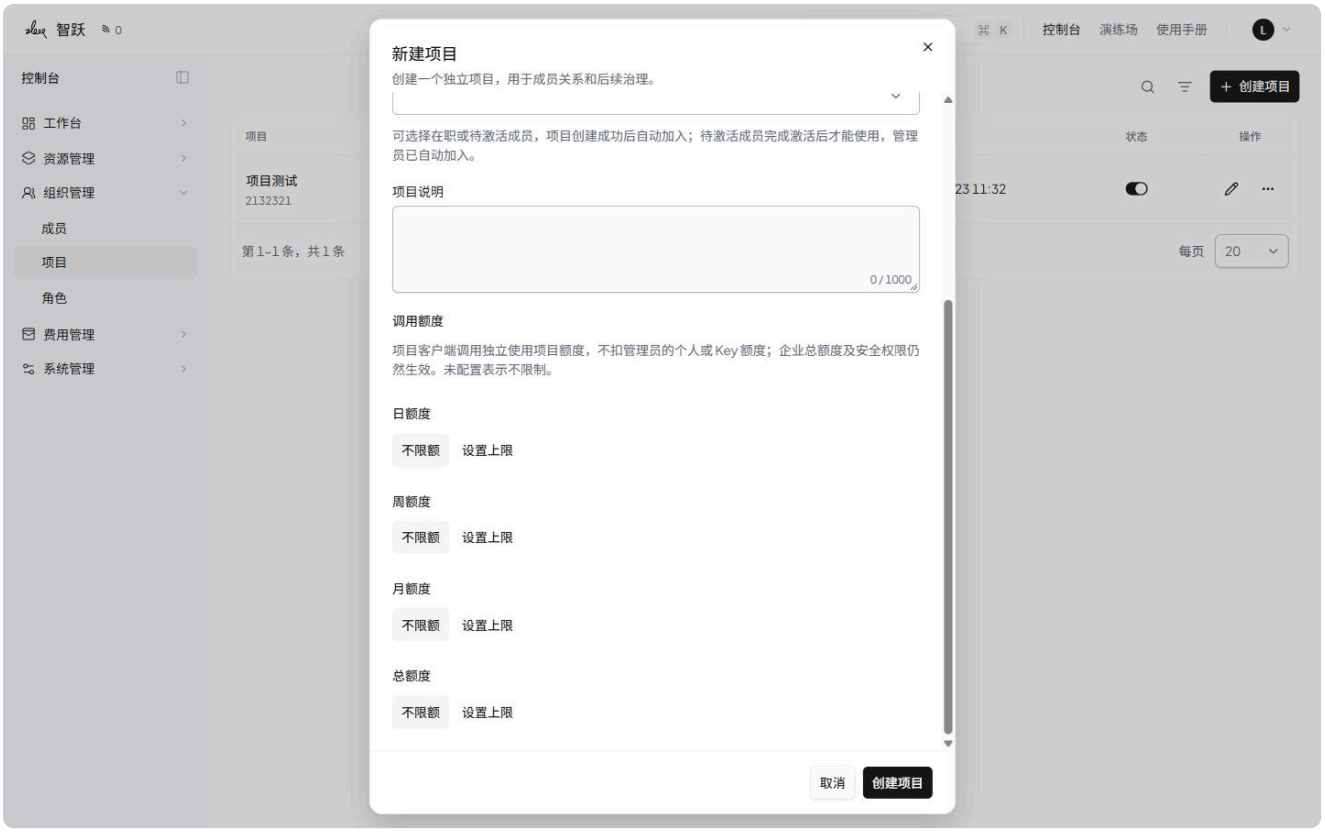
路径：【控制台】→【组织管理】→【项目】

创建项目

点击【创建项目】，填写名称，选择项目管理员。可同时邀请成员、填写说明；项目管理员会自动加入项目。



向下设置日、周、月和总额度，不限额时保留默认选项，再点【创建项目】。



管理项目成员

在列表点击项目名称，查看当前成员。点击【添加成员】选择参与者，修改后点【保存】。



连接客户端与设置额度

项目详情提供【读取客户端凭据】入口，读取后按项目使用范围交付。项目客户端调用使用项目额度，不扣管理员的个人或 Key 额度；企业总额度和安全权限仍然生效。

调整额度后点【保存】，可到预算查看已用金额。待激活成员需要先激活账号。

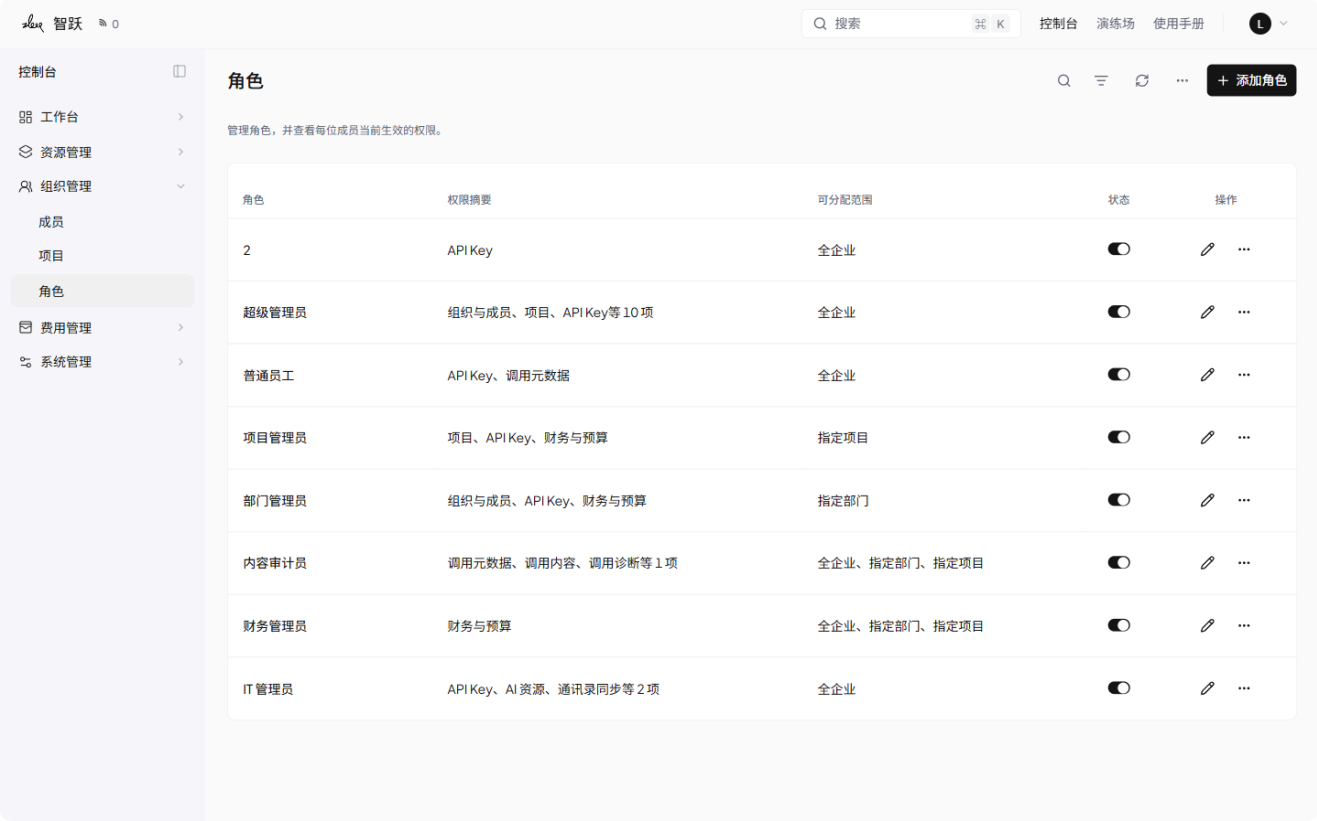
分配角色与权限

配置角色能力并向成员分配。

路径：【控制台】→【组织管理】→【角色】

查看现有角色

列表显示各角色的权限摘要和可分配范围。先看现有角色是否够用；需要新的权限组合时，点击【添加角色】。



填写名称与权限

在【新建角色】中填写名称，在【权限矩阵】为每项能力选择权限级别，例如【无权限】【查看】或【管理】。



选择可分配范围

向下选择【全企业】【指定部门】或【指定项目】，然后点【新建角色】。权限级别决定能做什么，可分配范围决定角色能分配到哪些范围。



给成员分配角色

打开成员，编辑目标成员的【角色】并保存。再用该成员的账号查看对应页面。



管理登录连接与通讯录来源

区分员工登录入口和唯一通讯录权威源。

打开【组织管理】→【成员】→【来源】。



添加登录连接

点击【新增登录连接】，选择飞书、钉钉、OAuth2 或 OIDC 等模板，再点击【下一步】。



在身份平台创建应用，使用页面显示的回调地址。填入应用 ID 和密钥，继续设置登录入口及高级协议参数，检查后保存。



登录连接供现有成员登录，不修改部门和在职状态；本地所有者的应急登录始终保留。

配置通讯录来源

在【通讯录权威源】中找到飞书或钉钉，点击【配置】，填写企业应用凭证。需要让该平台同步组织数据时，勾选【设为当前通讯录权威源】。



同一时间只能启用一个通讯录权威源。 已保存的密钥不会再次显示，每次保存配置都需重新填写。同步后查看页面下方的【同步记录】，再到成员和部门列表检查数据。

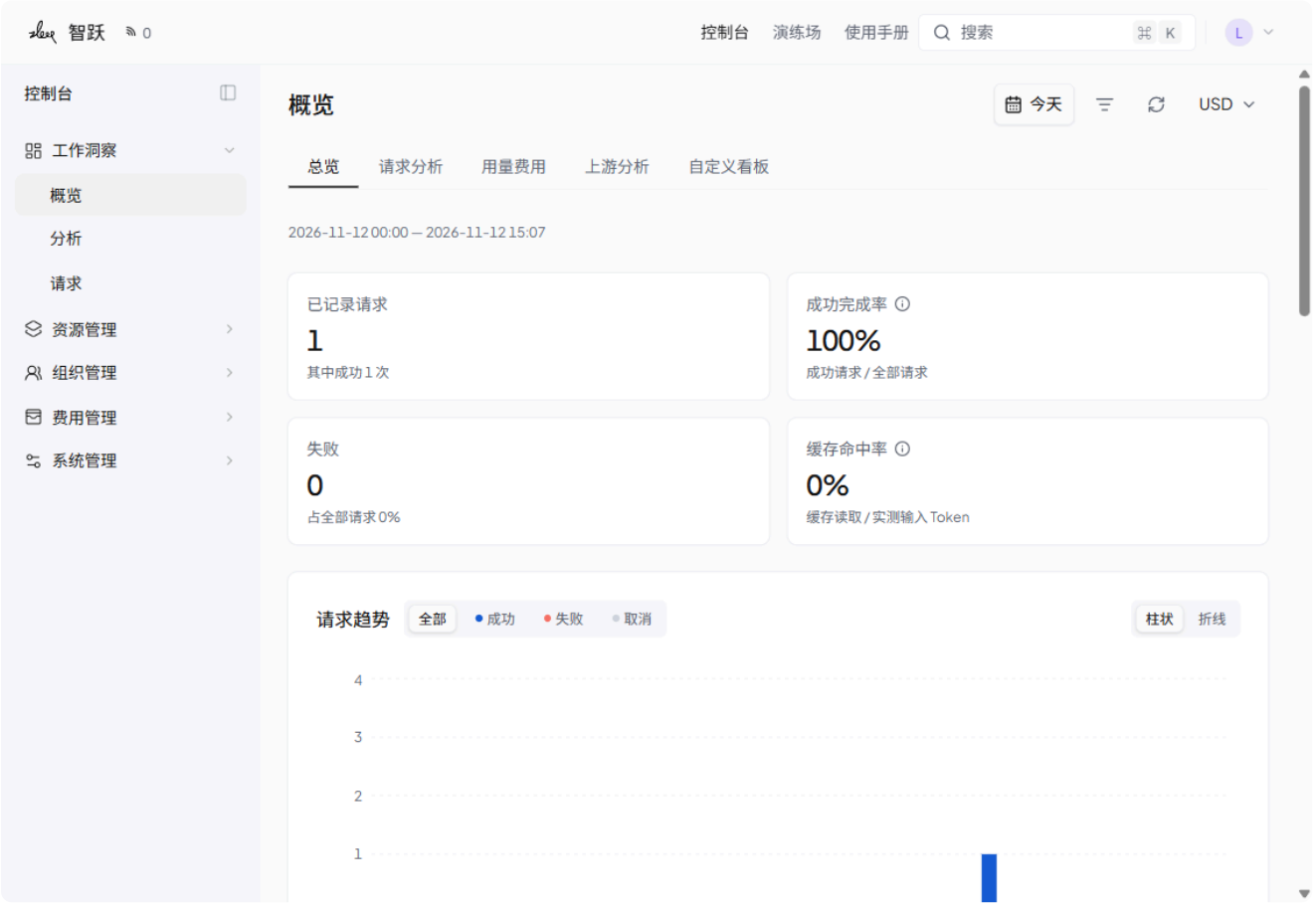
工作洞察与审计

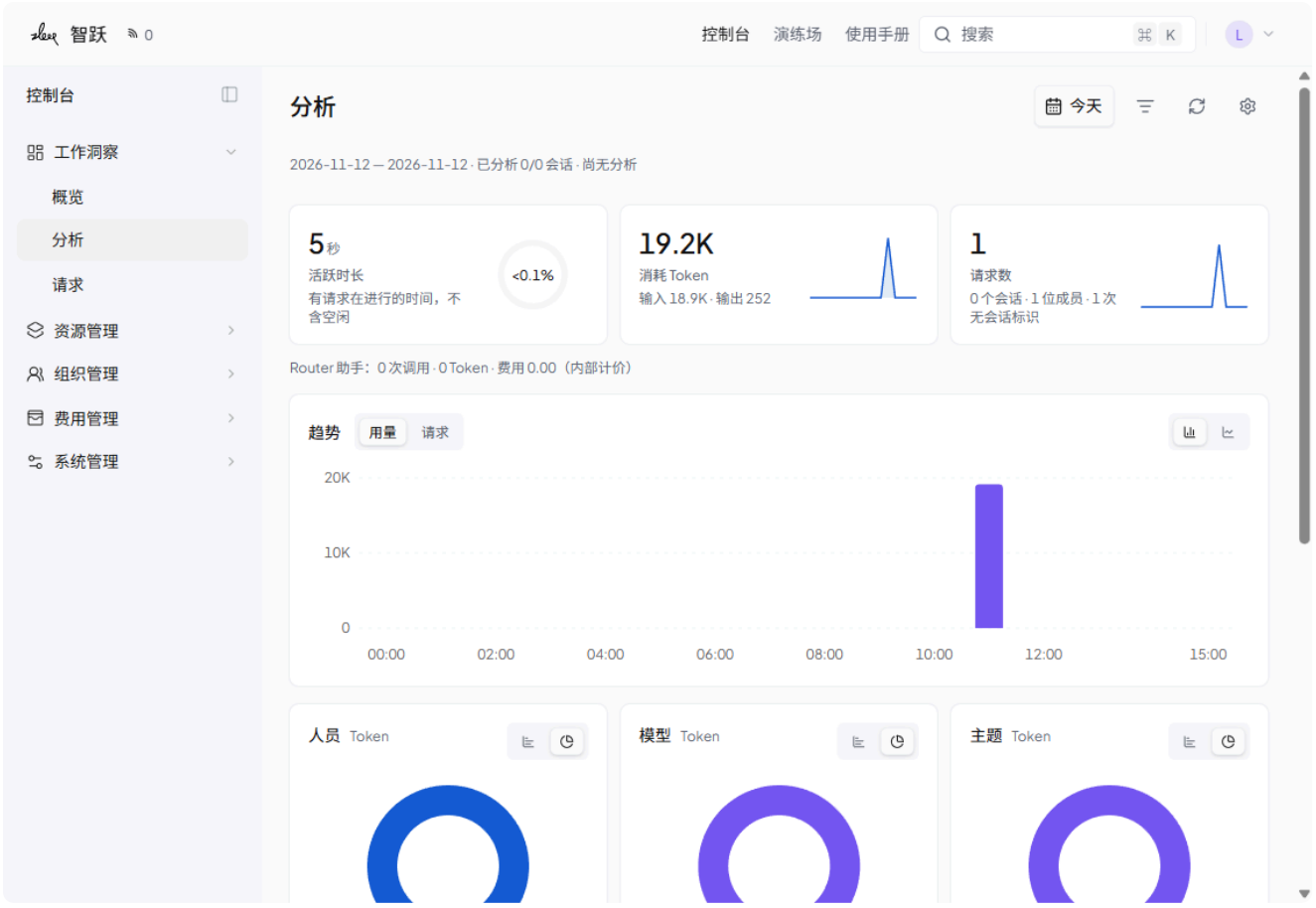
选择用量查看入口

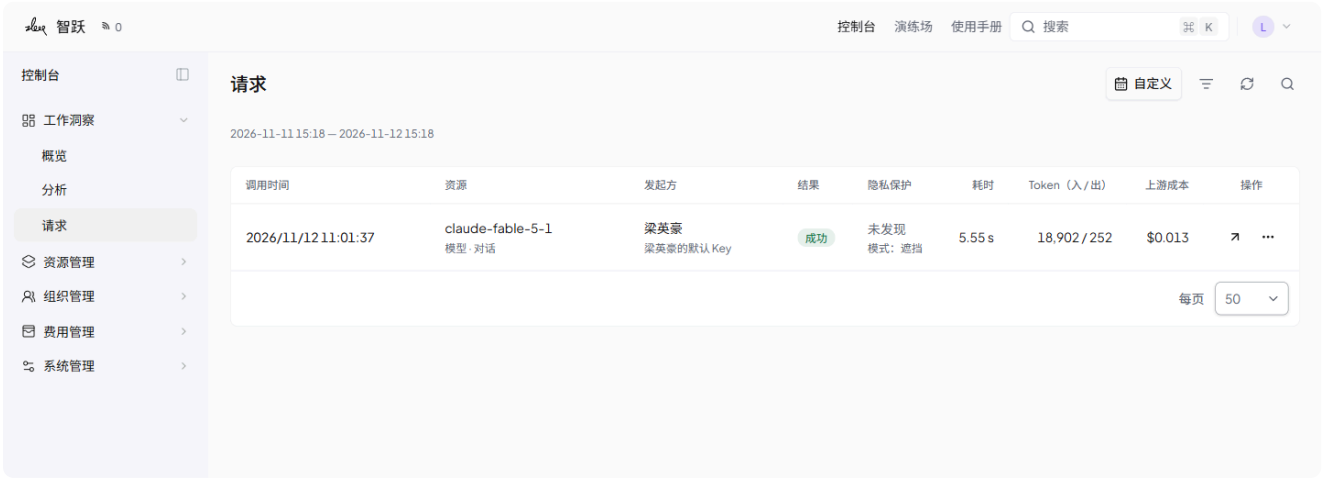
区分概览汇总、团队分析和逐条请求。

打开【工作洞察】，按目的选择入口。

想做什么	打开哪里
看请求量、费用趋势、模型和上游分布	【概览】，见概览与报表
看团队和成员的使用趋势、会话与 AI 洞察	所有者打开【分析】，见分析教程
排查一次失败、重试或费用异常	【请求】，见请求教程
核对管理员做过的配置操作	审计日志







对齐统计口径

比较数字前对齐时间、成员、项目、模型和币种。一次请求可能包含多次上游尝试；重试后成功仍可能产生多笔上游成本。内部费用按企业设置计量，不能直接当作供应商账单。

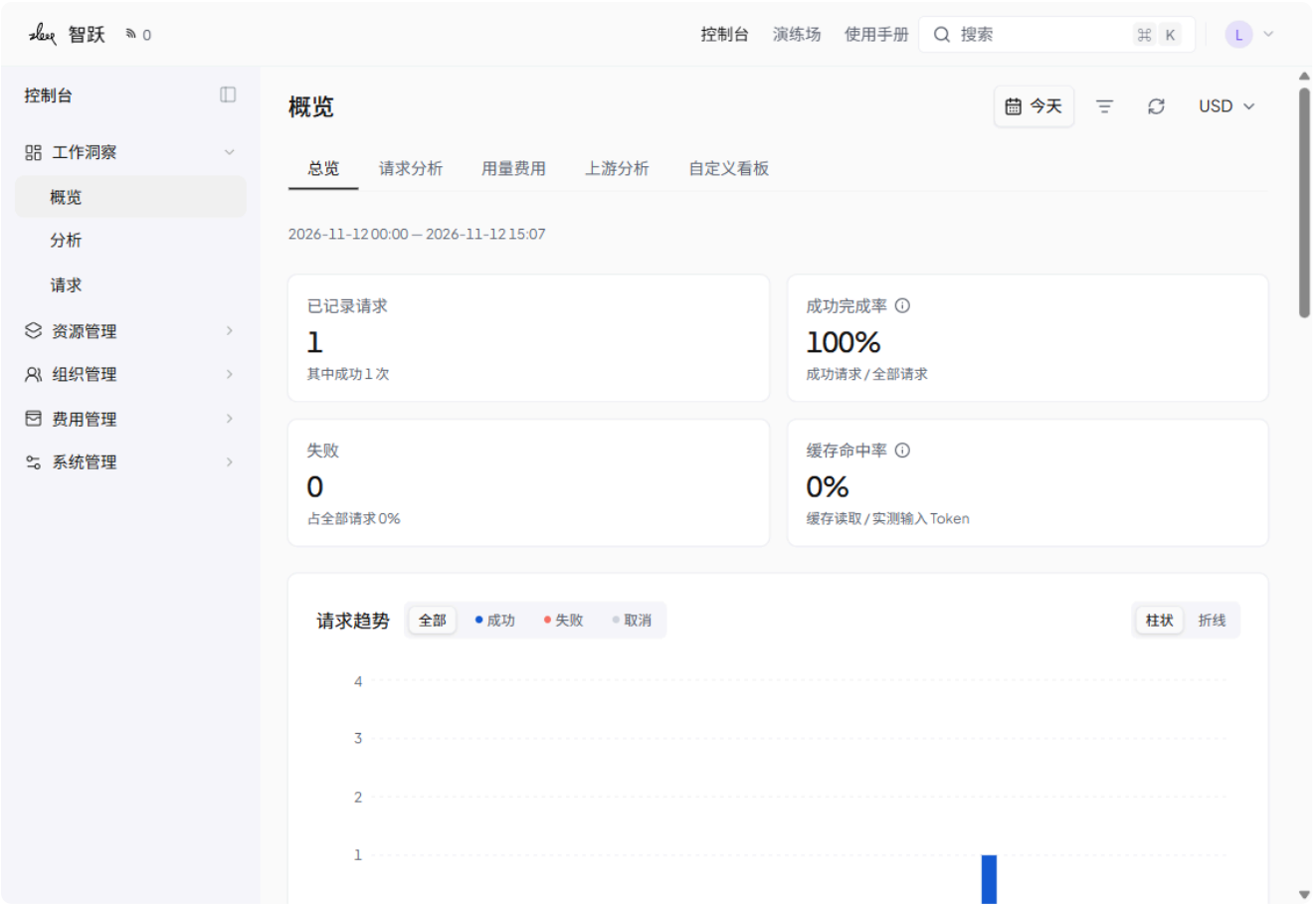
没有会话标识的请求仍可进入请求和 Token 统计，不应据此推断已形成完整会话分析。自动分析关闭时，基础统计仍可查看。

工作洞察与审计

查看概览与报表

在概览查看请求、用量费用、上游分析和自定义看板。

打开【工作洞察】→【概览】。页面按账号权限显示可查看的统计与数据范围；原报表地址会转到概览。



选择统计页签

页签	主要用途
【总览】	快速查看当前区间的汇总与趋势
【请求分析】	按请求结果及相关维度分析调用
【用量费用】	查看 Token 与费用分布
【上游分析】	核对上游尝试、表现和成本
【自定义看板】	按工作需要组织统计视图

先选择时间范围，再检查当前页支持的筛选与分组。记录异常时保留同一时间范围和筛选条件，便于在其他视图追溯。

追溯统计差异

请求量与上游尝试量不是同一口径。缓存读取与缓存写入 Token 分别统计，不应合并后套用同一种价格；支持搜索计费的请求还需核对相应费用。

需要逐条定位时进入请求；需要团队和成员洞察时进入独立的分析，不再按旧活动页的四个页签操作。

工作洞察与审计

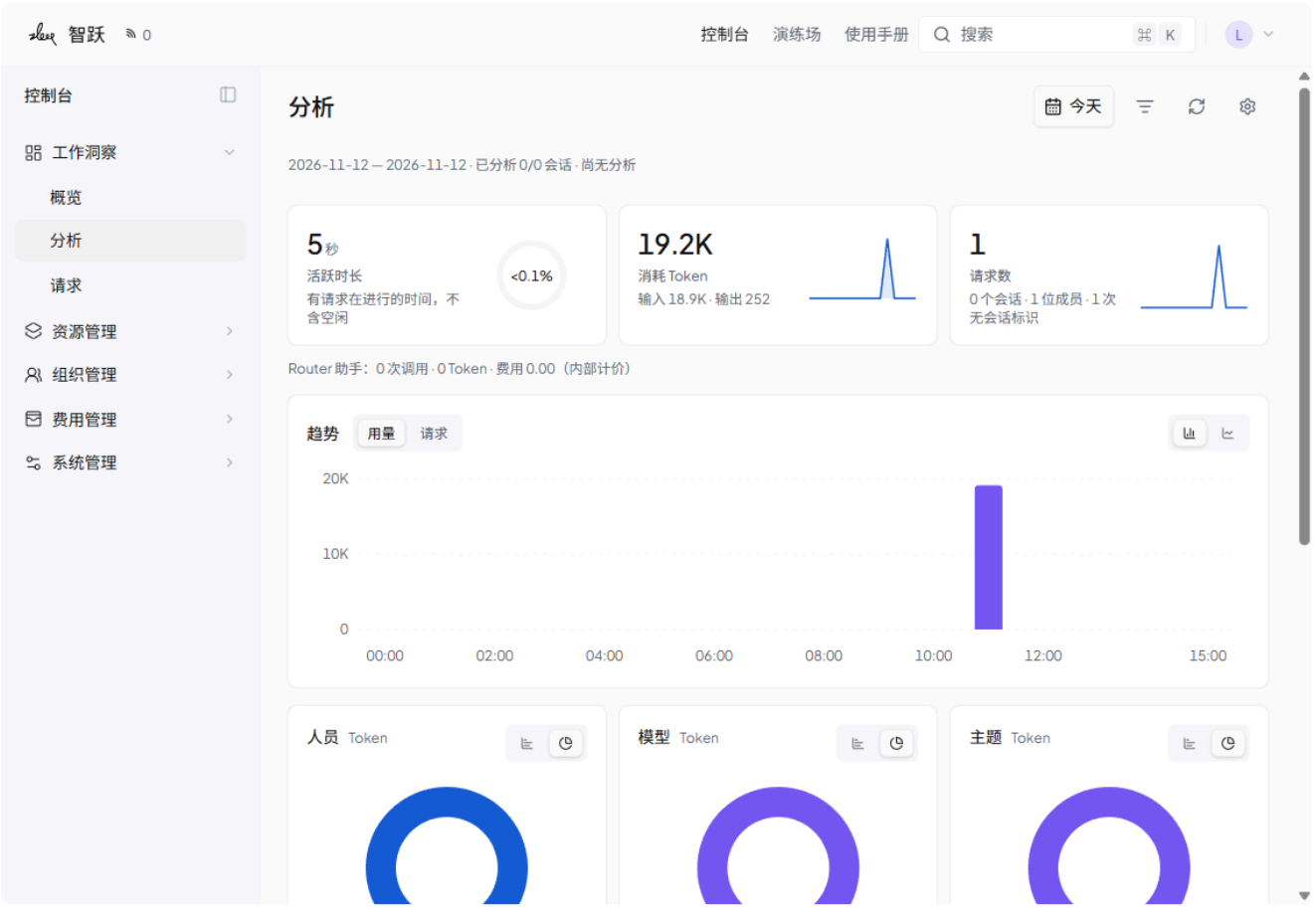
分析团队使用情况

所有者查看使用趋势，并按需启用模型分析。

打开【工作洞察】→【分析】。

查看基础统计

选择时间范围，查看活跃时间、Token、请求和会话，以及成员、模型和主题等维度。没有会话标识的请求会单独提示，不能直接当成完整会话数。



自动分析未启用时，Token 和会话统计仍可查看；进度及会话总结等 AI 洞察需要配置分析能力。

设置自动分析

1. 在分析页点击【前往设置】进入【分析设置】。
2. 在【配置】中开启【启用自动分析】，选择自动使用可用聊天模型或指定模型。
3. 设置更新节奏。安静时长决定内容停止新增后何时处理，最长等待时间限制持续有新增内容时的等待。
4. 按需填写【每日费用上限】，并检查运行时与并发设置，再保存。



开启后仅处理新增内容，不自动补算全部历史。 分析会产生模型用量，每日费用上限按内部价格计算；留空不设该上限。



检查提示词与用量

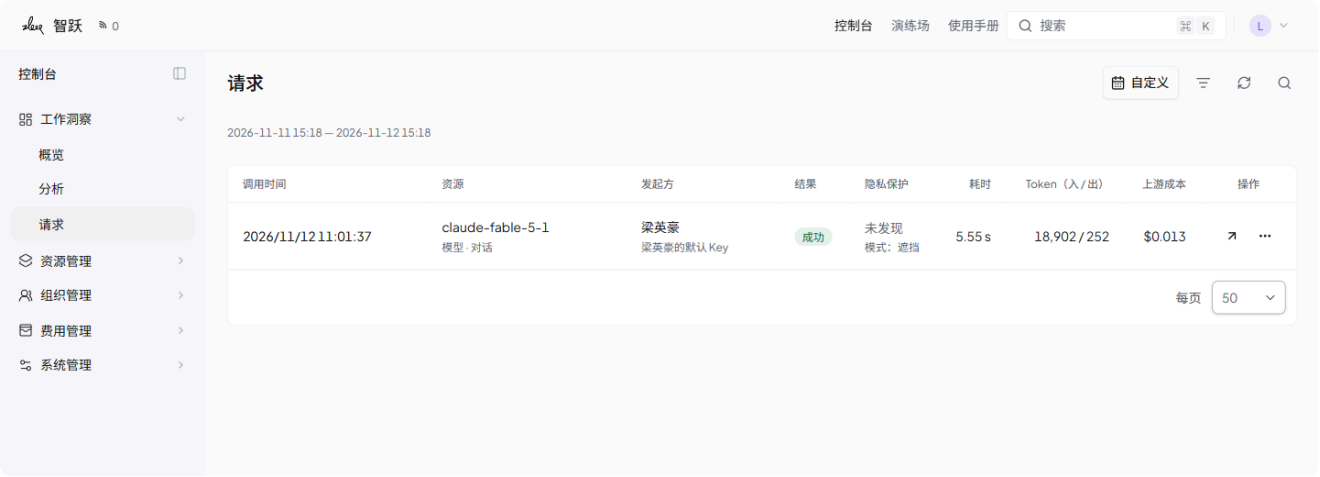
【提示词】用于检查和调整分析任务要求；使用试运行时先确认所选模型和输入。【用量】用于追踪分析处理的消耗。结果不足时先查是否启用、模型是否可用及是否满足处理节奏，不要把“尚未启用分析”当成统计丢失。

工作洞察与审计

查询请求与上游尝试

定位单条请求的结果、授权、费用和隐私处理。

打开【工作洞察】→【请求】，按页面提供的时间、成员、资源和结果条件缩小范围。



打开请求详情

选择记录，核对请求编号、时间、发起方、模型、归属和最终状态，再查看上游尝试及费用。请求正文和诊断内容取决于查看权限、内容留存及请求本身是否有对应数据。

一次请求可能重试多个上游。最终成功不代表只有一次尝试；排查超时或费用时，应检查各次尝试的结果与成本。

核对用量与安全结果

输入、输出、缓存读取和缓存写入是不同计量项；支持搜索计费的协议还需核对搜索费用。成本缺失或不可用不能一律当成零成本。

涉及隐私保护时，按详情中实际提供的处理结果核对规则和模式。原文访问仍受权限约束，不要为了排查问题向无权限人员转发敏感正文。覆盖边界见隐私保护。

排查调用失败

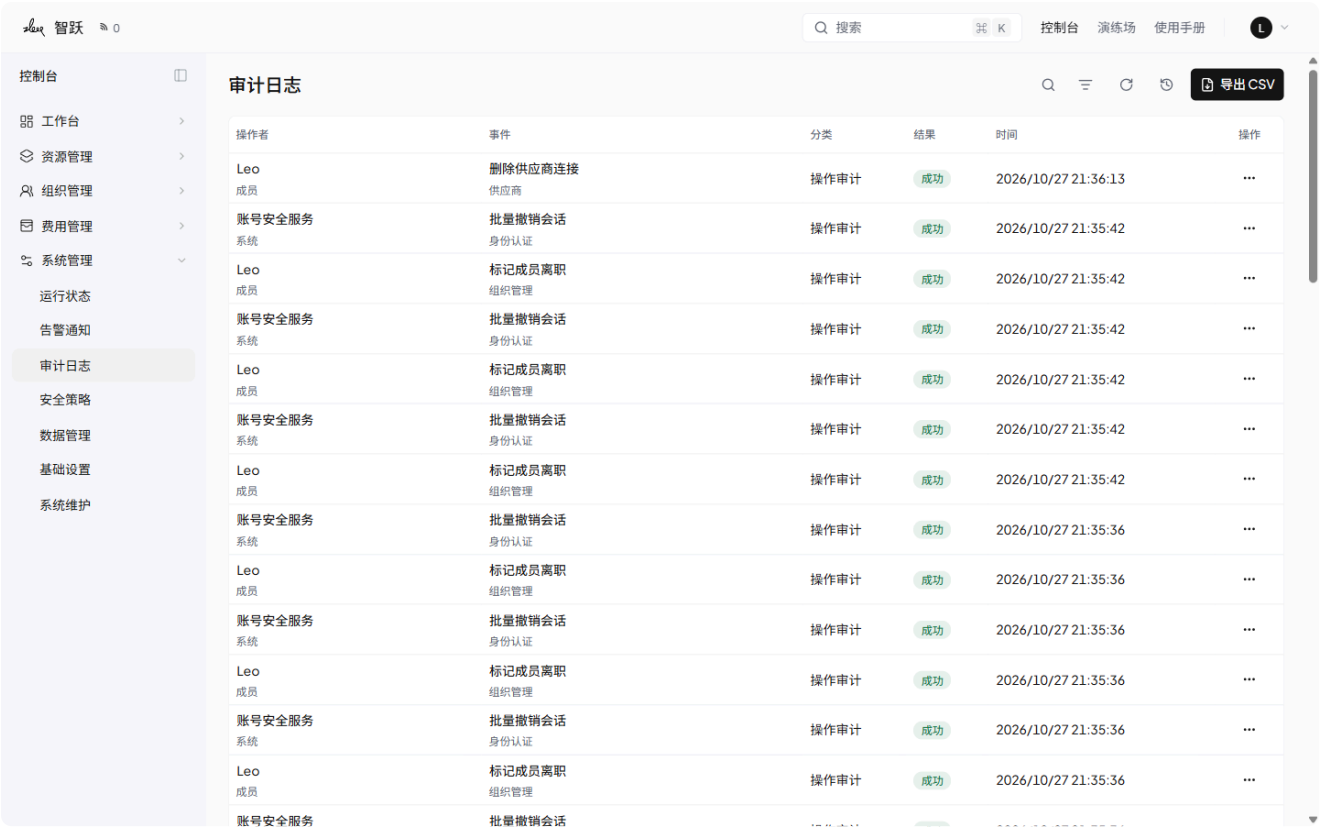
依次检查成员或 API Key 授权、预算、模型启用与上游协议，再看供应商响应。演练场内部请求与应用 API Key 请求的授权可能不同，前者成功不能直接证明后者有权调用。

工作洞察与审计

查询审计日志

查询配置变更与敏感操作记录。

打开【系统管理】→【审计日志】，按时间、操作者、事件分类和结果查找操作。



查看事件详情

打开记录，查看操作人、对象、动作和结果。事件分为【操作审计】【内容访问】【安全与授权】。



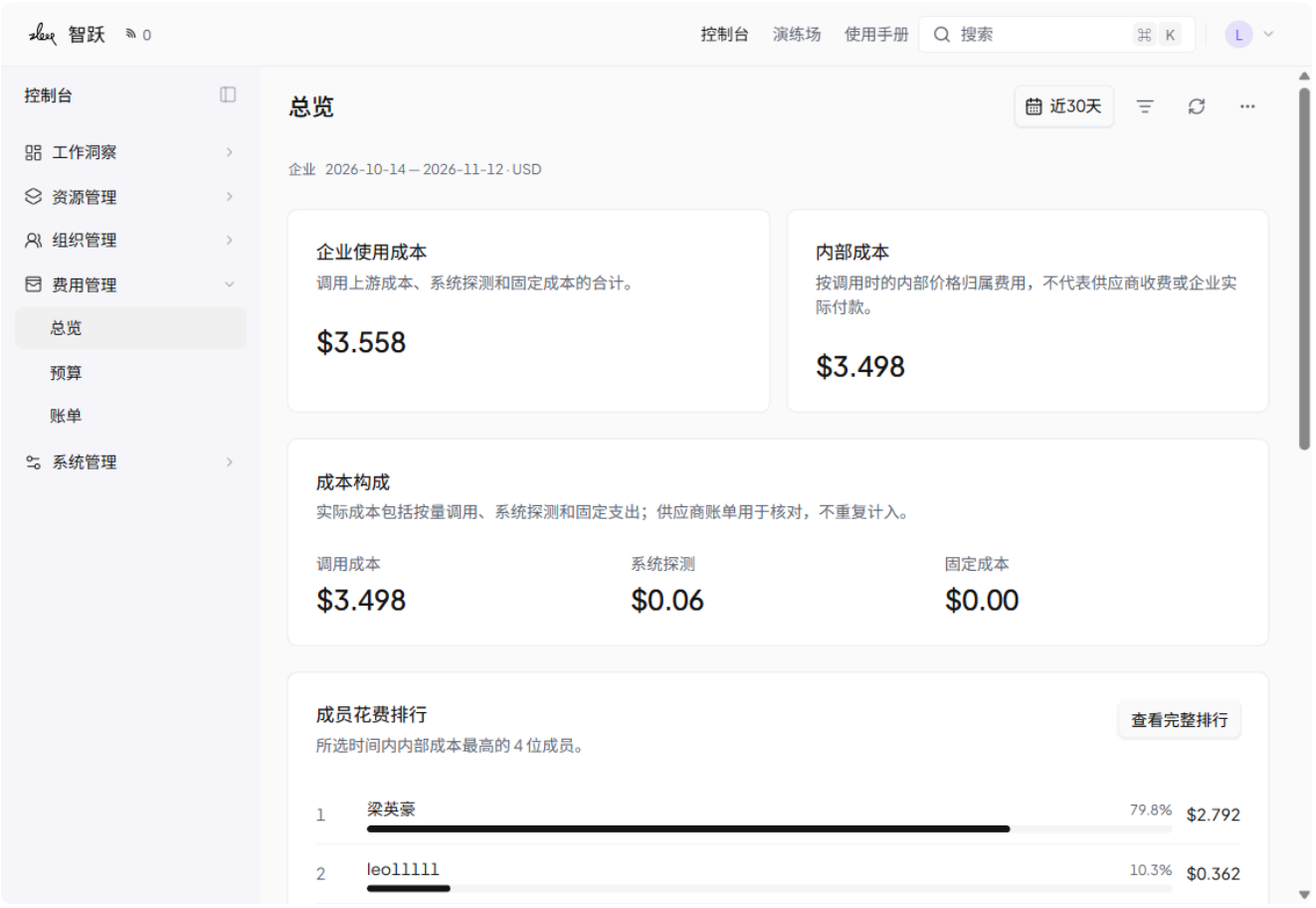
导出记录

选择导出范围后点击【导出 CSV】，按企业审计要求保存文件。排查模型调用时，结合请求日志继续查看。

查看费用总览

区分企业使用成本、内部成本与供应商账单。

路径：【控制台】→【费用管理】→【总览】。



选择统计范围

选择时间、归属范围与显示币种，核对页首区间，再阅读成本卡片和员工花费排行。

口径	含义
企业使用成本	调用上游成本、系统探测和固定成本的合计
内部成本	按调用发生时的内部价格归属，不代表供应商收费或企业实际付款
供应商账单	用于核对上游账单，不重复计入总支出

追溯费用

1. 在【费用记录】切换【调用成本】与【内部成本】。
2. 调用成本按上游尝试记录，包含重试，不代表供应商最终账单。
3. 点【分析明细】进入对应统计视图，继续按员工、资源或供应商核对。
4. 从更多菜单进入【汇率设置】或【历史固定费用】，查看对应配置与记录。

核对请求与计量口径

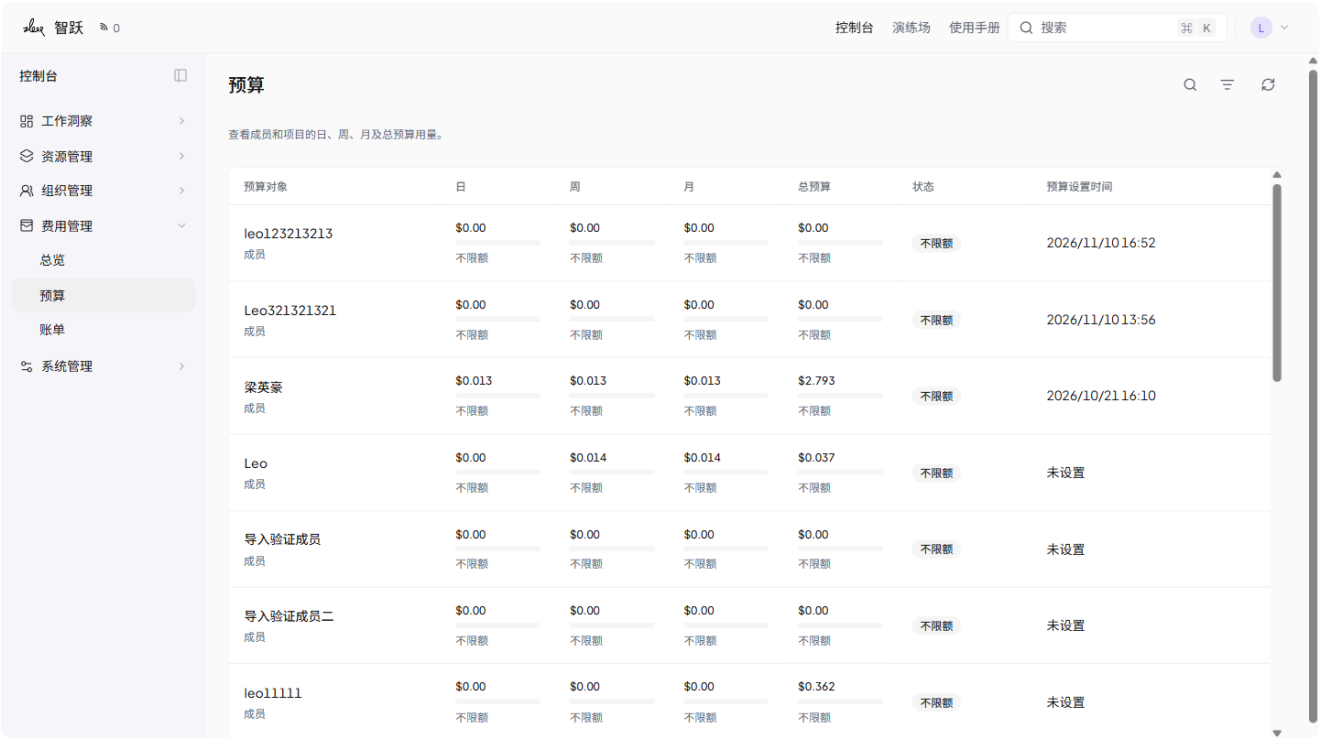
请求最终成功仍可能包含失败后重试的上游尝试，比较费用时应展开相应记录。缓存读取与缓存写入分别计量；支持搜索计费的请求还需核对搜索费用。成本不可用或缺失不能直接当成零。

内部价格用于成员预算和企业内部归属；供应商账单用于实际账单核对，两者不要求一致。

设置员工与项目预算

核对日、周、月及总预算用量。

打开【费用管理】→【预算】，查看员工和项目的日、周、月及总预算用量。



设置员工额度

在【组织管理】→【成员】中添加或编辑成员，设置对应周期的调用额度。新增成员也可在创建时设置。

设置项目额度

在【组织管理】→【项目】中添加或编辑项目，设置调用额度，保存后回到预算列表查看用量。

达到任一适用上限后，新的调用会受限。 如果把限额调到已用金额以下，后续调用可能立即停止。

核对预留与可用预算

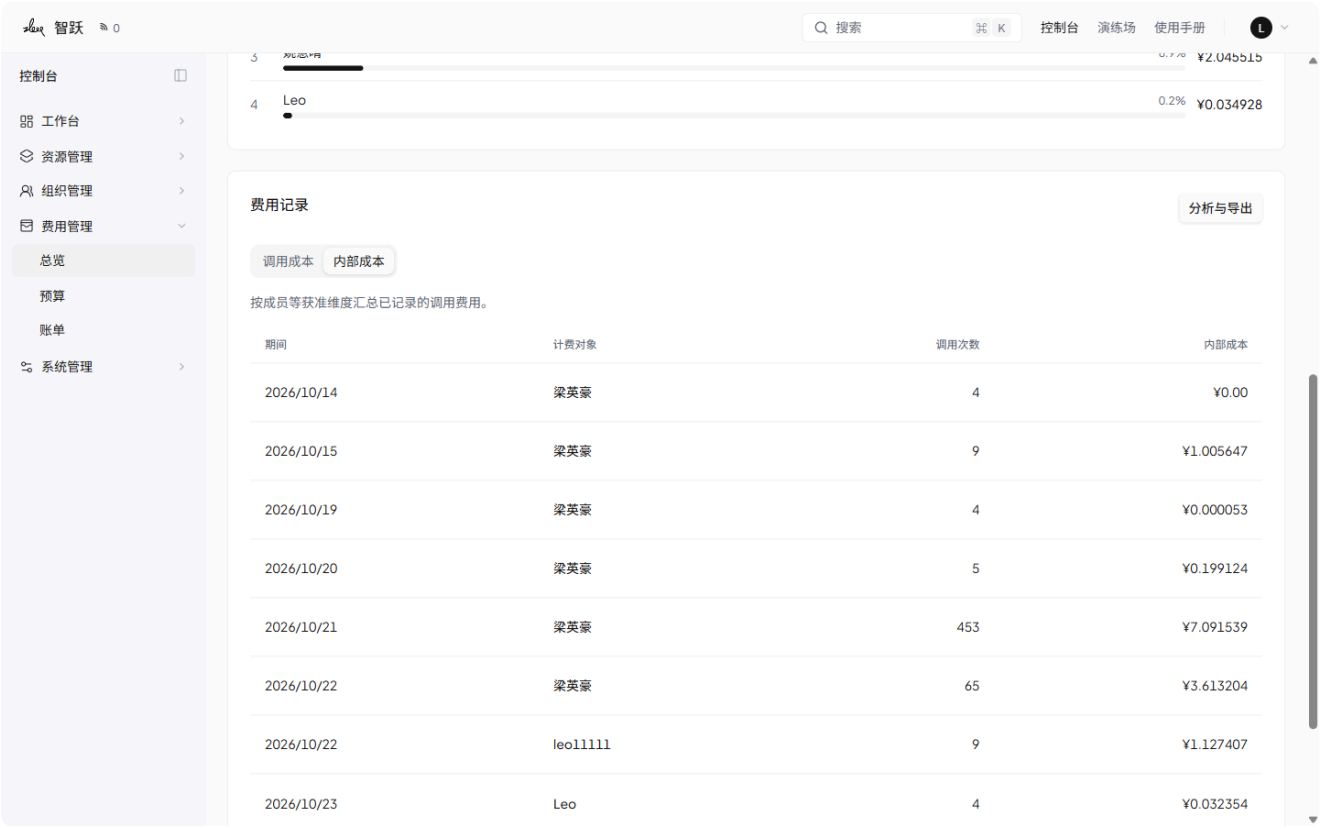
进行中的请求会预留预算，并在请求结算或结束时处理该预留。已用、预留与可用额度不是同一个数字；并发请求尚未结束时，可用额度可能先减少。

遇到额度不足，核对适用对象、预算周期、进行中的请求与最终结算，不能只看已经完成的请求费用。审批新成员时也要检查默认预算，关联已有成员时核对其原有设置。

核对内部成本

从费用总览查看按内部价格归属的费用。

打开【费用管理】→【总览】，选择时间、归属和显示币种，向下找到【费用记录】，切到【内部成本】。



内部成本按调用时的内部价格计算，不代表供应商收费或已付款金额。点击【分析明细】查看成员或项目用量；需要解释单笔费用时，在请求中查看价格与调用归属。

供应商收费见账单与对账。

费用管理

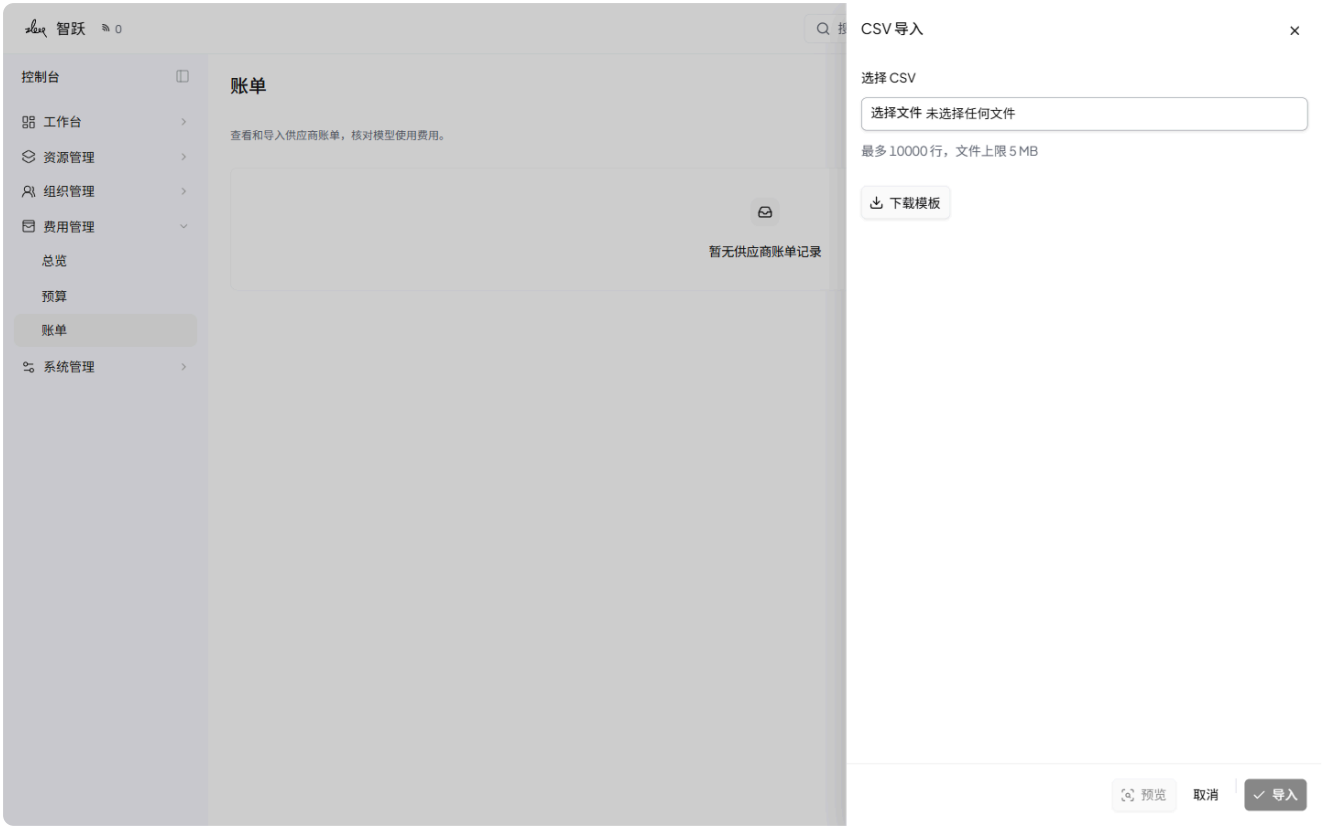
导入供应商账单

导入并查看上游供应商 CSV 账单。

路径：【控制台】→【费用管理】→【账单】

准备 CSV

点击【导入】，在【CSV 导入】窗口下载模板。按模板填写供应商、账期、币种和金额，保存为 UTF-8 CSV。文件最多 10,000 行、5 MB。

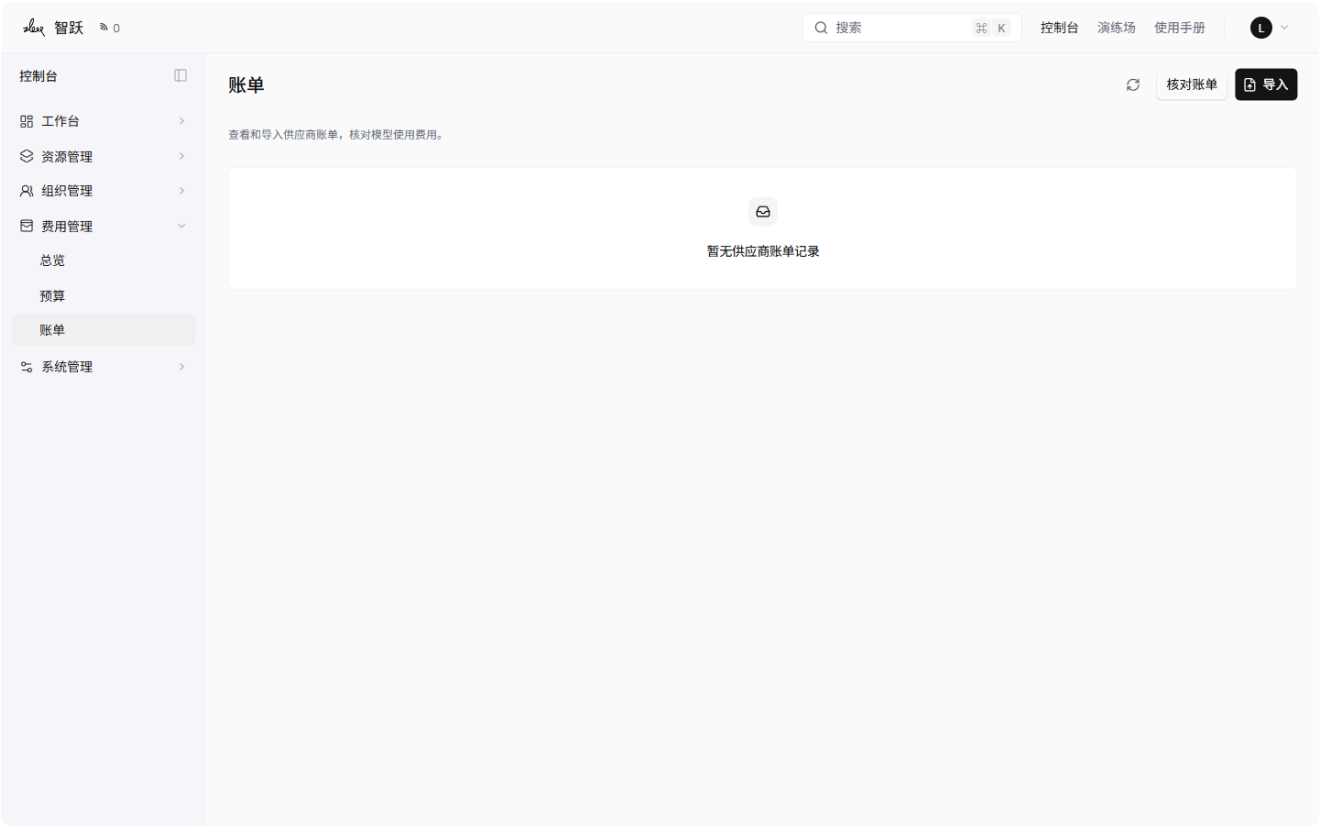


预览并导入

点击【选择 CSV】，先【预览】。有逐行错误时修改文件后重新预览，检查通过后再【导入】。

查看导入记录

回到账单列表查看本次导入的修订。需要比较供应商金额与管理后台记录时，点击【核对账单】。



导入账单不会替代请求的内部计费，也不应重复计入企业支出。比较金额时对齐账期、币种、汇率、请求与尝试数，以及缓存读写和搜索等适用费用。

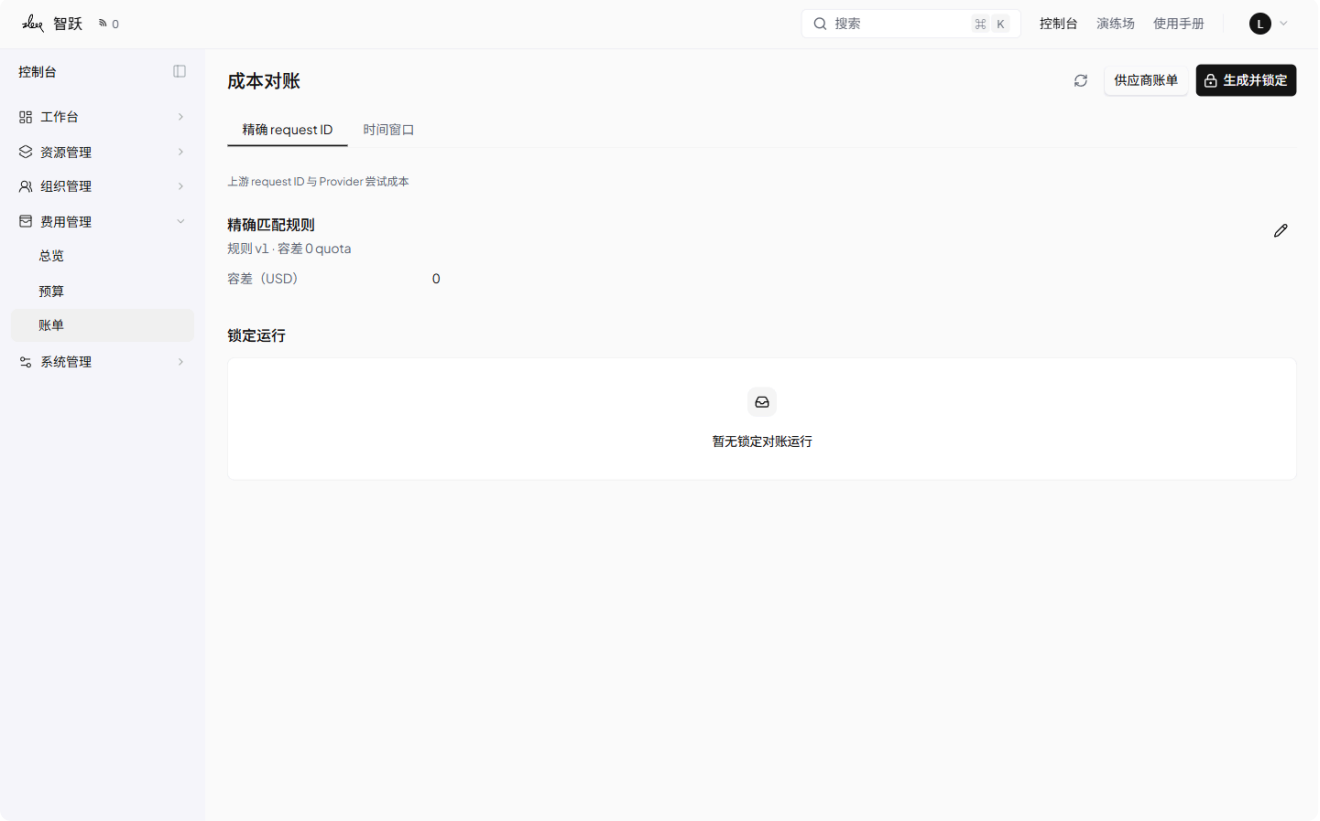
成本对账

比较供应商账单与内部上游成本。

打开【费用管理】→【账单】→【核对账单】，选择核对方式。

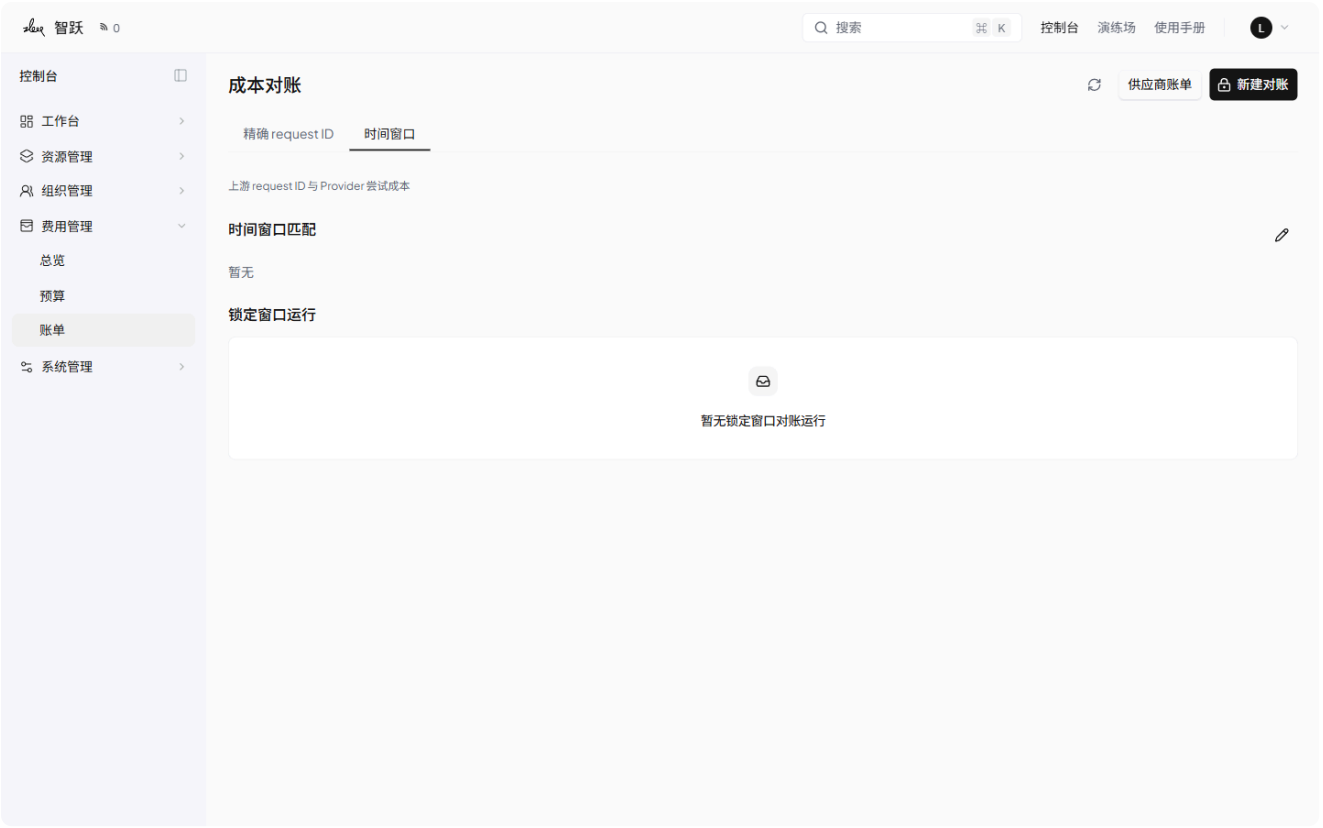
按请求标识核对

账单包含 request ID 时，选择【精确 request ID】，检查匹配规则和容差，点击【生成并锁定】保存此次结果。



按时间窗口核对

账单缺少请求标识时，切到【时间窗口】，点击【新建对账】，设置窗口匹配后比较汇总金额。



生成结果后查看匹配项、金额差异和未匹配项，需要时下载 CSV。锁定结果保留生成时的规则和账单；修改规则后需重新生成。

对账前还需对齐请求与上游尝试口径。缓存读写、搜索等费用应按实际支持的账单项比较，不能只用总 Token 乘单一价格；最终成功的请求也可能有多次上游尝试。

费用管理

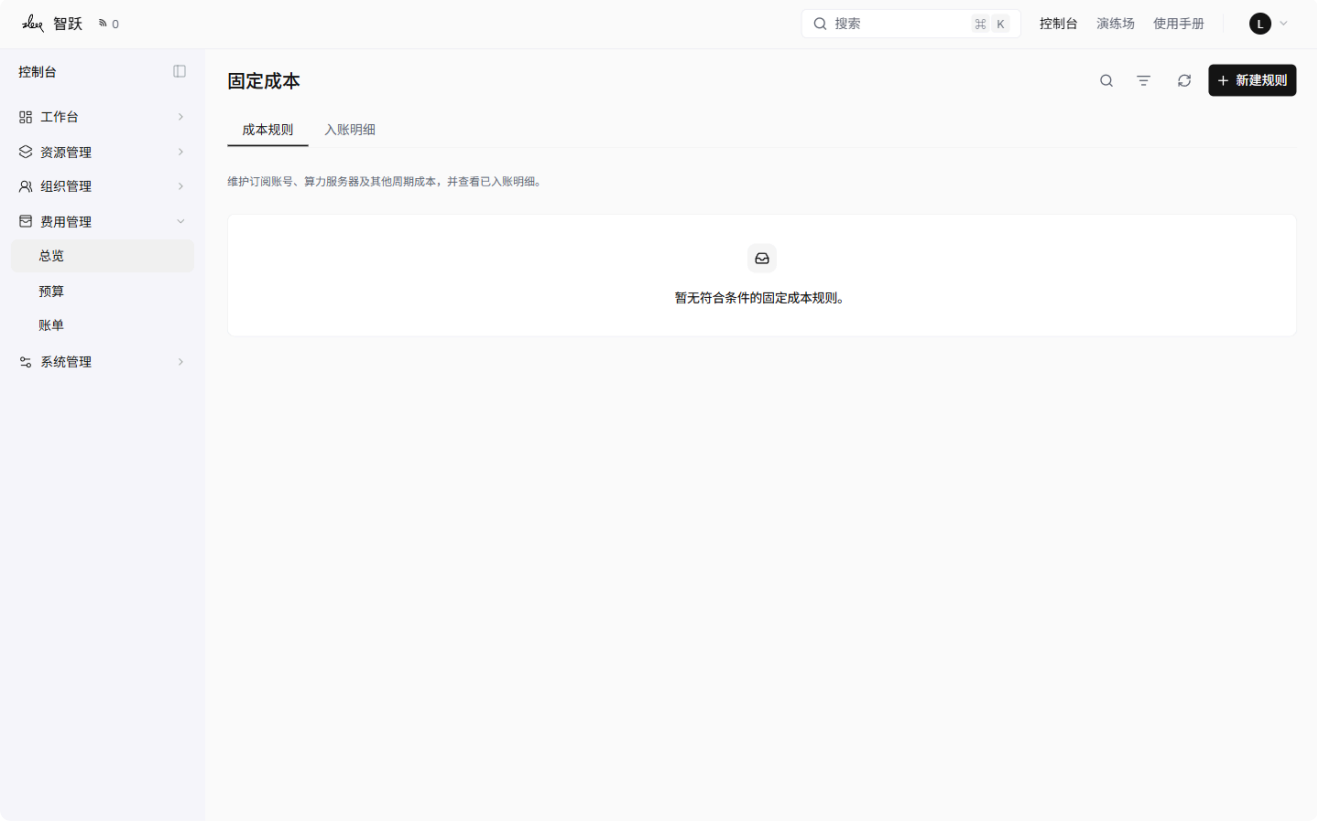
固定成本

维护不来自单次模型请求的费用。

路径：【控制台】→【费用管理】→【总览】→更多菜单【历史固定费用】

查看成本规则

【成本规则】列出订阅、算力等周期费用。点击【新建规则】添加一笔固定支出。



填写规则

填写名称、分类、金额、币种和计费周期，设置开始日期；没有结束日期可留空。需要关联供应商时选择对应连接，再点【保存】。



查看入账明细

保存后在【入账明细】查看已记录的支出。规则修改从当天起生效，不改写已经入账的数据。

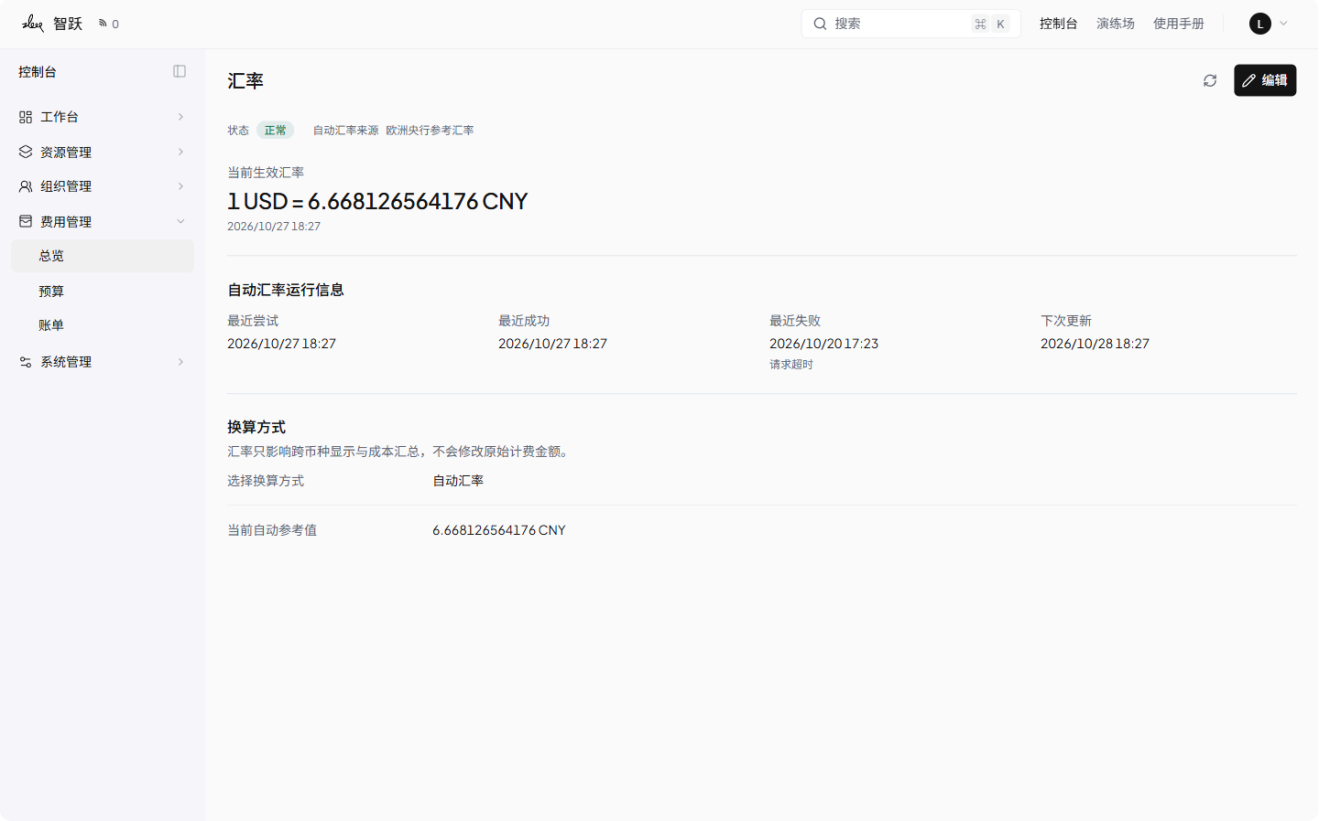
汇率

配置费用换算使用的汇率。

路径：【控制台】→【费用管理】→【总览】→更多菜单【汇率设置】

查看汇率

页面显示当前 USD/CNY 汇率、来源，以及最近成功和失败时间。



修改换算方式

点击【编辑】，选择【自动汇率】或【固定汇率】。使用固定汇率时填入换算值，再保存。



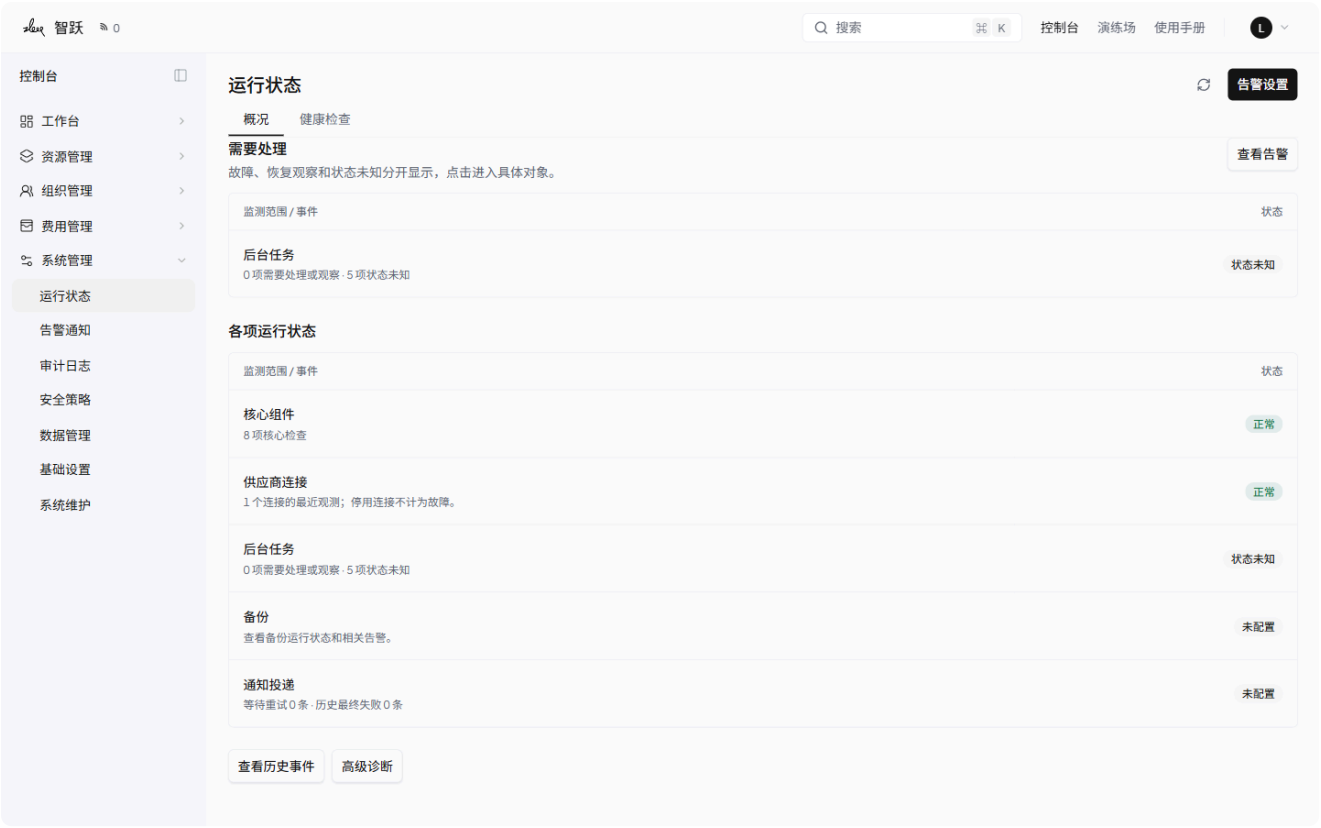
汇率只影响跨币种显示和成本汇总，不修改原始计费金额。

运行监控

检查运行状态

区分故障、恢复观察、状态未知及未配置。

路径：【控制台】→【系统管理】→【运行状态】→【概况】。



跟进状态

1. 先查看【需要处理】，打开对应对象确认故障或观察事项。
2. 查看【各项运行状态】，区分正常、需要关注、状态未知与未配置。停用的供应商连接不计为故障。
3. 进入【健康检查】检查依赖；进入【告警设置】维护规则与通知。
4. 修复后刷新相关对象，核对恢复状态；用户请求异常仍需结合请求。

运行监控

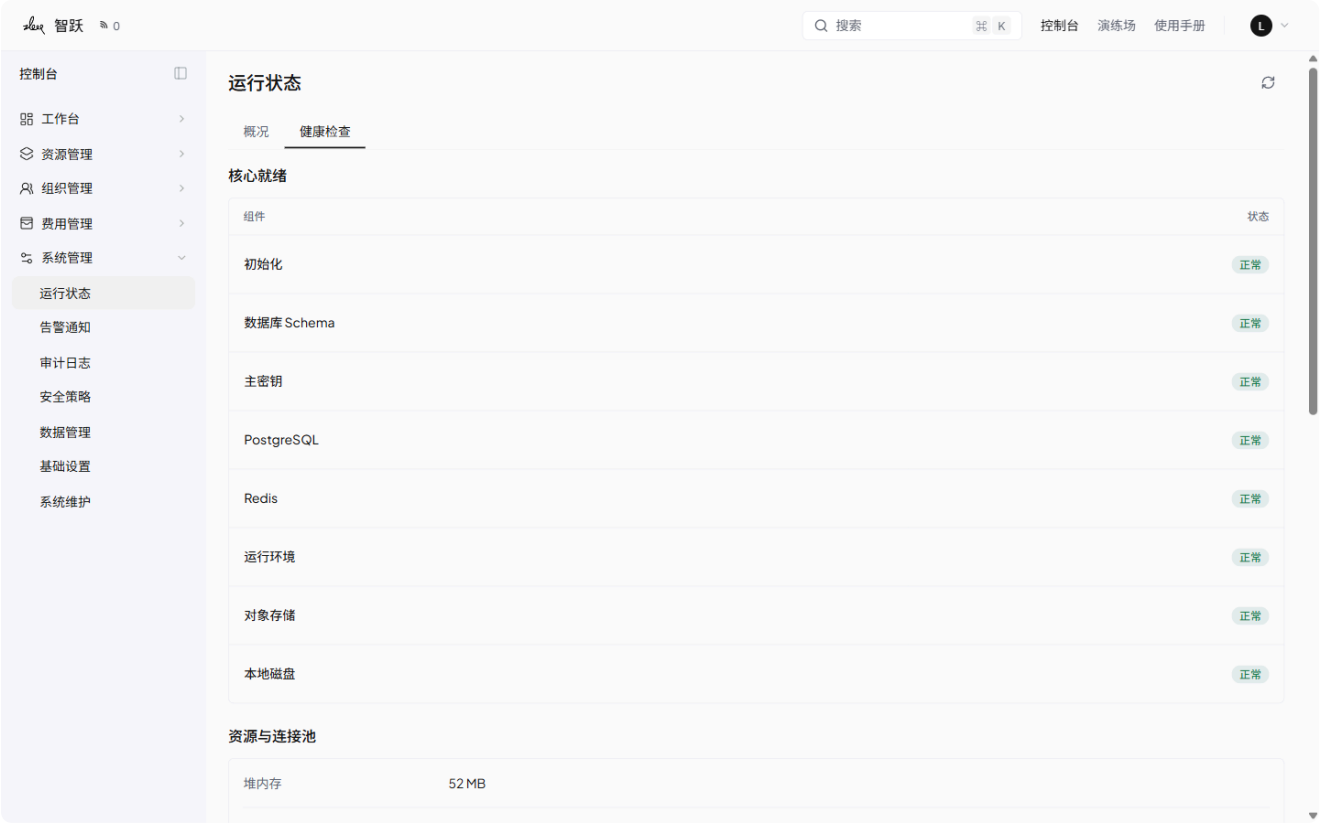
检查核心组件与依赖

在运行状态中查看健康检查并跟进失败项。

路径：【控制台】→【系统管理】→【运行状态】→【健康检查】。

查看组件状态

在【健康检查】查看初始化、数据库、Redis、对象存储等检查项。出现异常时，先按失败项排查对应服务和网络。



查看资源与后台任务

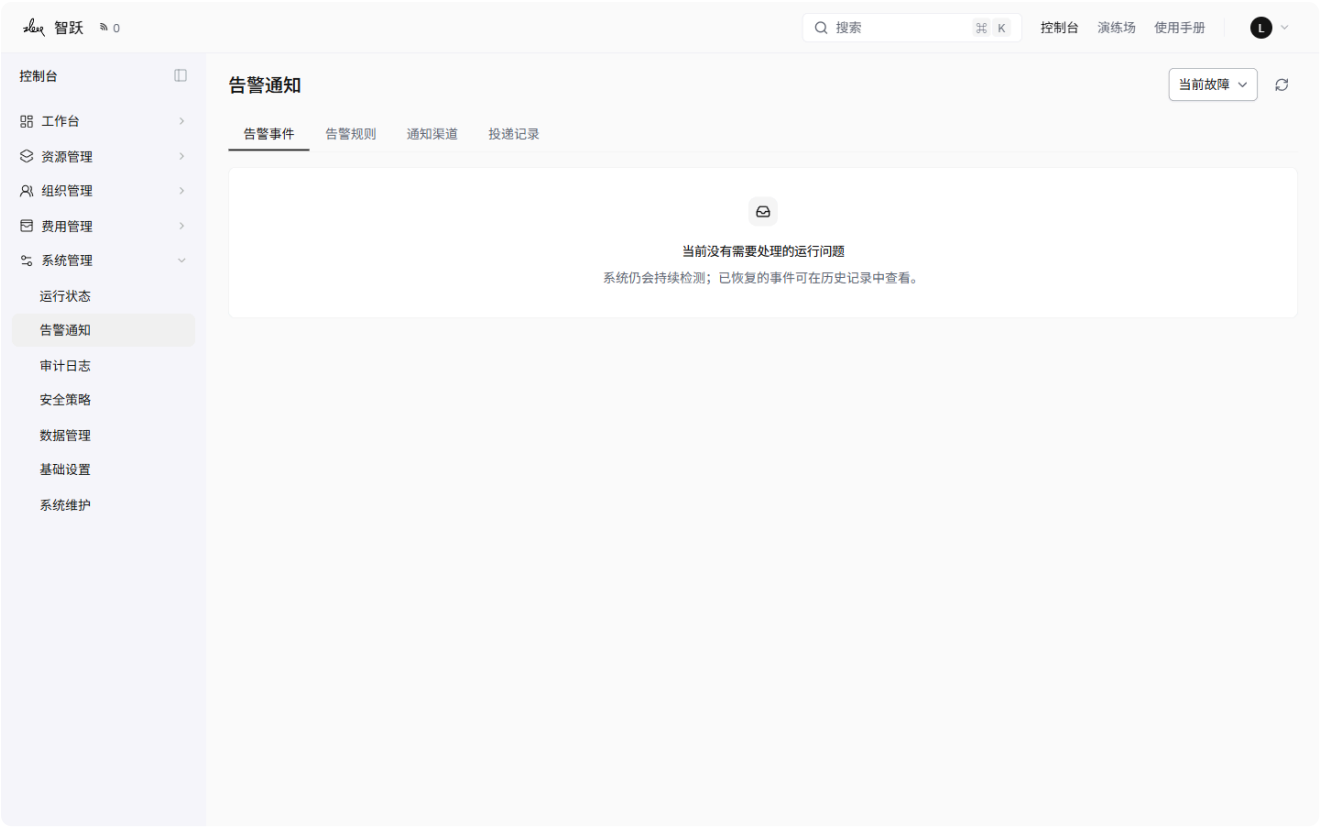
继续查看内存、连接池、磁盘空间和后台任务。修复后重新打开页面获取新结果；请求仍失败时，到请求查看具体错误。

运行监控

处理告警通知

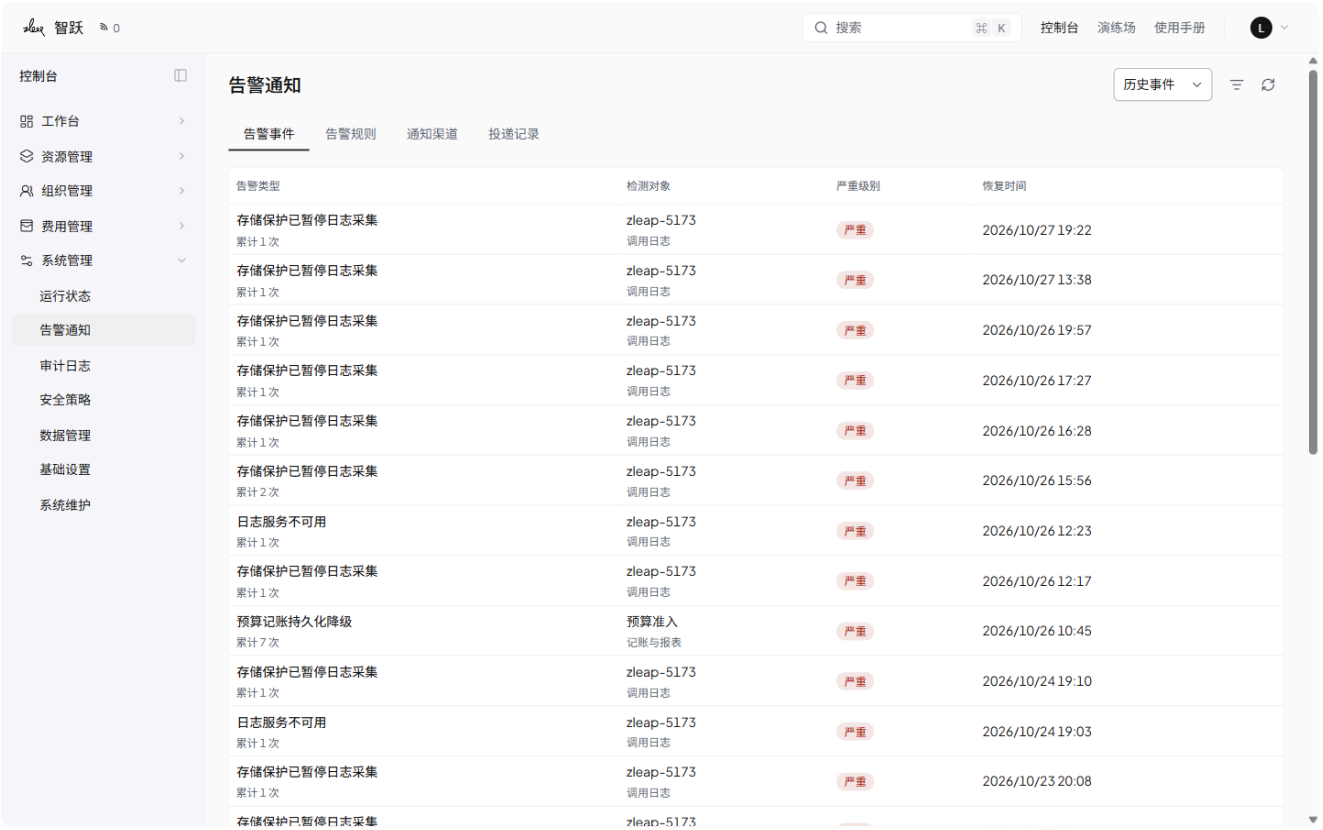
查看当前故障、历史事件、规则与投递记录。

打开【系统管理】→【告警通知】，查看当前故障及相关事件。



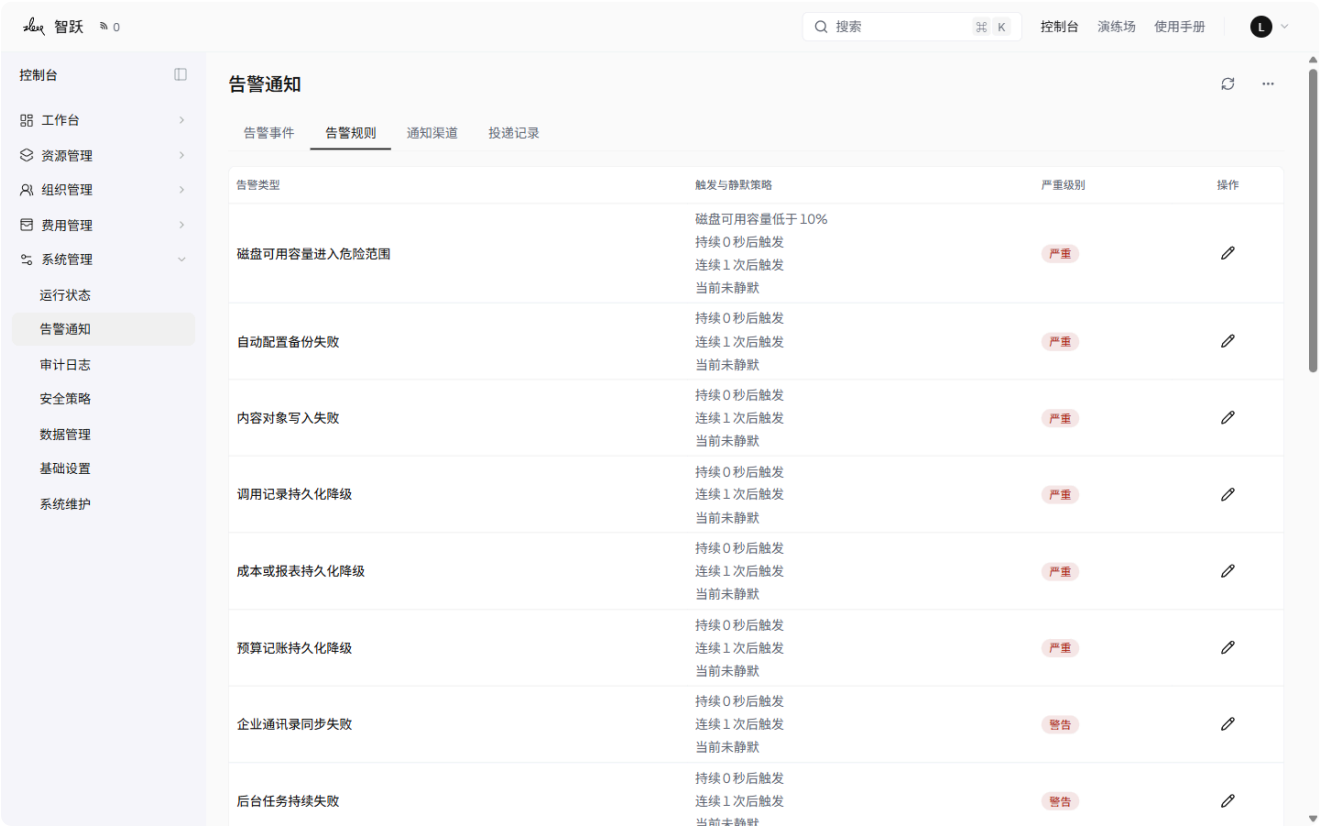
查看告警事件

在【告警事件】中用右上角的下拉框切换【当前故障】与【历史事件】，查看故障发生和恢复记录，再进入对应资源排查。



设置告警规则

切到【告警规则】，查看触发条件并调整需要启用的规则。



通知接收方式和发送结果见通知渠道与投递。修复后刷新故障状态；调用异常还可在请求日志中追查。

运行监控

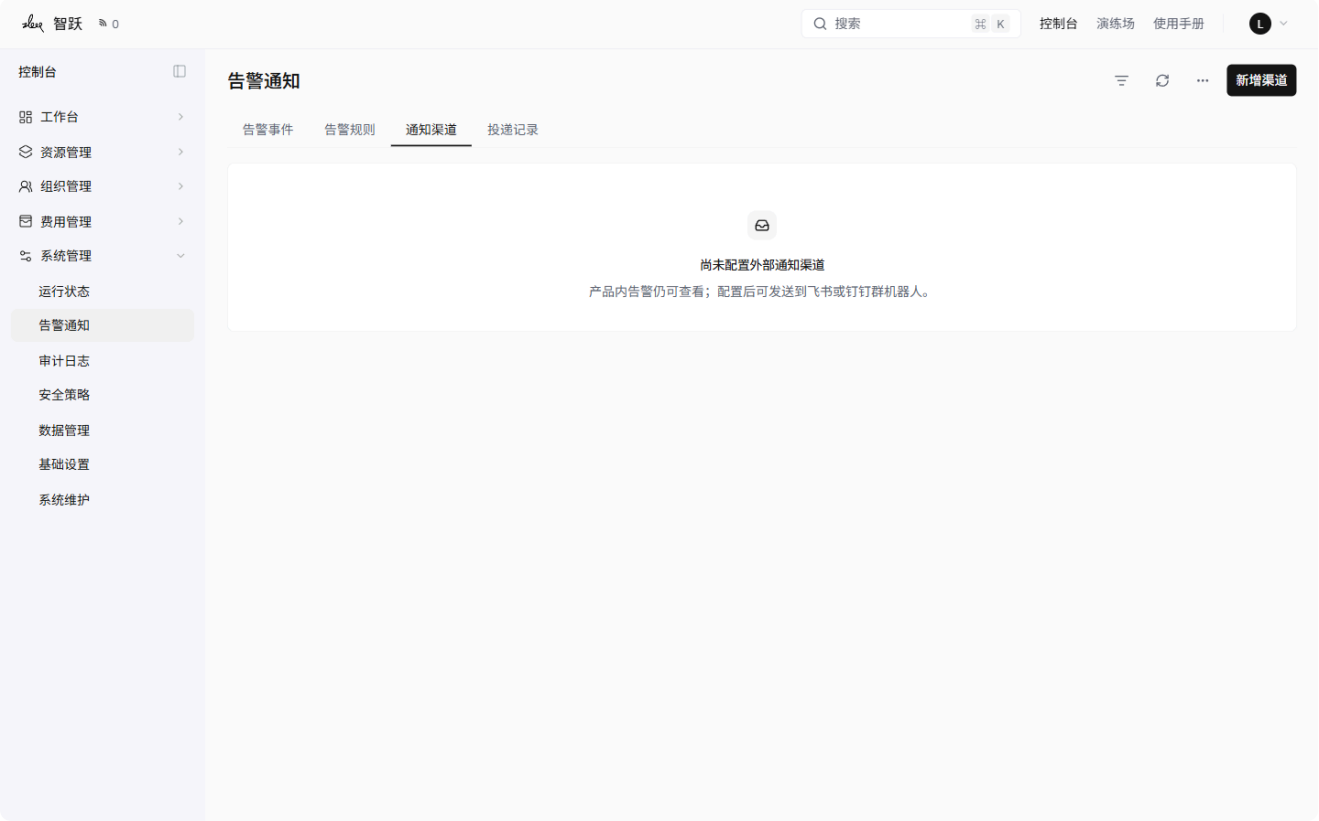
通知与投递

配置并查看运维通知投递。

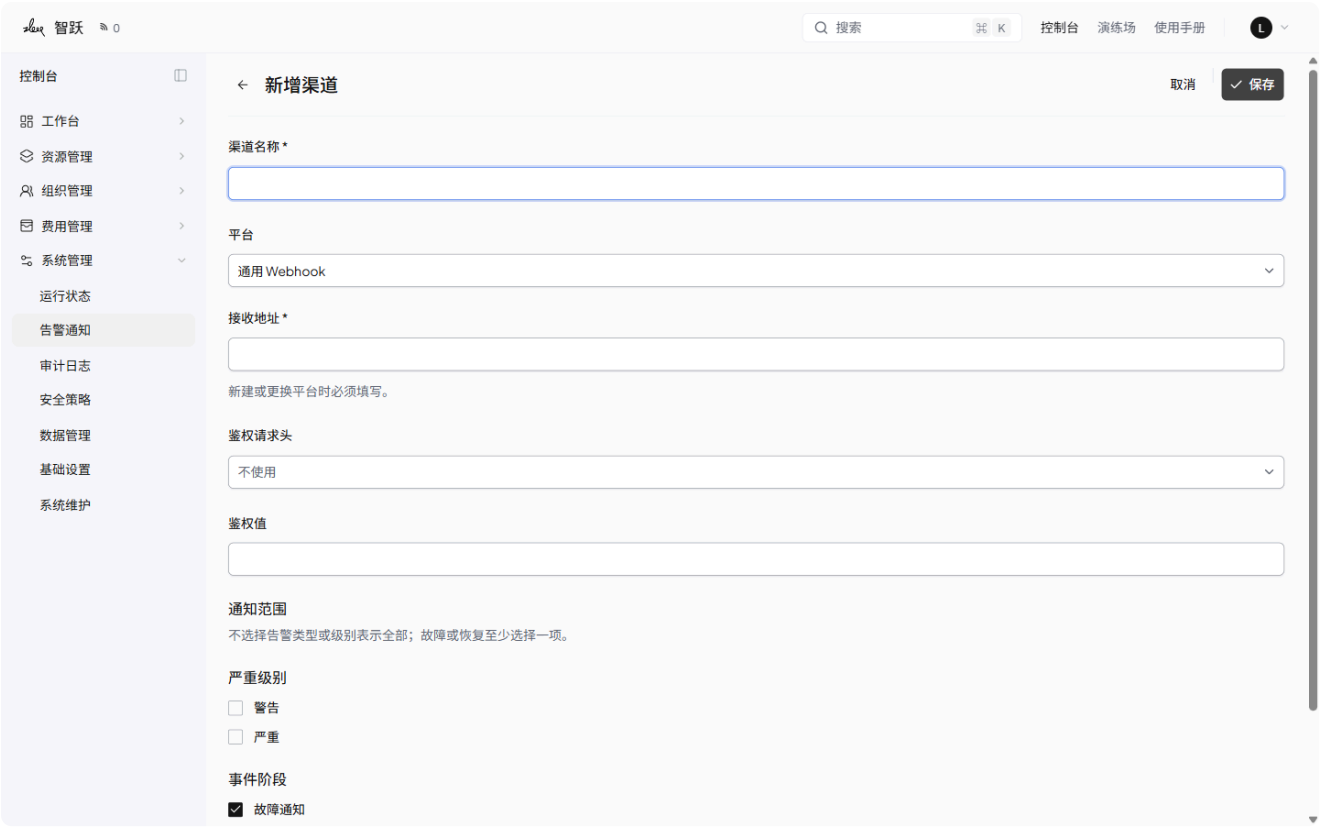
路径：【控制台】→【系统管理】→【告警通知】→【通知渠道】 / 【投递记录】

添加通知渠道

切换到【通知渠道】，点击【新增渠道】。



填写渠道名称、平台和接收地址。使用通用 Webhook 且需要鉴权时，再填写鉴权请求头和值。



选择严重级别、告警类型及【故障通知】或【恢复通知】，然后保存。故障和恢复至少选一项。

查看发送结果

在【投递记录】查看发送、重试和送达状态。没有记录时，先检查是否发生过符合通知范围的告警，以及渠道是否启用。

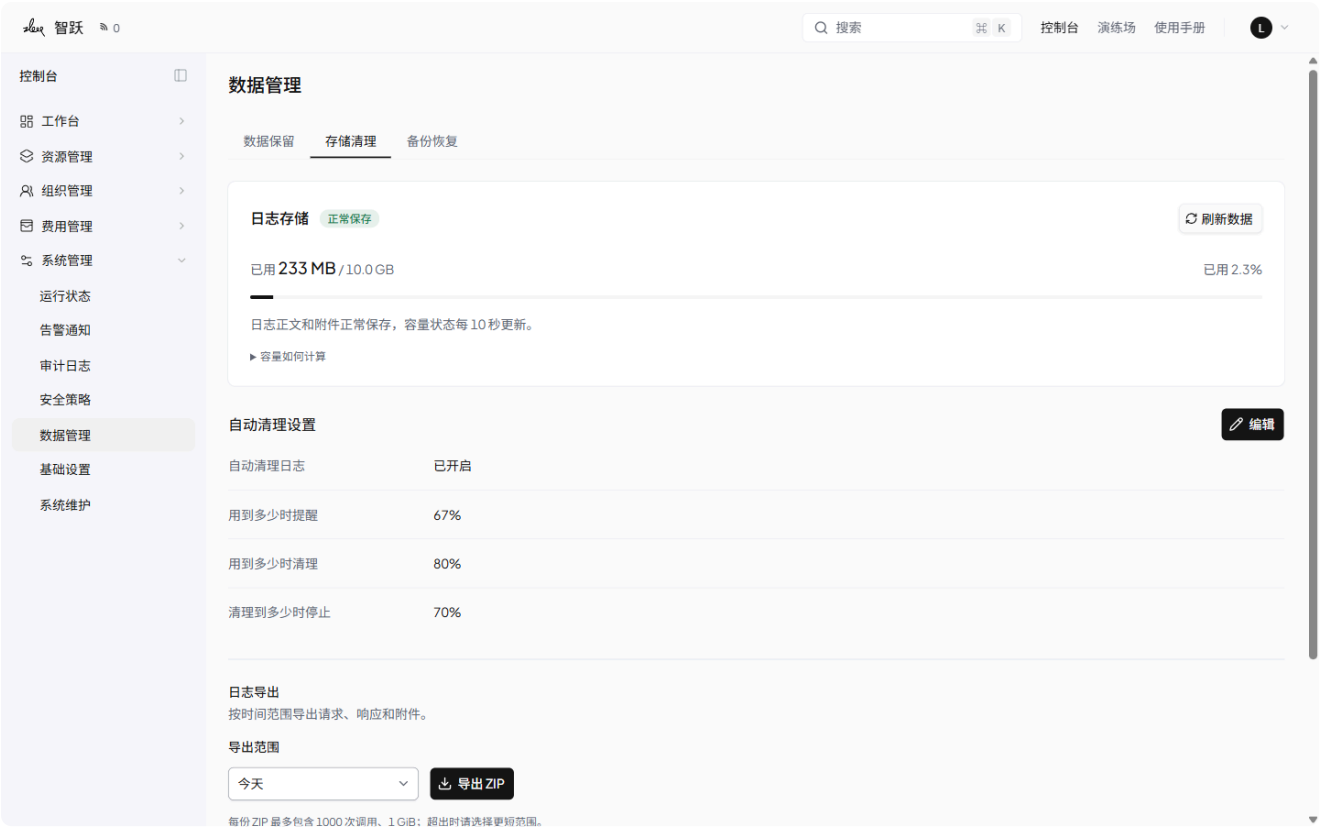


运行监控

管理日志容量与清理

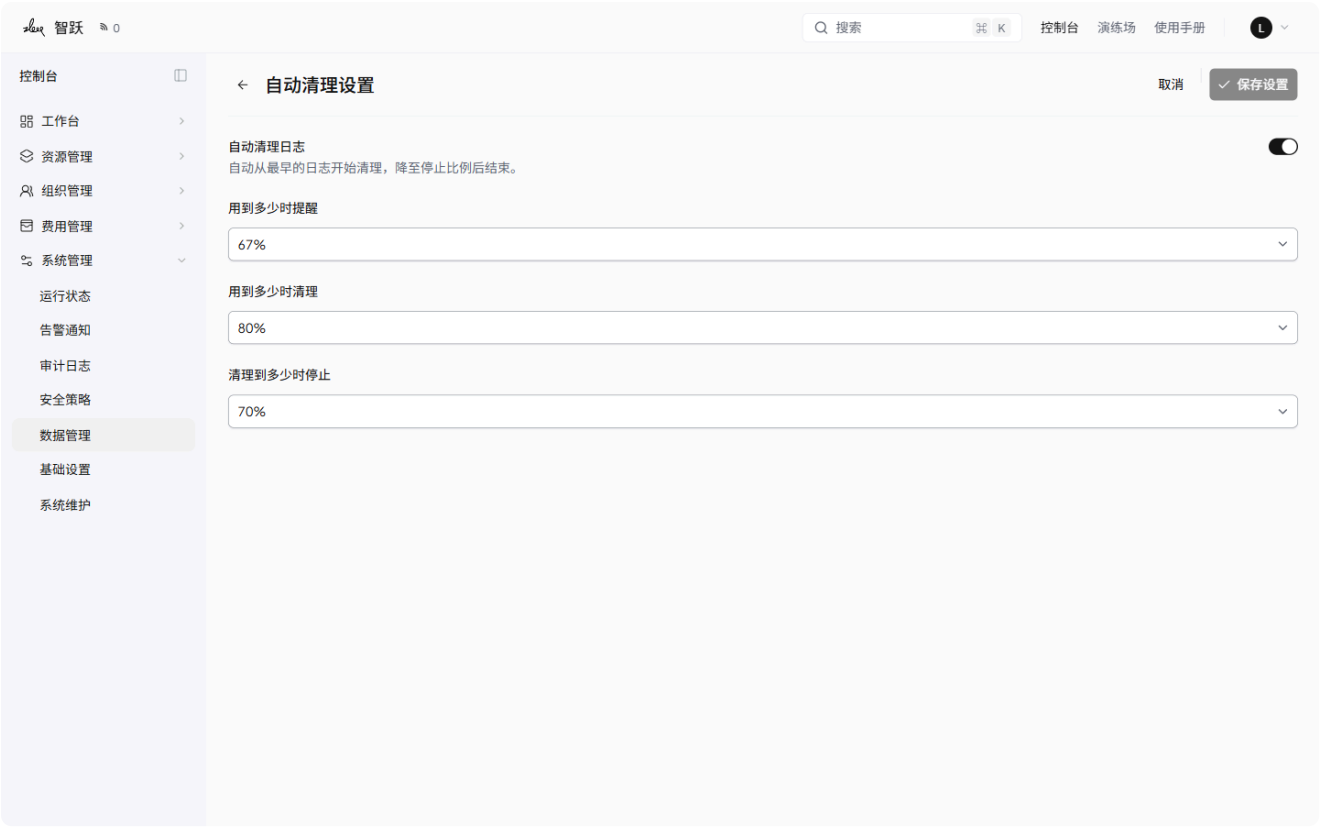
查看容量、导出日志并确认清理范围。

打开【系统管理】→【数据管理】→【存储清理】，查看已用容量和日志保存状态。容量状态每 10 秒更新。



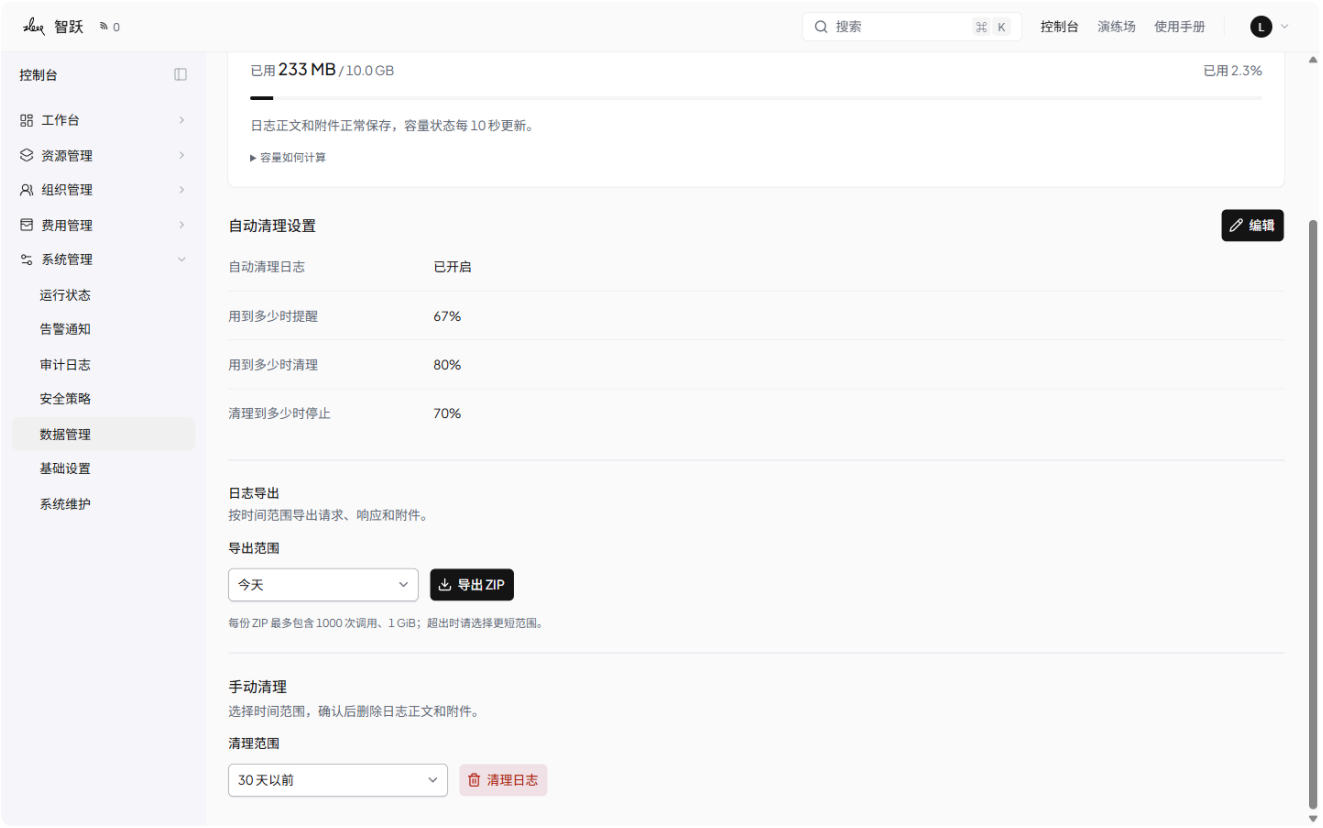
设置自动清理

点击【自动清理设置】中的【编辑】，设置提醒阈值、开始清理阈值和停止清理阈值，再保存。



导出或手动清理

在【日志导出】选择时间范围，点击【导出 ZIP】。每份最多包含 1000 次调用、1 GiB，超过限制时缩短时间范围。



需要手动清理时，选择时间范围，点击【清理日志】并确认。清理会删除日志正文和附件，操作前完成所需导出，并确认数据保留范围。

系统设置

设置隐私保护

选择观察、遮挡或阻断，并核对协议和内容覆盖范围。

打开【系统管理】→【安全策略】→【隐私保护】。审计查看能力按当前权限开放，不代表可读取全部原文。



选择处理模式

模式	请求如何处理
【关闭】	不执行隐私保护处理
【仅观察】	记录命中情况，仍向上游发送原文
【遮挡】	替换命中的敏感值，再转发；对对应回复执行还原
【阻断】	对高置信命中阻断，低置信命中转为遮挡

仅观察不会阻止敏感文本发送。 选择模式前先确认企业需要观察、替换敏感值还是阻止请求。

设置规则与例外

检查内置凭据、邮箱、电话、证件、银行卡和公网 IP 等规则，按需新增词语或正则规则。逐条核对启用状态、作用范围及允许值或例外。内网地址经常出现在代码和日志中，不要将未开启的地址规则误认为已生效。

点击【测试】，使用虚构数据检查命中和处理后的文本；验证允许值不会被误处理后，再按界面保存设置。测试通过仅证明该样例在所选规则下的结果。

核对覆盖边界

当前保护范围为 OpenAI Chat、OpenAI Responses（含 compact）、Anthropic Messages 的文本。不覆盖图片、文件附件、Gemini、向量及其他协议。即使客户端允许图片输入，也不能据此推断图片已经脱敏。

在【请求】中查看可见的隐私处理结果；排查异常时同时核对协议、规则范围和模式，避免将未覆盖内容当成漏拦截。

系统设置

维护企业信息与接入地址

配置企业标识、管理后台接入信息及网络出口。

打开【系统管理】→【基础设置】，按权限查看或编辑【企业信息】【功能设置】【网络出口】。

维护企业标识

在【企业信息】点击【编辑】，修改企业名称或 Logo 后保存。企业时区一旦锁定便不能修改，预算和报表按企业时区统计，不能用浏览器时间替代。

检查管理后台接入配置

核对自定义名称、接入地址、资源地址与额外下载源。对外接入地址应是员工设备实际可访问的地址；不要直接把管理员本机的 localhost 发给员工。

云目录登记用于辅助发现，登记成功不代表客户端已经能访问管理后台。改地址后让员工检查发现结果、原连接和模型目录，必要时按新地址重新连接。

设置网络出口

需要上游网络出口配置时，在相应页签核对后保存；这里的出口设置与宿主网络中的 Wi-Fi 连接不是同一项。更改后通过真实上游请求确认结果。

旧基础设置截图已撤下；新版企业信息与接入地址编辑图待补。

系统设置

连接与管理宿主网络

所有者在支持的设备上管理 Wi-Fi，并处理切网后的重连。

打开【系统管理】→【网络】。

检查设备能力

页面用于管理管理后台宿主设备的网络，不是修改当前员工电脑的 Wi-Fi。支持的设备可扫描无线网络、连接、断开及管理已保存网络；容器或不具备相应宿主能力的部署可能提示不支持 Wi-Fi 设置。

连接公司 Wi-Fi

在支持的设备上扫描并选择网络，填写密码后确认连接，核对连接状态。设备会保存网络并在开机后尝试重新连接。断开或删除已保存网络前，确认仍有可用管理路径。

切网可能中断当前管理页面和 Token路由客户端连接。 等宿主网络恢复后，用新网络中可达的地址重新打开管理后台；地址变化时同步检查基础设置中的接入地址与客户端连接。

排查发现但无法访问

发现或云目录登记仅帮助识别实例。还需确认客户端与宿主之间的路由、防火墙和访问地址可达，不能把目录登记成功当成跨网连通。

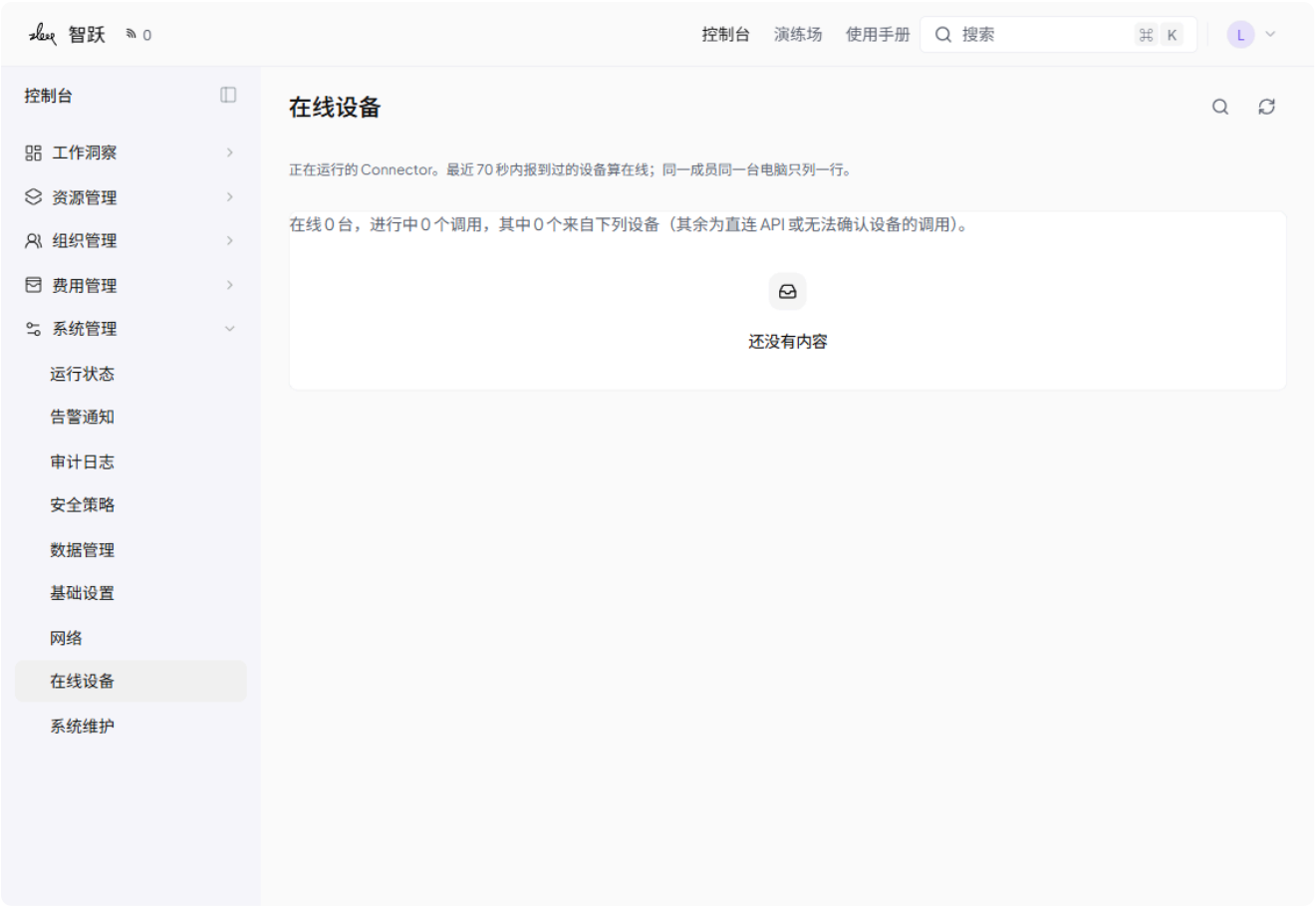
当前取证实例不支持 Wi-Fi 设置，未执行切网；扫描、连接与保存网络的实机过程图待补。

系统设置

查看在线客户端设备

核对 Token路由客户端报到设备及进行中的调用。

打开【系统管理】→【在线设备】。



查看设备在线情况

页面列出正在运行并上报的 Token路由客户端。最近 70 秒内报到过的设备算在线，同一成员同一电脑只列一行。按列表可见信息核对成员、设备及状态。

理解调用数量

页面同时显示进行中的总调用，以及其中归属于下列设备的调用。两者不一定相等：直连 API 或无法确认设备的调用可能没有设备归属。

没有在线设备不等于企业没有请求，也不代表成员授权被删除。先确认 Token路由客户端是否运行、连接到同一实例并保持联网；旧客户端缺少能力时以实际显示提示为准。

本次截图为真实空列表，已连接设备行及归属调用的结果图待补。

系统设置

设置请求限流

为员工设置跨 Key、模型与协议共享的请求限制。

打开【系统管理】→【安全策略】→【请求限流】，查看当前开关、统计窗口和请求数上限。



修改请求限制

点击【编辑】，设置统计窗口、总请求数及成功请求数上限，保存后检查显示值。



同一员工的所有 API Key、模型和协议共用此限制。费用额度另外在预算中设置。

需要处理正文敏感值时，继续配置隐私保护。请求限流只限制请求量，不代替隐私保护。

系统设置

管理留存与数据清理

设置留存周期、导出日志，并区分备份与调用内容。

打开【系统管理】→【数据管理】→【数据保留与清理】。修改留存和清理设置需要相应管理权限。



检查留存范围

按页面说明分别检查调用内容、请求元数据与审计数据的周期。当前页面提示请求元数据至少保留一个自然年，审计数据保留三年，审计数据清理按 UTC 整月规则执行。不要将内容留存天数套用到全部数据。

需要延长留存时，同时检查存储空间和容量阈值。页面显示的是实例当前设置，不应将示例实例中的阈值当成所有部署的默认值。

导出并清理

在【日志导出】选择时间范围，按提示导出 ZIP。单次上限为 1,000 条调用、1 GiB；超出时缩小范围分批处理。

【手动清理】按所选范围删除对应数据，执行前检查范围并保存确需保留的资料。运行日志的维护还需查看系统维护页支持的部署能力。

区分留存、备份和恢复

备份恢复保存配置及相应报表，不包含调用内容、生成媒体、演练场记录、日志、审计流水及临时导出。完成备份不能作为这些内容已归档的依据；清理后也不能期待用配置备份找回它们。

系统设置

备份并恢复配置与报表

创建备份，在控制台上传校验并确认恢复。

打开【系统管理】→【数据管理】→【备份恢复】。



创建并保存备份

点击【立即备份】，等待任务结束并检查记录。需要定期备份时在【备份计划】点击【编辑】，设置计划后保存，留意实例时区及执行结果。用【导出配置】取得需留存的文件，并保存在管理后台之外的可靠位置。

明确备份范围

包含	不包含
组织、权限、接入资源和凭据、授权、预算和费用规则	调用正文与内容
用量、预算、费用报表等记录	生成的图片、音频、视频
配置恢复所需的相应数据	演练场记录、运行日志、审计流水、临时导出

这不是全部业务文件或全部审计数据的备份。 需要保留排除项时，在清理前分别按企业要求归档。



上传并校验

1. 在【从备份恢复】选择 `.302aieb` 文件；旧加密文件需要时按提示输入密码。
2. 点击【上传并校验】，等待文件检查，阅读校验结果。
3. 确认文件来源、适用情况及恢复影响，再决定是否继续。

上传校验本身不会改动当前数据；文件选中或上传完成都不代表恢复已经执行。

确认恢复并等待重连

安排维护时段，先保存当前所需备份，并通知使用者。确认恢复将替换实例相应数据并重启服务，当前管理页面和客户端可能断开。

服务回来后重新连接，查看原恢复任务状态，核对组织、资源、授权与报表，再验证一次真实请求。中断或失败时按任务提供的恢复提示处理，不重复点击来启动新的恢复。

初始化导入和离线恢复仍适用于对应部署场景；控制台已提供上传恢复，不再要求所有恢复都到服务器操作。本轮只核对上传入口和源码流程，未对运行实例执行恢复；校验成功、确认及恢复完成的过程图待补。

系统设置

核对许可证状态

分别检查本机授权和外部验证结果。

打开【系统管理】→【系统维护】→【许可证】，查看本地授权、最近成功验证时间及下一次验证时间。



本地授权和外部验证结果分别显示。验证服务暂时不可用不代表本地授权已经失效；只有授权服务明确返回停用后，实例才会停止服务。需要处理授权问题时，将页面状态交给负责许可证的管理员。

部署与更新

升级并确认运行版本

检查可用版本、预检与操作回执，并在断线后继续核对。

打开【系统管理】→【系统维护】→【版本更新】。

检查版本与部署能力

查看当前版本、可用版本、更新说明和历史记录。能否在线升级取决于发布源、验签配置及安装更新器等部署条件；按页面诊断处理，不要把某台实例的“无更新源”或“未安装更新器”当成产品永久限制。

备份配置及需要单独保留的数据，安排维护时段，确认服务重启可能影响当前请求和客户端连接。

预检并确认升级

1. 选择目标版本并检查变更说明。
2. 执行页面提供的预检，处理阻止升级的项目；日志大小警告应结合磁盘和维护需求处理。
3. 阅读影响提示，由所有者确认并提交。
4. 保留本次操作信息，等待操作回执。

预检通过不等于升级完成。关闭尚未受理的升级界面，也不能理解成可以随时撤销已开始的更新。

等待回执与服务重连

升级期间服务可能重启，页面可能暂时无法连接。等待服务恢复后查询原操作回执，不反复提交新的升级。 重连后核对回执状态、当前版本和历史记录，再验证模型调用。

仅看到“等待重连”不能证明升级成功；应以完成回执和实际运行版本共同确认。

处理中断与恢复

遇到失败或中断时，记录原操作标识、失败阶段及页面恢复提示。按更新器提供的继续或恢复入口处理；无法在线处理的部署交由维护人员按对应部署方案恢复，不手工改版本号来代替修复。

离线部署由维护人员使用对应发行包或镜像，保留数据卷、配置和备份，更新后同样检查实际运行版本与请求结果。

本轮未执行实例升级。旧环境的失败截图已撤下；预检、回执、重连及恢复状态的真实过程图待补。

排查常见问题

按接入阶段、授权、预算、安全和维护状态定位问题。

找不到菜单或分析入口

先检查当前账号权限及功能开关。【分析】为所有者入口；【概览】看统计，【请求】查单次记录。旧活动与日志名称已调整，见使用入口。

已有成员为什么还显示后台未开通

后台账号与 Token路由客户端接入分别管理。员工可先申请企业接入，经批准领取连接，不以控制台登录激活为前提。已有成员申请时应关联原记录，见成员管理。

Token路由客户端发现不到管理后台

确认自动发现的平台与系统服务前提；使用管理员给出的地址或连接码尝试接入。云目录只辅助发现，不能绕过网络不可达、防火墙或地址错误。Linux 不能直接套用 Windows/macOS 自动发现步骤。

申请已批准但还没有连接

保持 Token路由客户端联网，检查原申请与连接状态。关闭窗口不会自动取消已提交申请；恢复网络或重启后先等待原申请领取，不重复提交。管理员同时核对关联成员和授权。

网络和目录正常但模型仍不能调用

地址可达、模型目录可读与实际调用成功是三个阶段。检查成员授权、预算、模型启用状态、上游路由和协议；到请求看实际失败原因。没有请求记录时，先确认调用是否到达正确实例。

企业方案更新后，员工为什么仍使用原配置

发布不等于自动覆盖个人配置。员工需要获取更新，核对个人修改并确认应用到目标工具；同时检查模型授权。软件下载、方案采用与权限分配也分别处理。

AI 洞察为空但 Token 有统计

基础用量不依赖自动分析。所有者在分析设置检查开关、模型及处理节奏；开启后只分析新增内容，不自动补算全部历史。没有会话标识的请求也不代表形成了完整会话。

隐私保护为何没有处理图片或附件

当前覆盖指定协议的文本，不覆盖图片、附件、Gemini、向量及其他协议。【仅观察】仍发送原文。先核对覆盖范围与处理模式，不能用图片可上传推断图片已脱敏。

预算尚有余额却提示不足

除已结算用量外，进行中的请求会预留预算。检查成员、项目及各周期的适用限额与进行中请求，再核对结算；不要通过换 Key 绕过预算。

报表与供应商账单金额不同

先对齐时间、时区、币种、汇率、请求与上游尝试口径，再核对内部价格、缓存读取/写入和搜索等费用。重试后成功仍可能有多笔上游成本。详见成本对账。

API Key 丢失或可能泄露

到API Key检查归属和状态；不能安全找回时停用并替换。排查或截图不粘贴完整凭据。

演练场可用而应用失败

演练场使用当前登录账号授权，应用使用 API Key 的归属和范围。对照两次请求的模型标识、权限、预算与网络限制。

备份后为什么没有恢复调用正文

配置备份包含配置及相应报表，不包含调用内容、生成媒体、演练场记录、日志、审计流水和临时导出。需要的内容应清理前单独归档，见备份范围。

升级或恢复后页面断开

服务重启可能暂时断开页面。等重连后查询原操作回执和实际运行版本或恢复状态，不反复提交。等待重连不等于成功，也不要删除数据卷来处理连接问题，见升级流程。

更新日志

按版本查看管理后台与 Token路由客户端的功能变化和使用影响。

本页按产品版本记录变化。部署版本与源码版本使用不同编号：已从镜像构建信息确认，管理后台部署版 1.0.56 对应源码 v1.2.10；部署版 1.0.35 对应源码提交 92da8c87。尚无逐个部署镜像的对应证据，因此不把下面的源码版本擅自标成 1.0.36～1.0.55。

管理后台 v1.2.6 依据 main 的 CHANGELOG；其余源码版本结合相邻标签间的提交差异整理。客户端依据 GitLab main 的 CHANGELOG。版本号说明功能所属范围，不代表目标实例已经升级或全部平台均已验收。

管理后台 1.0.56（源码 v1.2.10）

相对源码 v1.2.9：

- 用量报表拆分【缓存读】【缓存写】Token，便于分别核对缓存使用；报表默认查看今天，并记住筛选条件。
- 模型列表点击模型名称或整行直接进入编辑页；只有查看权限时进入只读详情。
- 修复上游先拒绝、重试后成功的请求成本展示，避免将这种请求一律显示为成本不可用。
- 调整公共上游地址的 DNS 连接处理，跳过不允许使用的解析结果。
- 升级时，日志数据库结构未变化可保留日志数据库与队列，减少不必要的迁移。
- 移除语义拦截实现。首次接入与安全教程改按当前隐私保护流程说明。

管理后台源码 v1.2.9

相对源码 v1.2.8：

- 数据管理支持在控制台上传备份并恢复；备份和恢复入口统一整理。
- 隐私保护统一规则作用范围，展示调用处理结果；阻断只针对能够确认的命中，改善拒绝原因说明。
- 宿主网络补齐 Wi-Fi 连接控件与配套组件安装；是否可用仍取决于部署设备和宿主支持。

- 企业发现支持向云目录报告入口；接入申请可关联已有成员并发放新的连接码，避免重复建号。
- 自动分析改为主动开启，升级后不要默认认为历史记录会自动补分析。
- 供应商导入保留手工设置的成员价格，避免重复折扣；保存后保留折扣显示，改变计费类型时清除不适用的折扣。
- 费用支持相应协议的供应商联网搜索计费；进行中请求预留预算，结束结算后释放预留；修正中断请求及厂商特有 Token 的统计。
- 用量增加按模型 Token 的面板；升级前展示日志体量提醒，并补充慢等待、取消与恢复提示。
- 软件资源页移除配置入口；管理员发布组合应进入【资源管理】→【配置】。

管理后台源码 v1.2.8

相对源码 v1.2.7：

- 分析统一处理节奏，可配置安静判定、最长等待与并发；运行参数可查看和调整，补充进度和停止任务的重试入口。
- 改善自动选模型、推理模型超时、异常输出修复及分析费用约束；分析来源时间按企业时区显示。
- 模型价格和供应商导入新增折扣控制。
- 接入申请支持仅填写手机号；局域网发现统一发布信息，手工入口与发现入口可并存。
- 升级保留日志正文，补充宿主升级日志，并完善中断恢复和只读检查；等待检查完成后再按实际回执处理。

管理后台源码 v1.2.7

相对源码 v1.2.6：

- 隐私保护允许图片通过，并避免将工具数据误判为附件；图片本身不等于已经完成文本隐私检查。
- 撤回后的组合可删除，已删除模型从组合选择中移除。

- 调整 Claude 推理配置与模型能力的匹配；修复部分上游拒绝推理参数的问题。
- 延迟探测不再发送不适用的温度参数。
- 安装流程可在许可证激活前完成，修复空机器恢复及待激活身份的识别。

管理后台源码 v1.2.6

源码 CHANGELOG 标注发布日期为 2026 年 11 月 9 日，相对 v1.2.5：

- 员工可在客户端自主申请接入；管理员单个或批量审批、拒绝，或关联已有成员，并设置部门、可用模型与额度。
- 团队和个人分析支持按成员、时间查看报告及来源，增加 AI 汇总、洞察、提示词编辑和试运行。
- 企业工具配置按组合创建、发布、更新和撤回；员工取得方案后仍需在客户端应用。
- 支持加密配置导出、空机器初始化和备份恢复；配置备份不能替代完整业务数据备份。
- 新增宿主 Wi-Fi 管理候选入口，仍需硬件实机验收。
- 改善模型默认选择、供应商模型勾选、接入申请中断处理与单路由故障恢复；后台登录状态与资源使用授权分别管理。

管理后台源码 v1.2.0～v1.2.5

以下补充到 v1.2.6 之前的主要变化；v1.2.0 以部署版 1.0.35 的已确认源码提交为起点，其余版本比较相邻标签。

源码版本	主要更新与使用影响
v1.2.5	供应商【更新模型】入口放入页头，按模型类型选择默认请求格式；模型使用统一编辑页，返回箭头回到模型列表；新增在线设备页、成员创建后的授权码及连接信息展示；修复角色切换和待激活成员设备计数。
v1.2.4	向已授权成员展示模型目录；企业方案原模板模型不可用时匹配可用项；资源拒绝时说明员工授权或网络限制；修复隐私规则对版本号、密钥、纯文本文件和策略缓存的处理。
v1.2.3	已保存的连接入口不依赖局域网发现，改善入口地址与慢发现处理；待激活成员的有效凭据可使用；升级前校验实际生效的更新源；供应商候选模型导入和软件权限操作给出更清楚的结果提示。
v1.2.2	改善表单标签布局与企业方案中已删除模型的提示；修复演练场诊断读取和升级结果的审计查询。
v1.2.1	日志服务启动观察阶段不再误报容量告警。
v1.2.0	相对本次部署起点，补充控制台资源连接与接入域名设置、模型图片输入能力与工具方案推理能力传递；改善不受限凭据编辑、成员选择和安装中断后的继续处理。此版本跨度较大，表内列出与手册操作直接相关的主要变化。

管理后台 main 尚未归入版本的变化

main 的 CHANGELOG 仍有【未发布】段，其中部分条目已经被后续标签包含，本页已按标签差异归类，不重复算作新功能。标签 v1.2.10 之后的 main 提交另有个人资料修改、新建 302.AI 供应商默认国内连接及 Claude 工具参数兼容修复；没有确认发布版本前，不将它们算入部署版 1.0.56。

Token路由客户端 0.1.20 测试版

版本	更新内容与使用影响
0.1.20-beta.3	修复 Windows 构建筛选器引号兼容问题，属于构建修复。
0.1.20-beta.2	桌面构建与命令行单文件构建分开，属于交付流程调整。
0.1.20-beta.1	新增命令行版本，提供登录、代理运行、环境、状态和更新命令；修复版本比较。Claude Desktop 未填写角色显示名称时显示实际模型名称，需重新应用方案并重启工具后生效。

本手册客户端截图适配 0.1.20-beta.6，但本次读取的 GitLab main CHANGELOG 仅单列到 beta.3，没有 beta.4～beta.6 的独立说明。 截图适配版本不能代替发行记录，这几个版本暂不编写推测性的优化清单。

Token路由客户端 0.1.19

源码发行说明将其标记为正式发布候选，平台验收完成前不切换正式推荐。

- 包含局域网和云端辅助发现、设置中的 Beta 更新开关；发现成功仍要求实际网络可达。
- 申请接入支持邮箱或手机号，审批后自动连接；失效连接码可重新申请。
- 新组织路由提示企业方案，切换前说明影响，可稍后处理。
- 完成的申请记录可删除后重新申请，保留已有路由与凭据；修复窗口失焦影响申请能力读取的问题。
- 发行说明仍列有 Windows 未签名及安全软件实机验收待确认事项。

0.1.19 测试版的主要变化

版本	更新内容
0.1.19-beta.4	扫描时增加云目录查询，辅助跨网段发现，并核对设备身份、合并重复入口；云查询仅在打开扫描列表或手动扫描时触发。
0.1.19-beta.3	设置增加【接收 Beta 更新】，默认关闭；兼容旧版发现公告。
0.1.19-beta.2	局域网持续发现，多网卡合并；系统发现服务不可用时给出提示，仍可手动连接；修复隐私状态提示。Linux 不含自动发现。

Token路由客户端 0.1.18

- 企业组合按实际配置内容判断是否与管理员方案一致；同名组合以组织区分。
- Codex 模型未声明推理档位时提供 low、medium、high，默认 medium。
- 汇入企业方案、组织申请、概览模型状态、软件下载引导及 Windows 安装检测和配置保存改进。

0.1.18 测试版的主要变化

版本	更新内容
0.1.18-beta.7	企业方案扩展到 14 类工具，修复协议、模型名称或能力变化未同步的问题；旧管理后台按支持的协议回退，工具目录不代表全部工具均完成实机调用验收。
0.1.18-beta.6	调整重启后的连接验收测试，不改变应用行为。
0.1.18-beta.5	修复 Windows 系统辅助程序的发布准备，不改变应用功能。
0.1.18-beta.4	新增接入申请进度、概览模型状态、企业软件下载及诊断 ZIP；改善连接领取和 Windows 操作。
0.1.18-beta.3	改善 Windows 配置保存与安装识别，探测失败时保留已有结果。
0.1.18-beta.2	改善技能仓库扫描、路由删除后的恢复引导、发现警告与 Codex 应用状态。

Token路由客户端 0.1.17

- 延长组织扫描的公开接口验证时间，改善瞬时故障与重叠子网漏发现。
- 接入错误提供可关联诊断日志的编号，便于定位失败阶段。
- 改善 WorkBuddy 应用配置后的模型选择，并提示需要手动切换的情况。

Token路由客户端 0.1.16 内部测试版

main CHANGELOG 没有单列 0.1.16 正式版，以下均为内部测试：

- 0.1.16-beta.13：修复多种工具的图片声明、接入错误提示、组合配置和安装识别；不兼容路由标明原因。
- 0.1.16-beta.12：统一工具与组合入口；保存配置定义后，仍需点击【应用】才写入工具；企业组合更新由员工确认，保留个人修改。

Token路由客户端 0.1.15

- 修复 macOS 更新检查缓存及安装记录目录权限不一致导致无法安装的问题。
- 保留企业模型推荐、软件模板应用及自定义配置保护。

查看更新依据

- 管理后台 main 更新日志
- Token路由客户端 main 更新日志

这些链接需要能够访问企业 GitLab。实际版本以目标实例和客户端的版本信息为准。