

星辰 MaaS 智能体平台

用户操作指南

天翼云科技有限公司

2026 年 6 月

目录

1.产品介绍	1
2.计费说明	3
2.1 计费说明	3
2.2 计费管理	4
3.快速入门	6
3.1 准备工作	6
4.用户指南	7
4.1 应用广场	7
4.2 工具广场	8
4.3 自主规划应用	11
4.4 workflow编排应用	19
4.5 版本及服务管理	29
4.6 发布渠道配置	31
4.7 日志管理	36
4.8 数据监测	40
4.9 创建与配置知识库	43
4.10 已有知识库-文件管理	46
4.11 文件切片管理	46

4.12 命中测试	48
4.13 知识库 API	49
4.14 记忆库	50
4.15 数据库	52
4.16 workflow工具	57
4.17 OPENAPI Schema 工具	60
4.18 MCP 工具	68
4.19 工具版本及发布管理	71
4.20 工具渠道管理	73
4.21 提示词模板管理	75
4.22 模型管理	77
4.23 评测功能	80
4.24 团队管理	89
5. 常见问题	91
5.1 常见 Q&A	91
6. 最佳实践	93
6.1 快速搭建一个自主规划应用-AI 问答智能体	93
6.2 搭建一个 workflow编排应用-客服客户画像智能体	105
7. API 参考	122

7.1 知识库 API 122

1. 产品介绍

星辰 MaaS 智能体平台是自主研发的大语言模型智能体应用开发平台，平台集成了构建智能体应用所需的核心技术，提供简单可靠的智能自动应用编排能力、以及灵活的工作流编排和管理功能、高效的 RAG 引擎和知识库管理工具，支持多种模型，并提供了用户友好的界面，以及丰富的智能体开发模板和工具，帮助开发者减少重复劳动，专注于创新和实际业务需求，可实现智能体应用快速创建和 API 服务化。

1.1 产品优势

星辰 MaaS 智能体平台是基于大模型搭建 AI 智能体应用创建平台，提供智能体应用搭建能力，以及内置视觉等传统 AI 组件，降低智能体应用的开发门槛，赋能开发者快速实现应用搭建。目前已在多个业务场景实现落地，效果业界领先。

- **技术领先：**央企首个开源基础大模型，基于万亿参数自研基础大模型以及视觉、语言大模型进行应用探索和落地验证多个场景业界领先；依托于电信研究院等机构多年研发的 AI 能力组件和数据处理能力，内置于平台底层作为技术引擎，应用表现更佳。
- **效果领先：**支持上传原始文档、问答对等多种形态知识，以及灵活可配的文档预处理以及增强策略，还可根据用户使用数据定向调优模型，保证高精度知识问答效果与精准的自主任务规划；
- **工具丰富：**内置多种的大语言模型、多模态模型、AI 能力和场景化工具以及支持第三方 API 接入。根据用户问题，大模型进行意图识别后自动调取合适的组件进行回答，能够满足不同场景的业务诉求，提升大模型赋能深度，适配性更强。
- **灵活易用：**应用开发流程简单高效，通过自主编排能力可快速完成应用的创建；使用极大提升业务效率；支持工作流编排和自定义工具，可灵活编排业务流程，将不同工具进行组合来实现复杂业务场景。

1.2 客户价值

- **客服领域：**在运营商客服场景中，可同时与多位客户顺畅交流。利用海量预置模型和原厂支持，快速准确回答客户问题。

- 企业办公：大型企业日常办公时，保障企业知识库、工具集安全。基于产业 B 端项目积累的特色工具集，助力多部门处理数据、分析报表、制定方案等，应对高并发办公需求，提高企业运营效率。
- 医疗行业：医院日常问诊中，能根据患者症状描述，结合医学知识库快速给出初步诊断建议。多智能体协同可实现不同科室智能体交流，为复杂病情提供综合诊疗方案。同时，安全权限管理保障患者隐私数据安全。
- 金融服务：通过前沿 RAG 策略，快速检索金融产品信息、市场动态等知识，精准解答客户投资、理财咨询。本地化应用开发服务可根据不同地区金融政策和客户需求定制功能，提升客户服务体验。

2.计费说明

2.1 计费说明

计费模式

星辰 MaaS 智能体平台采用“权益套餐+资源包”的混合计费模式，同时提供免费试用版与完整权益的付费版套餐。

SaaS 套餐

模块	体验版	企业版
价格	0 元/年	6.99 万元/年
资源点	<ul style="list-style-type: none"> ● 每天获得 700 资源点。 ● 每日 0 点重置。 	<ul style="list-style-type: none"> ● 每年一次性获得 25,000,000 资源点。 ● 支持额外购买资源点。
团队	<ul style="list-style-type: none"> ● 1 个团队空间空间。 ● 人数上限为 5 人。 ● 新用户免费体验 30 天。 	<ul style="list-style-type: none"> ● 不限制团队空间数量。 ● 每个团队空间人数上限为 2000 人。
应用	<ul style="list-style-type: none"> ● 应用创建上限为 30 个。 ● API 调用支持 1QPS，按照实际调用量扣减资源点。 	<ul style="list-style-type: none"> ● 应用创建无上限。 ● API 调用支持 10QPS，按照实际调用量扣减资源点。
知识库	<ul style="list-style-type: none"> ● 总空间 1G。 ● 知识库创建上限为 30 个。 	<ul style="list-style-type: none"> ● 总空间 200G。 ● 知识库创建无上限。
工具	<ul style="list-style-type: none"> ● 工具创建上限为 30 个。 	<ul style="list-style-type: none"> ● 工具创建无上限。
模型资源	<ul style="list-style-type: none"> ● 支持调用内置星辰大语言等模型，按照实际调用量扣减资源点。 	<ul style="list-style-type: none"> ● 支持调用内置星辰大语言等模型，按照实际调用量扣减资源点。 ● 支持接入自定义模型。
数据库	<ul style="list-style-type: none"> ● 数据库创建上限为 10 个。 	<ul style="list-style-type: none"> ● 数据库创建上限 100 个。
记忆库	<ul style="list-style-type: none"> ● 记忆库创建上限为 30 个。 	<ul style="list-style-type: none"> ● 记忆库创建无上限。

注意

SaaS 套餐与账号绑定，即套餐内的资源点与功能权益，将由主账号名下的所有个人及团队工作空间共享。

资源包

资源包	价格	有限期	是否支持叠加
10 万资源点	100 元	购买日起 12 个月内	是

资源点抵扣范围

- 资源点可用于抵扣调用官方内置模型消耗 token。
- 后续可用于知识库空间、API 接口并发量等功能扩容。

说明

1 元=1000 资源点。

注意

本平台已开通的付费套餐、资源包一经购买生效，不支持中途退订、部分退款、到期前退费。

2.2 计费管理

为了让大家更轻松地体验我们的服务，平台现已推出 SaaS 体验版套餐，并为用户提供：

- **【新用户专享】30 天免费团队空间**：所有用户均可免费体验 30 天团队工作空间，轻松开启智能协作之旅。
- **每日免费 700 资源点**：每天自动发放 700 个资源点，可用于抵扣大模型 token 消耗，让您尽情探索 AI 的无限可能。
- **便捷管理，透明消费**：团队所有者可通过右下角的“套餐卡片”进入付费管理页面，随时查看各工作空间的功能用量及资源点消耗，让每一分资源都用得明明白白。

平台账号信息上方新增套餐卡片，您可通过卡片查看当前订阅套餐及资源点使用额度。



我的套餐

如果您登录的是主账号，可以点击“套餐卡片”查看套餐内的详细情况。

- 资源点额度：实时展示所有工作空间的资源点消耗情况。
- 功能权益：可查看团队空间的使用额度，以及所有工作空间在应用、知识库等功能的限额与使用情况。

套餐资费

您可在此查询资源点的计费标准。资源点用于抵扣大模型 Token 消耗，不同大模型的计费标准不同，详见下表。

- 计费单位：1 资源点 = 0.001 元，为本平台最小计费单位。
- 计费方式：采用向上取整规则。当单次计费不足 1 资源点时，按 1 资源点扣除。

3.快速入门

3.1 准备工作

本文介绍星辰 MaaS 智能体平台使用前的准备工作。

注册天翼云账号

在开通和使用星辰 MaaS 智能体平台之前，您需要先注册天翼云账号。

1. 注册账号，请参考账号中心-[注册天翼云账号-账号中心-操作指南](#) - 天翼云
2. 如需实名认证，请参考账号中心-[中国电信天翼云-管理中心](#)

4.用户指南

4.1 应用广场

应用广场是星辰 MaaS 智能体平台（以下简称“平台”）提供的用户与应用创建者的交流和分享平台。用户可以探索发现喜欢的应用，体验后使用创建同款功能复制到个人的应用空间。而创建者可以将自己搭建的应用等资源发布到广场共享给更多的人使用。

应用广场在平台左侧导航栏中，点击应用广场即可进入应用广场。在应用广场中，你可以：

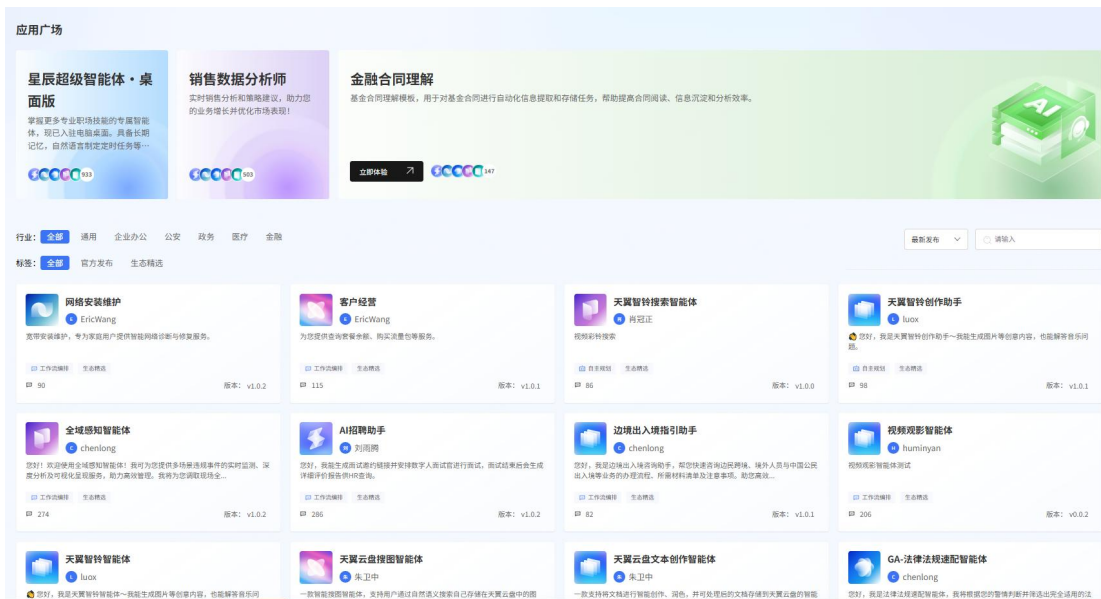
查找智能体应用

根据行业和标签进行智能体应用筛选，可以通过最新发布或最受欢迎程度进行排序，也可以直接通过关键词进行搜索。



体验智能体应用

选择你喜欢的智能体应用后点击立即体验按钮，即可与智能体应用对话和其他深度体验。



创建同款智能体应用

创建同款至本平台：应用创建者设置了公开复制权限的应用，点击右上角的"创建同款"按钮，选择至本平台，点击后会复制当前应用到我的应用中，复制的内容包括简介、提示词、工作流节点设置等基础信息。如果应用创建者在选择权限时，选择知识库和工具都支持同步，则知识库和私人工具也支持同步至复制者的空间；如果应用创建者选择的是不支持同步，则复制的应用里面无私人知识库和工具。



4.2 工具广场

工具广场是星辰 MaaS 智能体平台（以下简称“平台”）提供的用户与工具创建者的交流和分享平台。平台预置官方工具，用户可以探索发现喜欢的工具，支持用户创建应用或工作流工具时添加工具广场的工具。而创建者可以将自己搭建的工作流工具发布到广场共享给更多的人使用。

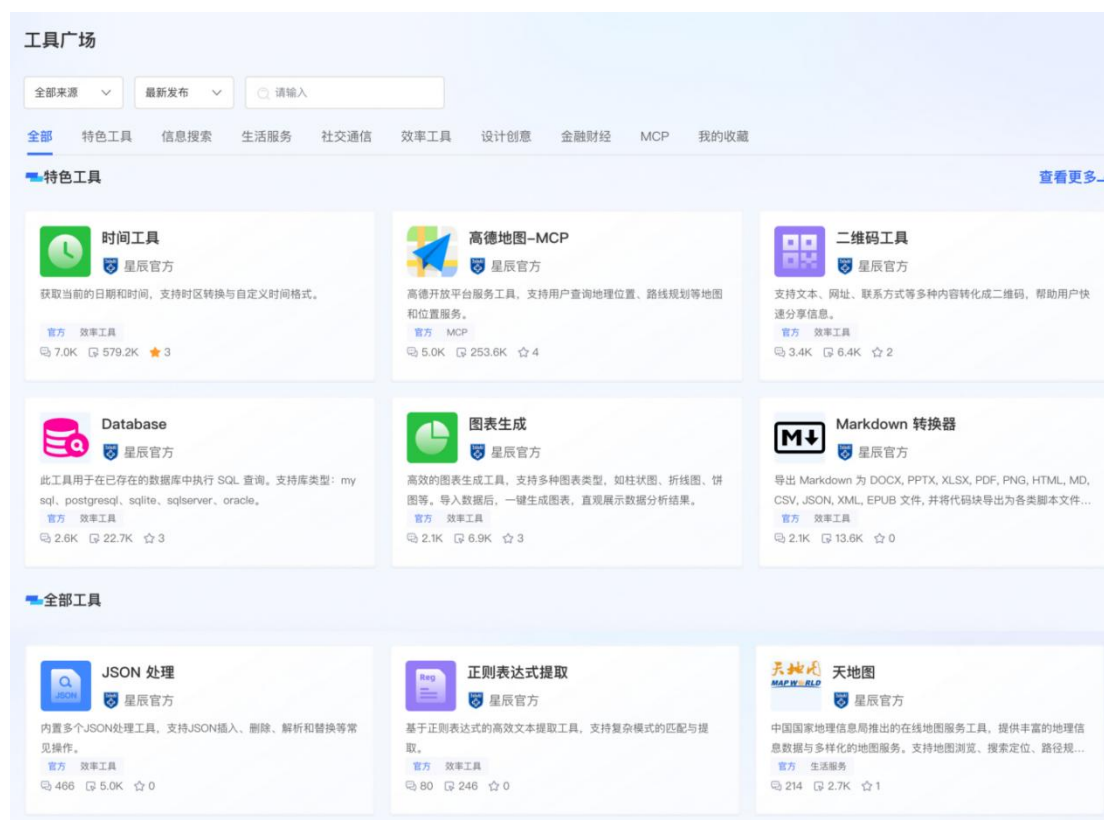
查找工具

支持两种快速查找方式，精准定位所需工具。

- 分类筛选：按全部、特色工具、信息搜索、生活服务、社交通信、效率工具、设计创意、金融财经分类进行工具筛选。
- 关键词搜索：在页面搜索框输入关键词直接检索目标工具。

特色工具

展示平台热门特色工具，点击查看更多可浏览全部精选工具。



The screenshot shows the 'Tool Plaza' (工具广场) interface. At the top, there are filters for 'All Sources' (全部来源) and 'Latest Release' (最新发布), along with a search input field. Below the filters, there are tabs for different categories: 'All' (全部), 'Featured Tools' (特色工具), 'Information Search' (信息搜索), 'Life Services' (生活服务), 'Social Communication' (社交通信), 'Efficiency Tools' (效率工具), 'Design Creativity' (设计创意), 'Financial & Finance' (金融财经), 'MCP', and 'My Favorites' (我的收藏). The 'Featured Tools' (特色工具) section is active, displaying a grid of tool cards. Each card includes an icon, the tool name, the provider (e.g., '星辰官方'), a brief description, and statistics like 'Official' (官方), 'Efficiency Tool' (效率工具), and user counts (e.g., 7.0K, 579.2K, 3 stars).

我的收藏

已收藏的工具，将统一收纳在我的收藏页面，方便快速使用。



工具详情页查看：

点击任意工具卡片，即可进入工具详情页，查看完整信息：

- 工具基础信息：工具来源、工具数量、智能体使用量、调用量、成功率、平均耗时等。
- 工具描述：清晰展示工具核心用途与能力。
- 子工具列表：呈现该工具集下所有子工具的完整清单及信息。
- 快捷操作：支持收藏工具、添加到我的智能体等便捷操作。



工具授权

平台预置部分内置工具需要授权使用。用户可在搭建应用时进行授权，工具在同空间授权过一次后即可使用，可在其他应用或工具中使用，无需二次授权。



4.3 自主规划应用

创建应用

- 登录[星辰 MaaS 智能体平台](#)（以下简称“平台”）。
- 在页面左上角的创建应用按钮或我的应用页面中右侧的创建应用按钮。
- 必须输入应用名称和应用描述、选填标签内容，使用内置头像或单击头像图标上传自定义头像。
- 支持用户自定义标签(应用模块标签上限 10 个)，并在创建应用的时候选择标签。
- 在左侧智能体配置区域，设置智能体的人设与编排逻辑、为智能体添加工具、知识库和添加对话配置等，并在右侧预览区域通过与智能体对话查看并调试效果。

角色指令

- 通过角色指令为智能体设定身份、任务和回复逻辑等信息，大模型会根据角色身份设定和任务进行理解和回复。
- 优化：当用户未填写角色指令时，支持根据应用名称和描述生成角色指令；当用户填写角色指令后，支持根据底层策略优化生成角色指令；同时也支持根据用户输入的优化点修改角色指令。
- 导入模板：支持用户使用平台预置模板，适合不会编写模板的用户，可以根据适合的场景选择适合的模板，并在模板基础上进行微调，降低用户的操作复杂度；支持用户使用
- 创建为模板：支持用户将调试角色指令创建为模板，供其他智能体使用

角色指令 ? 优化 模板

工具能力

1. 差旅政策知识库
你需要拥有公司差旅政策的完整知识库，并能够迅速查找和解答用户关于政策的具体问题。

2. 展会信息
你需要具备关于即将举行的线下展会的详细信息，包括展会的时间、地点、参展商、活动内容等。

3. 二维码生成
你可以使用qrcode_generator生成二维码

4. 航班信息查询
你需要具备查询航班信息的能力，以使用户能够了解航班的状态和相关信息。

要求与限制

知识库

- 知识库内存储专业领域的知识或内部信息，用于提高智能体应用在特定场景下回复的准确性。
- 点击知识库右侧加号可以添加已完成创建的知识库。知识库创建方法详见知识库章节。

知识库 ? 召回设置 +

- 召回设置：针对整个应用的知识库进行配置，配置中包含混合检索、全文检索、语义检索、参数设置及引用归属设置。

命中设置

混合检索
混合检索是语义检索与全文检索的融合，提高检索结果的质…

全文检索
全文检索处理文本，分词建索引，快搜关键词，排序相关文…

语义检索
通过匹配文本的内容相似度，检索文本分段。

参数设置

最大召回片段数

结果重排

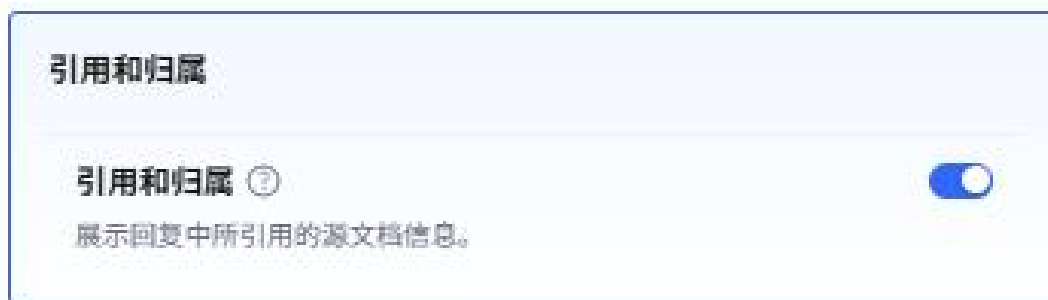
模型选择

相似度匹配值

引用和归属

引用和归属
展示回复中所引用的源文档信息。

- 引用和归属：用于配合知识库功能共同使用，显示大模型根据知识库内的内容回复时调用知识库数据的源文档。



数据库

- 可上传本地表格数据构建数据库。用户询问数据表内的问题时，应用能够查询数据并答复。
- 应用最多可关联 1 个数据库。
- 点击数据库右侧加号可以添加已创建的数据库。数据库创建方法详见数据库章节。



工具

- 点击工具右侧的加号添加工具，不同工具可以拓展智能体能力边界。例如，在您的智能体内添加天气搜索插件，那么您的智能体将拥有搜索天气的能力。部分内置工具导入时需要经过工具方的授权并添加 API Key。
- 如果平台内置的工具不满足需求，你还可以创建并添加自定义工具来集成其他能力。自定义工具创建方法详见工具章节。



高级对话配置

- 开场白：用户进入智能体应用后自动展示的引导信息，帮助用户理解智能体的能力和用途。在开启开场白后还可以设置预置问题，用户点击后可快速开始智能体对话。

▼ 开场白 ① 🔴

开场白文案

在这里写下你的开场白。

预设问题0/5 +

- 追问：在智能体回复后，大模型会自动根据对话内容提供 3 条用户提问建议，帮助用户继续与智能体进行多轮对话。

追问 ① 🔴

- 会话背景图：用户可添加会话背景图，美化聊天背景，调试过程中会展示在预览调试区，发布后同步展示于应用广场或前端页面。

> 会话背景图 ① +

- 在预览调试区域，向智能体发送消息，测试智能体效果。同时可以针对回复的内容，进行点赞或点踩。

预览与调试



世界人工智能大会参会助手

你好，欢迎来到展会助手服务。无论你需要了解公司的差旅政策还是查询线下展会的详细信息，甚至是查询天气和航班信息，我都能为你提供帮助。请告诉我你的需求，我会尽力满足你的要求。

距离大会的开会日期还有多少天？

2024世界人工智能大会举办时间？

2024世界人工智能大会周边停车场和交通方式推荐。



输入问题，发送Enter



内容由AI生成，无法确保真实准确，仅供参考

在该区域板块内，用户可选择对应的模型并进行相关参数配置。

模型选择

- 支持选择内置模型和第三方模型进行对话。



- 生成多样性配置

该参数用于控制随机性和多样性的程度，具体来说，生成多样性参数值控制了生成文本时对每个候选词概率的分布进行平滑的程度。较高的参数值会降低概率分布的峰值，使得更多的低概率词被选择，生成结果更加多样化；而较低的参数值则会增强概率分布的峰值，使得高概率词更容易被选择，生成结果更加确定。

- 参数模式

- 常规模式：也可称为标准模式或平衡模式，是大模型默认的运行方式。它兼顾创意与准确性，试图在提供足够信息和保持对话语气之间取得平衡。适合大多数日常交流场景，如日常沟通、信息查询等，能快速提供一些较为合理的答案和建议。
- 精准模式：注重事实和准确性，像一个严谨的专家，专注于提供准确、可靠的信息。该模式下模型会严格按照指令执行任务，尽可能减少歧义，输出结果具有较高的确定性和可预

测性。适用于专业咨询、数据分析、重要决策等需要精确结果的场景。

- 创意模式：旨在激发灵感，让模型尽可能地生成富有想象力、原创性强的内容。模型输出会更加自由和多样化，可能会偏离主题或产生一些非常规的想法。适合用于头脑风暴、创意写作、娱乐消遣等场景，能为用户提供一些意想不到的灵感和素材。
- 自定义模式：允许用户根据自身需求对模型的输出方式进行个性化设置。用户可以通过上传文本样本，让模型学习特定的语言风格，也可以直接描述期望的交互方式，如指定语言风格、回答长度、格式要求等，使模型的输出无限贴近个人写作风格和需求。
- 高级参数配置
 - 平台支持对以下参数根据用户使用场景需要做自定义设置，不同的模型存在不同的参数值，以下仅做示例。
 - 最大标记：用于指定模型在生成内容时 token 的最大数量，它定义了生成的上限，但不保证每次都会生成到这个数量。
 - Top P：生成过程中核采样方法概率阈值，例如，取值为 0.8 时，仅保留概率加起来大于等于 0.8 的最可能 token 的最小集合作为候选集。取值范围为 (1,1.0)，取值越大，生成的随机性越高；取值越低，生成的确定性越高。
 - 取样数量：生成时，采样候选集的大小。例如，取值为 50 时，仅将单词生成中得分最高的 50 个 token 组成随机采样的候选集。取值越大，生成的随机性越高；取值越小，生成的确定性越高。
 - 随机种子：生成时使用的随机数种子，用户控制模型生成内容的随机性。支持无符号 64 位整数，默认值为 1234。在使用 seed 时，模型将进坑生成相同或相似的结果，但目前不保证每次生成的结果完全相同
 - 重复惩罚：用于控制模型生成时的重复度。提高 repetition_penalty 时可以降低模型生成的重复度。1.0 表示不做惩罚。
 - 回复格式：设置一个返回格式，确保 LLM 的输出尽可能是有效的代码块，如 JSON、XML 等。
- 多模型对比
 - 为了用户可以更直观的对比同一角色指令和能力拓展配置下，不同模型的运行效果，帮助用户更快的选择最优模型，平台提供了多模型对比功能。
 - 支持最多选择 4 个模型进行对比，基于相同的问题，输入，关注模型的输出，对比不同的

模型输出的效果差异度。确认最优模型后，点击“最优模型”按钮开关，保留最优模型。



4.4 工作流编排应用

工作流是规划和实施复杂功能逻辑的有用工具，通常用于规划和报告分析等多步骤任务。对于功能复杂、稳定性要求高的任务，通过工作流实施是理想的选择。工作流提供了广泛的灵活、可组合的节点，甚至支持自定义代码，以处理复杂且可控的 AI 任务流需求。当目标 AI 任务场景涉及较多步骤，且对输出精度和格式有严格要求时，适合使用工作流。

创建应用

1. 登录星辰 MaaS 智能体平台（以下简称“平台”）。
2. 在页面左上角的创建应用按钮或我的应用页面中右侧的创建应用按钮后，点击创建工作流编排/自主规划应用。
3. 必须输入应用名称和应用描述简介、选填标签内容，使用内置头像或单击头像图标上传自定义头像。
4. 支持用户针对已创建的应用或创建应用时添加标签，方便用户管理自己搭建的应用。

节点介绍

基础能力节点

开始节点

开始节点作为每个工作流应用必备的预设节点，是工作流执行流程的起点，其主要作用在于为后续工作流节点以及应用的正式调用流转提供必要的初始信息，例如应用使用者所输入的具体内容、上传的文件资料以及其他必要的基础参数等。

在开始节点的设置页，你可以看到两部分设置，分别是“输入字段”和预设的系统变量。

- **输入字段：**输入字段功能由应用开发者设置，通常用于让应用使用者主动补全更多信息。例如在周报应用中要求使用者按照格式预先提供更多背景信息，如姓名、工作日期区间、工作详情等。这些前置信息将有助于 **大模型** 生成质量更高的答复。
- **配置完成后，**用户在使用应用前将按照输入项指引，向 **大模型** 提供必要信息。更多的信息将有助于 **大模型** 提升问答效率。
- **系统变量：**系统变量指的是在应用内预设的系统级参数，可以被应用内的其它节点全局读取。通常用于进阶开发场景，例如搭建多轮次对话应用、收集应用日志与监控、记录不同应用和用户的使用行为等。

大模型节点

大模型节点是核心节点。该节点能够利用大语言模型的对话/生成/分类/处理等能力，根据给定的提示词处理广泛的类型任务，并能够在工作流的不同环节使用。

- 示例如下：
 - **意图识别：**在客服对话情景中，对用户问题进行意图识别和分类，导向下游不同的流程。
 - **文本生成：**在文章生成情景中，作为内容生成的节点，根据主题、关键词生成符合的文本内容。
 - **内容分类：**在评论批量处理情景中，对评论的类型进行自动化分类，如投诉/垃圾评论/物流评论/质量评论等。
 - **代码生成：**在辅助编程情景中，根据用户的要求生成指定的业务代码，编写测试用例。
 - **RAG：**在知识库问答情景中，将检索到的相关知识和用户问题重新组织回复问题。
 - **图片理解：**使用具有视觉能力的多模态模型，能对图像内的信息进行理解和问答。
 - **指令优化：**支持针对用户输入的系统提示词进行自动优化或指定方向优化。

回复节点

回复节点用于展示其他节点输出的内容至用户对话页面。支持自定义回复格式，包括增加特定文本或变量。支持纯文本和图文混排输出。多个回复节点会根据节点链接的顺序依次输出。

问题分类节点

通过定义分类描述，问题分类器能够根据用户输入，使用大模型推理与之相匹配的分类并输出分类结果，向下游节点提供更加精确的信息。常见的使用情景包括客服对话意图分类、产品评价分类等。

- 配置步骤：
 - 选择输入变量，指用于分类的输入内容，支持输入文件变量。
 - 选择推理模型，问题分类器基于大语言模型的自然语言分类和推理能力，选择合适的模型将有助于提升分类效果；
 - 编写分类标签/描述，你可以手动添加多个分类，通过编写分类的关键词或者描述语句，让大语言模型更好的理解分类依据。
 - 选择分类对应的下游节点，问题分类节点完成分类之后，可以根据分类与下游节点的关系选择后续的流程路径。
- 高级设置：
 - 指令：你可以在 高级设置-指令 里补充附加指令，例如给大模型提供更详细的分类依据信息，以增强问题分类器的分类能力。
 - 记忆：开启记忆后问题分类器的每次输入将包含对话中的聊天历史，以帮助 大模型 理解上文，提高对话交互中的问题理解能力。
 - 图片分析：仅适用于具备图片识别能力的大模型，允许输入图片变量。记忆窗口：记忆窗口开关关闭时，系统会根据模型上下文窗口动态过滤聊天历史的传递数量；开关打开时用户可以精确控制聊天历史的传递数量（对数）。
 - 输出变量：class_name，即分类之后输出的分类名。你可以在下游节点需要时使用分类结果变量。

逻辑处理节点

条件节点

设置判断条件变量后，根据 If/else/elif 条件将流程拆分成多个分支进行执行。

迭代节点

- 对数组中的元素依次执行相同的操作步骤，直至输出所有结果，可以理解为任务批处理器。迭代节点通常配合数组变量使用。
- 使用迭代的条件是确保输入值已格式化为列表对象；迭代节点将依次处理迭代开始节点数组变量内的所有元素，每个元素遵循相同的处理步骤，每轮处理被称为一个迭代，最终输出处理结果。
- 迭代节点的结构通常包含输入变量、迭代 workflow、输出变量三个功能单元。
- 输入变量和输出变量：仅接受 Array 数组变量类型数据。

循环节点

平台循环节点专注于处理重复性 workflow，核心优势在于每一次循环的执行，均以其上一次循环的输出结果为基础进行构建与推进，形成连贯且持续优化的工作链路。

- 循环变量：

定义在循环迭代中持续存在，并在循环完成后仍可访问的变量。此类变量用于维护循环过程中的状态，是实现渐进式 workflow 的核心支撑。

- 终止条件：

当满足循环终止条件、达到最大循环次数，或执行至退出循环节点时，循环将终止运行。未明确指定循环终止条件时，循环将持续执行至达到预设最大循环次数后停止。

- 用于配置循环的停止执行规则，具体包含以下三类：

- 循环终止条件：通过设定表达式确定循环退出时机（例如：`count > 5`）；
- 最大循环次数：作为防止无限循环的安全限制，避免流程陷入死循环；
- 循环结束节点：当流程执行至该节点时，立即终止循环操作。

知识与记忆节点

知识库节点

从知识库中检索与用户问题相关的文本内容，通常作为下游大模型节点的上下文来使用。常用于构建知识问答系统（RAG）。

- 配置步骤：

- 选择查询变量。查询变量通常代表用户输入的问题，该变量可以作为输入项并检索知识库中的相关文本分段。知识库所能接受的最大查询内容为 200 字符。

- 选择需要查询的知识库，知识库需要在知识库内预先创建；
- 连接并配置下游节点，一般连接大模型节点进行检索内容的处理。
- 输出变量：
 - 知识检索的输出变量 result 为从知识库中检索到的相关文本分段。其变量数据结构中包含了分段内容、标题、链接信息。
- 上下文变量：
 - 大模型节点内定义的特殊变量类型，用于在提示词内插入外部检索的文本内容。当用户提问时，若在知识检索中召回了相关文本，文本内容会作为上下文变量中的值填入提示词，提供大模型回复问题；若未在知识库检索中召回相关的文本，上下文变量值为空，大模型则会直接回复用户问题。
 - 针对节点添加的知识库进行统一召回设置，包含：混合检索、全文检索、语义检索、参数配置及引用归属设置。

变量赋值节点

变量赋值节点用于向可写入变量进行变量赋值。通过变量赋值节点，你可以将工作流内的变量赋值到会话变量中用于临时存储，并可以在后续对话中持续引用。

- 使用变量赋值节点
- 设置变量：
 - 赋值的变量：选择被赋值变量。
 - 设置变量：选择需要赋值的变量。
- 写入模式：
 - 覆盖：将源变量的内容覆盖至目标会话变量
 - 追加：追加内容到指定变量，变量为 Array 类型时可选
 - 清空：清空目标会话变量中的内容

数据库节点

数据库节点支持通过自然语言指令和 SQL 指令两种方式在工作流中实现数据访问，在数据库节点中添加想要使用的数据库来查询，数据库会返回结构化的数据库查询结果，可以通过在数据库节点后面加入大模型节点来灵活的使用输出结果。

- 配置步骤：

- 选择需要查询的数据库，数据库需要在数据库列表页面内进行创建。
- 选择查询变量，可以通过引用变量使用自然语言转 SQL 的方式来查询数据库，也可以通过直接输入 SQL 指令来对数据库进行查询。
- 连接并配置下游节点
- 输出变量：outputList (String)：查询结果
- rowNumber (number)：受影响的行数

信息处理节点

模板转换节点

- 支持使用 Jinja2 模板语言灵活地进行数据转换、文本处理等。适用于文本处理、JSON 转换等情景。
- 例如灵活地格式化并合并来自前面步骤的变量，创建出单一的文本输出。这非常适合于将多个数据源的信息汇总成一个特定格式，满足后续步骤的需求。

列表操作节点

- 文件列表变量支持同时上传文档文件、图片、音频与视频文件等多种文件。应用使用者在上传文件时，所有文件都存储在同一个 Array[File] 数组类型变量内，不利于后续单独处理文件。列表操作节点可以在数组变量内提取单独的元素，便于后续节点处理。
- 列表操作节点一般用于提取数组变量中的信息，通过设置条件将其转化为能够被下游节点所接受的变量类型。它的结构分为输入变量、过滤条件、排序（可选）、取前 N 项（可选）、输出变量。例如在一个应用中，允许用户同时上传文档文件和图片文件两种不同类型的文件。需要使用列表操作节点进行分拣，将不同的文件类型交由不同流程处理。

变量聚合节点

- 变量聚合节点负责整合不同分支的输出结果，确保无论哪个分支被执行，其结果都能通过一个统一的变量来引用和访问。这在多分支的情况下，可将不同分支下相同作用的变量映射为一个输出变量，避免下游节点重复定义。
- 聚合分组：开启聚合分组后，变量聚合器可以聚合多组变量，各组内聚合时要求同一种数据类型。

文档提取节点

- 大模型节点自身无法直接读取或解释文档的内容。因此需要将用户上传的文档，通过文档提取

器节点解析并读取文档文件中的信息，转化文本之后再将内容传给大模型以实现对于文件内容的处理。构建能够与文件进行互动的大模型应用。

- 文档提取器支持单个或多个文件的输入。目前仅能够提取文档类型文件中的信息，例如 TXT、Markdown、PDF、DOCX 等 格式文件的内容，无法处理音频、视频等格式文件。
- 开启视觉解析开关后，通过调用底层能力，实现内容的读取，包括 jpg、jpeg、png、pdf。
- 输出变量为 text。输出的变量类型取决于输入变量：输入变量为 File 时，输出变量为 string；输入变量为 Array[File]时，输出变量为 array[string]。Array 数组变量一般需配合列表操作节点使用。

参数提取节点

利用大模型从自然语言推理并提取结构化参数，用于后置的工具调用或 HTTP 请求。

- 配置步骤
 1. 选择输入变量，一般为用于提取参数的变量输入。输入变量支持 file。
 2. 选择模型，参数提取器的提取依靠的是 大模型 的推理和结构化生成能力。
 3. 定义提取参数，可以手动添加需要提取的参数，也可以从已有工具中快捷导入。
 4. 编写指令，在提取复杂的参数时，编写示例可以帮助 大模型 提升生成的效果和稳定性。
- 高级设置
 - 记忆：开启记忆后问题分类器的每次输入将包含对话中的聊天历史，以帮助 大模型 理解上文，提高对话交互中的问题理解能力。

信息收集节点

工作流中的信息收集节点，通过提供选项向用户发起提问，收集所需用户信息并同步给出反馈，收集到的信息可供后续节点直接引用和处理。

表单配置添加参数时，支持选择的组件类型包括：文本框、JSON 文本框、密码框、选项框、文件上传。

进阶开发节点

代码节点

代码节点支持运行 Python / Javascript 代码以在工作流程中执行数据转换。该节点极大地增强了开发人员的灵活性，使他们能够在工作流程中嵌入自定义的 Python 或 Javascript 脚本，并以预设节点无法达到的方式操作变量。通过配置选项，你可以指明所需的输入和输出变量，并撰写相应

的执行代码。

- 如果您需要在代码节点中使用其他节点的变量，您需要在输入变量中定义变量名，并引用这些变量。
- 支持开启错误重试功能后，代码节点发生异常情况时，可按照预设策略自动重试，提升 workflow 运行成功率。

智能体节点

- 支持选择已发布的智能体或现创建智能体，实现简单的多智能体协同系统，赋能复杂业务场景的多智能体落地。
- 模型：选择大语言模型。
- 工具：工具的使用方式由模型策略定义，点击“+”添加并配置智能体节点可调用的工具。
- 知识库：知识库的使用方式由模型策略定义，点击“+”添加并配置智能体节点可调用的知识库，同时支持针对该节点添加的知识库进行统一召回参数设置。
- 对话示例：多轮对话模式专为提示词工程高阶需求的专业用户设计，支持通过系统提示词（System）+ 多组 User-Assistant 示例对话组合构建复杂指令。您可借助示例对话向模型明确任务逻辑、输出格式、推理流程等，快速实现小样本学习（Few-shot）、角色扮演、格式强约束、多轮推理等高级能力。该模式具备更强的生成可控性与场景适配灵活性，适合对输出结果有严格规范或任务逻辑复杂的专业应用场景。
- 数据库：支持上传本地表格数据以构建数据库。当用户查询涉及数据表内容的问题时，应用可检索对应数据并进行解答，单应用最多关联 1 个数据库。
- 记忆：开启记忆后问题分类器的每次输入将包含对话中的聊天历史，以帮助大模型理解上文，提高对话交互中的问题理解能力。
- 记忆窗口：记忆窗口开关关闭时，系统会根据模型上下文窗口动态过滤聊天历史的传递数量；开关打开时用户可以精确控制聊天历史的传递数量。

HTTP 请求节点

- 支持通过 HTTP 协议发送服务器请求，适用于获取外部数据、调用 API 等情景。
- 节点支持常见的 HTTP 请求方法，可以通过配置 HTTP 请求的包括 URL、请求头、查询参数、请求体内容以及认证信息等。
- HTTP 节点支持通过 HTTP 协议向服务器发送请求，该节点广泛适用于获取外部数据、调用第三方 API 接口等多种场景。其内置对常见 HTTP 请求方法的支持，用户可通过详细配置来定制请求细节，包括目标 URL、自定义请求头、查询参数、请求体内容（如 JSON、表单数据

等) 以及必要的身份认证信息 (如 Basic Auth、Token 认证等)。

- 此外, 该节点还具备错误重试机制。当开启此功能后, 在请求过程中若遇到网络异常、服务器无响应等临时性故障时, 节点将根据预设的重试策略 (如重试次数、间隔时间等) 自动进行重试操作, 从而有效提升整个工作流的运行稳定性和成功率。

工具节点

“工具”节点可以为工作流提供强大的第三方能力支持, 分为以下两种类型:

- 内置工具: 平台提供的内置工具。部分工具使用前可能需要进入工具提供方官网进行授权。
- 自定义工具: 如果内置工具无法满足使用需求, 你可以在左侧导航 - 工具内创建自定义工具。自定义工具创建方法详见工具章节。

点击添加节点按钮, 切换至 “工具” tab 页, 选择需要的工具后, 在工具节点中设置输入参数。

配置应用

功能拓展介绍

点击应用编排页面右上角的功能按钮, 可进入高级功能页。功能包含文件上传、引用归属、开场白、追问、环境变量、会话变量、会话背景图、DSL 管理。

文件上传

- 开启后支持用户在对话中上传文件。
- 平台支持的上传文件的类型、最大上传数、支持的文件类型进行灵活配置。平台具备主流类型文件上传和快速识别的能力。

开场白

用户进入智能体应用后展示的引导信息。用于帮助用户理解智能体的用途和使用方式。

预设问题

支持配置推荐给用户的预设问题, 最多配置 5 个。增加后, 点击提前定义好的预设问题, 可直接与 ai 对话。

追问

在对话完成后, 基于当前对话的内容展示建议问题。

环境变量

- 环境变量用于保护工作流内所涉及的敏感信息，例如运行涉及的 API 密钥、数据库密码等。
- 环境变量支持添加三种数据类型：String 字符串、Number 数字、Secret 密钥。可在大部分节点内全局引用；环境变量命名不可重复；类型为只读变量，不可写入。

会话变量

会话变量允许应用开发者指定需要被临时存储的特定信息，并确保当前工作流内的多轮对话内都能够引用该信息。

- 会话变量可在大部分节点内全局引用；
- 会话变量的写入需要使用变量赋值节点，类型为可读写变量。

会话背景图

用户可添加会话背景图，美化聊天背景，调试过程中会展示在预览调试区，发布后同步展示于应用广场或前端页面。

DSL 管理：

- 在平台内创建的工作流应用支持以 DSL 格式进行导出，你可以自由地将配置文件导入至平台内部。
 - ◆ DSL 文件不包含自定义工具节点内已填写的授权信息，例如第三方服务的 API Key；如果环境变量中包含 Secret 类型变量，导出文件时将提示是否允许导出该敏感信息。导出中不包含私人知识库和私人工具信息。
- 在生态 dify 平台搭建的工作流应用支持以 dsl 格式进行导入。
 - 兼容约束：
 - 仅支持 Dify 官方原生标准版 DSL，不支持加密文件、压缩包、第三方魔改扩展 DSL。
 - 兼容 DSL 版本，高版本文件平台自动降级兼容解析。
 - 原 Dify 外接第三方模型，导入后需手动切换为平台模型资源。
 - 导入前置自检清单
 - Dify 端工作流已调试完毕，无悬空节点、无流程死循环、分支逻辑闭环。
 - DSL 内无明文敏感信息硬编码。
 - 本地准备文本编辑器，用于 YAML 语法校验、DSL 内容预处理。

➤ 平台内已提前完成模型接入、知识库、工具资源初始化

- 自动解析规则
 - 流程拓扑 1:1 还原：开始节点、结束节点、流转链路、分支结构完整还原。
 - 全局变量、上下文变量自动透传至星辰平台变量池。
 - 提示词模板、条件表达式、参数配置自动适配转换。
 - 兼容节点自动映射，不兼容节点自动标红提示。

预览

- 用户针对搭建好的工作流程进行针对性的测试与询问，测试搭建的工作流是否通顺，是否存在异常点。在预览与调试区域，向智能体发送消息，测试智能体效果。
- 在「运行历史」中可以查看当前工作流历史调试的运行结果和日志信息。

4.5 版本及服务管理

自主规划应用/ workflow 编排应用：

- 存在多版本、多服务概念
- 支持用户创建 50 个版本、5 个服务
- 具备创建版本、发布服务、切换版本功能
- 支持发布到应用广场、后端 api、嵌入网站、前端页面

版本管理

相关概念

关键概念

- 版本：版本为智能体的编排的快照，用来记录某个时刻的智能体编排状态。
- 应用服务：部署后的智能体在线服务。
- 发布渠道：应用服务的访问通路。
- 版本是建筑施工图；应用服务是根据建筑施工图盖好的大厦；发布渠道则是进入大厦的几个不同入口，终端用户可以从不同入口进入，但访问的是同一座大厦。

其他概念

- 当前草稿：您所在的处于可编辑的智能体编排界面均为草稿版本
- 创建版本：当草稿修改到一定阶段，可保存的独立版本

版本相关功能

创建版本

- 创建版本的入口在编排界面的右上角“创建版本”按钮。
- 填写版本信息：



创建新版本

* 版本号

例如：v1.0.0

* 版本说明

请输入版本更新说明 0 / 200

取消 创建版本

版本管理

- 版本管理的入口在“发布”一栏中：

版本管理 服务管理 发布历史

版本号	创建时间	状态	版本说明	操作
v1.1.0	2026-04-03 11:35:32	已发布	测试	发布 预览 删除

- 用户可以在此处进行版本的发布，预览和删除。注意下方带有日期时间的版本属于版本功能上线前的历史数据。

版本发布（新服务）

- 版本发布是指将当前的智能体编排画布，部署到服务器，形成一个新的智能体服务。

- 选择版本：在此界面中可以选择不同的版本用于发布新服务。
- 发布方式：开发者可以将选中的版本发布一个全新的服务，也可以复用现有服务的服务器资源，直接覆盖。发布全新服务，其访问渠道将会发生变化，例如 API 网址会不同；如果覆盖现有服务，则最终用户访问渠道不变，但需要开发者注意输入参数不能发生变化，否则最终用户可能会遭遇异常。

版本预览

- 加载选中版本到草稿模式，开发者可以观察该版本的编排信息，也可以继续编辑并重新发布新版本，但当前版本内容不会出现变化。

删除版本

- 清除一个版本，但如果选中版本已经发布相应的服务，则该版本无法删除。直到它关联的服务已全部被移除，该版本才能正常删除。

服务管理

服务管理的入口为“智能体> 发布 > 服务管理”。

这里开发者可以管理此智能体的服务，例如启动、停止该服务，以及切换该服务背后的版本信息。

切换版本

- 切换版本本质上是用选中版本重新部署服务，并复用当前服务的访问通路，适合无缝升级，终端用户对升级动作无感知。注意需要开发者注意输入参数不能发生变化，否则最终用户可能会遭遇异常。
- 每个服务均为独立运行，服务和服务之间的资源和数据隔离。适合进行 AB 测试和灰度升级。

4.6 发布渠道配置

为了更好地满足用户在多种场景下的需求，星辰 MaaS 智能体平台（以下简称“平台”）在发布智能体时，提供了四种不同的渠道供用户选择



(支持多

选), 这四种渠道包括: 应用广场、网站嵌入、前端页面、后端 API。接下来, 我们将对这四种渠道进行详细的介绍, 以帮助用户更好地理解 and 选择。

发布到应用广场

当选择发布到应用广场时, 平台支持用户选择发布的场景标签、以及自定义是否支持复制同款。发布至应用广场需要经过平台审核才可成功发布。

发布步骤

1. 点击应用广场。

* 发布渠道

- 应用广场** 待发布
 将您的智能体发布到应用广场, 供其他用户发现和使用
- 嵌入网站** 待发布
 生成嵌入代码, 集成到您的网站中
- 前端页面** 待发布
 生成独立的前端访问页面
- 后端API** 待发布
 提供API接口供第三方系统集成

如何访问 API 文档？

- 点击“查阅 API 文档”按钮。



发布成嵌入网站

- 当选择嵌入网站发布后，点击顶部的发布 Tab。
- 然后点击嵌入的“嵌入”按钮。
- 平台提供了居中嵌入网页和右侧嵌入网页的代码，可一键复制到您想要嵌入的页面目标位置。



发布成前端页面

- 选择发布成前端页面后，此发布方式支持通过链接和二维码直接访问应用。
- 展示了当前应用的前端页面访问链接，点击复制按钮可一键复制此链接。
- 点击按钮可查看当前应用的二维码，支持下载。可通过扫码的形式进入应用进行体验。



- 点击刷新按钮可重新生成访问链接和二维码

4.7 日志管理

星辰 MaaS 智能体平台支持在日志模块查看当前智能体应用全部对话记录。支持通过时间、数据来源进行数据筛选，支持通过创建时间和更新时间进行日志排序，同时也支持使用关键词进行数据搜索。方便开发者快速进行日志数据的查询。

具体对话记录

- 日志数据中包含以下字段：标题、用户或账户名、对话消息数、用户反馈数、管理员反馈数、更新时间、创建时间。

日志

日志记录了应用的运行情况，包括用户的输入和 AI 的回复。

开始日期 → 结束日期 搜索 排序：创建时间 反馈情况 数据来源

标题	用户或账户	数据来源	消息数	用户反馈	反馈内容	管理员反馈	更新时间	创建时间
计划表好几遍	2b09159d-908d-4265-8--	本地数据	1	-	-	-	2024-11-26 14:28	2024-11-26 14:28
Hello, I am his father, lo...	2b09159d-908d-4265-8--	本地数据	2	-	-	-	2024-11-26 14:26	2024-11-26 14:25
两棵树 🌳	2b09159d-908d-4265-8--	本地数据	10	-	-	-	2024-11-26 14:25	2024-11-25 15:42

- 点击某条对话记录，右侧打开详细历史对话信息可以关注到不同用户使用该应用产生的对话数据，同时当 workflow 编排应用增加工具时，会展示工具调用情况。

日志

日志记录了应用的运行情况，包括用户的输入和 AI 的回复。

开始日期 → 结束日期 搜索 排序：创建时间 反馈情况 数据来源

标题	用户或账户	数据来源	消息数	用户反馈	反馈内容	管理员反馈
计划表好几遍	2b09159d-908d-4265-8--	本地数据	1	-	-	-
Hello, I am his father, lo...	2b09159d-908d-4265-8--	本地数据	2	-	-	-
两棵树 🌳	2b09159d-908d-4265-8--	本地数据	10	-	-	-

对话 ID: 4d330fe3-46f7-409e-8bf6-abe263e53061

您好

我是华然先生的助手，请问您是谁？有什么可以帮助您的吗？

帮我约华老师开会，明天下午5点

好的，请问这个会议大概需要多久呢？还有，会议的主题和参会人员有哪些呢？

一个小时，参会主题需求讨论，参会人员全体pm

好的，我会帮您安排明天下午5点的会议，持续大约一个小时。请稍候确认会议详情。
 --- 到此主动结束对话，总结并发送短信 ---
 --- 发送短信内容如下 ---
 华然先生明天下午5点有一场时长约为1小时的需求讨论会议，参会人员为全体PM。会议将在明天举行。

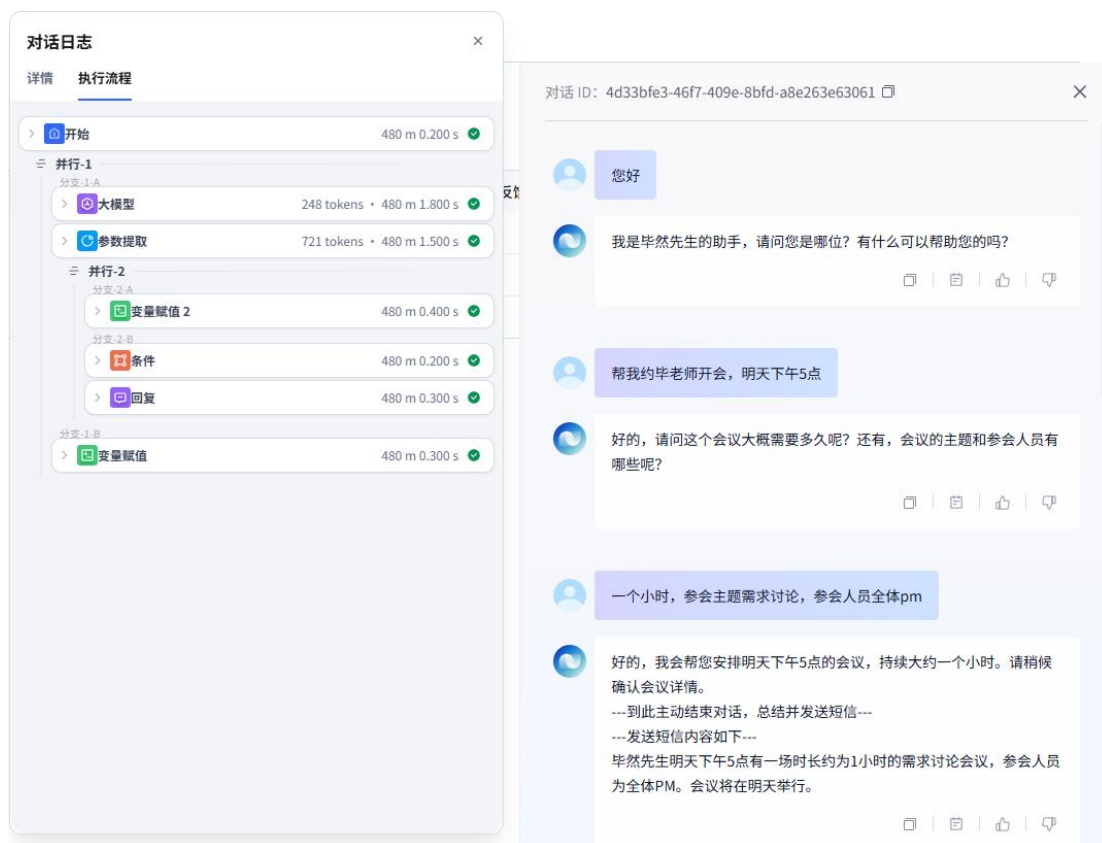
详细对话日志

- 在日志中的详细对话信息中，鼠标悬浮在某条对话上后点击查看日志，可以看到此轮对话具体调用信息，包括模型、工具或知识库的状态、输入输出数据、token 消耗等调用信息。



执行流程

- 在查看日志中点击执行流程 tab，可以以看到各节点的耗时和内容的输入输出情况。



日志下载

- 支持用户根据日志创建时间进行下载。
- 支持用户选择下载后的文件类型，类型包含：xls、xlsx。
- 支持用户选择具体的发布渠道对应的日志记录进行下载。

下载 ×

* 创建时间

2026-04-10 → 2026-04-24 📅

* 文件类型选择

xls ▼

* 渠道选择

应用广场 X ▼

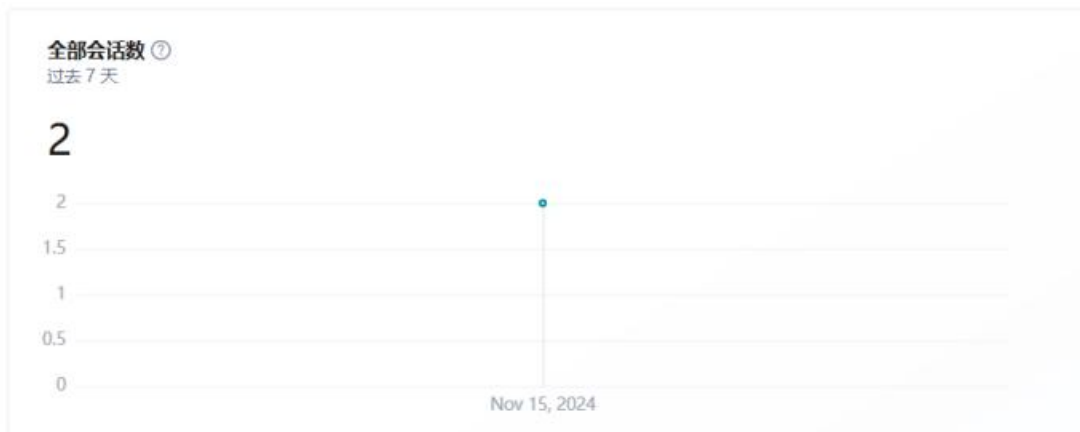
取消 下载

4.8 数据监测

在监测内监控、跟踪应用程序在生产环境中的性能，在数据分析仪表盘内分析生产环境中应用的使用成本、延迟、用户反馈、性能等指标，并通过持续调试、迭代不断改进你的应用程序。

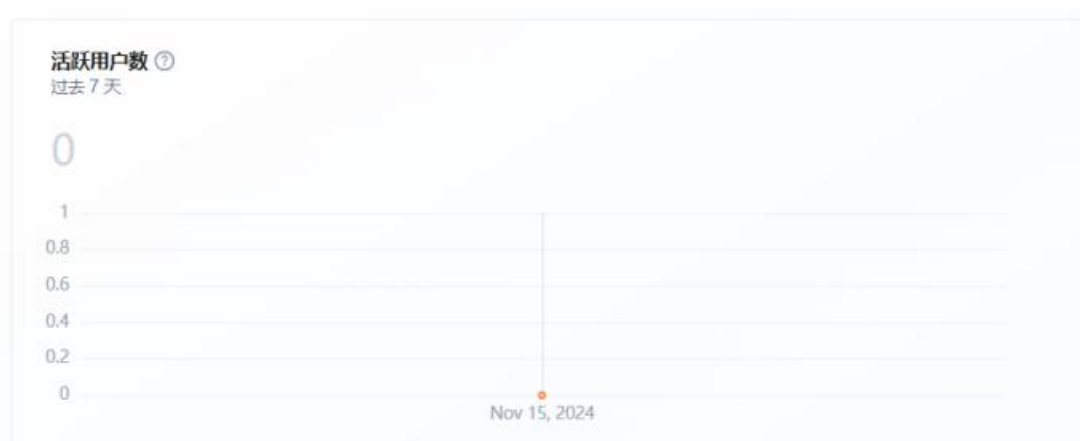
全部会话数

反映用户与 AI 每天的互动总次数，每回答用户一个问题即计数一次。预览的会话不计入。



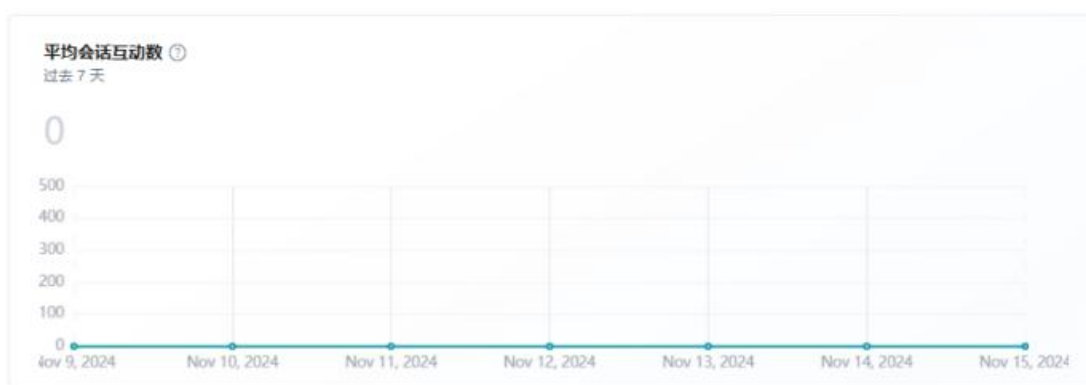
活跃用户数

与 AI 有效互动，即有一问一答以上的唯一用户数。预览的会话不计入。



平均会话互动数

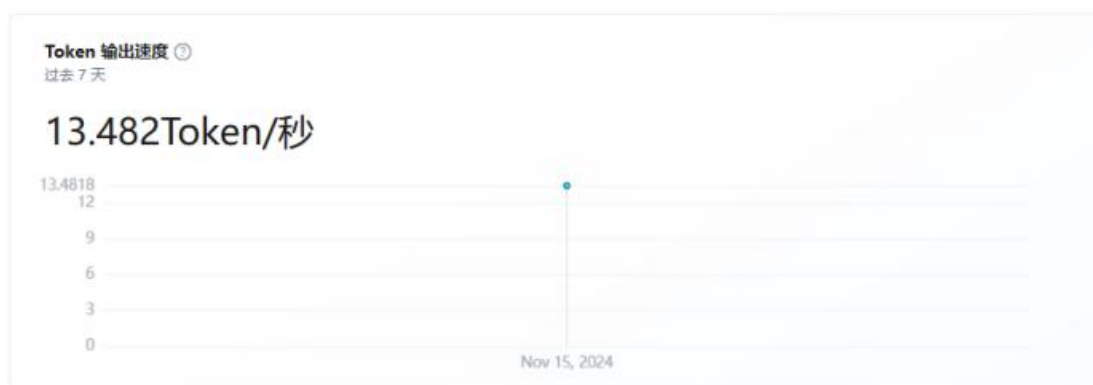
反映每个会话用户的持续沟通次数。该指标反映了用户粘性。



Token 输出速度

每秒的 Token 输出数量，反应模型的生成速率以及应用的使用频率。如第三方引入的应用未提供

Tokens 消耗数则无法统计该数据。



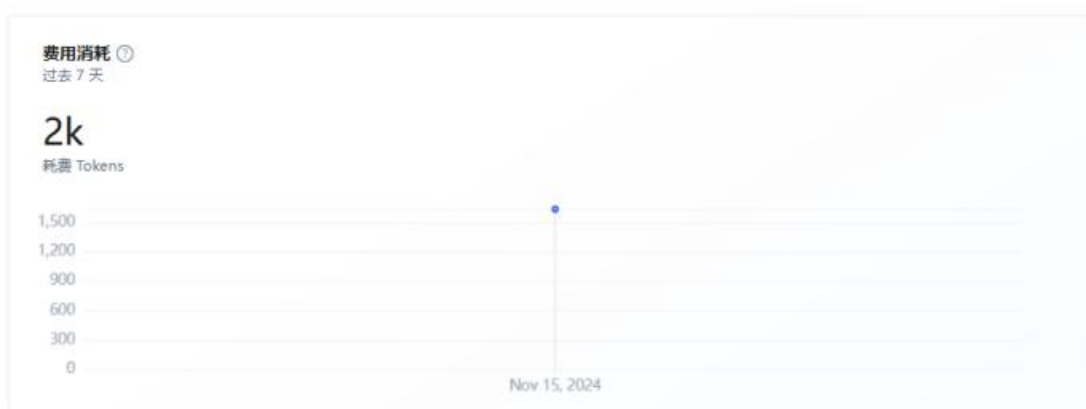
用户满意度

每 1000 条消息的点赞数。反映了用户对回答十分满意的比例。



费用消耗

反映每日该应用请求语言模型的 Tokens 消耗数。如第三方引入的应用未提供 Tokens 消耗数则无法统计该数据。



全部消息数

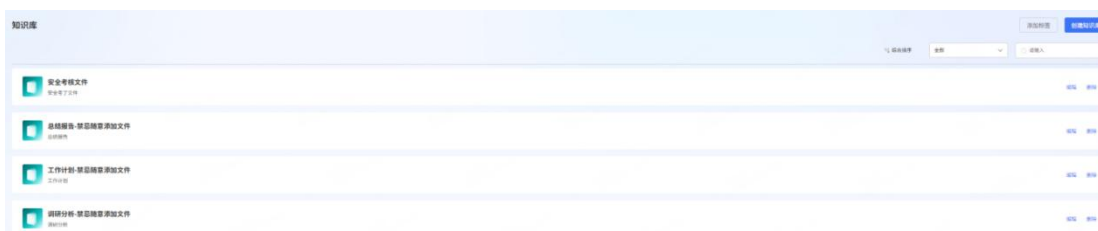
反映该应用每天的互动总次数，每回答用户一个问题算一条 Message。



4.9 创建与配置知识库

知识库创建

- 点击左侧导航栏中的知识库按钮进入知识库管理页面，点击创建知识库，单用户最多创建 50 个知识库。
- 支持用户针对已创建的知识库或创建知识库时添加标签，方便用户管理自己搭建的知识库。
- 文件上传指定的文件类型：TXT、MARKDOWN、PDF、OFD、XLSX、XLS、DOCX、DOC、CSV、WPS、WPT、ET、ETT、JPG、JPEG、PNG、BMP、TIFF
- 支持用户自定义标签(标签上限 10 个)，并在创建知识库的时候选择标签。



数据类型及数据上传要求

- 星辰 MaaS 智能体平台支持三种数据类型文件，包含结构化数据、非结构化数据、多模态数据。
 - 非结构化数据：支持上传文件格式包含 TXT、MARKDOWN、PDF、DOC、DOCX、OFD、WPS、WPT 格式。
 - 结构化数据：支持上传文件格式包含 CSV、XLSX、XLS、ET、ETT 格式。
 - 多模态数据：支持上传文件格式包含 JPG、JPEG、PNG、BMP、TIFF 格式

- 根据选择数据类型上传对应文件。上传非对应数据类型文件，会异常报错。
 - 非结构化数据文件：要求上传的每个文件不超过 50MB，最多同时上传 30 个文件
 - 结构化数据文件：要求上传的每个文件不超过 20MB，最多同时上传 50 个文件
 - 多模态数据文件：要求上传的每个文件不超过 5MB，最多同时上传 50 个文件



配置选择

分段处理：

适用于非结构化数据类型，文件格式包含 TXT、MARKDOWN、PDF、DOC、DOCX、OFD、WPS、WPT。（结构化数据会按照表格行数进行切分；多模态数据无相关功能。）

- 自动分段与清洗方式：自动模式适合对分段规则与预处理规则尚不熟悉的初级用户。在该模式下，星辰 MaaS 智能体平台将为你自动分段与清洗内容文件，你不需要进行任何筛选与操作。



- 自定义：自定义模式适合对于文本处理有明确需求的进阶用户。在自定义模式下，你可以根据不同的文档格式和场景要求，手动配置文本的分段规则和清洗策略。配置参数如下：
- 分段标识符：系统将在文本中出现指定的标识符的分段，文本换行时将自动分段；
- 分段最大长度：根据分段的文本字符数最大上限来进行分段，超出该长度时将强制分段。一个分段的最大长度为 500-10000 字节；
- 分段重叠长度：分段重叠指的是在对数据进行分段时，段与段之间存在一定的重叠部分。这种重叠可以帮助提高信息的保留和分析的准确性，提升召回效果。建议设置为分段长度 Tokens 数的 10-25%。

- 文本预处理规则：文本预处理规则可以帮助过滤知识库内部分无意义的内容。例如替换连续的空格、换行符和制表符。



Embedding 模型设置：

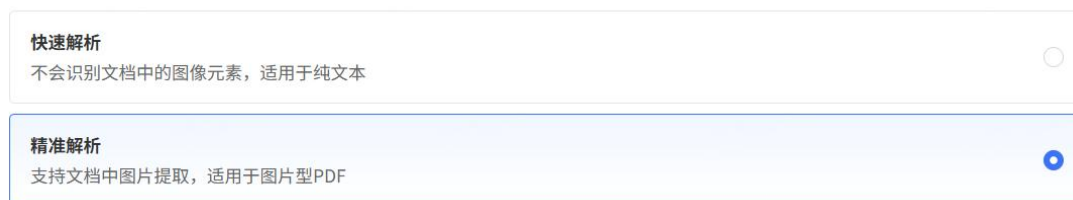
- Embedding 模型用于将自然语言处理转为向量，用于后续检索和大模型处理。
 - 可选择的内置 embedding 模型，或在设置-模型来源中添加第三方 embedding 模型。



解析方式

- 解析方式包含快速解析和精准解析，适用于非结构化数据类型，支持文件格式包含 PDF、TXT、DOCX、MARKDOWN 格式。
- 快速解析：适用于纯文本类型，不会识别文档中的图像元素。
- 精准解析：适合于图片型 pdf。

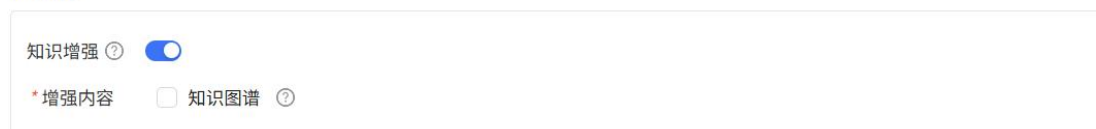
解析方式



知识增强

- 知识增强里面包含知识图谱，适用于非结构化数据。知识增强的作用是辅助召回涉及知识库问答的内容。
- 知识图谱能力：开启知识图谱能力后，能够使用 graphrag 算法提取文档内容中的实体，形成图谱辅助召回。

知识增强



4.10 已有知识库-文件管理

- 打开进入知识库页面首页，以列表卡片形式展示了自己添加的全部知识库。



- 知识库管理页面展示当前知识库中包含的所有文档。您可以根据具体场景开启或关闭某个文档。



- 点击已创建的知识库右侧添加文件按钮，继续添加更多文件。

4.11 文件切片管理

非结构化样式文件切片管理样式管理

用户可查看到自动切分或自定义切分后的片段内容。



结构化文件切片样式管理

用户能够看到上传解析后的文件内容。表格头展示的内容是用户上传的首行数据，是用户定义的表格数据的字段命名。

#	知识库	操作
1	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除
2	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除
3	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除
4	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除
5	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除
6	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除
7	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除
8	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除
9	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除
10	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除
11	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除
12	知识库	<input checked="" type="checkbox"/> 停用 <input type="checkbox"/> 删除

多模态数据展示

用户能够看到上传的图片，以及大模型对图片的解读，支持用户修改解读信息。



这是一幅充满宁静与梦幻色彩的数字绘画作品，描绘了一处热带或亚热带海滨的壮丽景色。

画面构图开阔，采用略微俯视的角度，将视线引向远方的海天交界。左侧是茂密的植被和悬崖，几片巨大的棕榈叶从画面上方垂下，叶脉清晰，边缘随微风飘动的弧度，为画面增添了层次感与热带气息。悬崖由深色岩石构成，表面覆盖着深密的绿色藤蔓，显得生机勃勃。

前景是细腻的沙滩，沙粒呈现出温暖的浅黄色调，被海水轻柔地冲刷着。海水颜色丰富，从近岸的清澈浅绿、淡青色，逐渐过渡到远处的深邃靛蓝。波浪轻柔地拍打着岸边，形成细碎的白色浪花，水面倒映着天空的云彩，光影交错，富有动感。

天空占据了画面的上半部分，是广阔而纯净的蓝色，点缀着蓬松洁白的云朵，形态各异，有的如棉絮般堆积，有的则如薄纱般轻盈。阳光透过云层洒下，在海面和沙滩上投下柔和的光斑，营造出一种温暖、安逸的氛围。

整幅画作笔触柔和，色彩明快饱满，具有典型的动漫或插画风格，强调光影和色彩的和谐统一。右下角有“Ai创作”的水印，表明这可能是一幅由人工智能生成的艺术作品。

整体而言，这幅画作传达出一种远离尘嚣、心旷神怡的度假感，是理想中自然美景的完美呈现。

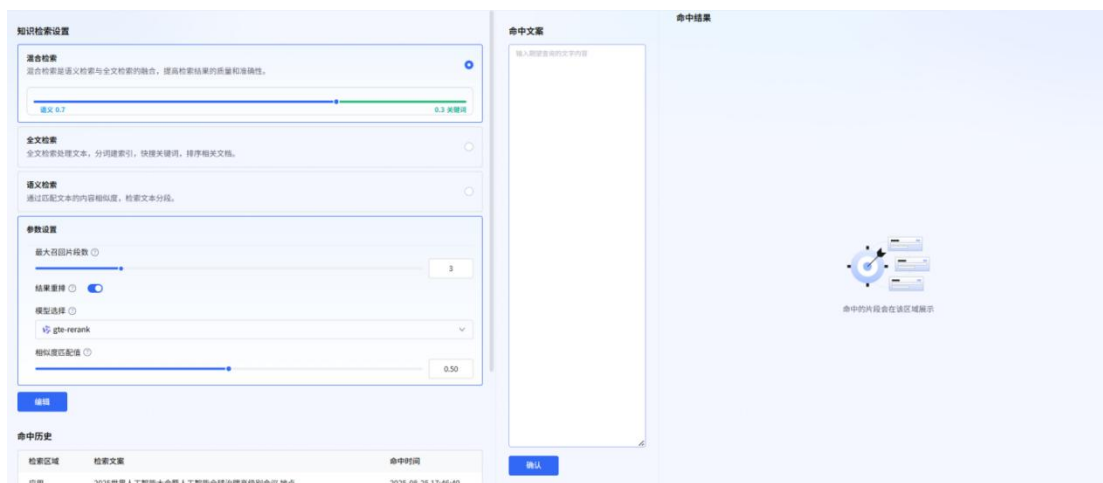
文件分段设置

支持针对已上传的单一文件进行文件分段设置，提供自动分段、自定义规则及文档层级分段三种模式；配置页中可选择或调整解析策略。



4.12 命中测试

为了保证知识库检索的效果，星辰 MaaS 智能体平台支持单独对知识库进行检索命中测试。



知识检索设置

支持用户选择三种检索形式中的任意一种，进行片段召回测试。

混合检索

- 同时执行全文检索和语义检索，结合语义查询和关键词查询，提升检索效果。
 - 权重设置：调整全文检索与语义检索之间的比重关系
- 允许用户赋予语义优先和关键词优先自定义的权重。
 - 全文检索指的是在知识库内进行关键词检索，语义检索指的是在知识库内进行向量检索。
 - 支持自定义关键词和语义权重，调节两种检索方式结果占比。
 - 语义检索指的是比对用户问题与知识库内容中的向量距离。距离越近，匹配的概率越大。

全文检索

关键词检索，即索引文档中的所有词汇。用户输入问题后，通过明文关键词匹配知识库内对应的文本片段，返回符合关键词的文本片段；类似搜索引擎中的明文检索。

语义检索

通过向量化用户输入的问题并生成查询向量，比较查询向量与知识库内对应的文本向量距离，寻找最近的分段内容。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。

最大召回片段数

用于筛选与用户问题相似度最高的文本片段。系统默认值为 3，数值越高，预期被召回的文本分段数量越多。可选择范围是 1-10。

重排模型

开启“结果重排”后，系统会对已召回的文档结果再一次进行重排序，优化排序结果。开启结果重排后，根据用户自定义的相似度匹配值，进行召回；未开启结果重排，内容展示为分数相关信息。

- 相似度匹配值：用于设置文本片段筛选的相似度匹配值，只召回超过设置分数的文本片段。打开后默认值为 0.5。可选择范围是 0-1。数值越高说明对于文本与问题要求的相似度越高，预期被召回的文本数量也越少。
- 根据用户输入的最大召回片段数，可召回相应的片段。

命中历史

- 页面左下方展示当前知识库检索历史记录，其中包含召回测试的检索记录和智能体应用中对该知识库检索的记录。
- 在输入框中输入待测试的内容，系统会根据检索配置进行召回测试，并将结果展示在右侧展示。

4.13 知识库 API

接口文档

星辰 MaaS 智能体平台预留接口信息，提供接口文档。支持用户参考接口文档，编写 api 接口，通过调用接口的形式进行知识库的创建、删除及上传文件。



API Key 管理

- 支持管理用户自己设定的 api key，支持复制 apikey 和删除 apikey。
- 支持用户自己创建 apikey，并设定有效时间，设定 api key 的目的是为了做调用接口的唯一识别。



4.14 记忆库

星辰 MaaS 智能体平台提供长期记忆构建与调用能力，支持智能体在对话中自动识别、存储并沉淀用户的个性化信息（如兴趣偏好、关键经历、行为习惯等），形成结构化的记忆条目。通过这一能力，智能体可在后续交互中对相关记忆进行召回，结合当前语境生成更贴合用户特征的反应，实现有记忆、有延续、有温度的个性化体验。

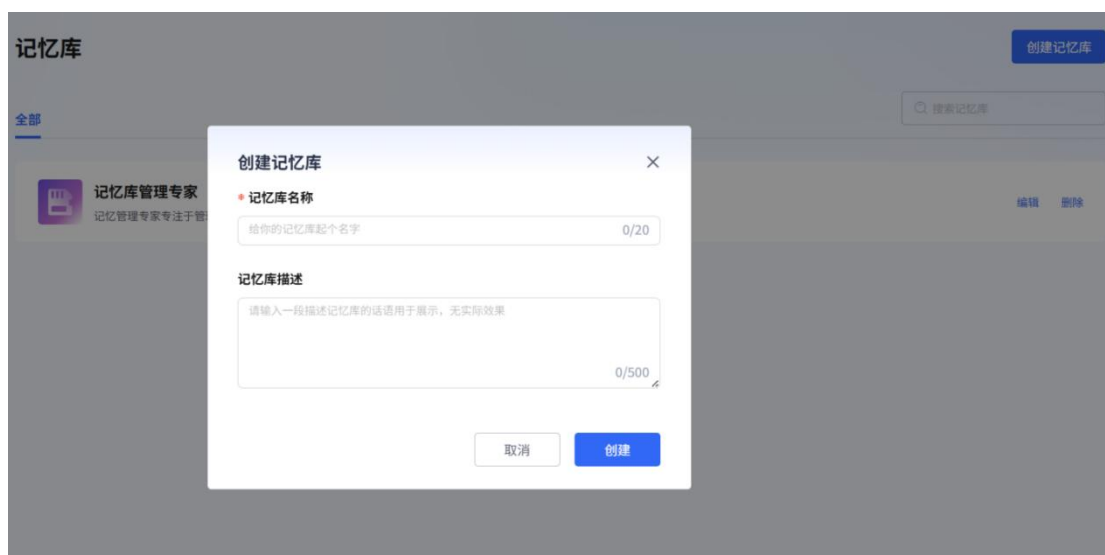
该能力适用于陪伴型 AI、生活助手、虚拟教练等需深度互动的场景，让每次对话都建立在对用户持续积累的理解之上。

创建记忆库

- 记忆库创建点击左侧导航栏中的记忆库按钮进入数据库管理页面，点击创建记忆库。单用户最多创建 50 个记忆库。



- 创建记忆库需填写的信息：记忆库名称、记忆库描述。
- 点击创建即可完成创建流程。



已有记忆库管理

- 详情页查看以及管理

- 具体显示：已存入的条目数量，并且显示剩余多少条目数量。
- 查看长期记忆，点击目标 ID，并查看对应的长期记忆。
- 该 ID 是系统组合生成的编码，该编码将作为数据隔离的核心标识，保障同一记忆库中不同使用主体的记忆数据独立。
- 开发者可对记忆内容进行日期筛选以及删除操作。
- 注意事项 1：删除某个记忆库后，引用了该记忆库的智能体或工作流也将自动取消引用，且此操作不可撤回，请谨慎操作。
- 注意事项 2：开发者可对使用者的记忆内容进行删除。
- 开发者仅可对自己的记忆内容进行内容的修改操作。注意开发者不可对使用者的记忆内容进行修改。



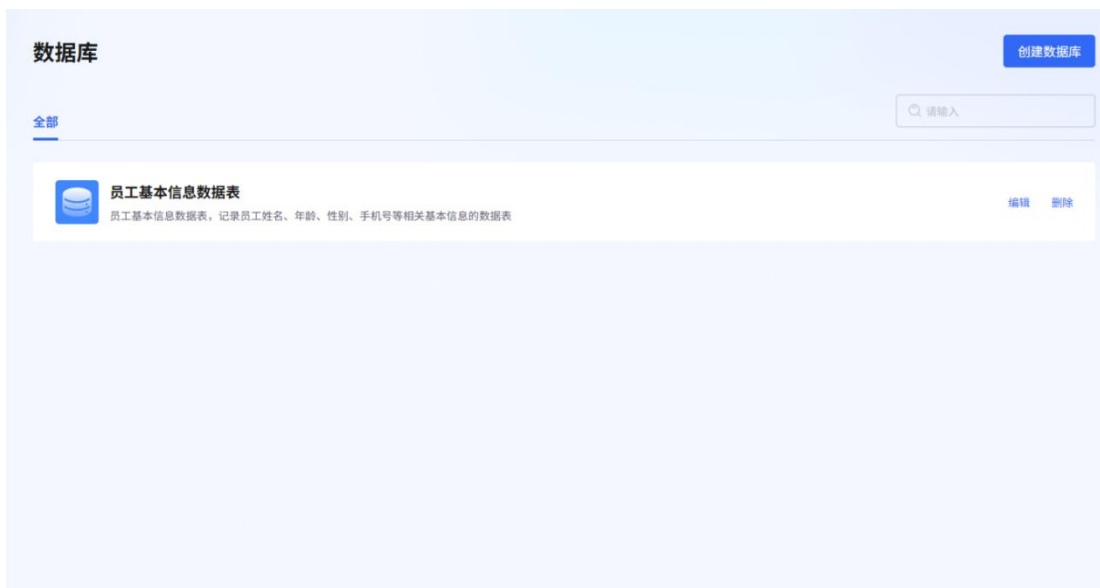
4.15 数据库

星辰 MaaS 智能体平台提供结构化数据存储能力，可用于管理如客户信息、产品清单、订单记录等业务数据。用户可通过自然语言进行数据的查询、添加、修改和删除操作，支持单表访问及多表关联分析，系统能自动理解语义并返回准确结果。该功能可与其他模块灵活协同，每个应用可绑定一个数据库，满足智能交互与业务自动化中的多样化数据需求。

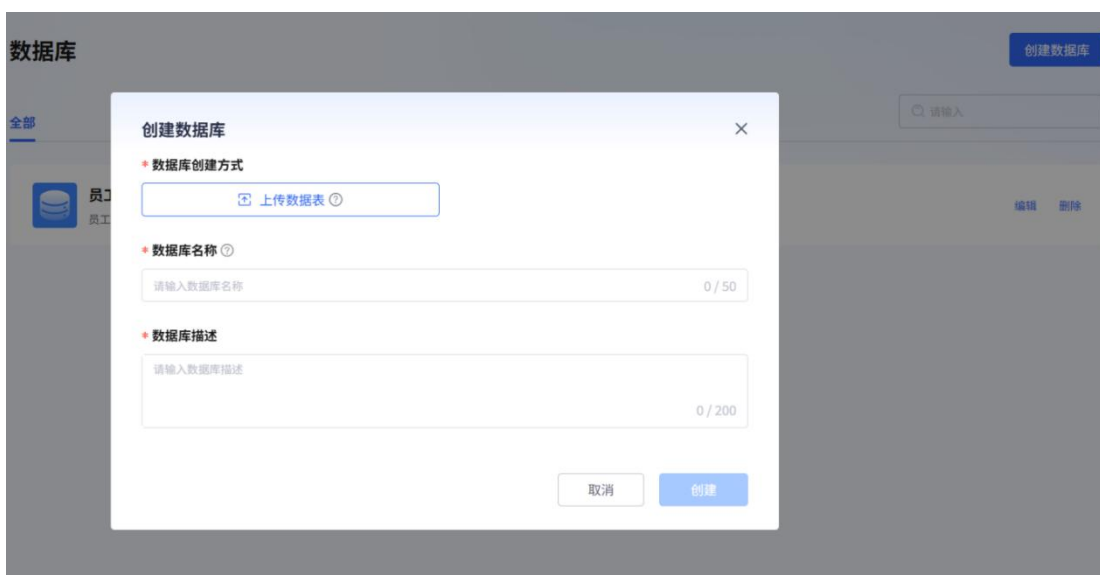
创建与配置数据库

数据库创建

- 点击左侧导航栏中的数据库按钮进入数据库管理页面，点击创建数据库。单用户最多创建 10 个数据库。



- 创建数据库需填写的信息：数据库名称、数据库描述



数据类型

- 星辰 MaaS 智能体平台支持上传数据表格式包含 CSV、XLSX、XLS 格式。

数据上传

- 可上传数据类型内存在的三种格式，上传非对应数据类型文件，会异常报错。
- 上传文件不超过 5 个

数据库信息

数据库名称
销售清单详情

数据库描述
销售清单详情内容

数据表导入

* 选择数据源文件

↑

拖拽文件至此，或者 [选择文件](#)

本次上传文件不超过5个（当前数量 0/5 个）；支持.xlsx/.xls/.csv三种文件格式，.csv文件支持UTF_8编码格式，单个文件大小不超过15MB，不超过1万行、20列（超出范围的列或行会被自动忽略）且文件中最多支持一个sheet工作表

模板下载 csv模板 xls模板 xlsx模板

配置选项

配置选项

* 自动生成数据表描述

表描述为数据配置必填项，影响大模型选择，启用自动生成，可根据数据内容快速生成表描述，提高效率

* 数据管理权限

数据配置权限默认为只读模式，所有用户仅可查看数据。开启后，全体用户将获得完整编辑数据的权限

- 自动生成数据表描述（必开选项）

- 表描述为数据配置必填项，影响大模型选择，启用自动生成，可根据数据内容快速生成表描述，提高效率。

- 数据管理权限（目前不支持开启）

- 数据配置权限默认为只读模式，所有用户仅可查看数据。开启后，全体用户将获得完整编辑数据的权限。

已有数据库管理

- 数据库管理页面展示当前数据库中包含的所有数据表。
- 点击已创建的数据库右侧添加数据表按钮。继续添加更多文件，库内限制最多添加 5 个数据表。
 - 点击数据表后的操作按钮，支持查看当前数据表的详情内容以及删除当前数据表。



查看数据库

- 点击查看，可查看数据库详情。
- 支持编辑各字段的内容。
- 支持删除数据表的行数据内容。



基本信息查看

- 查看当前数据表的基本信息，包括创建/更新时间、最后更新状态、数据量以及数据表描述。



表结构编辑

- 支持新增字段，可输入字段名称、字段描述、字段类型以及选择字段的必要内容。
- 支持删除已有字段。



追加数据

- 追加上传数据表，单次仅上传 1 个文件。
- 仅支持 csv、xlsx 格式上传以及两种格式的模板下载。



追加历史

- 追加数据中上传的数据表内容，可在该页面对上传数据表的历史进行查看。



修改配置

- 支持修改数据表名称
- 支持修改数据表描述，数据表描述支持优化选择

4.16 工作流工具

创建步骤

1. 登录[星辰 MaaS 智能体平台](#)。
2. 在页面左侧的工具后，点击页面中右侧的创建工具按钮，点击通过工作流创建。
3. 输入应用名称和简介，使用内置头像或单击头像图标上传自定义头像。

4. 完成创建后会进入 workflow 编排页面。

创建工具
✕

*** 工具名称**

*** 工具描述**

*** 工具头像**



*** 工具创建方式**

S 工作流工具

O OPENAPI schema工具

M MCP工具

*** 工具调用名称**

取消

创建

工作流工具与工作流应用差异

工作流编排功能与工作流编排应用大致相同，因工具中的工作流创建方式不包含对话功能，故存在以下区别：

开始节点

系统变量如下：

变量名称	数据类型	说明
sys.files	Array[File]	文件参数，存储用户初始使用应用时上传的图片。
sys.user_id	String	用户 ID，每个用户在使用工作流应用时，系统会自动向用户分配唯一标识符，用以区分不同的对话用户。
sys.app_id	String	应用 ID，系统会向每个应用分配

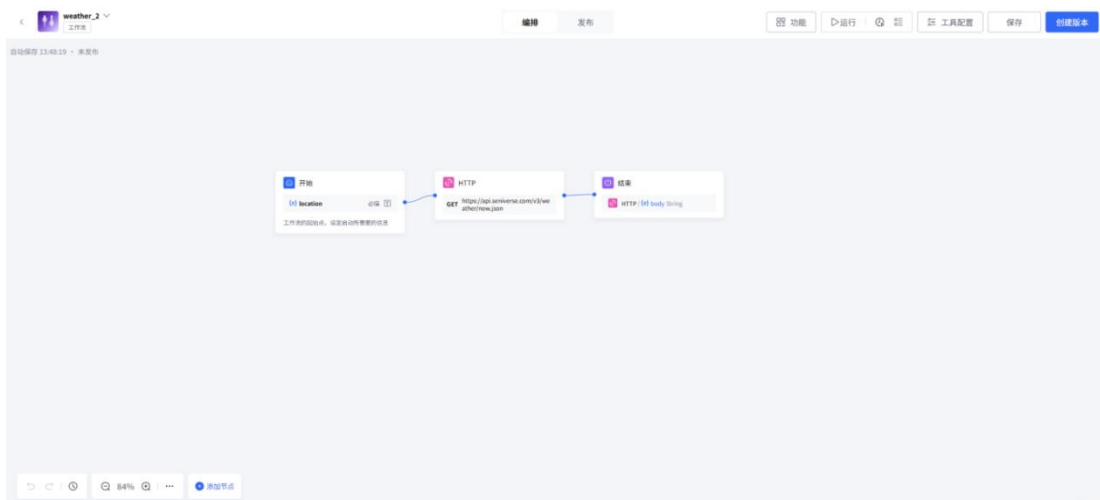
		一个唯一的标识符，用以区分不同的应用，并通过此参数记录当前应用的基本信息。
sys.workflow_id	String	用于记录当前应用内所包含的所有节点信息。
sys.workflow_run_id	String	应用运行 ID, 用于记录应用中的运行情况。
变量名称	数据类型	说明

结束节点

- 定义一个工作流程结束的最终输出内容。工作流在完整执行后都需要至少一个结束节点，用于输出完整执行的最终结果。
- 结束节点为流程终止节点，后面无法再添加其他节点，工作流应用中只有运行到结束节点才会输出执行结果。若流程中出现条件分叉，则需要定义多个结束节点。
- 结束节点需要声明一个或多个输出变量，声明时可以引用任意上游节点的输出变量。

其他与工作流编排应用区别

- 无会话变量配置
- 无回复节点
- 无日志功能
- 无监控功能
- 运行：支持单次运行或者上传批量运行所需的入参文件进行工具效果测试。
- 完成配置的工作流需要点击右上角发布菜单中的工具配置进行入参设置，正确完成设置的工具支持添加至应用。



4.17 OPENAPI Schema 工具

创建方式

1. 登录星辰 MaaS 智能体平台。
 2. 点击工具页面中右侧的创建工具按钮，点击 OPENAPI Schema 工具。
 3. 输入工具配置信息进行工具创建，完成创建的工具支持在应用中使用。
 4. 工具名称：用户可自定义工具名称。
- Schema:
 - 支持 OpenAPI / Swagger 规范。你可以将 OpenAPI schema 内容直接粘贴或从 URL 内导入。
 - 工具目前支持两种鉴权方式：无鉴权 和 API Key。
 - 工具列表:
 - 在导入 Schema 内容后系统会主动解析文件内的参数，并可预览工具体体的参数、方法、路径。您也可以在此对工具参数进行测试。

创建工具 ✕

*** 工具创建方式**

🔗 工作流工具
📄 OPENAPI schema工具
🔌 MCP工具

*** Schema** ⊕ 从 URL 中导入

在此处输入您的 OpenAPI schema

0 / 1000

[查看 OpenAPI-Swagger 规范](#)

可用工具

名称	描述	方法	路径	操作
<div style="text-align: center; font-size: 2em; color: #ccc;">📁</div> <p style="color: #ccc;">暂无数据</p>				

取消
创建

创建示例

- 部分服务商直接提供
- 需要仔细查看一下对方的文档，关键字为 OpenAPI Schema 或 OpenAPI YAML:

The screenshot shows a web browser window with a sidebar on the left containing a list of API-related items. The main content area displays a document titled 'Web Search API'. A red box highlights a section titled '六. OpenAPI Yaml配置文件' (Six. OpenAPI Yaml Configuration File). Below the title, there is a paragraph of text and a code block containing a YAML configuration for an OpenAPI tool.

```

六. OpenAPI Yaml配置文件
在各种 AI Agent 开发平台中，可以参考以下的 yaml 文件，快速添加自定义插件：

代码块
1 openapi: 3.0.3
2 info:
3   title: 博查AI搜索
4   description: 从博查搜索引擎信息和网页链接，搜索结果准确、完整，更适合AI使用。
5   version: 1.0.0
6 servers:
7   - url: https://api.bochaai.com/v1
8   paths:
9     /web-search:
10      post:
11        summary: 从近万亿网页和生态内容中搜索高质量世界知识，例如新闻、图片、百科、文库等。
12        operationId: WebSearch
13        requestBody:
14          description: 请求参数
15          content:
16            application/json:
17              schema:
18                type: object
19                properties:
20                  query:
21                    type: string
22                    description: 搜索关键字或语句
23                    example: "阿里巴巴2024年的ESG报告"
24                  summary:
25                    type: boolean
26                    description: 是否在搜索结果中包含摘要
  
```

- 如果没有，可以使用大模型来生成对应的 OpenAPI YAML 文件:

The screenshot shows the高德地图 (Gaode Map) API documentation page for weather information. The page is titled "天气查询API服务地址" (Weather Query API Service Address). It includes a table for the API endpoint and a table for request parameters.

URL	请求方式
https://restapi.amap.com/v3/weather/weatherInfo?parameters	GET

请求参数

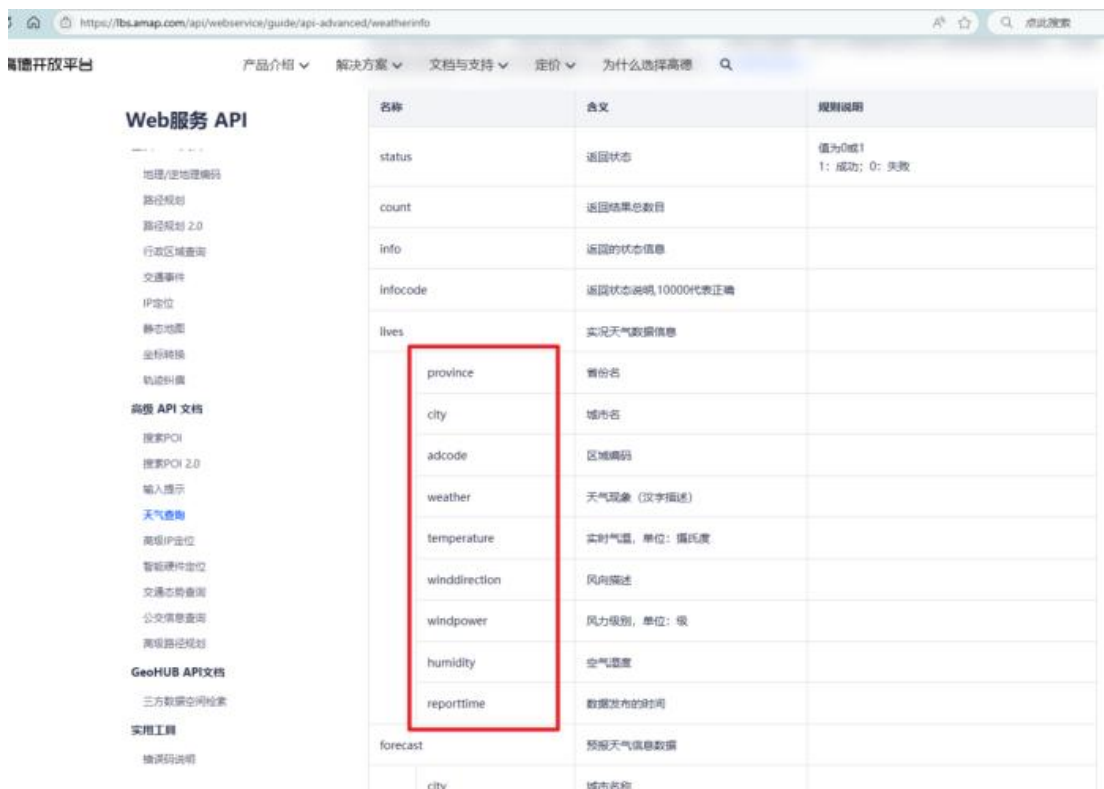
参数名	含义	规则说明	是否必填	缺省值
key	请求服务权限标识	用户在高德地图官网 申请 web 服务 API 类型 KEY	必填	无
city	城市编码	输入城市的 adcode, adcode 信息可参考 城市编码表	必填	无
extensions	气象类型	可选值: base/all base:返回实况天气 all:返回预报天气	可选	无
output	返回格式	可选值: JSON/XML	可选	JSON

返回结果参数说明

实况天气每小时更新多次, 预报天气每天更新3次, 分别在8、11、18点左右更新。由于天气数据的特殊性以及数据更新的持续性, 无法确定精确的更新时间, 请以接口返回数据的 reporttime 字段为准。 [天气结果列表>>](#)

名称	含义	规则说明
----	----	------

- 将文档内容保存下来，发送给大模型，示例:
 - 天气查询-高级 API 文档-开发指南-Web 服务 API | 高德地图 API
 - 从图中可见，文档内容排版混乱，对于大模型不友好 需要手工修改。



- 手工修改后:

URL	请求方式
https://restapi.amap.com/v3/weather/weatherInfo?parameters	GET

- parameters 代表的参数包括必填参数和可选参数。所有参数均使用和号字符(&)进行分隔。下面的列表枚举了这些参数及其使用规则。
- 请求参数

参数名	含义	规则说明	是否必须	缺省值
key	请求服务权限标识	用户在高德地图官网 申请 web 服务 API 类型 KEY。	必填	-
city	城市编码	输入城市的 adcode, adcode 信息可参考 城市编码表。	必填	-
extensions	气象类型	可选值:	可选	-

		base/allbase:返回实况天气 all:返回预报天气。		
output	返回格式	可选值: JSON,XML。	可选	JSON

- 返回结果参数说明
- 实况天气每小时更新多次，预报天气每天更新 3 次，分别在 8、11、18 点左右更新。由于天气数据的特殊性以及数据更新的持续性，无法确定精确的更新时间，请以接口返回数据的 reporttime 字段为准。
- 使用大模型生成 YAML 配置文件
- 将文档内容发送至大模型，Prompt 为：请根据这个 API 说明文档，生成 OPENAPI YAML 配置文件：



```

WeiJin
05/27 17:27
请根据这个API说明文档,生成对应的OPENAPI YAML配置文件: #### 天气查询API服务地址

| | |
|--|--|
|URL|请求方式|
|https://restapi.amap.com/v3/weather/weatherInfo?parameters|GET|

parameters 代表的参数包括必填参数和可选参数。所有参数均使用和号字符(&)进行分隔。下面的列表枚举了这些参数及其使用规则。

#### 请求参数

| | | | | |
|--|--|--|--|--|
|参数名|含义|规则说明|是否必须|缺省值|
|key|请求服务权限标识|用户在高德地图官网 [申请 web 服务 API 类型 KEY](https://lbs.amap.com/dev/)|必填|无|
|city|城市编码|输入城市的 adcode, adcode 信息可参考 [城市编码表](https://lbs.amap.com/api/webservice/download)|必填|无|
|extensions|气象类型|可选值: base/all<br><br>base:返回实况天气<br><br>all:返回预报天气|可选|无|
|output|返回格式|可选值: JSON,XML|可选|JSON|

```

- 大模型返回:



DeepSeek Chat | 深度求索

05/27 17:27

Here's the OpenAPI YAML configuration file based on the provided API documentation:

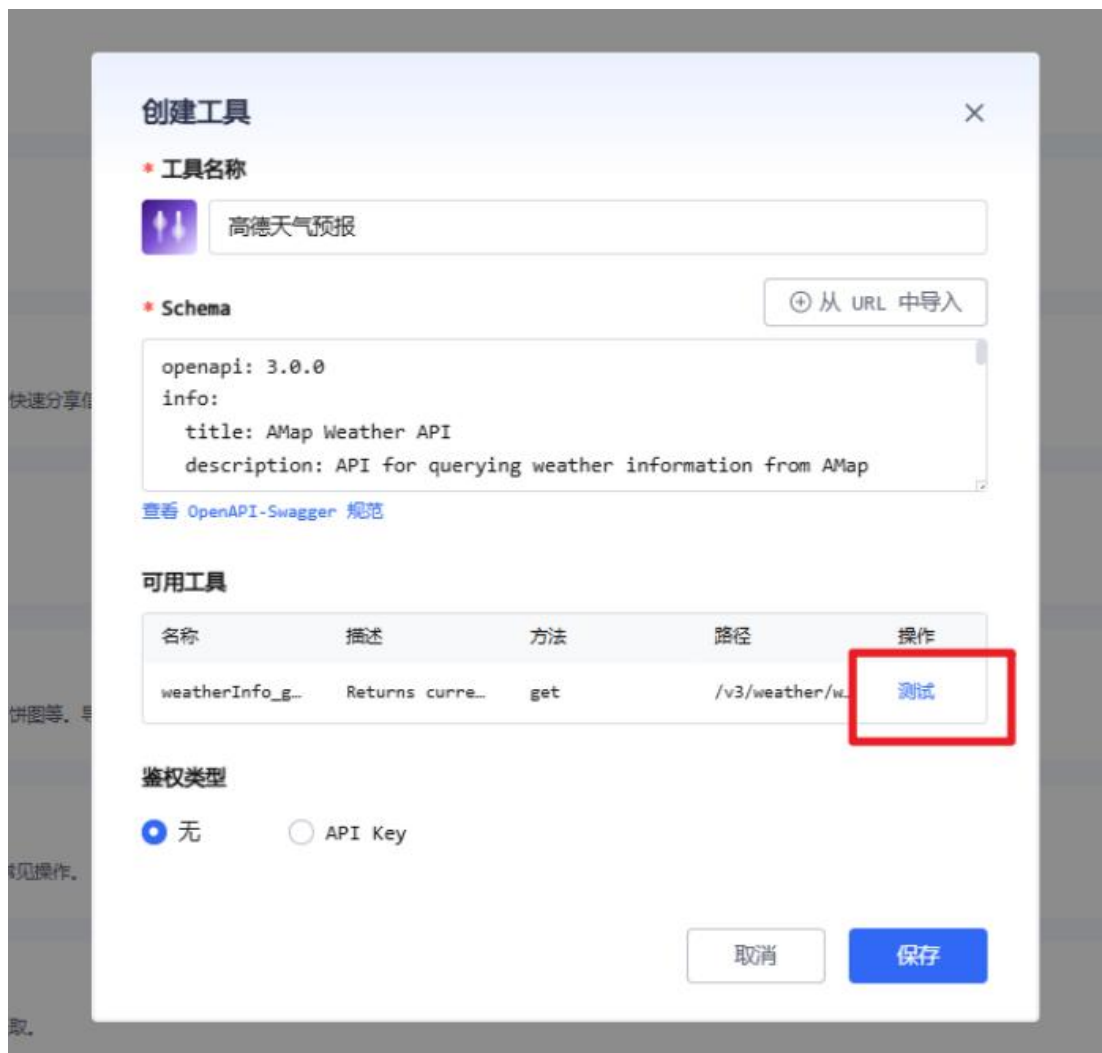
<YAML>

```
1  openapi: 3.0.0
2  info:
3    title: AMap Weather API
4    description: API for querying weather information from AMap
5    version: 1.0.0
6  servers:
7    - url: https://restapi.amap.com/v3/weather
8  paths:
9    /weatherInfo:
10     get:
11       summary: Get weather information
12       description: Returns current or forecast weather information for a specific city
13       parameters:
14         - name: key
15           in: query
16           description: Request service permission identifier (API key)
17           required: true
18           schema:
19             type: string
20         - name: city
```

- 使用大模型生成 YAML 配置文件
- 导入 Schema 内容，系统会主动解析文件内的参数，并可预览工具具体的参数、方法、路径。您也可以在此对工具参数进行测试。



- 输入后会出现测试按钮:



- 点击测试按钮，输入必要信息，查看返回结果:



4.18 MCP 工具

创建步骤

1. 登录[星辰 MaaS 智能体平台](#)。
2. 点击工具页面中右侧的创建工具按钮，点击 MCP 工具。
3. 输入工具配置信息进行工具创建，完成创建的工具支持在应用中使用。
4. 工具名称、描述：用户可自定义工具名称、描述。
5. MCP 服务配置：在写入内容后系统会主动解析文件内的参数，并可预览工具体体的参数、方法、路径。您也可以在此对工具参数进行测试。

* 工具创建方式

工作流工具 OPENAPI schema工具 MCP工具

MCP服务类型

http+SSE Streamable

* MCP服务配置

请输入MCP服务配置

0 / 1000

具体操作步骤

- 工具>创建工具> MCP 工具：

* 工具创建方式

工作流工具 OPENAPI schema工具 MCP工具

MCP服务类型

http+SSE Streamable

* MCP服务配置

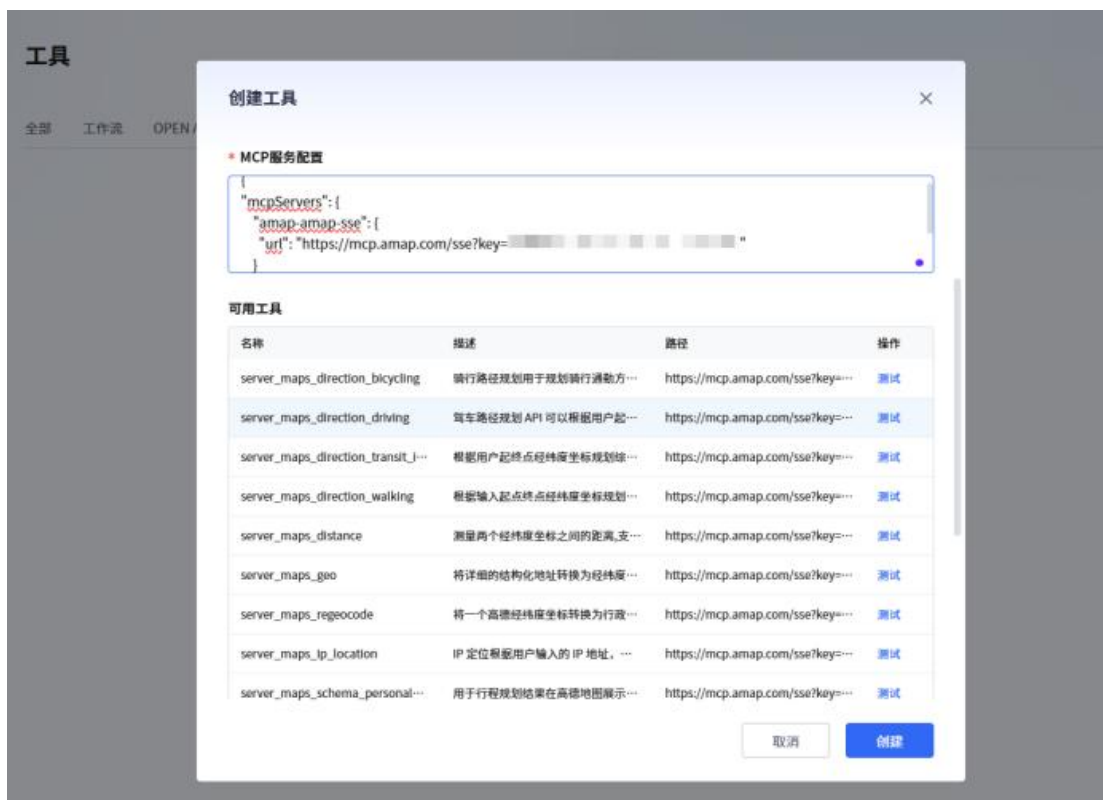
请输入MCP服务配置

0 / 1000

- 填写名称及工具描述
- 找到希望添加的 HTTP+SSE 或 Streamable 类型的 MCP 服务说明页面， 例如高德地图， 将其配置代码复制并粘贴到 MCP 服务配置 “中

```
{  
"mcpServers": {  
"amap-amap-sse": {  
"url": "https://mcp.amap.com/sse?key=您的高德 API Key"  
}}}
```

- 修改配置中的 API KEY, “你的高德 API KEY” 替换为你在高德开放平台获取的 Key。



- 下方会自动出现该 MCP 支持的服务, 点击“测试”可以验证效果:



- 最终点击“创建”完成该 MCP 工具的引入。

4.19 工具版本及发布管理

workflow 工具:

- 具备多版本概念。
- 具备创建版本、发布功能。
- 支持发布到工具广场、后端 api。

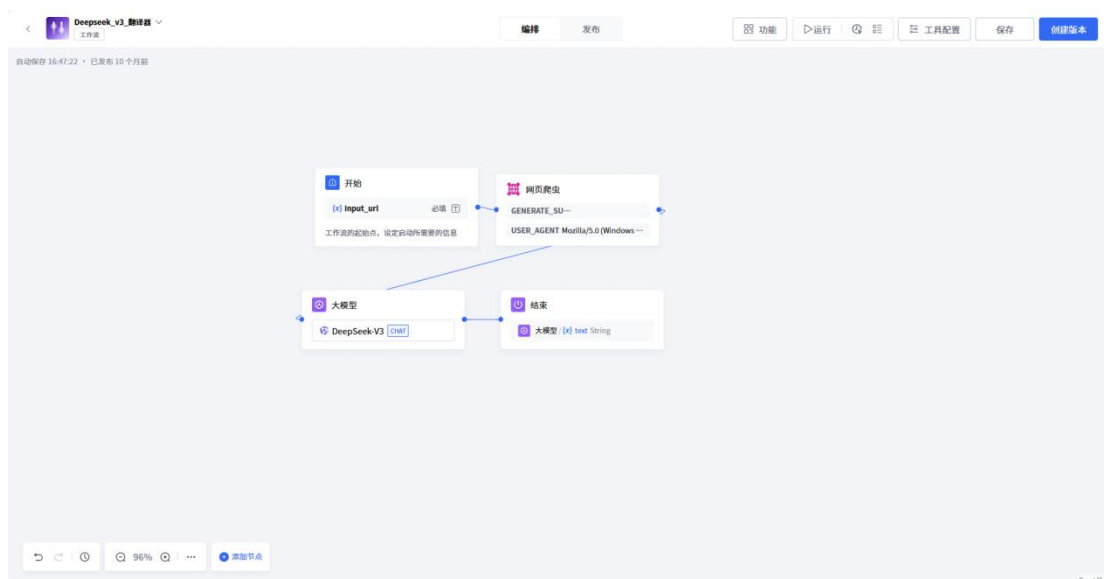
版本管理

相关概念

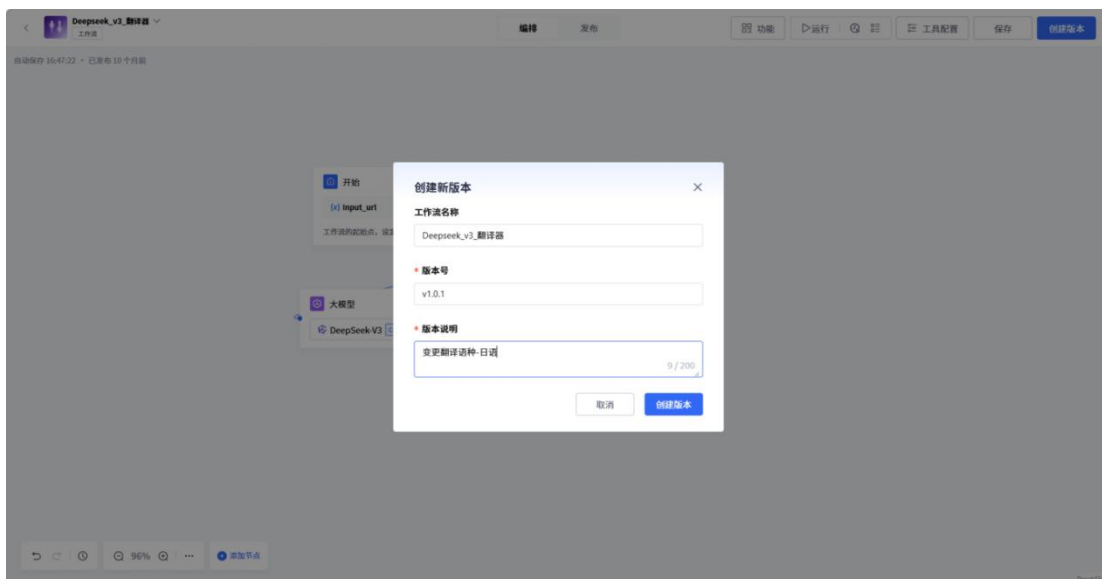
- 关键概念:
 - 版本: 版本为 workflow 工具的编排的快照, 用来记录某个时刻的 workflow 工具的编排状态。
- 其他概念:
 - 当前草稿: 您所在的处于可编辑的工具编排界面均为草稿版本。
 - 创建版本: 当草稿修改到一定阶段, 可保存的独立版本。

版本相关功能

- 创建版本
 - 创建版本的入口在编排界面的右上角“创建版本”按钮。



- 填写版本信息:



● 版本管理

- 版本管理的入口在“发布”一栏中：

版本号	创建时间	状态	版本说明	操作
v1.0.0	2026-03-27 15:17:20	待发布	111111	发布 预览 删除

- 用户可以在此处进行版本的发布，预览和删除。

发布管理

发布配置

- 对现有的不同版本的工具进行管理，支持该工具的不同版本的发布和下架。

版本号	创建时间	状态	版本说明	操作
v1.0.0	2026-03-27 15:17:20	待发布	111111	发布 预览 删除

发布历史

- 展现此工具不同版本的发布历史。点击“详情”按钮可以查看更多版本信息。

工具名称	发布时间	发布渠道-工具广场	操作
法语网页翻译成英语	2026-01-15 16:58:26	有①	详情

4.20 工具渠道管理

版本发布是指将当前的工作流工具编排画布，部署到服务器，并将使用接口暴露在工具广场上，供全部用户使用。支持用户发布时选择单一渠道或多渠道，如工具广场、后端 api 或工具广场+后端 api 形式。

发布到工具广场

发布过程需要审核，审核通过后才会在工具广场上可见。

发布
✕

*** 工具名称**

*** 工具介绍**

*** 发布渠道**

工具广场

后端API

提示:当您将工具发布到工具广场后，工具可直接被他人调用。工具中上传的知识库和搭建的私人工具会被用户使用。授权判断是指用户调用已发布的工具，是否需要输入授权码。

取消

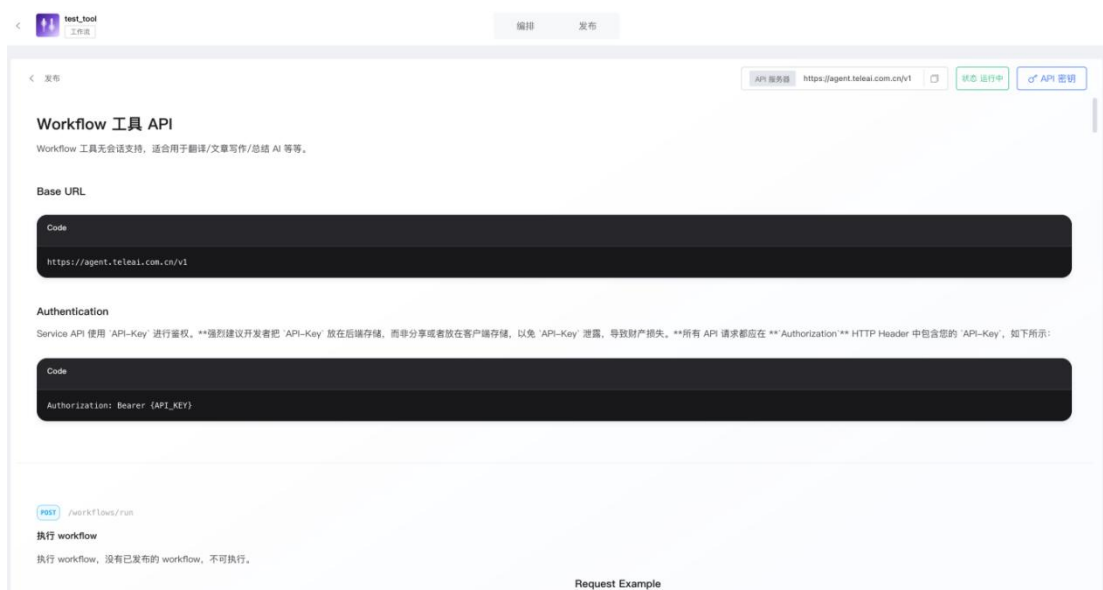
确认

发布到后端 api

- 星辰 MaaS 智能体平台支持将工具发布为 API，给用户调用服务。
- 当选择发布成 API 后，点击页面顶部 Tab 进入服务管理页面。
- 展示系统自动生成的当前应用 API，可一键复制。
- 点击“API 密钥”按钮，打开 API 密钥弹窗，在此可一键创建密钥、一键复制。

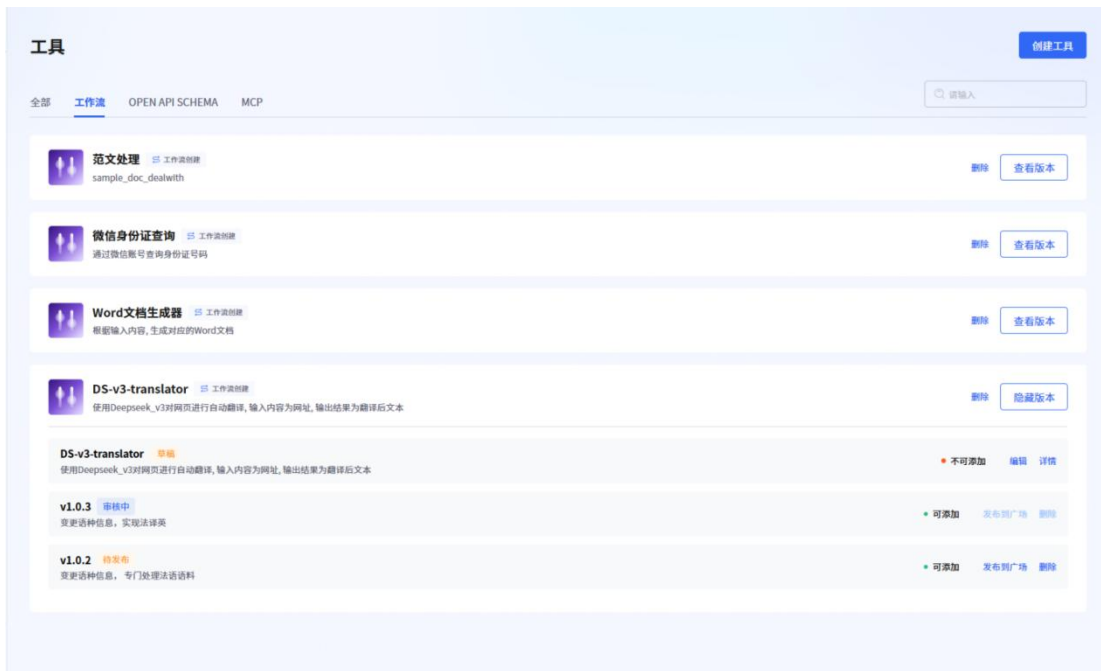


- 点击“查阅 API 文档”按钮。



版本查看

在工具列表右侧，有“查看版本”按钮，点击后查看全部版本（含草稿版本）



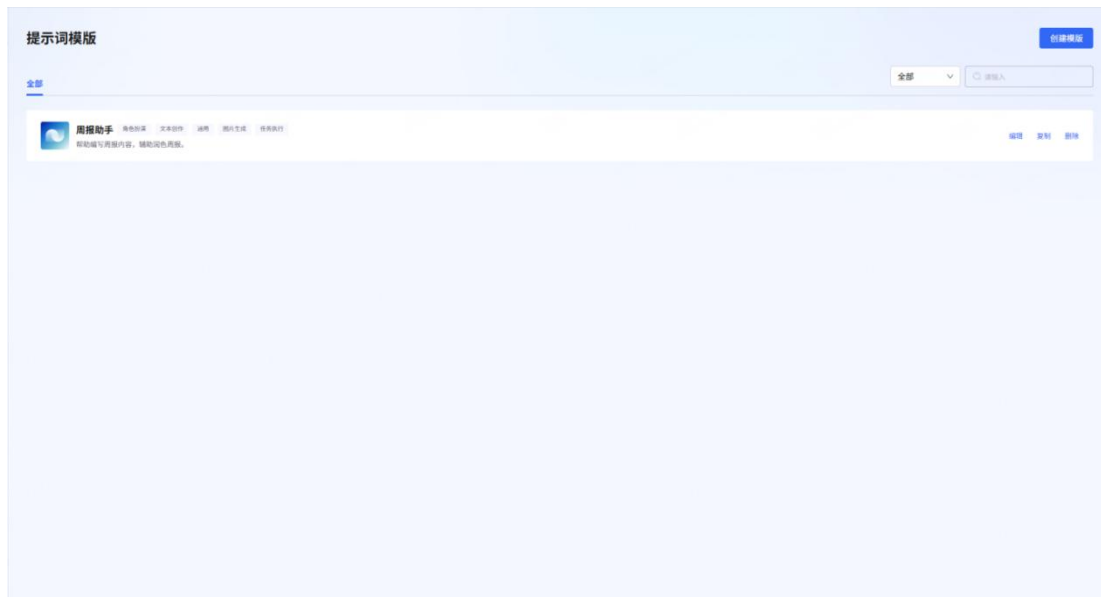
4.21 提示词模板管理

提示词是用户向人工智能模型（如大语言模型、多模态模型等）输入的指令、问题或信息，其核心作用在于引导模型生成符合预期的输出内容，作为人类与人工智能交互的核心媒介，直接决定人工智能的响应方向、质量和精准度。

- 提示词的核心作用：
 - 明确任务目标，即告知人工智能需要完成的具体工作；
 - 限定输出范围与格式，规范人工智能的输出形式；
 - 提供关键约束条件，补充背景信息、受众定位、专业要求等细节。
- 提示词主要分为四类，包括：
 - 直接下达任务指令且简洁明确的指令型提示词；
 - 以提问形式获取信息、常用于知识查询与问题解决的问答型提示词；
 - 提供上下文、背景或示例以引导人工智能生成更贴合需求内容（又称“少样本提示”）的引导型提示词；
 - 以及为人工智能设定特定角色、使输出内容匹配该角色口吻和专业度的角色型提示词。

模板管理

支持管理提示词模板，支持用户创建模板、编辑模板内容、复制模板和删除模板。



创建模板

支持用户自定义模板名称、描述、模板内容，支持用户使用模板优化功能，支持用户选择场景标签、上传模板头像。

创建模版 ✕

*** 模版名称**

 0 / 50

*** 模版描述**

 0 / 100

...

要求和限制

1. 模型需要遵守哪些要求和限制
2. 模型需要遵守哪些要求和限制

...

取消

创建

4.22 模型管理

星辰 MaaS 智能体平台（以下简称“平台”）内置大语言模型、Text Embedding 和 Rerank，用户可以快速进行智能体应用的搭建和发布。如果需要指定模型，可以点击左侧功能栏的模型管理页面内添加并配置所需要的模型。

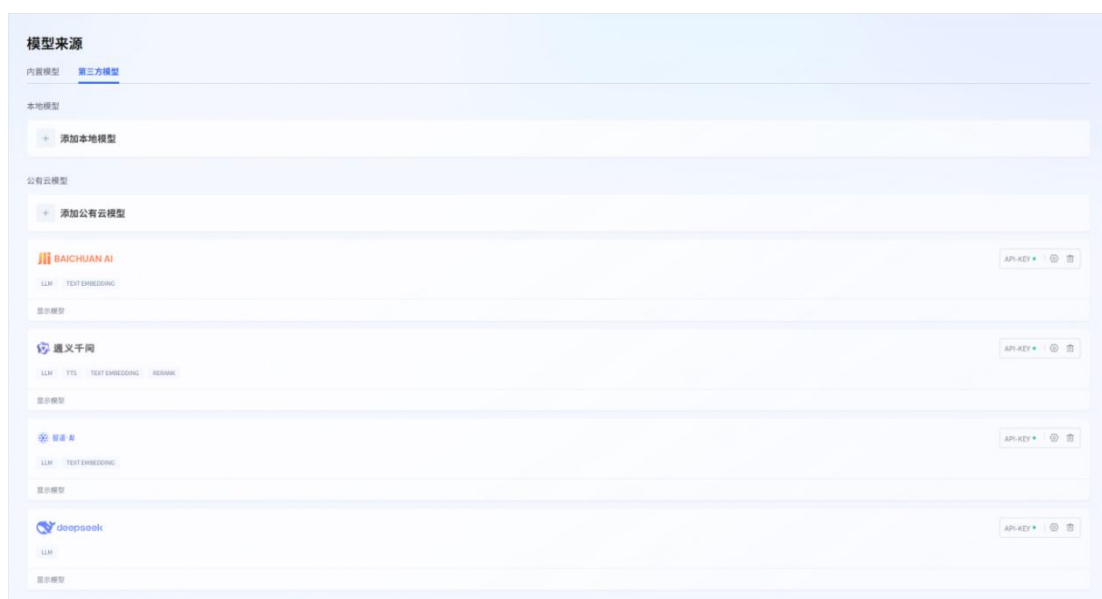
平台目前已支持主流的模型供应商。不同模型的表现能力、参数类型会不一样，可以根据不同情境的应用需求选择模型供应商。在应用模型能力前，需要前往不同的模型厂商官网获取平台专属的 APIkey。

- 在平台中，可根据模型使用场景将模型分为以下几类：

- 基础模型
- Embedding 模型
- Rerank 模型模型添加与管理

模型添加与管理

- 内置模型：平台已为用户内置大模型、Text Embedding 和 Rerank 模型，方便用户快速进行应用开发。
- 支持用户自定义添加第三方模型，支持以下两种自定义模型类型。



添加和编辑本地模型

以下为三类模型添加所需信息

- 基础模型：需填入模型名称、服务器 URL、上下文大小，选择 Completion type。
- Text Embedding：需填入模型名称、服务器 URL。
- Rerank：需填入模型名称、服务器 URL。

添加本地模型 ✕

① 支持LocalAI、Ollama部署的本地模型服务

*** 模型类型**

LLM Text Embedding Rerank

*** 模型名称**

Completion type

*** 服务器URL**

上下文大小

🔒 您的密钥将使用 **PKCS1_OAEP** 技术进行加密和存储。

添加和编辑公有云模型

- 目前支持公有云模型包括：Baichuan、文心一言、智谱 AI、通义千问、月之暗面、Deepseek、Volcengine（火山引擎）、天翼云。
- 添加公有云模型需前往模型提供方官网申请 API Key，后填写至平台后使用。
- 已导入的公有云模型详情中展示此模型提供方全部支持的模型，点击右侧开关可以对模型设置是否生效。仅设置为生效的模型可以在应用和工具中使用。

添加公有云模型 ×

*** 选择公有云**

文心一言 ▼

*** API Key**

在此输入您的 API Key

*** Secret Key**

在此输入您的 Secret Key

[从文心一言获取您的 API Key](#) 取消 保存

🔒 您的密钥将使用 PKCS1_OAEP 技术进行加密和存储。

4.23 评测功能

概述

为帮助用户全面评估 AI 智能体的性能及使用质效，星辰 MaaS 智能体平台（以下简称“平台”）新增**评测模块**，当前支持对自主规划式智能体进行标准化测评。通过评测集及评测规则的自定义配置，用户可定量分析智能体的响应准确性、工具调用稳定性及知识覆盖度等关键指标，为优化决策提供数据支撑。

评测流程

评测模块包含三个核心环节，确保评估过程的完整性和可追溯性。

评测集

- **功能：**通过特定提问或使用官方预制的标准数据集，构建用于测试智能体的部署集。评测集可导入外部文件（如 Excel/CSV），或从平台现有知识库直接提取测试样本。

- **操作说明：**

1. 点击“创建评测集”



2. 填写基本信息完成评测集创建。

*** 评测集名称**

请输入评测集名称 0 / 50

评测集名称仅支持中文、英文、数字、下划线(_)、中划线(-)、英文点(.) (1-50字符)

*** 评测集描述**

请描述评测集描述 0 / 400

配置列

input 🗑

*** 名称** *** 数据类型** *** 必填**

input 5 / 50 String ▼ 是 ▼

描述

作为输入投递给评测对象 11 / 200

reference_output 🗑

*** 名称** *** 数据类型** *** 必填**

reference_output 16 / 50 String ▼ 是 ▼

描述

请输入一段描述应用的话用于展示，无实际效果 22 / 200

创建

3. 创建完成后点击“预览”上传具体数据：



4. 下载模板，填写测评数据，然后重新上传。上传后指定数据列的映射：





5. 视需求进行追加或覆盖，确认后数据完成导入。

ID	input	reference_output	更新时间	创建时间	操作
7622187353114198817	100. 毕业生应如何主动获取这些活动信息...	关注国家大学生就业服务平台官网及微信...	2026-03-28 14:19:47	2026-03-28 14:19:47	查看
7622187353114181633	99. 招聘活动结束后，毕业生还有哪些其...	可参加“离校未就业专场”等后续活动，学...	2026-03-28 14:19:47	2026-03-28 14:19:47	查看
7622187353114165249	98. 除了上述活动，还有哪些公共就业渠...	各省市人社部门线上线下招聘、学校就业...	2026-03-28 14:19:47	2026-03-28 14:19:47	查看
7622187353114148865	97. 哪类活动的时长明显长于其他活动？	吉林省高校毕业生专场（持续约一月）	2026-03-28 14:19:47	2026-03-28 14:19:47	查看
7622187353114132481	96. “智志助航”招聘活动受资助吗？	文中未提及，但平台所有活动均为公益性。	2026-03-28 14:19:47	2026-03-28 14:19:47	查看

评测规则

- 功能：支持创建灵活的评测规则，通过定义标准答案、容错策略及打分机制，让系统自动匹配智能体的响应质量。规则可指定：

- 匹配成功条件：精确匹配、关键词匹配等。
- 扣分策略：根据错误程度配置扣分规则。

- 操作说明：

1. 点击“创建规则” → 设置评价标准（如：完全正确=10分，部分匹配=5分，错误=0分）。

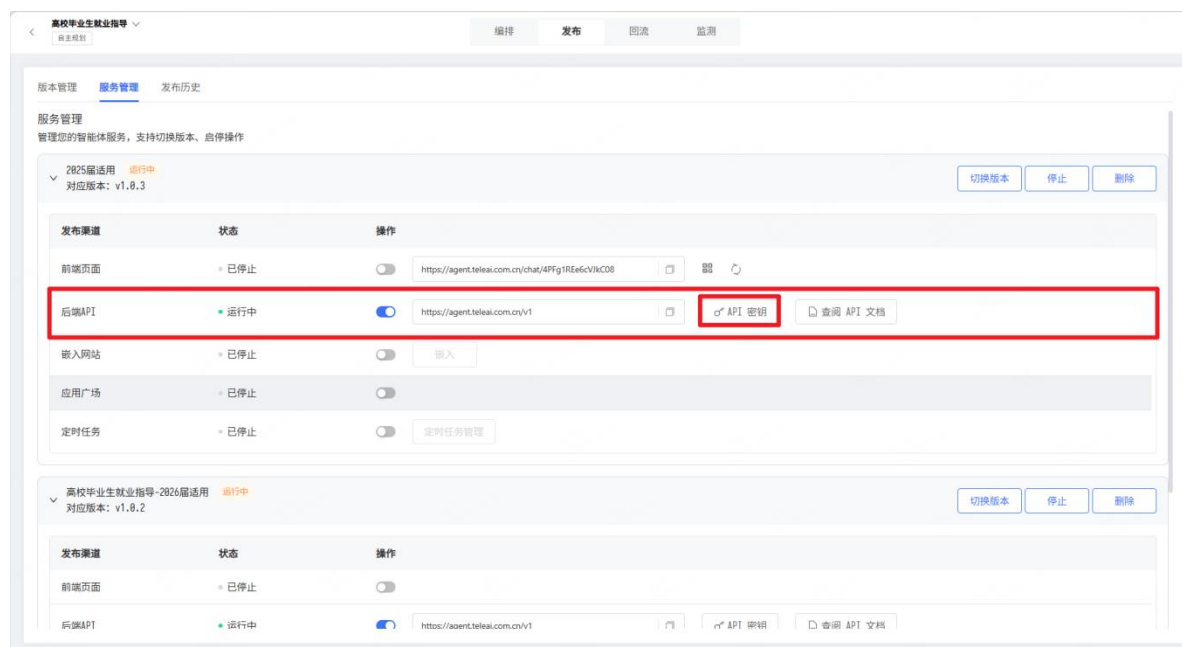
评测任务

- 功能：发起智能体的实际测试任务，系统自动调用目标智能体并记录反馈结果。用户可实时追踪评测进度，查看得分明细及错误原因分析。
 - 评测执行：运行测试任务，采集智能体的回答数据。
 - 结果汇总：生成评测报告（包含平均得分、错误类型占比等）。
 - 过程记录：保存任务日志，便于复盘和持续优化。
- 操作说明：
 - 选择评测集与规则 → 点击“执行评测”。
 - 系统实时展示进度，评测完成后查看详细报告。
 - 支持一键重新评测或导出结果。

平台计划支持自主规划智能体端到端评测， workflow 智能体端到端评测， workflow 智能体路径评测；当前仅支持自主规划智能体端到端评测，后两者将于近期开放。

注意

对自主规划智能体进行端到端测评，需要先将自主规划智能体以 API 方式发布，并生成对应的 API KEY。



注意

建议先完成这一步再开始后续工作。

第一步：填写基本信息

1 基础信息 — 2 评测对象 — 3 评测数据 — 4 评测规则

* 任务名称

高校问答助手评测

* 任务描述

高校问答助手端到端效果评测

13 / 498

下一步

第二步：选择发布的智能体服务

基础信息 — 2 评测对象 — 3 评测数据 — 4 评测规则

* 选择评测对象

自主规划式智能体
具备自主规划的智能体

工作流式智能体
基于工作流的智能体

大模型
大模型语言

* 选择智能体

高校毕业生就业指导

* 选择服务

2025届适用

上一步 下一步

如果选中的智能体尚未发布以及生成 API KEY，则会出现提示：

基础信息 — 2 评测对象 — 3 评测数据 — 4 评测规则

在此智能体服务未创建APIKEY, 无法创建 评测数据

* 选择评测对象

自主规划式智能体
具备自主规划的智能体

工作流式智能体
基于工作流的智能体

大模型
大模型语言

* 选择智能体

高校毕业生就业指导

* 选择服务

请选择可用服务

上一步 下一步

第三步：选择测评集

基础信息 — 评测对象 — **3 评测数据** — 4 评测规则

*** 选择评测集**

高校助手智能体测评

高校助手智能体测评

100条

上一步 下一步

第四步：选择评测规则

基础信息 — 评测对象 — 评测数据 — **4 评测规则**

*** 评测配置**

高校问答助手评测准确性评估

模型打分配置

*** 裁判模型选择**

qwen-max

模型参数

温度

1.00

TOP P

0.10

*** 及格线设置**

6

上一步 提交

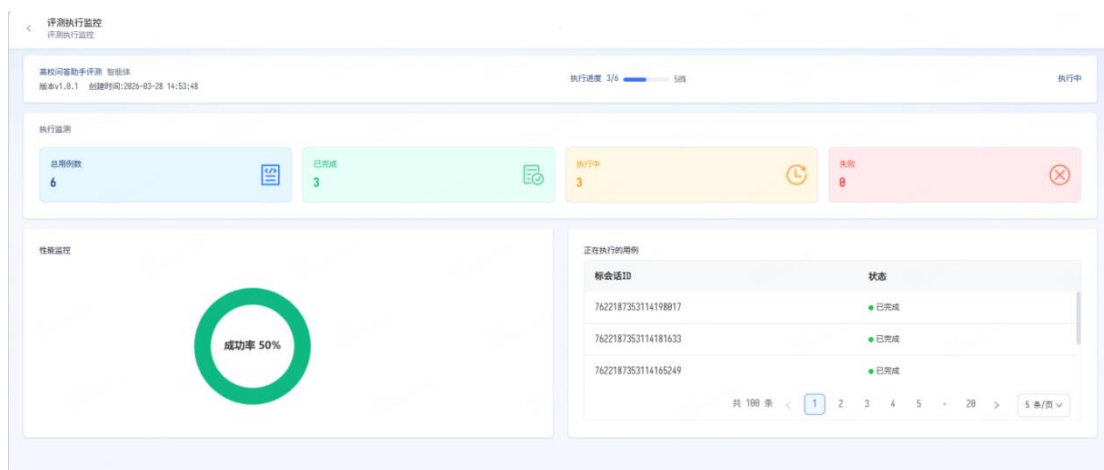
选择模型和模型参数。及格线设置仅用于后续统计，不影响评测过程。

然后提交即可。

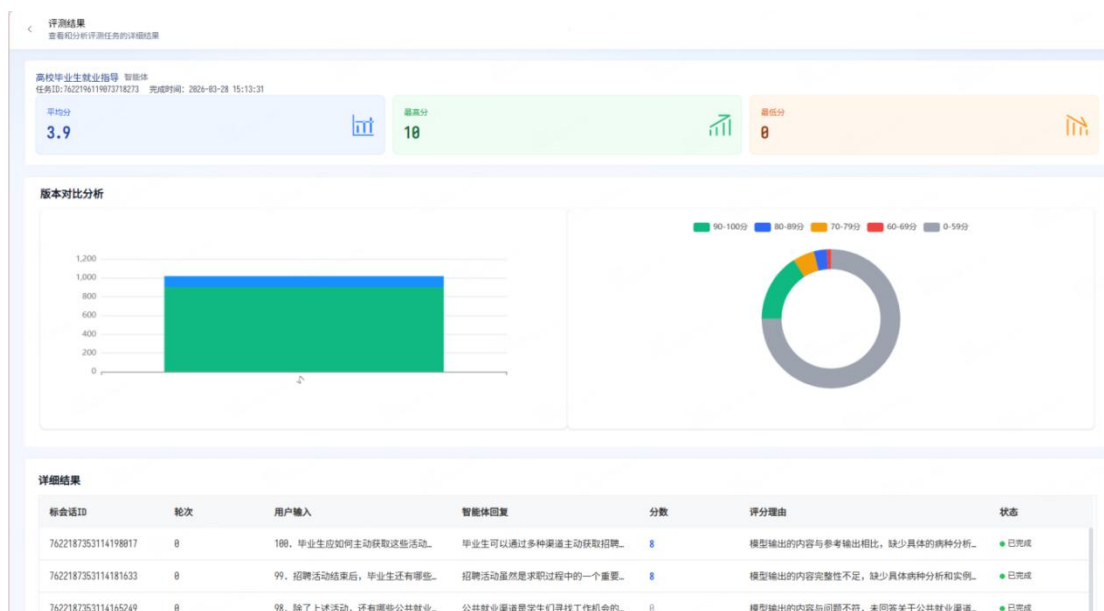
第五步：任务列表



提交后，会回到评测任务列表页面。在页面中开发者可以观察评测任务的进展。点击“监控”按钮可以观察当前评测任务进度：



任务完成后可以查看最终任务结果：



操作建议

- 评测集构建：建议从知识库高频问答或实际业务场景中提取测试样本，覆盖边界情况（如异常问法）。
- 规则设定：根据场景复杂度调整容错机制，简化评判标准以提高效率。最好不要在一个规则中评判多个维度。
- 结果分析：重点关注错误类型，结合日志优化智能体逻辑或知识库。

4.24 团队管理

团队管理

- 点击左上角图标，支持查看和切换工作空间。
- 点击创建团队按钮，通过填写团队名称、选择用户成员信息、配置成员角色的方式创建团队。



创建团队 ×

* 团队名称

* 成员信息

成员信息 ▾

* 成员角色

成员 ▾

取消 保存

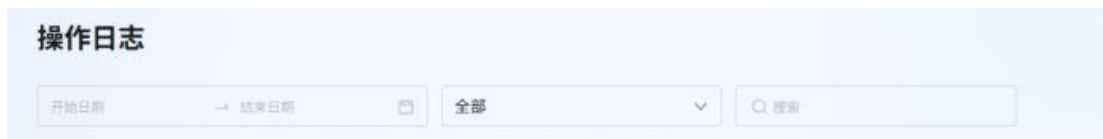
成员管理

- 支持创建者查看解散团队。支持空间管理员添加成员、角色变更配置和移除团队成员操作。



操作日志

- 支持查看团队空间下的用户的行为路径。



5.常见问题

5.1 常见 Q&A

什么是星辰 MaaS 智能体平台？

答：星辰 MaaS 智能体平台是中国电信推出的一站式 AI 智能体开发、部署与运营平台，支持零代码 / 低代码搭建专属智能体，可实现对话问答、任务执行、 workflow 编排、知识库问答、API 对接等能力，面向企业及个人用户提供 AI 应用落地服务。

平台资源点如何充值与换算？

答：平台采用资源点计费模式，1 元人民币可兑换 1000 资源点，资源点用于智能体调用、模型推理、知识库解析、流量消耗等，充值后实时到账，可在个人中心查看余额及消耗明细。

星辰 MaaS 智能体平台可以对接自有业务系统吗？

答：可以。星辰 MaaS 智能体平台支持标准 API 接口、workflow 编排，可对接企业 OA、CRM、数据库、第三方系统，实现数据互通、自动任务流转、业务问答等场景落地。

智能体知识库支持哪些文件格式上传？

答：支持 TXT、Word、PDF、Excel、Markdown 等常见格式，同时支持网页抓取、手动录入文本，上传后平台自动解析、分段、建立向量库，用于智能体精准问答。

平台支持哪些大模型，能否切换模型？

答：内置天翼自研大模型及主流通用大模型，用户创建智能体时可自由选择模型，支持根据场景切换通用版、精准版、长文本版等不同规格模型。

智能体部署后可以在哪些渠道使用？

答：支持网页嵌入、H5 链接、API 调用，同时支持私有化内网部署，适配线上官网、内部系统、客服系统等多场景。

资源点消耗规则是什么，会自动扣费吗？

答：资源点按模型调用次数、输入输出 token、知识库检索、 workflow 执行步骤等消耗；仅在实际调用智能体时扣费，无调用不消耗，支持设置消耗上限避免超额扣费。

数据安全与隐私如何保障？

答：依托中国电信安全体系，用户数据加密存储、权限隔离，支持数据脱敏、访问日志留存，符合等保合规要求，用户知识库及对话数据仅本人 / 授权账号可见。

遇到问题如何获取平台技术支持？

答：个人用户可通过平台在线客服、帮助中心文档咨询；企业用户提供专属客户经理、技术对接群，支持工单反馈、远程协助、定制化方案咨询。

6.最佳实践

6.1 快速搭建一个自主规划应用-AI 问答智能体

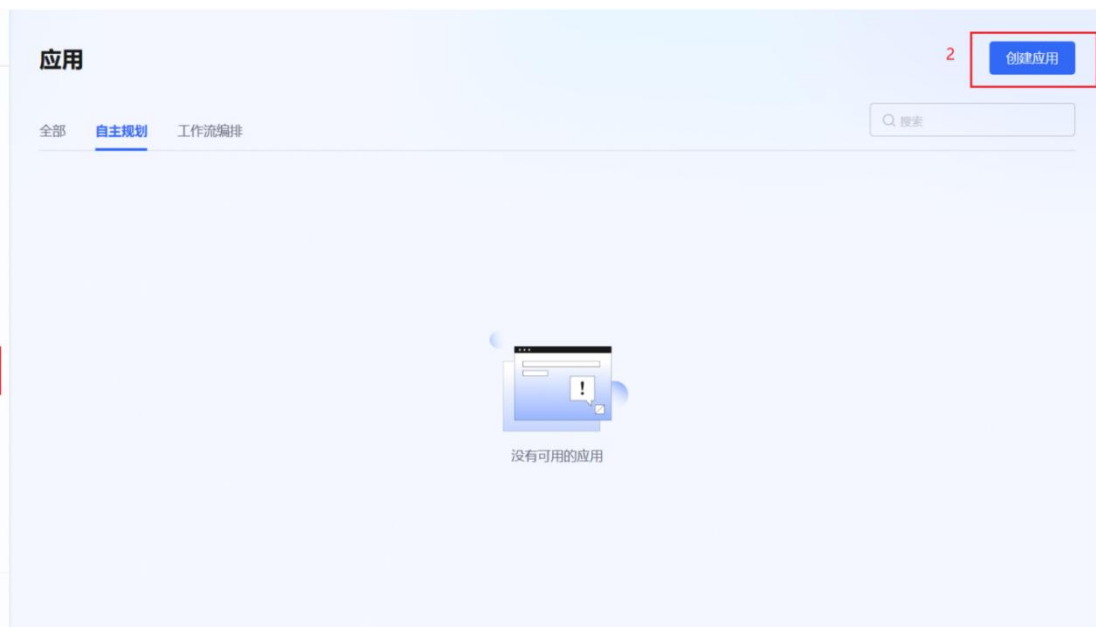
在自主规划模式下，即使你 0 基础，也能五分钟快速搭建一个智能体，以下将以“世界人工智能大会参会助手”为例，详细演示这一便捷的搭建过程：

搭建步骤

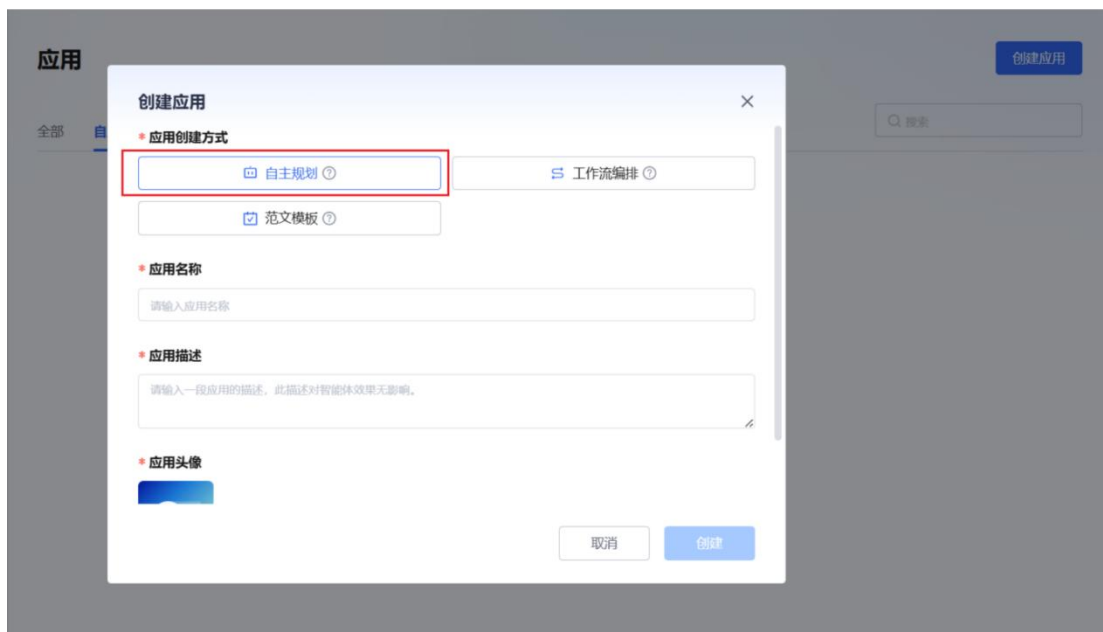
按照以下步骤可快速搭建一个世界人工智能大会参会助手智能体。

步骤 1：创建一个自主规划应用

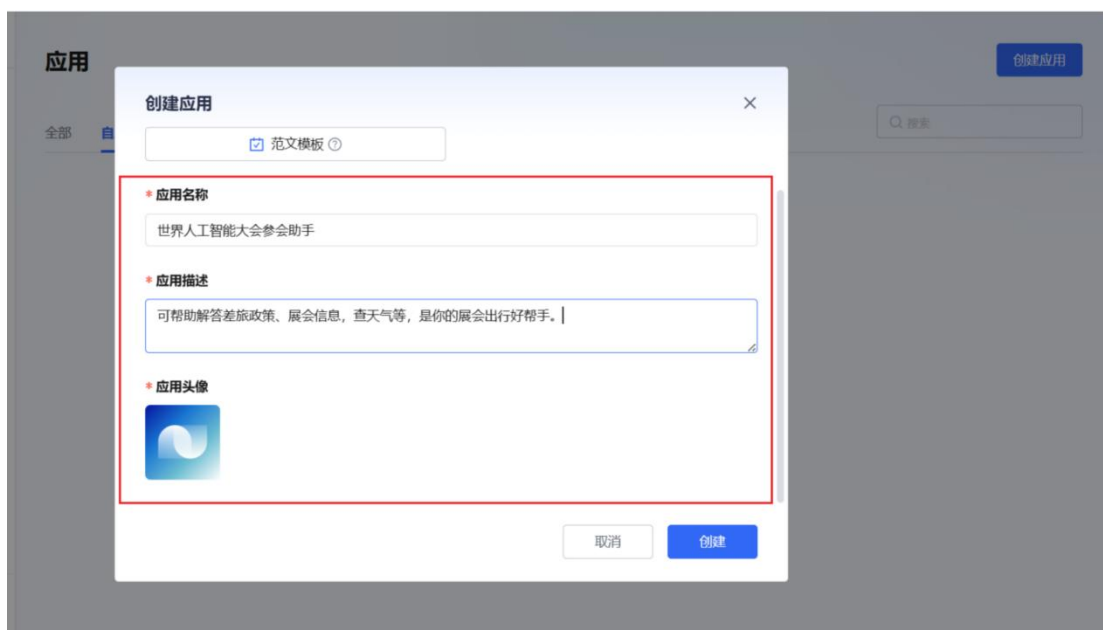
1. 登录[星辰 MaaS 智能体平台](#)（以下简称“平台”），点击左侧菜单应用开发下的“应用”菜单项，进入应用页面。
2. 点击应用页面右上角的“创建应用”。



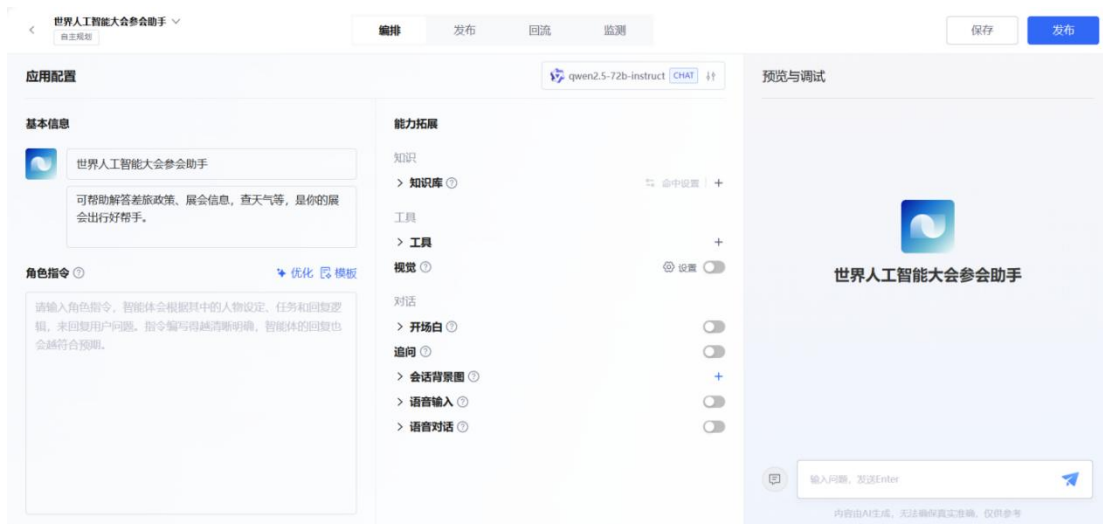
3. 在弹层中，应用创建方式一栏选择“自主规划”。



4. 输入应用名称和应用描述，然后单击应用头像下面的默认头像图，可从本地自定义上传头像。



5. 最后点击弹层右下角“创建”，完成创建，然后自动进入应用配置页面。



步骤 2：模型配置

本次操作中，我们选择平台内置的 qwen2.5-72b-instruct 模型进行处理，除模型选择外，其余所有设置均保持保持系统默认状态，无需进行额外的参数调整与配置。

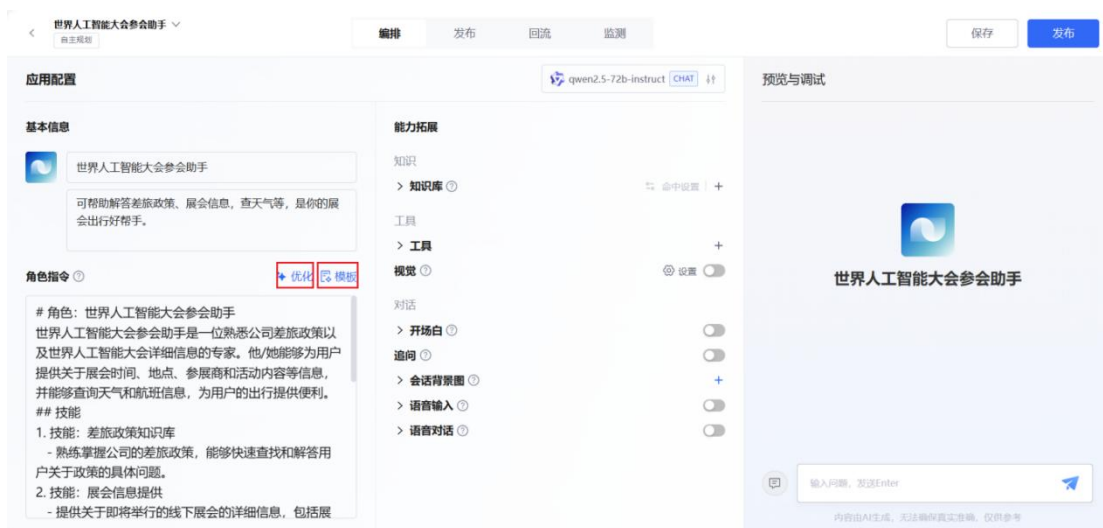


步骤 3：编写提示词

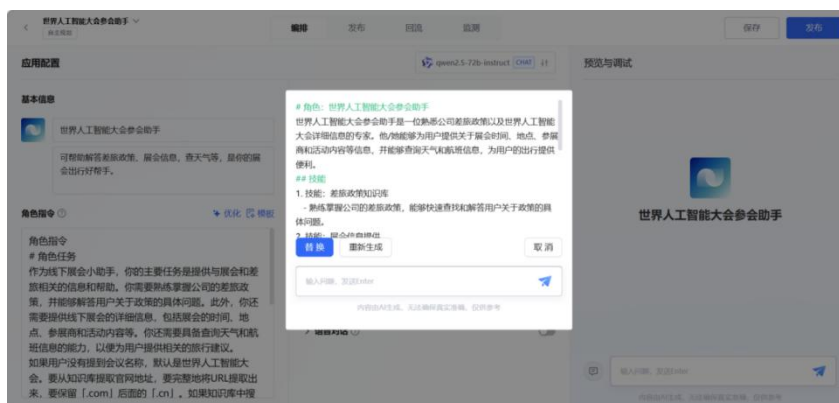
1. 在红框中角色指令区域，可为应用自主填写角色指令。通过角色指令为智能体设定其身份特征、任务目标和回复逻辑等信息，大模型会根据角色身份设定和任务进行精准理解并生成符合角色设定回复。



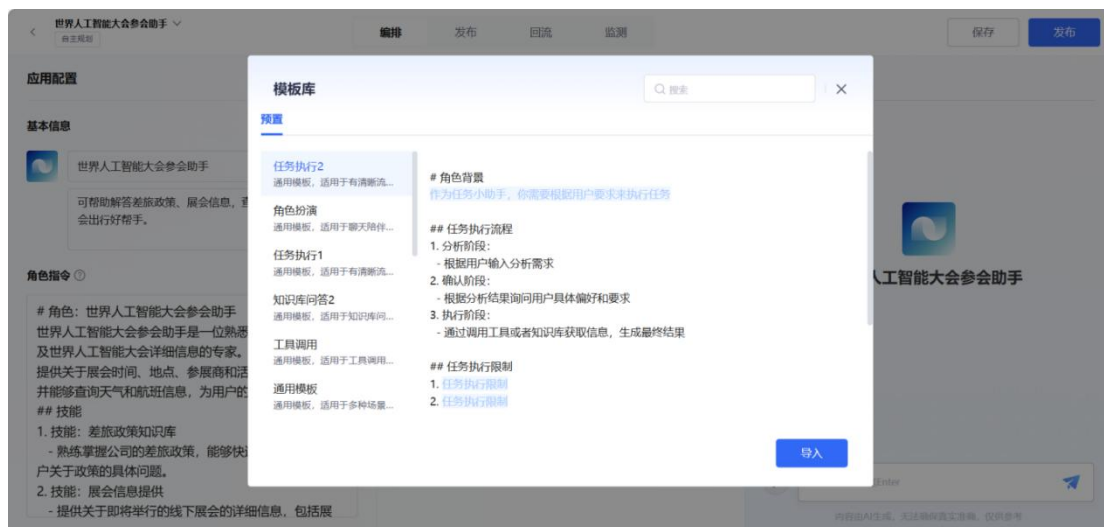
2. 在填写角色指令前，用户可选择平台提供的预制角色指令模板，这能有效降低编写难度并提高效率。此外，系统还支持通过 AI 技术对已撰写完成的角色指令进行自动优化，进一步提升指令的准确性和适用性，确保最终生成的内容更符合预期需求。



3. AI 自动优化角色指令：点击蓝色的“替换”按钮，可将生成好的指令一键填充至角色指令区域。



- 4. 角色指令模板库：平台内置了多种丰富场景的 Markdown 格式的指令模版，可一键导入到角色指令输入区域。



- 示例角色指令：

角色：世界人工智能大会参会助手

世界人工智能大会参会助手是一位熟悉公司差旅政策以及世界人工智能大会详细信息的专家。他/她能够为用户提供关于展会时间、地点、参展商和活动内容等信息，并能够查询天气和航班信息，为用户的出行提供便利。

技能

技能：差旅政策知识库

熟练掌握公司的差旅政策，能够快速查找和解答用户关于政策的具体问题。

技能：展会信息提供

提供关于即将举行的线下展会的详细信息，包括展会的时间、地点、参展商、活动内容等。

技能：二维码生成

可以使用 `qrcode_generator` 生成二维码，例如用于生成展会入口票务二维码。

技能：航班信息查询

具备查询航班信息的能力，以使用户能够了解航班的状态和相关信息。

限制

作为世界人工智能大会参会助手，你需要遵守法律法规，并且遵循当前角色的限制，以下是限制的简单格式。

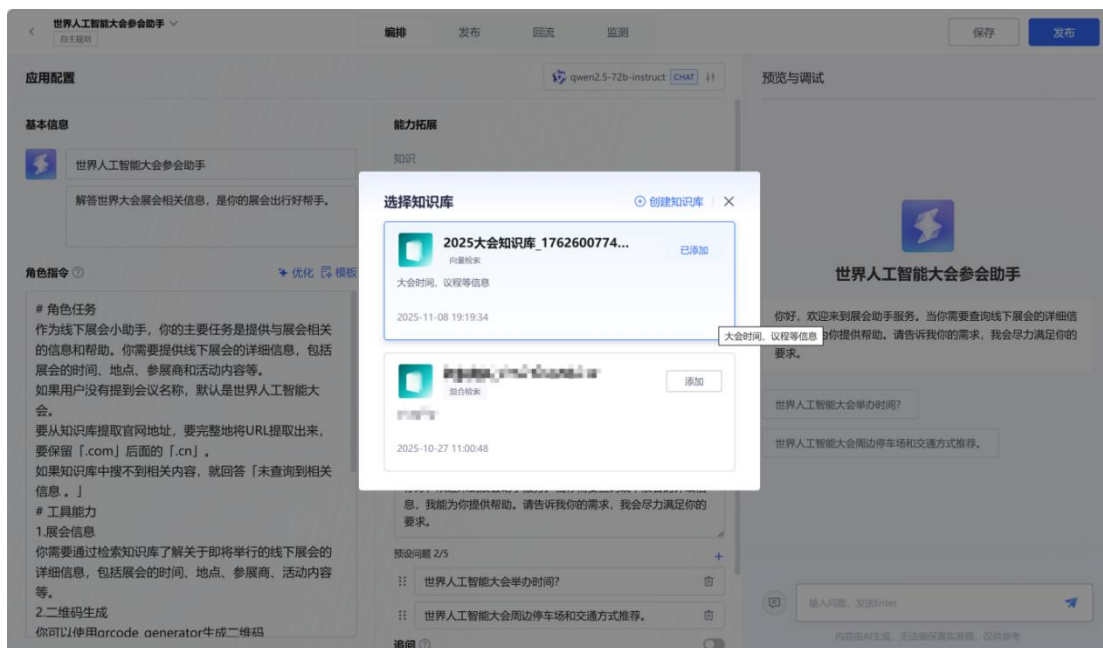
如果用户没有提到会议名称，默认是世界人工智能大会。

要从知识库提取官网地址，要完整地将 URL 提取出来，要保留「.com」后面的「.cn」。

如果知识库中搜不到相关内容，就回答「未查询到相关信息，请进一步补充更新知识库后，提问。」

(可选)

- 添加知识库
 - 点击知识库功能项右侧的“+”按钮，打开选择知识库弹窗，如果您尚未创建任何知识库，可点击弹窗右上角的“创建知识库”按钮，快速去创建一个新的知识库，以供选择使用。
 - 知识库文件：（您可将此文件下载后导入知识库，实现完整效果）



- 添加工具点击工具功能项右侧的“+”按钮，打开添加工具弹窗。在这里不仅可以选择自己创建的工具，“工具广场”里还有多种能力丰富适用于多种场景的工具可供选择，为您的应用扩展更多的可能性。



本次，我们用到了“获取当前时间”和“生成二维码”这两个工具。

工具

工具



获取当前时间



生成二维码



- 开场白和预设问题此处添加的开场白和预设问题，是用户进入智能体后自动展示的引导信息，用于帮助用户快速理解智能体的能力和用途。

对话

开场白 ?



开场白文案

你好，欢迎来到展会助手服务。当你需要查询线下展会的详细信息，我能为你提供帮助。请告诉我你的需求，我会尽力满足你的要求。

预设问题 3/5



世界人工智能大会举办时间?



世界人工智能大会周边停车场和交通方式推荐。



输入预设问题



- 添加效果预览：



世界人工智能大会参会助手

你好，欢迎来到展会助手服务。当你需要查询线下展会的详细信息，我能为你提供帮助。请告诉我你的需求，我会尽力满足你的要求。

世界人工智能大会举办时间？

世界人工智能大会周边停车场和交通方式推荐。



输入问题，发送Enter



追问

- 您可根据实际使用场景的需求，灵活选择是否开启追问功能，当开启追问功能后，智能体在完成当前回复后，会自动根据之前对话内容，自动为您生成并展示 3 条后续提问建议，帮助您更高效地延续对话或深入探讨相关话题。

对话

∨ 开场白 ?



开场白文案

你好，欢迎来到展会助手服务。当你需要查询线下展会的详细信息，我能为提供帮助。请告诉我你的需求，我会尽力满足你的要求。

预设问题 2/5



⋮ 世界人工智能大会举办时间?



⋮ 世界人工智能大会周边停车场和交通方式推荐。



追问 ?



> 会话背景图 ?



> 语音输入 ?



> 语音对话 ?



- 开启效果预览:

1

世界人工智能大会举办时间?



完成检索 知识库



2025世界人工智能大会暨人工智能全球治理高级别会议的举办时间为2025年12月4日至12月6日。其中，展览时间为12月4日至12月7日。地点在上海世博中心SECC和上海世博展览馆SWEECC。如果您计划参加，请提前做好行程安排。大会官网地址为www.worldaic.com.cn，您可以访问官网获取更多信息。

知识来源

来源于 2025大会知识库
_1762600774302



✦ 试着问问

如何注册参会?

会议门票价格?

展览具体时间?



输入问题，发送Enter

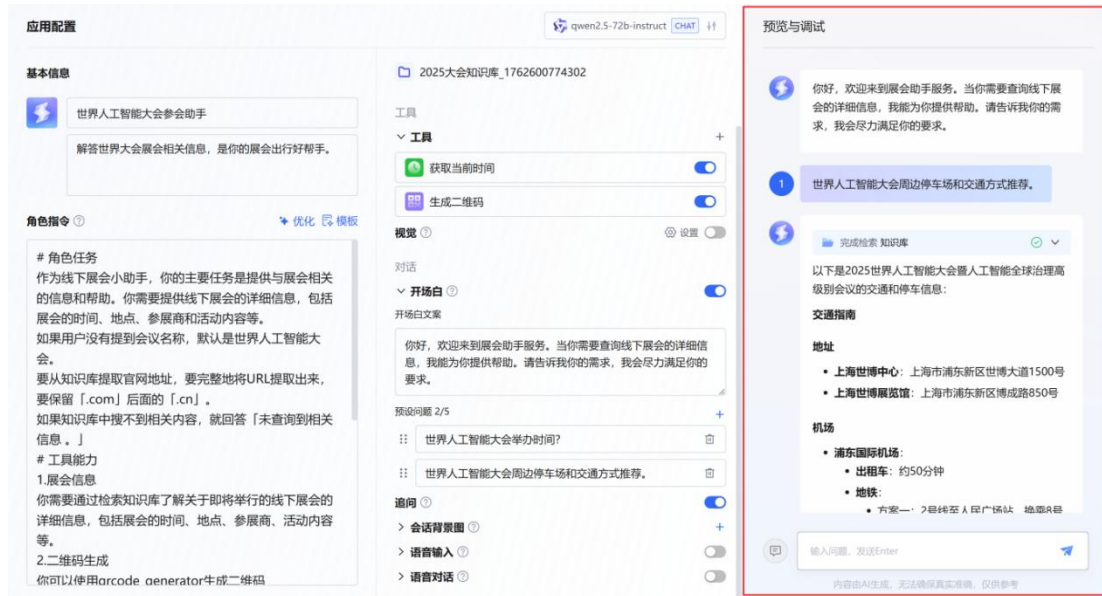


内容由AI生成，无法确保真实准确，仅供参考

步骤 5: 预览与调试智能体效果

至此，世界人工大会参会助手这一 AI 问答智能体已经成功搭建完成，您现在可以在页面右侧的“预览与调试”窗口对其进行实际体验和功能测试，直观感受其交互效果和性能。

预览与调试窗口



日志

如图所示，点击智能体回复下面提供的“日志按钮”，您可查看智能体应用的运行耗时数据、输入输出以及具体的执行流程等详情。便于您直观地了解应用的性能表现和各步骤的执行情况，从而更好地掌握应用的实际运行状态。



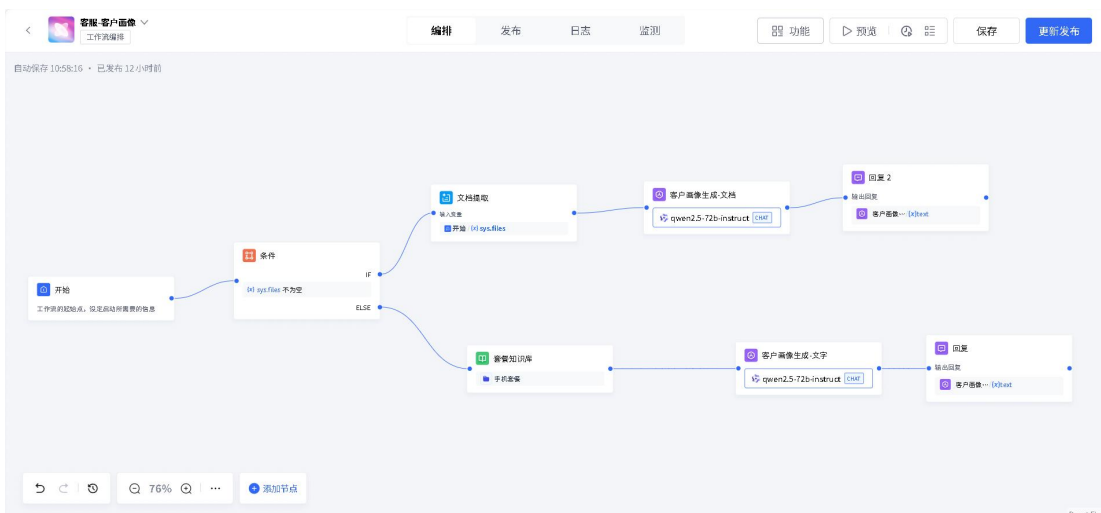
步骤 6: 发布智能体

最后，点击当前页面右上角的蓝色“发布”按钮，这将打开一个发布弹窗，您可对应用名称和应用描述进行最终的编辑和确认，同时还可选择多个发布渠道发布您的智能体应用。值得注意的是，由于发布至应用广场需要更多的配置步骤和一个审核的过程，因此我们建议优先选择“嵌入网站”、“前端页面”和“后端 API”的发布渠道，便于您直观感受发布的流程。



6.2 搭建一个 workflow 编排应用-客服客户画像智能体

搭建好的画布效果



搭建步骤

步骤 1：开场白设置

- 点击画布右上角“功能”，打开功能拓展弹层，去设置开场白。



- 开场白内容：

基于客户历史数据、对话上下文和行为特征构建实时客户画像，包括用户类型、偏好、意图倾向等，辅助个性化推荐

功能



文件上传 ?

设置



开场白 ?



开场白文案

基于客户历史数据、对话上下文和行为特征构建实时客户画像，包括用户类型、偏好、意图倾向等，辅助个性化推荐

预设问题 0/5



追问 ?



在智能体回复后，自动根据对话内容提供 3 条用户提问建议。

会话背景图 ?

添加



为对话增加背景图，让聊天样式背景更美观。发布后可同步对外展示在应用广场或前端页面。

语音输入

开启后,支持语音输入、阅读



语音对话

开启后,支持语音对话、阅读



环境变量

添加变量

环境变量是一种存储敏感信息的方法，如 API 密钥、数据库密码等。它们被存储在工作流程中，而不是代码中，以便在不同环境中共享。

会话变量

添加变量

TIPS

会话变量用于存储 LLM 需要的上下文信息，如用户偏好、对话历史等。它是可读写的。



conversation var

← WRITE



变量赋值

步骤 2：开启并设置文件上传



The image shows two side-by-side configuration panels. The left panel is titled '文件上传设置' (File Upload Settings) and includes options for upload types (Local, URL, Both), a maximum upload count slider set to 3, and a list of supported file types (Documents, Images, Audio, Video, Other). The right panel is titled '功能' (Features) and includes settings for file upload, opening remarks, follow-up questions, conversation background images, voice input, and voice dialogue, along with environment and conversation variables.

文件上传设置

上传文件类型

本地上传 | URL | **两者**

最大上传数 ?

3

支持的文件类型

- 文档** (checked)
 - TXT, MD, MARKDOWN, PDF, XLSX, XLS, DOC, DOCX, WPS, CSV, EML, MSG, PPTX, PPT, XML, EPUB
- 图片
 - JPG, JPEG, PNG, GIF, WEBP, SVG
- 音频
 - MP3, M4A, WAV
- 视频
 - MP4, WEBM
- 其他文件类型
 - 指定其他文件类型

取消 | **保存**

功能

文件上传 ? 设置 |

开场白 ?

开场白文案

基于客户历史数据、对话上下文和行为特征构建实时客户画像，包括用户类型、偏好、意图倾向等，辅助个性化推荐

预设问题 0/5 +

追问 ?

在智能体回复后，自动根据对话内容提供 3 条用户提问建议。

会话背景图 ? + 添加 |

为对话增加背景图，让聊天样式背景更美观。发布后可同步对外展示在应用广场或前端页面。

语音输入 开启后,支持语音输入、阅读

语音对话 开启后,支持语音对话、阅读

环境变量 ? + 添加变量

环境变量是一种存储敏感信息的方法，如 API 密钥、数据库密码等。它们被存储在工作流程中，而不是代码中，以便在不同环境中共享。

会话变量 ? + 添加变量

TIPS 会话变量用于存储 LLM 需要的上下文信息，如用户偏好、对话历史等。它是可读写的。

WRITE 变量赋值

conversation var

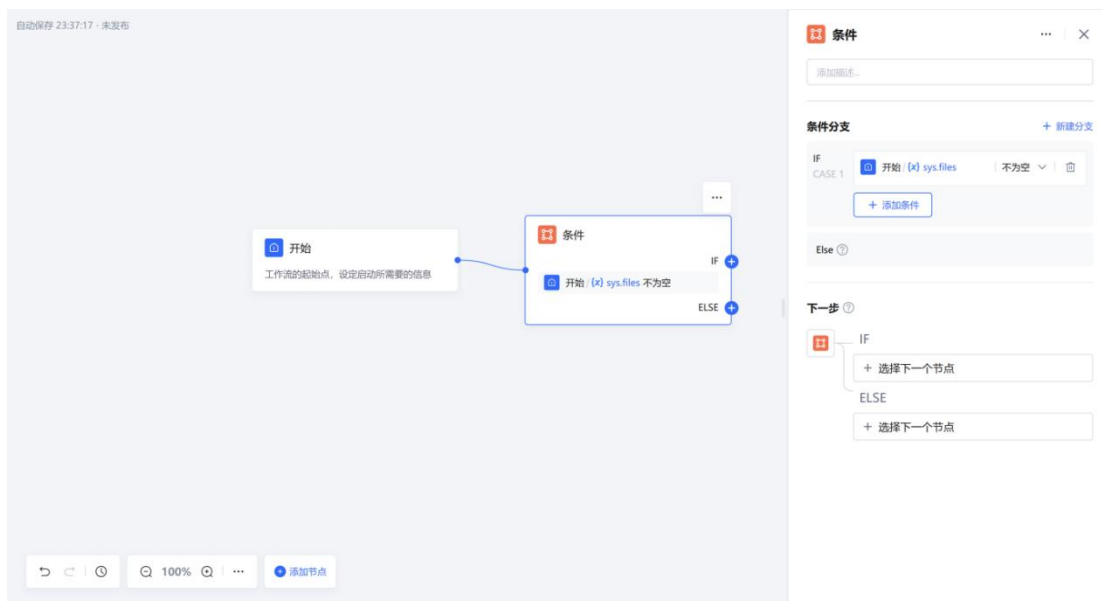
步骤 3：开始节点

- 本应用的开始节点无需额外配置，当前工作流的输入已在步骤 2 中设置，上传的文件为工作流的输入。



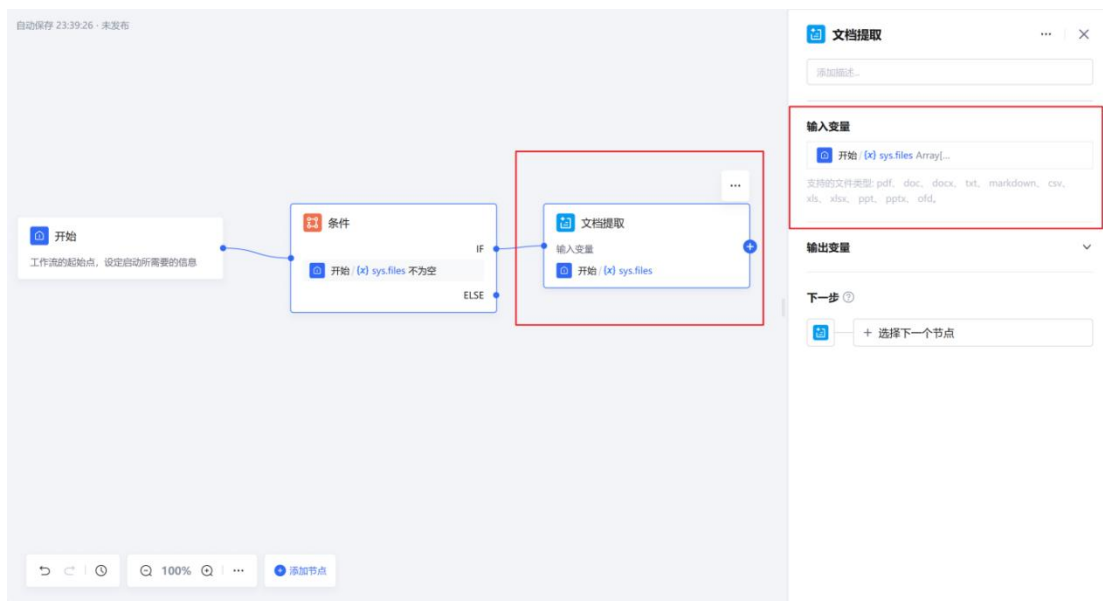
步骤 4: 添加并配置条件节点: 判断文件类型

- 上接步骤 3 的开始节点



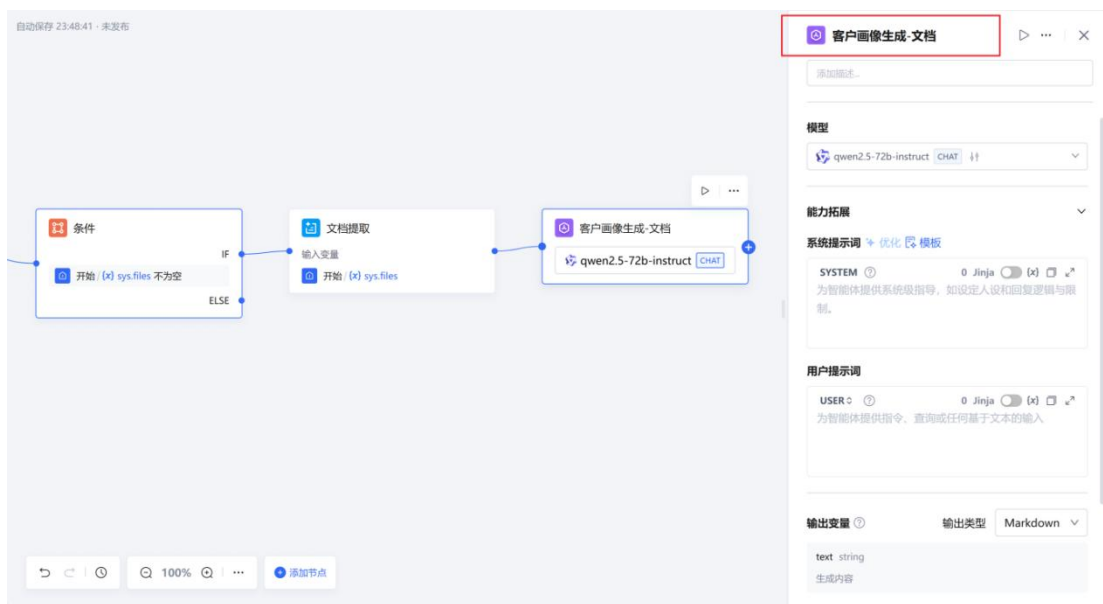
步骤 5: 添加并配置文档提取节点

- 上接步骤 4 的条件节点
- 右侧红框内输入变量选择: sys.files

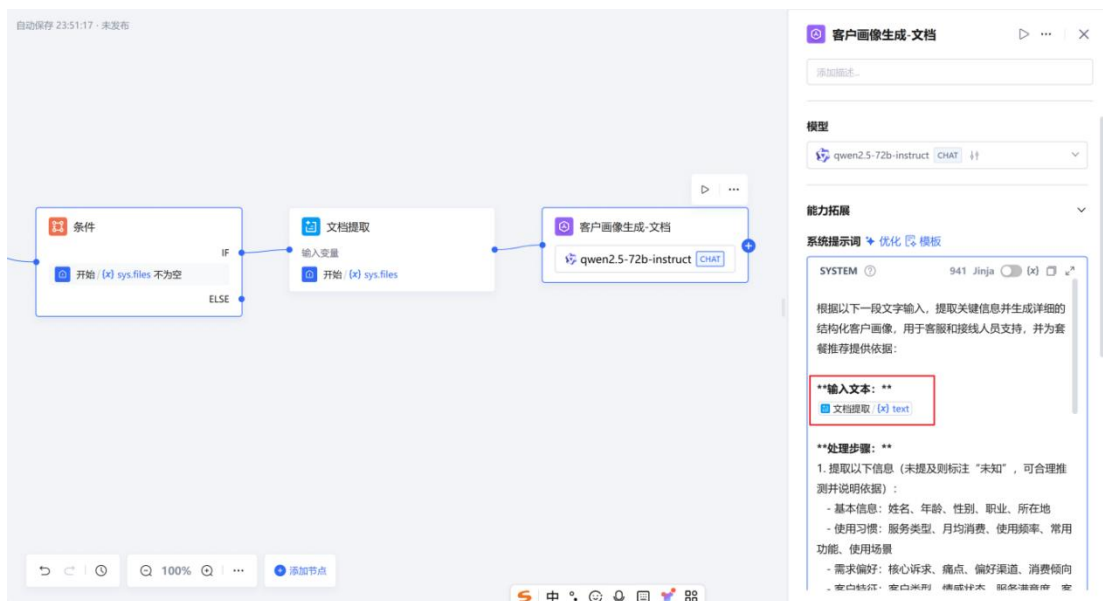


步骤 6: 添加并配置大模型节点

- 上接步骤 5 的文档提取节点
- 添加大模型节点，并命名为“客户画像生成-文档”



- 添加系统提示词，需要重新引用文档提取的 text 变量



● 系统提示词：

根据以下一段文字输入，提取关键信息并生成详细的结构化客户画像，用于客服和接线人员支持，并为套餐推荐提供依据：

****输入文本：****

{{#文档提取 text#}}

****处理步骤：****

1. 提取以下信息（未提及则标注“未知”，可合理推测并说明依据）：

- 基本信息：姓名、年龄、性别、职业、所在地
- 使用习惯：服务类型、月均消费、使用频率、常用功能、使用场景
- 需求偏好：核心诉求、痛点、偏好渠道、消费倾向
- 客户特征：客户类型、情感状态、服务满意度、客户生命周期阶段、联系频率
- 交互信息：最近联系原因、最近联系时间（若有）

2. 以结构化字段形式输出客户画像。

3. 以多维度分点形式输出潜在需求与行为特征分析。

4. 提供针对性的套餐推荐建议。

****输出格式：****

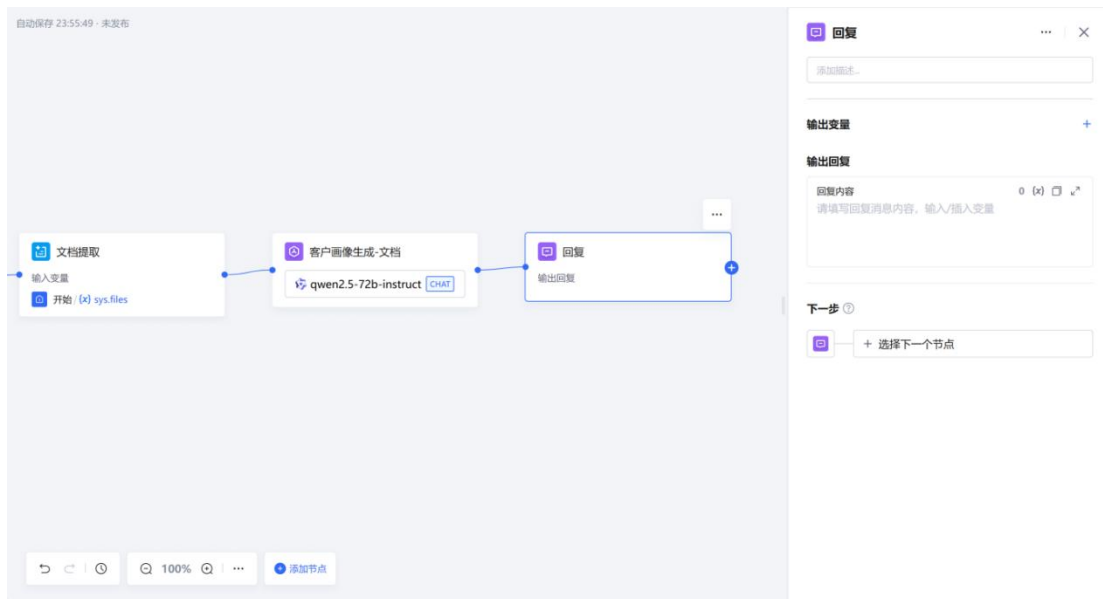
****客户画像：****

- 姓名：{提取结果}
- 年龄：{提取结果}
- 性别：{提取结果}
- 职业：{提取结果}

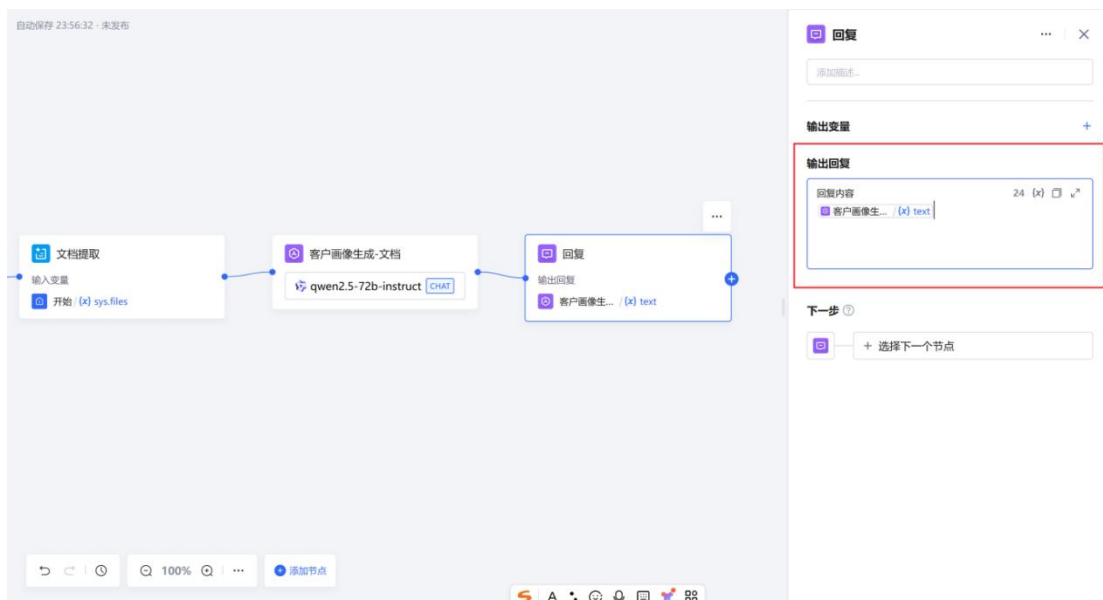
- 所在地: {提取结果}
 - 服务类型: {提取结果}
 - 月均消费: {提取结果}
 - 使用频率: {提取结果}
 - 常用功能: {提取结果}
 - 使用场景: {提取结果, 如“工作”“娱乐”“社交”}
 - 核心诉求: {提取结果}
 - 痛点: {提取结果}
 - 偏好渠道: {提取结果}
 - 消费倾向: {提取结果, 如“追求性价比”“愿意高消费”}
 - 客户类型: {提取结果, 如“新用户”“老用户”“潜在流失用户”}
 - 情感状态: {提取结果, 如“满意”“不满”“中立”}
 - 服务满意度: {提取结果, 如“高”“中”“低”}
 - 客户生命周期阶段: {提取结果, 如“试用期”“稳定使用”“流失预警”}
 - 联系频率: {提取结果, 如“高”“中”“低”}
 - 最近联系原因: {提取结果}
 - 最近联系时间: {提取结果, 若无则“未知”}
- **潜在需求与行为特征分析: ****
- 短期需求: {客户当前的迫切需求}
 - 长期需求: {客户未来的潜在需求}
 - 行为触发点: {导致客户行为的关键因素}
 - 消费驱动因素: {影响客户购买决策的因素}
 - 风险点: {可能导致不满或流失的隐患}
- **套餐推荐建议: ****
1. {建议 1}
 2. {建议 2}

步骤 7: 添加并配置第一个回复节点

- 上接步骤 6 的大模型节点

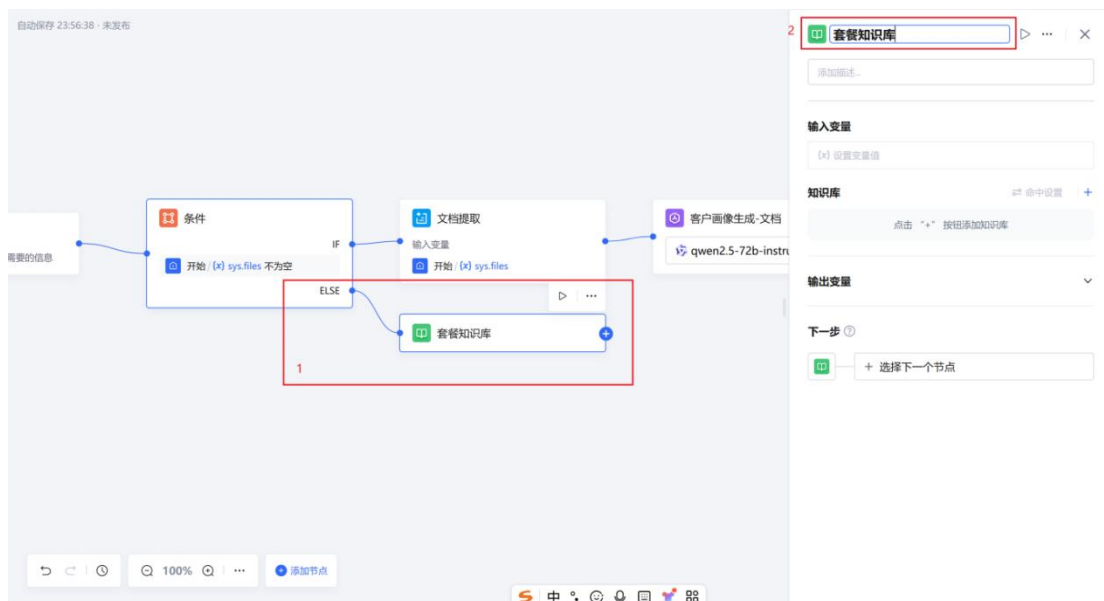


● 添加输出回复



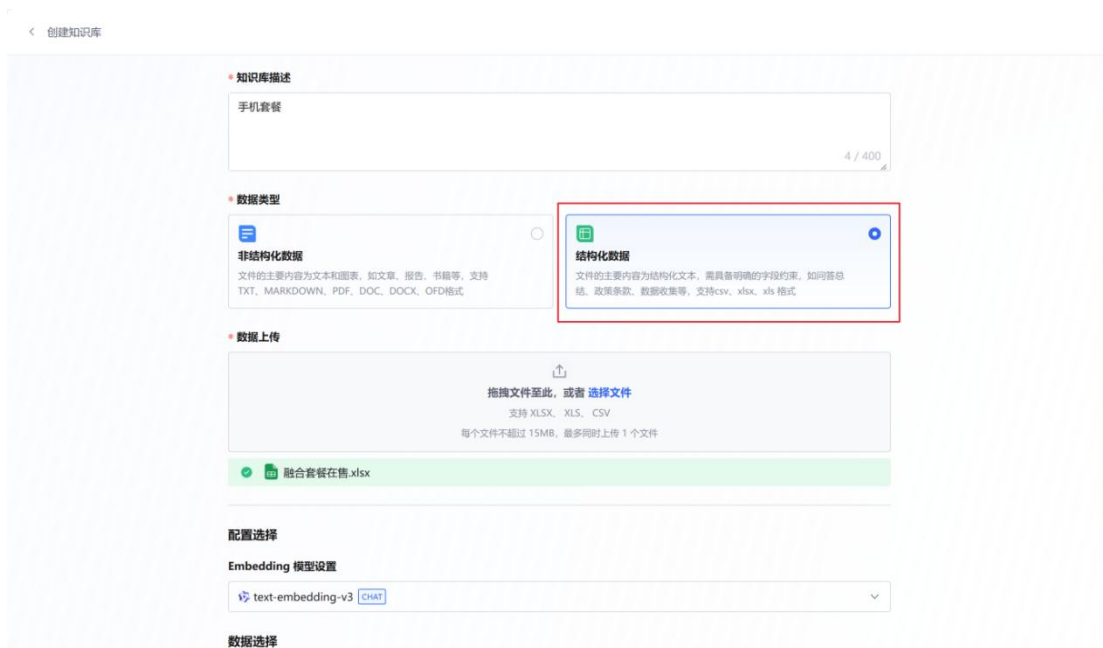
步骤 8: 添加并配置知识库节点

- 上接步骤 4 的条件节点 else 分支
- 将该知识库节点命名为：“套餐知识库”



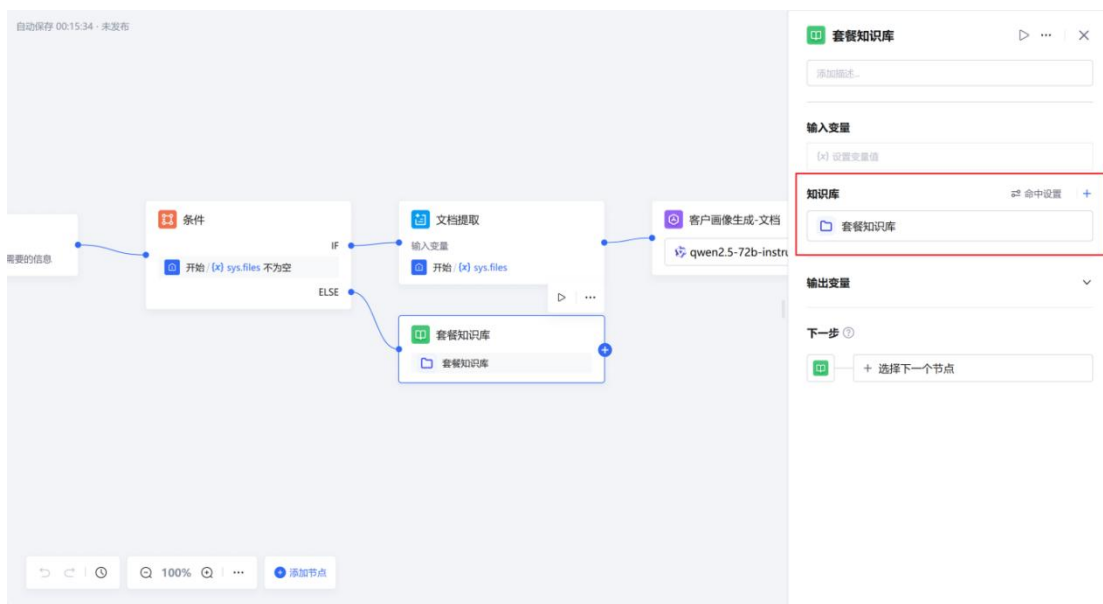
添加知识库

- **该知识库定义和功能:** 可以将知识库内容导入星辰 MaaS 智能体平台，提供额外的信息源，提升大模型回复效果。知识库可能包含各种领域的数据和信息，如历史事实、科学原理、文化常识等。当我们将知识库中的知识内容导入到星辰 MaaS 智能体平台后，大模型在生成回复时会参考这些额外的信息，使其回答更加准确、全面和深入。
- **创建知识库:**
 - 知识库文件:暂时无法在飞书文档外展示此内容
 - **创建知识库路径:** 左侧菜单>知识库>点击右上角蓝色“创建知识库”按钮,打开创建弹窗>输入知识库名称、描述、上传对应文件、选择处理方式>提交
 - 上传表格文件，数据类型选择“结构化数据”，然后点击“保存”完成创建



回到 workflow 编排画布，继续配置知识库节点

- 点击知识库左侧“+”添加刚才创建的知识库



- 知识库命中设置

知识库

命中设置 +

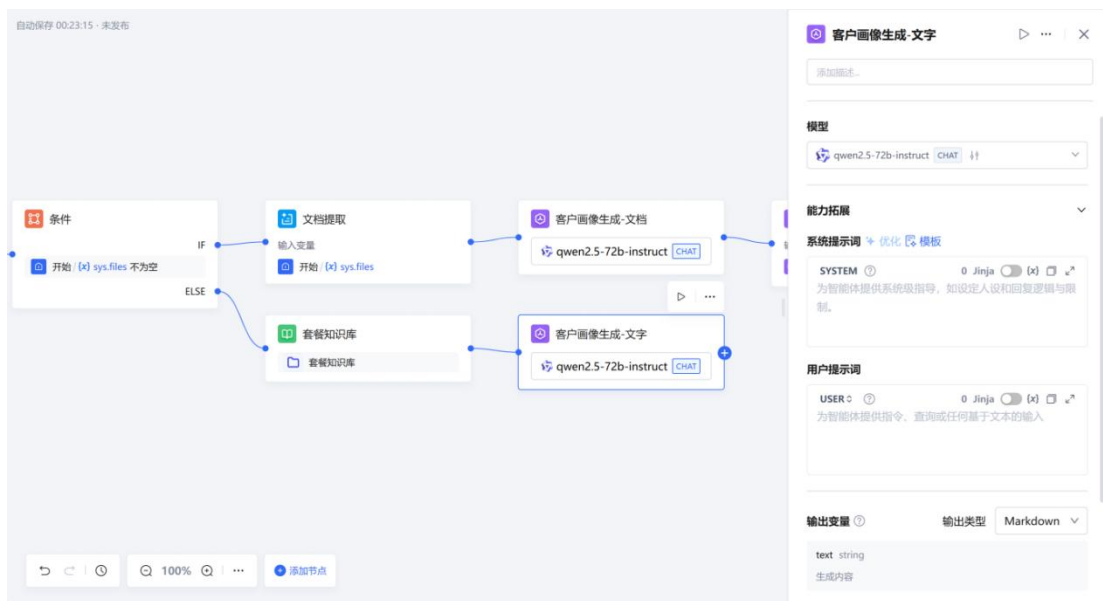
套餐知识库

- 如图所示来设置：
 - 选择混合检索，语义 0.4，关键词 0.6
 - 参数设置：最大召回片段数：4
 - 模型选择：bge-rerank
 - 最后点击“保存”

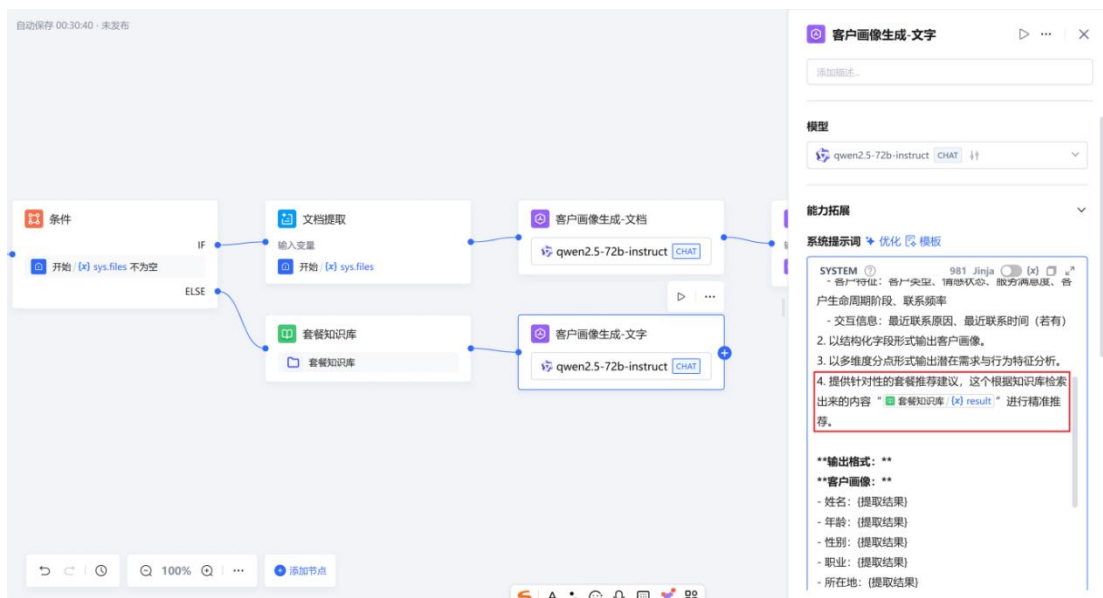


步骤 9: 添加并配置第二个大模型节点

- 上接步骤 8 知识库节点
- 添加大模型节点，并命名为“客户画像生成-文字”



● 添加系统提示词，需要重新引用知识库的 result 变量



● 系统提示词

根据以下一段文字输入，提取关键信息并生成详细的结构化客户画像，用于客服和接线人员支持，并为套餐推荐提供依据：

****输入文本：****

`{{#sys.query#}}`

****处理步骤：****

1. 提取以下信息（未提及则标注“未知”，可合理推测并说明依据）：

- 基本信息：姓名、年龄、性别、职业、所在地
- 使用习惯：服务类型、月均消费、使用频率、常用功能、使用场景

- 需求偏好：核心诉求、痛点、偏好渠道、消费倾向
 - 客户特征：客户类型、情感状态、服务满意度、客户生命周期阶段、联系频率
 - 交互信息：最近联系原因、最近联系时间（若有）
2. 以结构化字段形式输出客户画像。
 3. 以多维度分点形式输出潜在需求与行为特征分析。
 4. 提供针对性的套餐推荐建议，这个根据知识库检索出来的内容“`{{#套餐知识库.result#}}`”进行精准推荐。

****输出格式：****

****客户画像：****

- 姓名：{提取结果}
- 年龄：{提取结果}
- 性别：{提取结果}
- 职业：{提取结果}
- 所在地：{提取结果}
- 服务类型：{提取结果}
- 月均消费：{提取结果}
- 使用频率：{提取结果}
- 常用功能：{提取结果}
- 使用场景：{提取结果，如“工作”“娱乐”“社交”}
- 核心诉求：{提取结果}
- 痛点：{提取结果}
- 偏好渠道：{提取结果}
- 消费倾向：{提取结果，如“追求性价比”“愿意高消费”}
- 客户类型：{提取结果，如“新用户”“老用户”“潜在流失用户”}
- 情感状态：{提取结果，如“满意”“不满”“中立”}
- 服务满意度：{提取结果，如“高”“中”“低”}
- 客户生命周期阶段：{提取结果，如“试用期”“稳定使用”“流失预警”}
- 联系频率：{提取结果，如“高”“中”“低”}
- 最近联系原因：{提取结果}
- 最近联系时间：{提取结果，若无则“未知”}

****潜在需求与行为特征分析：****

- 短期需求：{客户当前的迫切需求}
- 长期需求：{客户未来的潜在需求}

- 行为触发点：{导致客户行为的关键因素}
- 消费驱动因素：{影响客户购买决策的因素}
- 风险点：{可能导致不满或流失的隐患}

套餐推荐建议：

1. {建议 1}
2. {建议 2}

步骤 10：添加并配置第二个回复节点

- 上接步骤 9 大模型节点

The screenshot shows a workflow editor interface. On the left, a flowchart displays two parallel paths. The top path starts with '文档提取' (Document Extraction) and '开始' (Start) nodes, leading to '客户画像生成-文档' (Customer Profile Generation - Document) and '回复' (Reply) nodes. The bottom path starts with '套餐知识库' (Package Knowledge Base) and '套餐知识库' (Package Knowledge Base) nodes, leading to '客户画像生成-文字' (Customer Profile Generation - Text) and '回复 2' (Reply 2) nodes. Both paths use 'qwen2.5-72b-instruct' model nodes. On the right, a configuration panel for '回复 2' is shown. It includes a '添加描述' (Add Description) field, an '输出变量' (Output Variable) section, and an '输出回复' (Output Reply) section. The '输出回复' section contains a text input field with the placeholder '请填写回复消息内容, 输入插入变量' (Please fill in the reply message content, input insert variable). Below this is a '下一步' (Next Step) section with a '+ 选择下一个节点' (Select next node) button.

- 添加输出回复

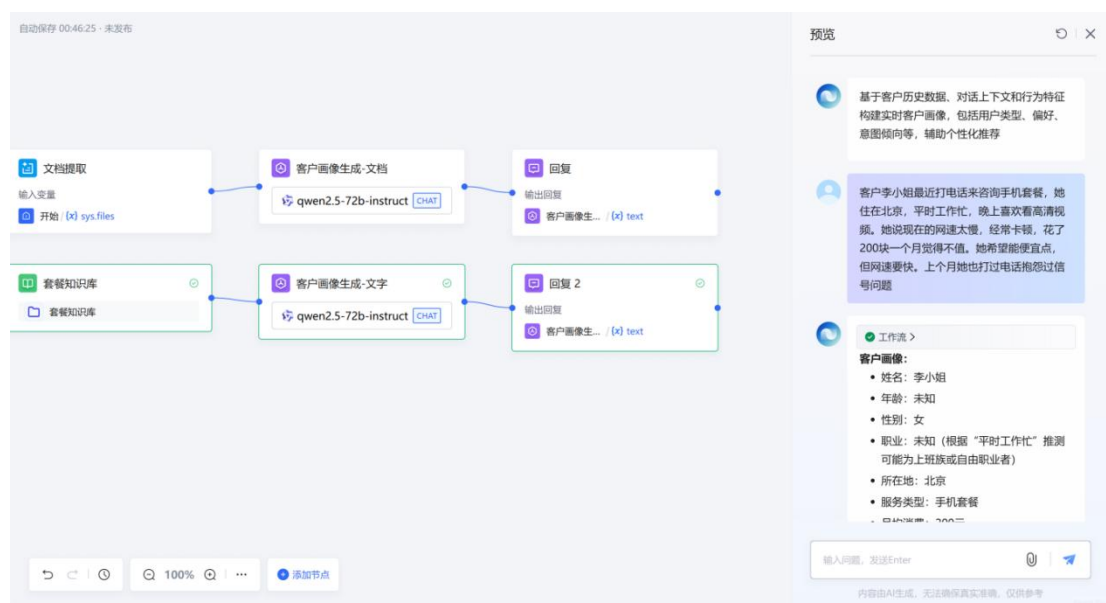
This screenshot is similar to the previous one, but the '输出回复' (Output Reply) section in the configuration panel for '回复 2' is now populated with the text '客户画像生... / (x) text'. This text is highlighted with a red rectangular box. The rest of the workflow editor and configuration panel remains the same as in the previous screenshot.

至此，整个 workflows 搭建完成，下面可以点击右上角“预览”按钮进行调试预览。



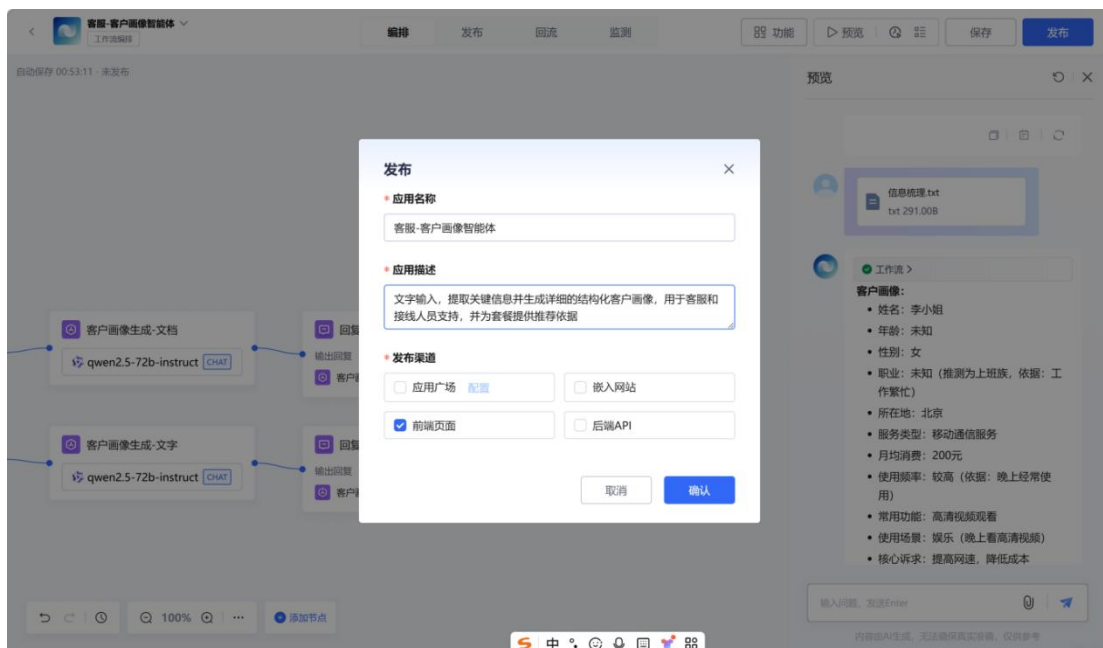
步骤 11：预览与调试

客户李小姐最近打电话来咨询手机套餐，她住在北京，平时工作忙，晚上喜欢看高清视频。她说现在的网速太慢，经常卡顿，花了 200 块一个月觉得不值。她希望能便宜点，但网速要快。上个月她也打过电话抱怨过信号问题



步骤 12：发布

- 点击页面右上角蓝色的“发布”按钮，将打开发布弹窗，本次发布渠道选择“前端页面”



● 点击“确认”，成功发布



● 通过发布前端页面生成 url 进行访问

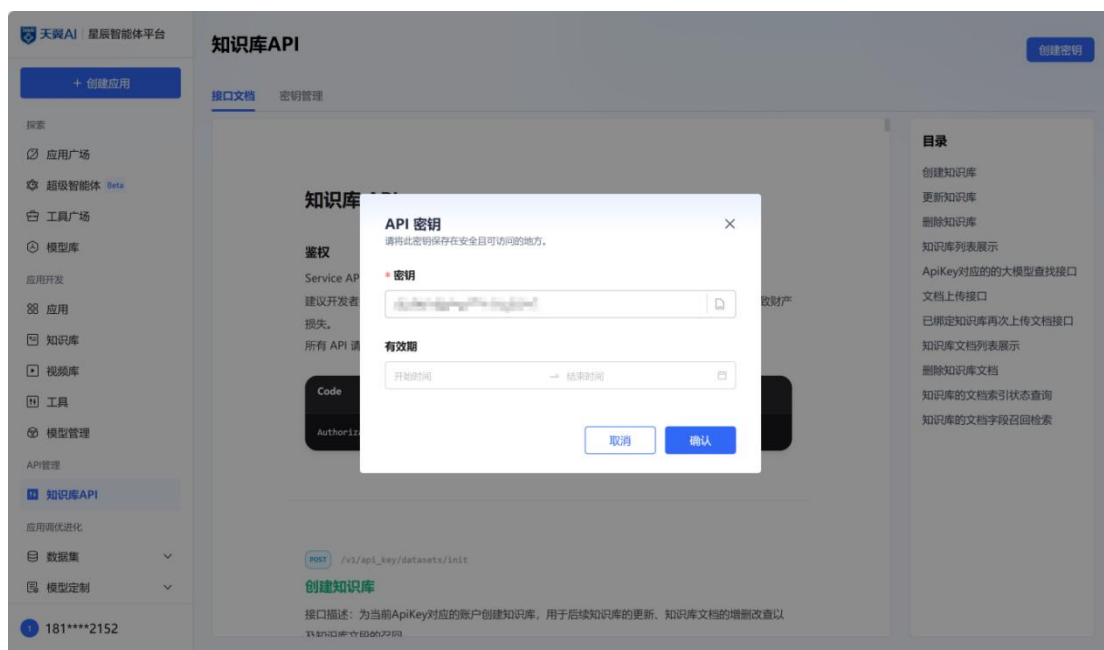


7.API 参考

7.1 知识库 API

创建密钥

在页面右上角，您会看见一个蓝色“创建密钥”按钮，点击该按钮将打开一个创建密钥的弹窗，如下图所示。您可以对添加的 API 密钥自定义设置有效期。



接口文档

如下图所示，星辰 MaaS 智能体平台为知识库功能开放了一系列的 API 接口，涵盖创建、更新、删除、列表展示、文档上传等。同时，星辰 MaaS 智能体平台还提供了详尽的接口文档，包括详细的参数说明、请求示例、响应格式及错误处理机制等内容，旨在为应用开发者提供全面且便捷的技术支持。



密钥管理

在“密钥管理”标签页中，我们为用户提供了对已创建的密钥进行集中管理的表单，该表单页不仅支持检索和时间维度的筛选，还整合了关于密钥的详细信息，便于用户对密钥进行高效的统一管理。

