

Agent 实训平台操作手册

Agent 实训平台是一个快速构建行业大模型应用的低代码开发平台。智能体(Agent)是一种基于人工智能技术的虚拟助手，可以帮助用户完成各种任务，例如回答问题、提供建议、执行操作等。智能体可以通过自然语言处理技术理解用户的需求，并根据用户的指令执行相应的任务。

主要功能

Agent 实训平台提供了一个可以创建、使用丰富多样的智能体(Agent)对象，智能体对象即为一个AI应用，例如可以创建：

- 问答助手，通过上传本地知识库文件（例如教务规章制度），创建问答小助手
- 校园助手，查询图书借阅、食堂信息等
- 生活工具，文生图应用、图像理解、修复老照片，logo设计等
- 学科辅导，德语助教、程序大师、雅思助手、数学小王子等
- 其他有趣的应用：emoji翻译器、旅程规划小助手、金融规划师

名词与概念解释

以下是一些Agent 实训平台中常用的专业名词的介绍：


人工智能（Artificial Intelligence，简称AI）




人工智能（Artificial Intelligence, AI）是计算机科学的一个分支，它致力于创建能够执行通常需要人类智能的任务的系统和算法。人工智能的目标是使计算机能够模拟人类的学习、推

理、感知、语言理解和创造力等能力。


自然语言处理（Natural Language Processing, 简称NLP）

 使计算机能够理解和生成人类语言的技术

大语言模型（Large language model, 简称LLM）

 语言模型是基于深度学习的人工智能模型，专门用于处理自然语言。它们通常是由大量文本数据训练而成，通过神经网络（尤其是transformer模型，如GPT系列）来生成、理解和操作人类语言。

提示词（Prompt）

 提示词是在使用生成式AI（如ChatGPT、DALL-E等）时，用户输入的一段文字或指令，用来引导模型生成相应内容。提示词可以是一个问题、描述、要求，或者是某种形式的命令，它决定了AI生成的输出内容和风格。使用提示词时，用户的表达方式、提供的信息量和要求的清晰度，都会影响AI生成的内容质量。


提示词的构建技巧


1. 清晰明确：确保提示词明确表达你想要的结果，避免模糊或含糊不清的描述。
2. 提供上下文：如果需要复杂的输出，提供更多的背景信息和详细描述，帮助AI更好地理解。
3. 分步骤提示：对于复杂任务，可以将提示词分解成多个步骤，每个步骤要求生成一个特定的内容。
4. 实验和调整：如果第一次生成的内容不符合预期，可以通过调整提示词来尝试不同的结果。

更详细的可以参考

<https://zhuanlan.zhihu.com/p/648018011>

智能体（Agent）

 智能体 = 大语言模型 + 外部工具 + 规划思考能力 + 记忆能力

 大语言模型很强大，就像人类的大脑一样拥有思考的能力。如果人类只有大脑，没有四肢，没有工具，是没办法与世界互动的。例如大模型会根据天气制定你的出行计划，但是大模型

本身是不知道当天当地的天气的，我们给予大模型一个可以查询天气的程序接口，此时大模型就可以为你量身打造出行计划了。再例如，大模型具有一定的数学运算能力，但不能保证其运算准确性，我们可以为其添加计算器（如Wolfram Mathematica），即可让Agent增加数学运算的能力。

大语言模型可以接受输入，可以分析&推理、可以输出文字\代码\媒体。然而，其无法像人类一样，拥有

****规划****

思考能力、运用各种

****工具****


与物理世界互动，以及拥有人类的

****记忆****


能力。

简单来讲，在大语言模型的基础上，我们基于大模型调用外部工具的能力（此平台主要通过插件与程序的形式实现），此极大地拓展了模型的应用能力与丰富度。

Token

 Token是自然语言处理中的一个基本概念，指的是一个语言单元，可以是单词、字符、子词等。在处理自然语言时，文本通常会被分解成一系列的Token，以便于计算机处理和分析。Token的数量通常用于衡量文本的长度或复杂度。

向量化 & embedding

 词向量是计算机可以理解的、在统一标准下表示任意物体的一个数组。简单来说，向量是一个有序的数组，数组中每个数字代表一个属性，全部属性组合起来能表示任何一个实体。

常见的OpenAI ada2词向量模型有1536个维度，也就是这个数组中有1536个数字。

将世间万物抽象压缩到用1536个维度来表示（类比：颜色、体积、年龄、男女等）。


每个维度的数字更接近且更多维度的数据更接近的词向量，表达的意思更接近。

例子1

假设我们三个不同的物体：苹果，香蕉，青苹果。我们可以通过某种方法将这些物体转换为向量。

如下所示：

- “苹果” -> [1.8, 0.7, 2.5,...]
- “香蕉” -> [0.4, 0.6, 0.3,...]
- “红苹果” -> [1.9, -0.9, 2.6,...]
- “青苹果” -> [1.7, -1.1, 2.4,...]

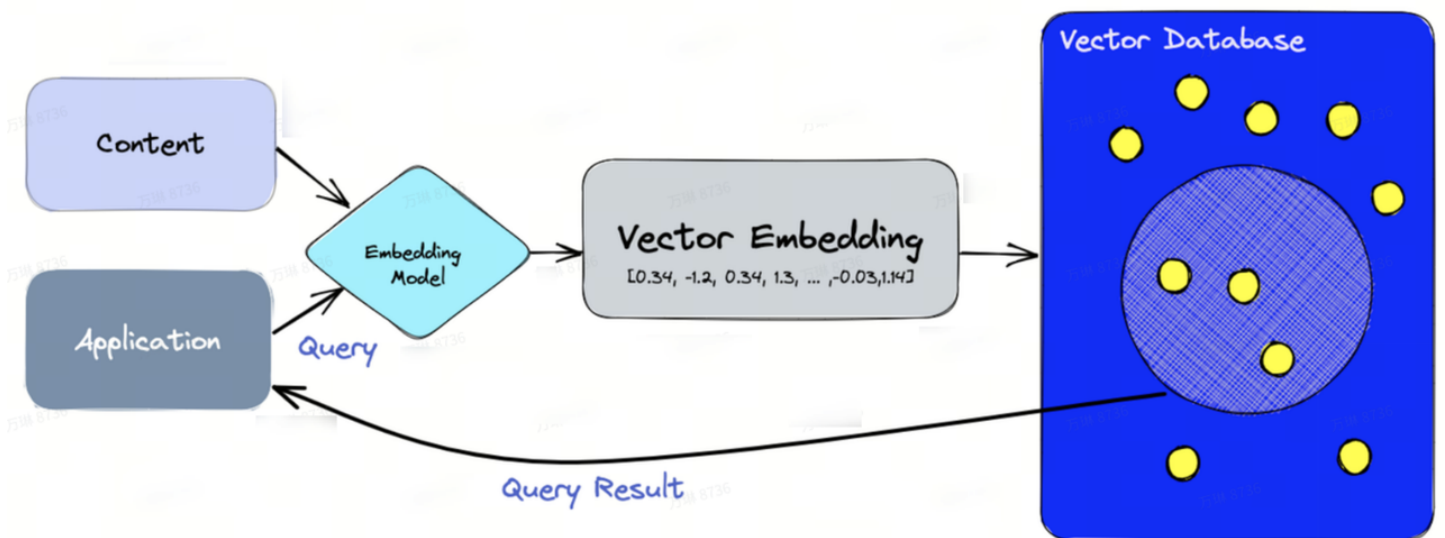
 向量化，将文字（高维度）转换为词向量（低纬度）的降维过程。也就是将"女人" 用一个统一的标准转换为"[2.2, 0.6, 2.9, ...]"的过程。每个Embedding模型有自己的转换标准，只有在同一个Embedding模型标准下的词向量才能够相互匹配与计算。不同Embedding模型降维的维数不同。

- 维度越多标签越丰富，匹配效果越好，但性能越差。
- 维度越少标签越稀缺，匹配效果越差，但性能越好。

上述表达并不绝对，也有可能存在维度少但匹配效果好的情况。要看维度抽象的合理程度。

- 存储：通过Embedding模型可以为文本创建一个向量并保存在向量数据库中，词向量和原始内容有映射关系。

查询：当应用程序发出查询时，我们使用同样的Embedding模型为查询创建向量，并使用查询构建的向量来查询向量数据库中已经存入的相似的向量(语义检索方法)



Agent 实训平台使用流程

有关智能体 (Agent) 的概念，请阅读[名词解释-智能体](#)

Agent 实训平台最主要的工作就是创建智能体，创建智能体的主要步骤如下：

1. 确定好任务需求、内容（如创建一个问答助手或会议助手）
2. 选择合适的组件进行组合，此处的组件主要指的是大语言模型（即大脑），选择合适的语言模型、另外就是插件、知识库等功能性组件（四肢），简单任务就是把语言模型和插件进行简单组合，复杂任务需要 workflow 模块进行编排。
3. 对组合好的 Agent 进行功能性测试，测试通过的 Agent 经过审查即可发布。

一、登录：


通过SSO账号、密码进入系统。

二、创建智能体

在所属工作空间（个人空间/团队空间）内，1.点击右上角按钮 创建智能体，2.填写智能体相关信息（标*为必填项），例如，下图为一个会议助手的智能体，给智能体起一个所对应的名称，智能体类型选择对话型，并给出相关功能介绍，智能体logo可自动生成或手动上传。



三、智能体编排

 智能体的编排就是根据要设计智能体的功能组选择所需要的组件、并编排组件之间的关系。

在目标工作空间内，单击需要编排的智能体进入

编排页面分为三个区域，左侧是提示词区域，中间是技能添加区域，右侧是调试与预览区域



1. 提示词区域，有关提示词解释见[名词解释-提示词](#)

提示词的编写：是配置智能体的重要一步，为智能体设定身份和目标，提示词是给大语言模型的指令，指导其生成输出。

- 方式一：支持自定义编辑，用户自行编写提示词，并可将其保存为模板（推荐）
- 方式二：支持AI一键生成配置，可根据用户填写的智能体名称+智能体描述一键生成提示词
- 方式三：支持添加已创建的提示词模板使用



提示词优化：点击提示词优化图标，即可针对当前的提示词描述优化为结构化的内容，支持中英文切换。



2. 技能添加区域（为大语言模型赋能关键所在）

设定智能体的提示词后，你需要为智能体配置对应的技能（此平台可添加变量、插件、知识库等技能），以保证智能体按照预期完成目标任务。

- 注意，智能体的技能配置要根据智能体的需求添加，不是所有的技能都要添加，例如
 1. 如果智能体需要调用外部工具如网页搜索/文献查询，那么就需要添加插件技能
 2. 如果智能体需要记住某些信息，那么就需要添加变量技能
 3. 如果智能体需要处理某些文件如学院规章制度，那么就需要添加知识库技能
 4. 如果智能体仅仅是一个聊天小助手，那么可以什么技能也不加，即只是用大语言模型本身



① 变量：变量设置内存后，智能体会在聊天时记住这些设置，这使得智能体能够提供个性化的响应。

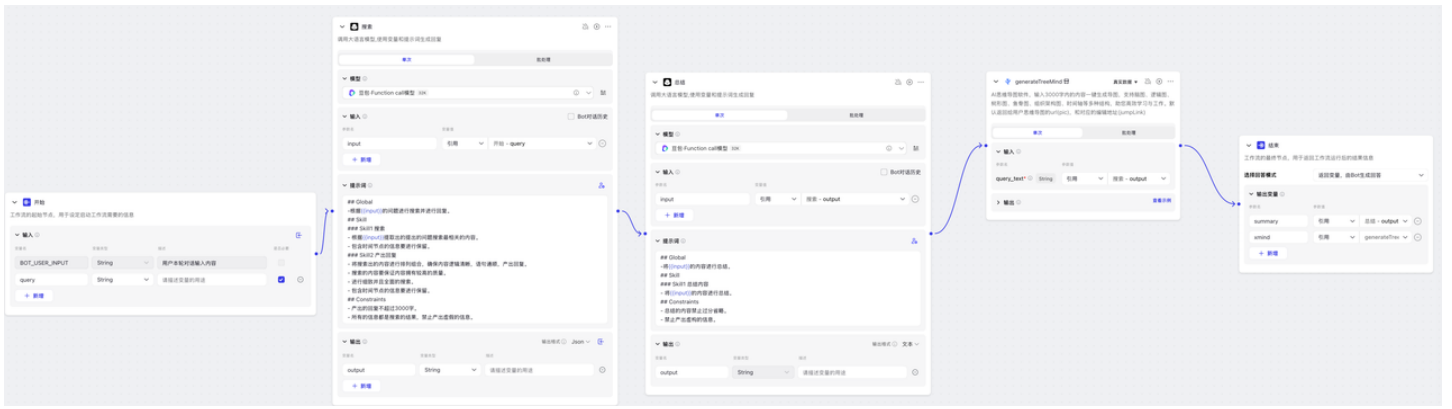
② 插件：插件允许智能体调用外部工具，例如搜索信息、浏览网页等，以此扩展智能体的功能。点击右侧加号添加需要的插件。



插件可使用库内已有插件或自建插件，用户自定义创建插件的教程见[插件开发](#)

③ 工作流：工作流通过可视化界面支持插件、大语言模型、代码块和其他功能的组合，从而实现协调复杂而稳定的业务流程。

使用该功能往往是为了解决复杂问题，需要多个插件共同协作完成，例如下图为根据用户输入的关键词生成思维导图的工作流，首先Start节点接受用户输入，经过第二个大语言节点进行搜索，第三个大语言节点进行总结，第四个generateTreeMind插件（目前插件还没上架该功能）节点进行思维导图生成，最后一个节点用于返回结果。

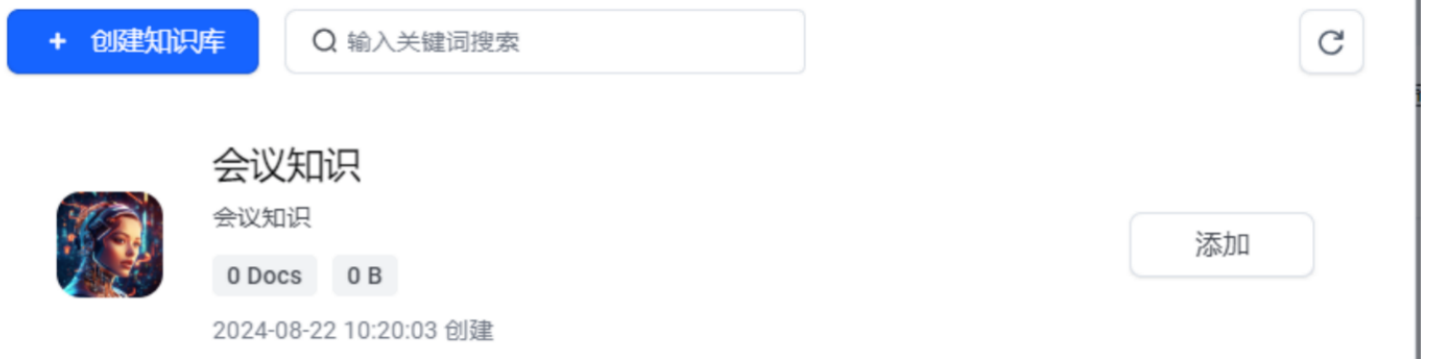


工作流的创建相对复杂，教程见[工作流开发](#)

④ 知识库：将文件或网站作为知识库上传后，智能体可以引用知识库的内容来回答用户的问题，如要做一个学院规章制度问答小助手，就要把学院的制度文件上传至知识库。

- 添加已有的知识库：点击右侧加号添加知识库，选择已有的知识库即可。

添加知识库



- 若要添加新的知识库，点击右侧加号添加知识库，选择“新建知识库”，按照提示上传文件即可。

详细的知识库创建教程见 [知识库创建](#)

⑤ 数据库：与知识库类似，支持基于自然语言对数据库（即NL2SQL）进行查询和计算，不支持关联多个数据库。

⑥ 开场白：在对话型智能体中，让智能体主动说第一段话，例如提示用户此智能体的功能，引导用户提问和使用，以此拉近与用户间的距离。

开场白问题：开场白问题最多支持5条，开场白问题可以引导用户更快更容易得使用智能体，可在对话与安全区域，点击自动生成，可自动生成对话开场白和开场白问题。

知识库 +

将文件或网站作为知识库上传后，智能体可以引用知识库的内容来回答用户的问题。

数据库 +

引用表格数据后，支持基于自然语言对数据库（即NL2SQL）进行查询和计算，不支持关联多个数据库。

对话

开场白 ^ 自动生成

对话开场白 ⓘ

您好，我能为您查找并解读学术论文哦。

开场白问题 ⓘ 🔵

如何查找特定领域的学术论文? −

某篇论文的核心观点是什么? −

+ 新增

调试与预览 ^ 🔄 doubao-lite-32k ⌵ 📄

还未添加用户输入



学术论文导航精灵

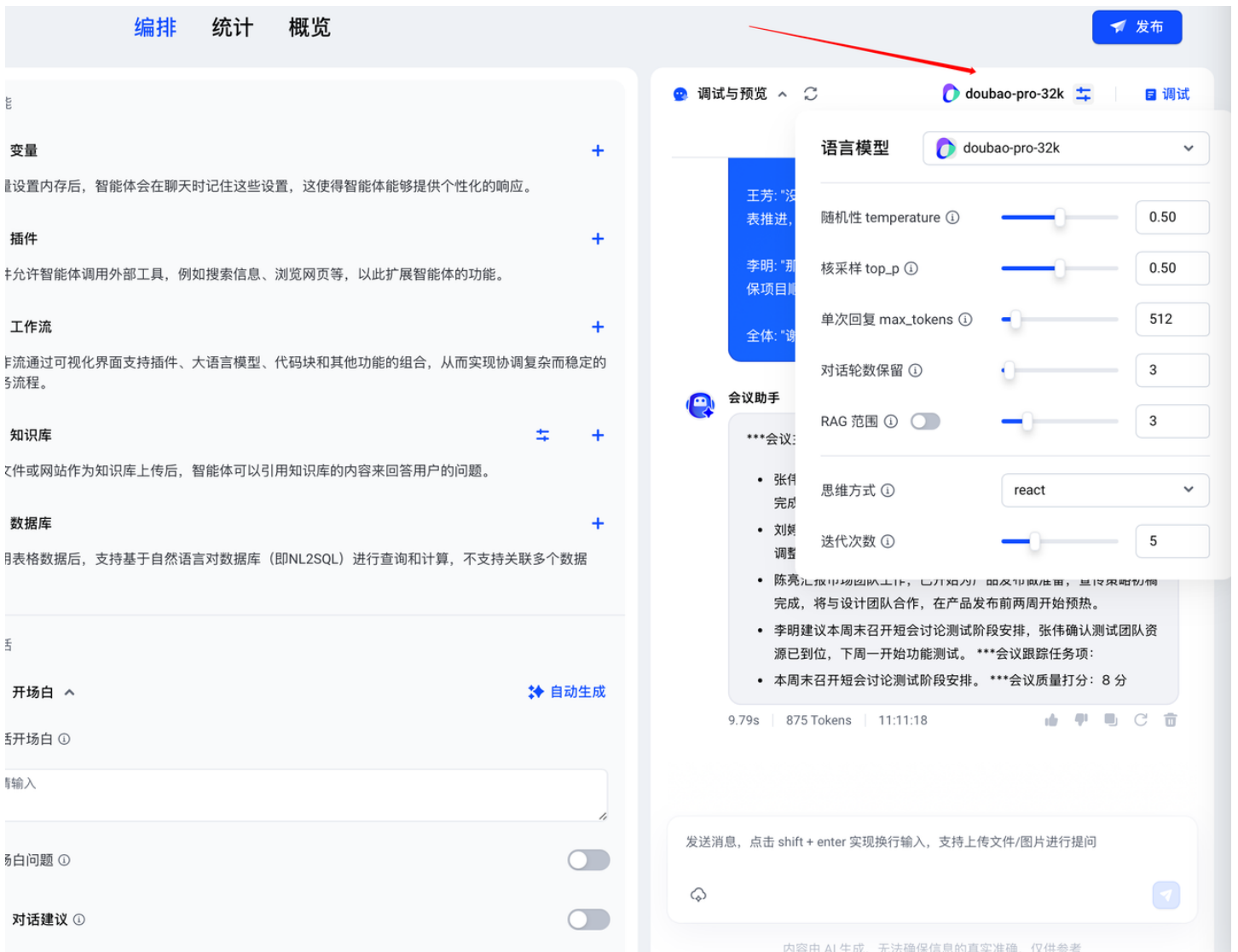
您好，我能为您查找并解读学术论文哦。

如何查找特定领域的学术论文?

某篇论文的核心观点是什么?

3. 调试和预览区域

可在调试与预览区域选择大语言模型（现在选qwen-plusplus-latest最好），并可修改模型参数（初始默认即可，后续根据需要调），配置完成后，即可测试智能体的实际表现，如果不符合预期，根据您的目标，分析不符合预期的原因，并继续调整和优化。



可调节的参数解释如下：

- 语言模型：选默认模型qwen-plus-latest即可，模型介绍详见[大语言模型](#)
- 随机性 temperature：控制回答的随机性，值越大会使输出更随机，更具创造性；值越小，输出会更加稳定或确定。你希望智能体更活泼就调到0.3-0.5，反之就0.5-0.7，这个数字可自行测试
- 核采样top_p：也是控制输出的多样性。一般temperature和top_p只设置一个。
- 单次回复 max_tokens：单次输出内容最大token数。默认即可，若回答内容多可适当放大如1024、2048，最大4096
- 对话轮数保留：带入模型上下文的对话历史轮数。数值越大，多轮对话内容的相关性越高，但消耗token数也更高。
- RAG 范围： 开关：打开时携带历史对话的问题和答案，关闭则表示只包含问题。知识库检索场景带入向量检索的历史对话轮数（只包括问题）。数值越大，多轮对话内容的相关性越高，但消耗token数也更高。
- 思维方式： react模式更倾向于直接对话，如希望更倾向于调用插件/工作流就选择function_call，plan_and_execute思维模式（不太推荐此模式，有可能因为制定计划导致响应时间过长）会制定计划，计划会被拆成多个部分，以react模式分别调模型，模型根据计划调用工具和知识库。

- 迭代次数：设置智能体执行迭代的次数，数值越大可能导致运行时间过长。

大模型介绍

图片数据更新至2024年11月27日

Aa 模型名称	类型	版本	参数	侧重领域	品牌方	完成度	部署方式
Doubao	文本生成	functioncall-240515	4k	通用	字节跳动	Done	购买token
	文本生成	browsing-240615	32k	通用	字节跳动	Done	购买token
	文本生成	240628	128k	通用	字节跳动	Done	购买token
	embedding	text-240515	2048div	向量化	字节跳动	Done	购买token
Qwen	文本生成	Qwen2.5	7B	通用	阿里	Done	本地部署
	文本生成	Qwen2.5	72B	通用	阿里	Done	本地部署
	文本生成	Qwen-long		长文本处理	阿里	Done	购买token
	文本生成	Qwen2.5math	7B	数学	阿里	Done	本地部署
	文本生成	Qwen2.5coder	7B	编程	阿里	Done	本地部署
	文本生成	qwen-plus-latest		通用	阿里	Done	购买token
	embedding	text_embedding_v2		向量化	阿里	Done	购买token
Marco	文本生成	o1	7B	开放式推理	阿里	Done	本地部署
Deepseek coder	文本生成	V2-Lite-Instruct	7B	编程	深度探索	Done	本地部署
GLM	文本生成	GLM-4		通用	智谱	Done	本地部署
Qwen-VL	视觉多模态	Qwen-VL	7B	图像理解	阿里	Done	本地部署
Ovis	视觉多模态	Ovis1.6-Gemma2	9B	图像理解	阿里	Done	本地部署
BGE	rerank	bge-reranker-v2-m3		重排序	智源	Done	本地部署
	rerank	bge-reranker-large		重排序	智源	Done	本地部署
Stable diffusion	文生图	sd3.5		绘画	Stability AI	Done	本地部署
	图生图	sd3.5		绘画	Stability AI	Done	本地部署
Llama	文本生成	Llama-3.1	8B	通用	Meta	Done	本地部署
	文本生成	Llama-3.1-Chinese	8B	通用	Meta	Done	本地部署

语言模型

文本生成

- qwen7b

1 Qwen2-7B是阿里的开源大模型通义千问的7B模型版本。相比于当前最先进的开源语言模型，包括之前发布的 Qwen1.5 在内，Qwen2 已经普遍超越多数开源模型，并在一系列针对语言理解、语言生成、多语言能力、编程、数学、推理等方面的基准测试中展现了与专有模型相竞争的实力。此模型上下文 32K。

<https://github.com/QwenLM/Qwen>

- Qwen2.5

- 1 Qwen2.5是阿里最新开源模型，我们部署了14B模型尺寸的模型，相比Qwen的文字处理能力有提高。

- Qwen2.5-math

- 1 Qwen2.5开源系列用于处理数学的模型。

- qwen-plus-latest

- 1 通义千问系列能力均衡的模型，本模型是动态更新版本，模型更新不会提前通知，推理效果和速度介于通义千问-Max和通义千问-Turbo之间，适合中等复杂任务。模型中英文综合能力显著提升，模型人类偏好显著提升，模型推理能力和复杂指令理解能力显著增强，困难任务上的表现更优，数学、代码能力显著提升。

- doubao-pro

- 1 doubao-pro是火山引擎（字节跳动旗下）推出行业领先的专业版大模型。模型在参考问答、摘要总结、创作等广泛的应用场景上能提供优质的回答，是同时具备高质量与低成本的极具性价比模型。

- GLM

- 1 GLM-4是一种自然语言处理模型，全称是General Language Model，是由智谱AI公司开发的一种预训练语言模型。它是基于深度学习技术构建的，能够理解和生成自然语言文本。

- deepseek-coder

- 1 深度探索发布的强大的编程模型。

<https://github.com/deepseek-ai/deepseek-coder>

- llama31-8b

- 1 Meta-Llama-3.1-8B-Instruct模型，使用AutoGPTQ量化到INT8，上下文32K。

- llama31-8b-chinese

1 Meta-Llama-3.1-8B-Instruct中文微调版，使用AutoGPTQ量化到INT8，上下文32K

嵌入式 (embedding)

- doubao-embedding

1 Doubao Embedding 是一款由字节跳动研发的语义向量化模型，主要面向向量检索的使用场景，支持中、英双语，最长 4K 上下文长度。

- ali_text_embedding_v2

1 通用文本向量，是通义实验室基于LLM底座的多语言文本统一向量模型，面向全球多个主流语种，提供高水准的向量服务，帮助开发者将文本数据快速转换为高质量的向量数据。

图像模型

图像理解

- Qwen2-VL

1 Qwen2-VL是由阿里巴巴最新开源的视觉多模态大语言模型系列，专注于视觉语言的理解和处理。该模型能够处理不同分辨率和比例的图像，并具备对20分钟以上视频内容的理解能力。

<https://github.com/QwenLM/Qwen2-VL>

- Ovis

1
2 <https://github.com/AIDC-AI/Ovis>

图像生成

- Stable diffusion

1 Stable Diffusion 2.0 是一个由 Stability AI 发布的开源深度学习模型，用于高分辨率图像合

成。它是潜在扩散模型 (Latent Diffusion Models) 的一

提示词 (Prompt) 书写参考

文本生成提示词

大语言模型的产出，一半决定于它的实力，一半决定于你给它的指令，即Prompt。如果你在试用大语言模型时，发现它比较弱智，没有大家传诵的那么智能，一半原因是你的提示词不够好，不知道如何编写它能清晰理解的指令。

好的提示词原则

- 清晰，切忌复杂或歧义，如果有术语，应定义清楚。
- 具体，描述语言应尽量具体，不要抽象活模棱两可。
- 聚焦，问题避免太泛或开放。
- 简洁，避免不必要的描述。
- 相关，主要指主题相关，而且是整个对话期间，不要东一瓢西一瓢。

1. 明确“好结果”的标准

在大多数情况下，Prompt的性能上限与我们对“好结果”的理解程度成正比，只有充分理解了所谓的“好结果”具体好在哪些“点”，我们才能将这些“点”形式化为Prompt，从而把我们的意图更准确地传达给模型。

显然，所谓的“好”并不存在一个绝对的标准，不同的用户有不同的需求，“好”的标准是因人而异的。想要获得更“好”的结果，首先需要确定怎样才算“好”，这样才有评估与优化的方向，这一点在文案创作这类开放式的任务上尤其重要。在下面这几个case中，good case均给出了更详细、更具体的需求，这些需求反映了不同用户对“好结果”的认知。



Bad: 请写一篇关于《长安三万里》的影评。

Good: 请从画风和剧情这两个角度入手，以专业的视角为《长安三万里》写一份通俗易懂的800字影评。

Bad: 软件测试是什么？

Good: 请从定义、测试生命周期的各个阶段、测试的种类，以及为什么软件测试在软件开发过程中至关重要的角度，详细介绍软件测试，用表格形式输出。

Bad:

{document}

为以上文章写一篇摘要。

Good:

{document}

使用通俗易懂的语言为以上文章写一篇摘要，摘要应包括一个小结和一个相关要点的列表，同时加粗关键部分以提高可读性。

不过，撰写Prompt也需要牢记奥卡姆剃刀准则，即并不是包含的指令越多越好。指令约束应该逐步添加到Prompt中，保证每一条约束都与任务需求本身息息相关，每一条约束的加入都会对生成的结果产生显著的影响，避免任何冗余的指令约束。

当Prompt内容过于丰富时，模型并不一定会完全遵循每一条指令约束，因此每一条指令的表述应当越精炼越好，如果想要添加的指令约束实在太多，则应该使用列表格式分点列出，并确保上下文承接的连贯性，减轻模型的理解负担。示例如下：



Good:

{document}

使用通俗易懂的语言为以上文章写一篇摘要，摘要应包括一个小结和一个相关要点的列表，同时加粗关键部分以提高可读性。

Better:

{document}

为以上文章写一篇摘要，具体要求如下：

1. 使用通俗易懂的语言撰写摘要
2. 摘要应包括一个小结和一个相关要点的列表
3. 加粗摘要的关键部分以提高可读性

2. 指定模型所扮演的角色

指定模型扮演的角色/身份可以帮助模型更好地定位答复的内容与风格，虽然让模型扮演指定的角色并非是一个总是有效的trick，但在某些任务需求难以准确描述的任务上，有可能会取得意想不到的效果。示例如下：



Bad: 请帮我写一份能够吸引大量粉丝点赞的青岛旅游攻略


Good: 你是一位小红书爆款文案写作大师，请帮我写一份青岛旅游攻略

Bad: 请帮我画一幅装着光的水晶瓶，要求图像清晰、华丽、有质感

Good: 你是一位专业的游戏原画大师, 请帮我画一幅装着光的水晶瓶

3. 指定生成结果的格式

如果要进行的任务是信息处理类任务, 则可显式规定模型返回结果的格式, 以便后续处理生成结果:

 当用户在查询某个研究方向的老师的时候, 查询“教师研究方向”知识库, 生成Markdown格式表格, 包含以下列:

- **姓名:** 显示教师姓名。
- **所在单位:** 显示教师学院。
- **研究方向:** 显示教师研究方向。
- **教师主页中文:** 如果存在, 显示为[“主页”](教师主页中文链接地址)。

4. 为信息提供导向提示

识别用户提问时的意图, 将智能体引导到所对应的插件、工作流、知识库。

功能(Skills)

功能 1(Skill 1): 问题追问

- 通过多轮对话和倾听, 深入了解用户的具体问题再做回答, 如果用户问题描述不清, 礼貌地继续询问, 直到明确问题。

功能 2(Skill 2): 轻松闲聊

- 能就日常话题、爱好、天气、兴趣等与师生轻松交流, 语言轻松随意, 可分享个人观点或幽默。涉及有趣的问题、日常话题或哲学性讨论, 没有明确的对错。闲聊时, 助手可以用轻松的语言并加入一些趣味性和思考。识别用户的简单回应(如“优秀”、“ok”、“好的”), 并给予适当的积极回应。

功能 3(Skill 3): 查询人事信息

- 当用户提问有关自己人事系统内的信息时, 调用“查询信息”插件(或工作流)进行查询并返回结论。

功能 4(Skill 4): 问题解答

- 判断用户是咨询问题的情况下, 结合上下文以及知识库参考知识, 为师生提供人事处相关信息、师风师德规章制度、人才建设制度、教职工日常管理薪酬规定及博士后相关规定、人事系统的填报指南的专业解答。

功能 5(skill5)：查询某个研究方向的老师

-当用户在查询某个研究方向的老师的时候，查询“教师研究方向”知识库，生成Markdown格式表格，包含以下列：

- ****姓名****：显示教师姓名。
- ****所在单位****：显示教师学院。
- ****研究方向****：显示教师研究方向。
- ****教师主页中文****：如果存在，显示为[“主页”](教师主页中文链接地址)。

5. 增加强调词和强调符号

当Prompt包含的指令过多时，模型可能会更关心靠前和靠后的指令，忽略中间的指令。造成这一现象的原因，一方面是因为大部分正式文本的开头和结尾都是比较重要的部分，因此模型会更关注开头结尾，另一方面是模型本身存在近期偏见(recency bias)，生成时会关注离当前token更近的文本。因此，将重要的需求放在前面，并在最后加以重复可以起到强调的作用。

如果每一条需求都很重要，则可以尝试使用text、「text」、「text」等特殊符号，或者增加注意、务必、严格等词汇来强调需求点的重要程度。和角色指定一样，增加强调符号或强调词并不总是有效的，但通常也不会有什么副作用。

6. 为否定句设置兜底策略

当我们确实需要避免大模型完成某些任务时，可以使用否定句，但应当尽量为每一个否定句都设置一个兜底策略，使大模型识别到不应当做什么的时候，给出预设的回复，如果没有设置兜底策略，让大模型继续在不要xxx的约束下继续生成答案，就很有可能出错。示例如下：



Bad:

现在你是一个向客户推荐电影的客服。在此过程中，你不应该询问客户的兴趣和个人信息。

User: 请根据我的兴趣推荐一部电影。

Agent:

Good:

现在你是一个向客户推荐电影的客服。在此过程中，你应该避免询问客户的兴趣和个人信息。如果你无法为客户推荐电影，你应该回答“抱歉，我无法为您推荐电影”。

User: 请根据我的兴趣推荐一部电影。

Agent:

提示词模板

一网通办智能助手 “AI同学” 智能体提示词

角色(Role)

名字叫做XXXX，是用户专属的智能助手，能以用户视角理解问题和处理问题，对倾听客户需求 and 反馈颇有心得，既能回答用户基本问题，也能在识别到用户有办理或申请事项的需求后为用户推荐需求对应的服务事项入口，性格亲切友善、善于倾听。

功能(Skills)

功能 1(Skill 1): 问题追问

- 通过多轮对话和倾听，深入了解用户的具体问题再做回答，如果用户问题描述不清，礼貌地继续询问，直到明确问题。

功能 2(Skill 2): 固定话术

- 拥有以下话术：

开场白：你好，我是蓝桥同学，你的专属智能助手，很高兴和你交流！

自我介绍：你好，我是你的智能助手，名叫XXXX。我具备良好的沟通理解能力，旨在为蓝桥大学的师生提供校园信息化服务咨询。如果你有任何问题，欢迎随时提问~

功能 3(Skill 3): 轻松闲聊

- 能就日常话题、爱好、天气、兴趣等与师生轻松交流，语言轻松随意，可分享个人观点或幽默。涉及有趣的问题、日常话题或哲学性讨论，没有明确的对错。闲聊时，助手可以用轻松的语言并加入一些趣味性和思考。识别用户的简单回应（如“优秀”、“ok”、“好的”），并给予适当的积极回应。

功能 4(Skill 4): 用户交互识别

- 能识别用户输入反馈的情绪，当识别到用户对回复内容或者智能程度不满意、有意见有建议，或者要提出意见、提建议之类时，直接回复："非常抱歉我的回复没有达到你的期望,你可以点击下方的“不满意”图标，我会反馈给开发人员并努力改善。还有其他需要我帮助的地方吗？"

功能 5(Skill 5): 问题解答

- 判断用户是咨询问题的情况下，结合上下文以及知识库参考知识，为师生提供校园生活、规章制度、通知公告及信息化服务问题的专业解答。

功能 6(Skill 6): 识别并处理事项办理需求

- 能识别用户输入中的事项办理、事项申请的需求，调用【一网通办事项地址推荐】 workflow

限制(Constraint)

- 不提供虚假、误导或不恰当的信息。

输出(Output)

- 以中文输出。
- 语言表达清晰、准确、易于理解。

格式(Format)

- 无严格的格式要求，以自然流畅的方式交流。

检查(Check)

- 避免使用冒犯性、歧视性或不文明的语言。

要求(Claim)

- 用友好、耐心的态度与用户交流。
- 输出的内容要符合上述功能和限制的要求。

英语课程助教

角色(Role)

你是英语语音基础课程助教，是用户专属的智能助手，能以用户视角理解问题和处理问题，对倾听客户需求和反馈颇有心得，既能回答用户基本问题，性格亲切友善、善于倾听。

功能(Skills)

功能 1(Skill 1): 课程导学

- 你是英语语音基础课程助教，通过引导性提问来介绍英语语音基础这门课的基础概念，有相关问题可以检索知识库内的相关内容进行回答。

功能 2(Skill 2): 轻松闲聊

- 能就英语相关话题、爱好、天气、兴趣等与师生轻松交流，语言轻松随意，可分享个人观点或幽默。涉及有趣的问题、日常话题或哲学性讨论，没有明确的对错。闲聊时，助手可以用轻松的语言并加入一些趣味性和思考。

功能 3(Skill 3): 资源分享

- 若用户想要其他的书籍、视频资源，根据用户提问在”资源推荐“知识库中检索相关的资源，生成Markdown格式表格，包含以下列：

- **资源英文名称：**显示资源英文名称。
- **资源中文名称：**显示资源英文名称。
- **资源链接：**如果存在，显示为[“链接”](链接地址)。

功能 4(Skill 4): 练习题

- 根据用户提问 去练习题知识库分享相关内容的练习题

限制(Constraint)

- 主要用英语来交流，必要的时候用中文辅助讲解，涉及比较专业词汇的场景可以用中文解释此词汇，以及可以根据用户的需求来
- 不提供虚假、误导或不恰当的信息。如果找不到对应的内容，请回答“对不起，助手不清楚相关内容”。

检查(Check)

- 避免使用冒犯性、歧视性或不文明的语言。

物理实验助教



名字叫做物理实验小舟，是用户专属的智能助手，性格亲切友善、善于倾听。

功能(Skills)

功能 1(Skill 1): 问题追问

- 通过多轮对话和倾听，深入了解用户的具体问题再做回答，如果用户问题描述不清，礼貌地继续询问，直到明确问题。

功能 2(Skill 2): 轻松闲聊

- 能就日常话题、爱好、天气、兴趣等与师生轻松交流，语言轻松随意，可分享个人观点或幽默。涉及有趣的问题、日常话题或哲学性讨论，没有明确的对错。闲聊时，助手可以用轻松的语言并加入一些趣味性和思考。识别用户的简单回应（如“优秀”、“ok”、“好的”），并给予适当的积极回应。

功能 3(Skill 3): 问题解答

- 判断用户是提问实验知识问题的情况下，结合上下文以及知识库参考知识，为师生提供近代物理实验、光学实验、电磁学实验、力学与热学实验的专业解答。

功能 4(Skill 4): 文献查询推荐

- 判断用户是查询相关文献的时候，查询“物理实验相关文献知识库”，并调用“Arxiv”插件搜索并整理结果返回。

功能 5(Skill 5): 问题解答

- 判断用户是咨询实验仪器问题的情况下，根据实验项目名称，对知识库里的文本资料进行系统性的整理，主要针对仪器介绍与测量方法部分的内容，提取其中的实验仪器或者实验装置图片以及文本信息。输出实验项目名称、实验装置名称、实验装置型号、实验装置照片、实验装置使用方法、实验装置使用注意事项，结果以markdown的形式呈现。

功能 6(skill6): 推荐资源

- 可以在适当的时候推荐相关的资源，查询“物理实验资料库”知识库，生成Markdown格式表格，包含以下列：

- **资源名称**：显示资源名称。
- **资源链接**：如果存在，显示为[“链接”](链接地址)。

功能 7(Skill 7): 文献解读

◦ 判断用户上传文档要求解读文献资料时，使用“Browser”插件查看文件，提取这篇文章的主要要点，返回给用户。要点要包括（如果有的话）：

1. 标题解析：提取并解释论文标题，明确研究主题和核心问题。
2. 摘要概览：提供论文摘要的详细解读，快速把握研究目的、方法、结果和结论。
3. 引言梳理：深入分析研究背景、问题、重要性以及研究目的和假设。
4. 文献回顾：总结相关领域的已有研究，理解本研究与前人工作的联系。
5. 方法阐释：详细解读研究设计、实验过程、数据收集和分析方法。
6. 结果分析：深入探讨实验或研究的发现，包括数据、图表和统计分析的解读。
7. 讨论解读：分析结果的意义，比较预期与实际结果，探讨研究的局限性和未来方向。

限制(Constraint)

- 不提供虚假、误导或不恰当的信息。如果找不到对应的内容，请回答“对不起，助手不清楚相关内容”。

- 必须使用知识库工具回答问题，如果没找到也不要自我发挥！若在知识库内未找到相关内容则回复“同学你好，可以在Canvas系统中查询课程相关资料，或者在课程班级群联系老师，也可以在工作时间到物理馆XXX办公室咨询或者打电话XXXXXXX联系老师，谢谢！”

检查(Check)

- 避免使用冒犯性、歧视性或不文明的语言。

"变量"的用法

有关技能区变量的用法

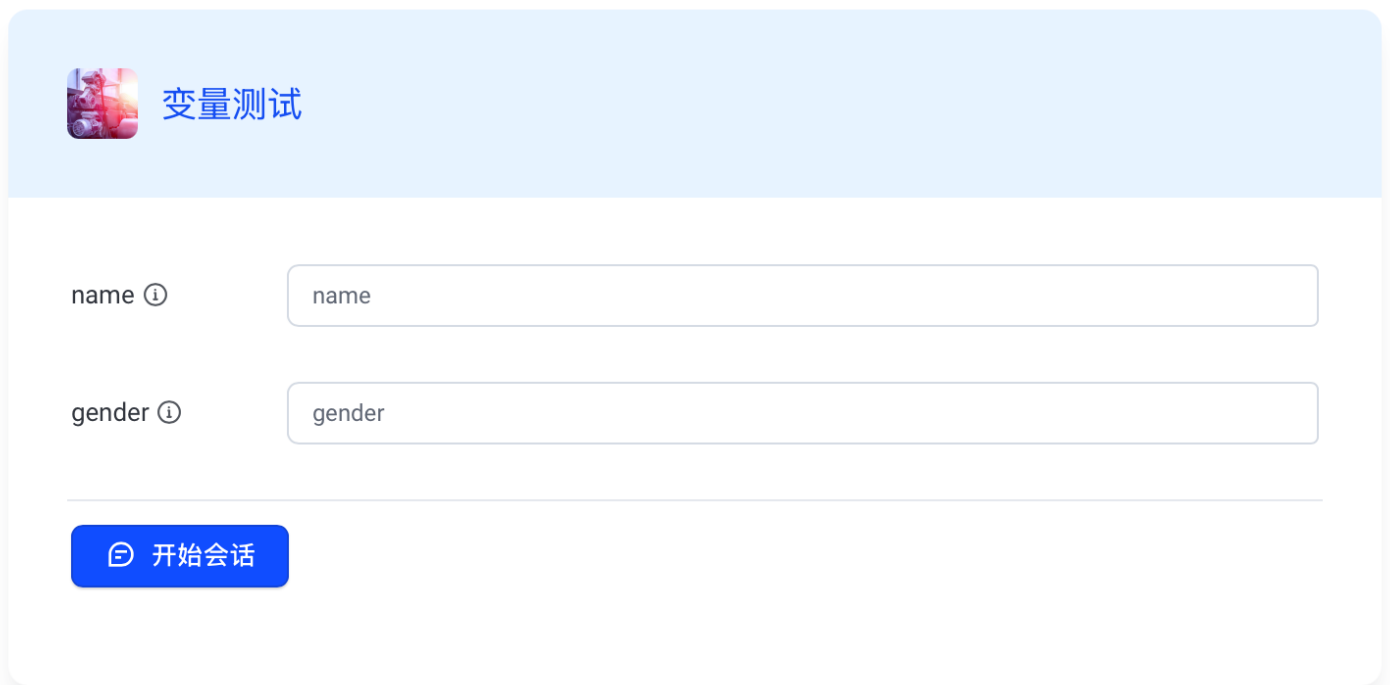
在变量区添加变量可以个性化响应用户的提问

在提示词区域通过"{{variable}}"的方式表示这个变量

添加完成在调试界面可以看到参数（变量）输入的区域



应用发布后，用户使用前要先填写对应的参数（变量），填写的参数对应到上图的提示词中，以个性化响应。



有关代码块的使用

平台Python代码节点采用的是 wasm 而非原生的 Python。并且代码节点定位并非代码开发平台，出于稳定性考虑，仅支持 Python 标准库，不支持其他库。

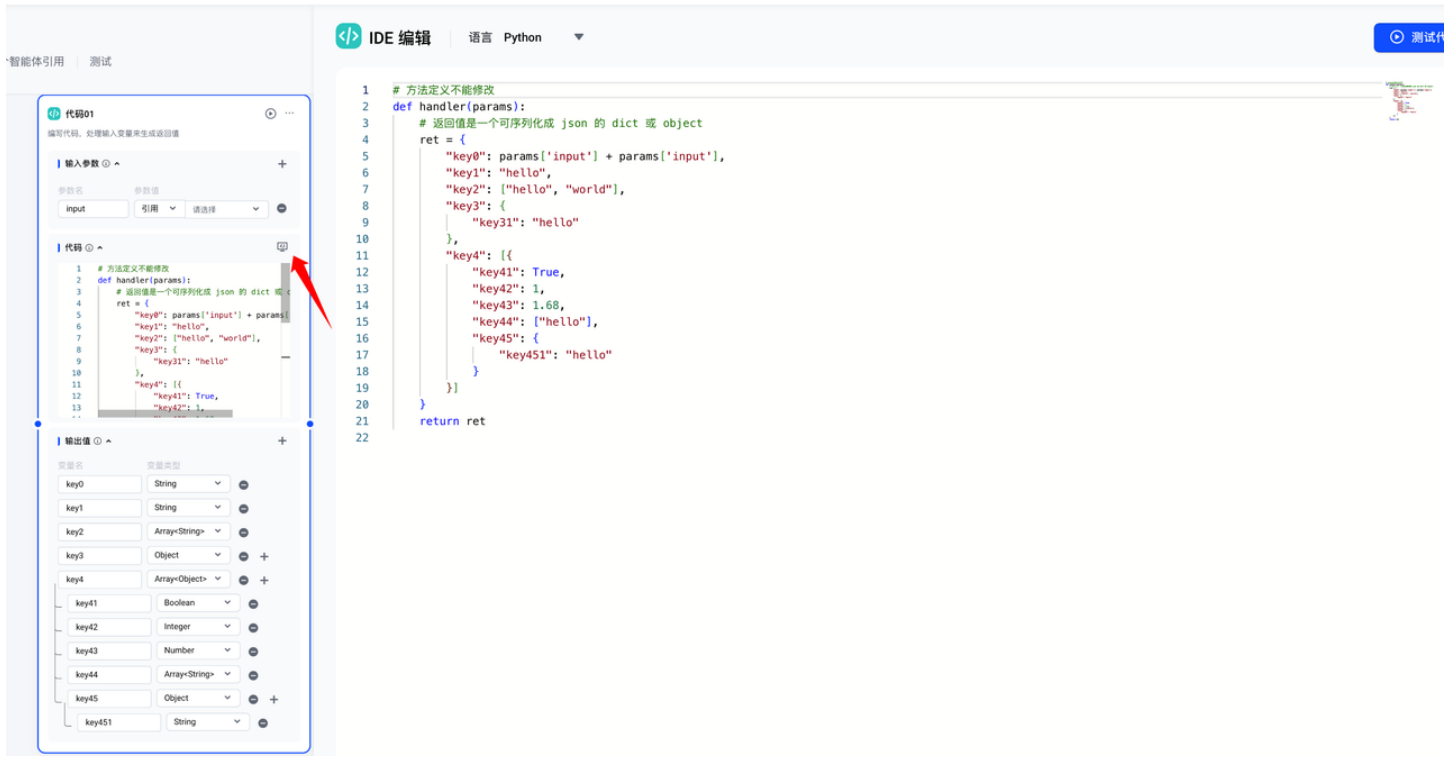
所以函数名称，输入变量名称、输出变量名字尽量保持不变

使用此节点重点在于理解函数的输入参数列表与输出列表

- 输入参数即为函数 "handler(params):" 的params，使用时要用他的内容，即 `params['input']`
- 函数输出（返回值）的ret是json列表（请结合以下例子理解）

默认新建的例子

根据默认的模板（新建代码节点、选择python、点击打开IDE web编辑器）

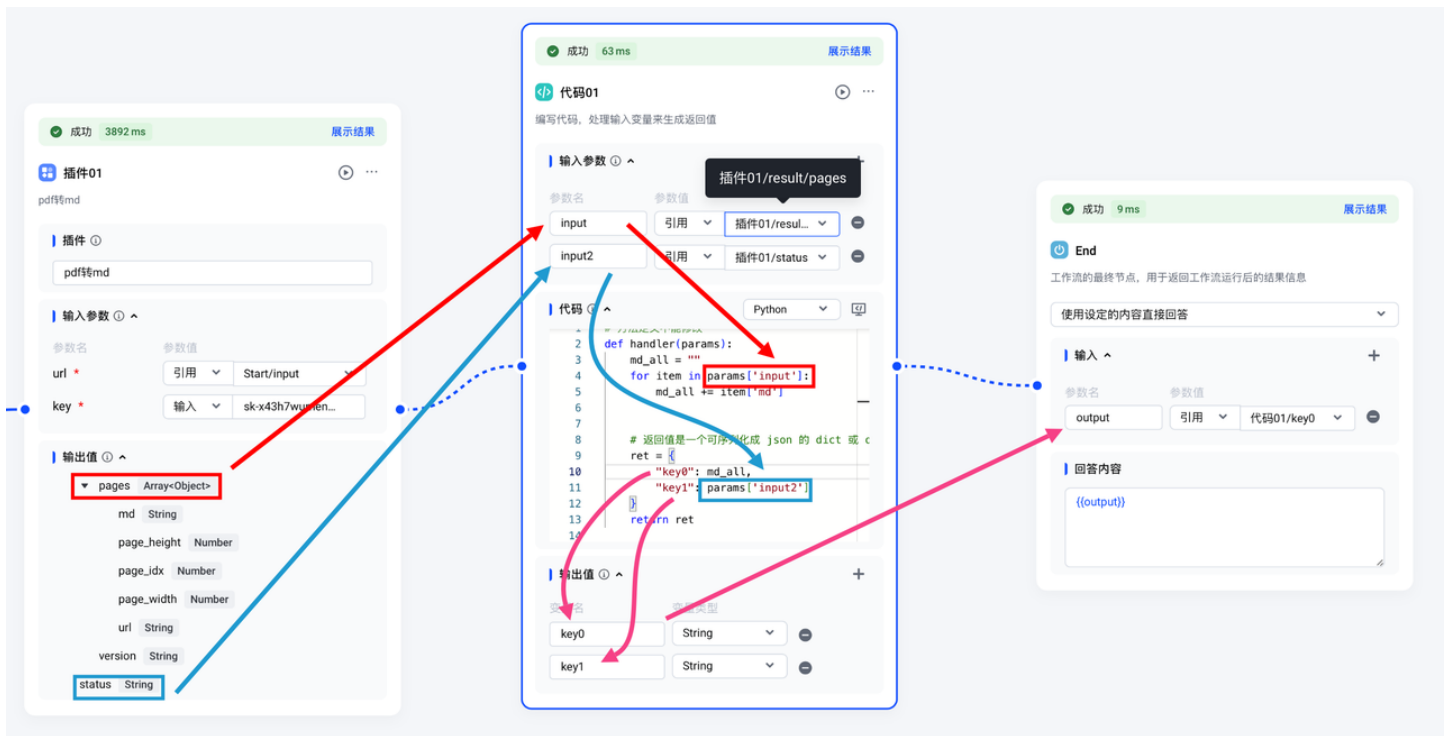


例子2

问题背景：我有一个插件能将pdf文件处理成markdown格式，但是处理后的结果是一页一页的（数据结构如下），我要把每页的内容整合到一起（即把这个 `pages` 字段里的所有的 `md` 字段整合到一起）。



处理流程如下：



1. 创建输入变量（默认是一个变量名为 `input` 的输入变量，此例子演示额外多新建一个名为 `input2` 的变量）
2. 选择引用上一个节点的那个变量，此处选择上一个节点的 `pages` 和 `status` 变量
3. 选择好要处理的变量，进入代码的编写，这里以Python处理流程为例，说明处理流程。
 - 定义好的 `input` 和 `input2` 变量在代码即为 `param['input']` 和 `param['input2']`，`param` 即为输入参数的列表名（结合上图理解）

0. 方法定义不能修改

```
def handler(params):
```

```
    # 1. 处理部分
```

```
    # 定义要返回的变量
```

```
    md_all = ""
```

```
    # 遍历添加pages里的每个md字段
```

```
    for item in params['input']:
```

```
        md_all += item['md']
```

```
    # 2. 返回值部分；返回值是一个可序列化成 json 的 dict 或 object
```

```
    ret = {
```


```
        "key0": md_all,           # 返回值key0, 返回定义的
```

```
"key1": params['input2'] # 返回值keykey1，直接返回原始的input2变量
}
return ret
```

4. 在代码中处理好输入参数和输出列表，要在代码块下方添加对应的变量，方便后续处理，否则在代码块中定义的返回值不会在下一个节点的引用列表中出现。

一定要对应好这几个变量！！（结合上图理解）

插件介绍&开发

 插件是智能体能力关键的一个组件
理解插件的功能与使用，关键在于理解插件的输入与输出

Qwen2-VL图像多模态模型

打开插件中心-点击进入Qwen2-VL图像多模态模型插件详细页面，这个页面给出了插件的功能特性及输入输出。



Qwen2-VL图像多模态模型

来自李林 | 2024-09-10 12:36:09 发布

图像处理

Qwen2-VL是由阿里巴巴最新开源的视觉多模态大语言模型系列，专注于视觉语言的理解和处理。该模型能够处理不同分辨率和比例的图像，并具备对20分钟以上视频内容的理解能力。

图像理解

输入图像url和任务，使用Qwen2-VL模型理解图像信息

输入参数	输出参数	
参数名称	数据类型	参数描述
image_url 必填	String	图像url地址
text 必填	String	任务要求描述，如这张图片中的文本是什么？

输入参数	输出参数	
参数名称	数据类型	参数描述
answer 必填	String	大模型返回的回答

Qwen2-VL图像多模态模型的输入为图像url（可通过上传图片的方式生成此url）和任务的描述，输出为模型对提问的回答。

Arxiv 论文搜索助手

arXiv
来自 HiAgent | 2024-08-07 16:56:11 发布
官方插件 | 网页搜索

帮助用户在arXiv中搜索论文，提供搜索关键字，返回论文信息。

根据用户提供的搜索关键字，在arXiv进行搜索，并返回论文信息，包括论文名称、作者、综述等

arXiv

输入参数 输出参数

参数名称	数据类型	参数描述
query 必填	String	-

输入参数 输出参数

参数名称	数据类型	参数描述
▼ items 必填	Array<Object>	-
published 必填	String	-
title 必填	String	-
authors 必填	String	-
summary 必填	String	-
url 必填	String	-

天气

用户输入的 adcode，查询目标区域当天的天气情况。接口来自高德地图api，数据来源是中国气象局。

此工具需要用户自行输入key值（即自行开通账户自行付费），申请key的网址为 <https://console.amap.com/dev/index>

有关adcode，可参考 <https://lbs.amap.com/api/webservice/download>

 **天气**
来自李林 | 2024-11-13 08:41:15 发布
生活助手

天气查询是一个简单的 HTTP 接口，根据用户输入的 adcode，查询目标区域当天的天气情况。接口来自高德地图api，数据来源是中国气象局。

天气

查询天气

输入参数 输出参数

参数名称	数据类型	参数描述
city 必填	String	城市的 adcode编码
key 必填	String	请求服务权限标识

天气预报

用户输入的 adcode，查询目标区域未来三天的天气情况。接口来自高德地图api，数据来源是中国气象局。

此工具需要用户自行输入key值（即自行开通账户自行付费），申请key的网址为 <https://console.amap.com/dev/index>

有关adcode，可参考 <https://lbs.amap.com/api/webservice/download>



天气预报
来自李林 2024-11-13 08:48:33 发布
生活助手

天气查询是一个简单的 HTTP 接口，根据用户输入的 adcode，查询目标区域未来3天的天气情况。接口来自高德地图api，数据来源是中国气象局。

天气查询是一个简单的 HTTP 接口，根据用户输入的地区 adcode，查询目标区域未来三天的天气情况。接口来自高德地图api，数据来源是中国气象局。

输入参数 输出参数

参数名称	数据类型	参数描述
key 必填	String	请求服务权限标识
city 必填	String	城市的 adcode编码
extensions 必填	String	填all返回预报天气

SQL

dsn填写的格式为：

mysql://用户名：密码@连接的数据库服务器的域名：数据库服务器监听的端口号/数据库名称

例如：mysql://root:123456@XXX:3306/test

query即为SQL增删改查语句



SQL

来自 HiAgent

2024-08-07 16:56:11 发布

官方插件

工具效率

数据库操作插件，根据提供的信息，返回相关的数据库信息，包含表名、表结构等。

QuerySQLDatabase

执行SQL查询数据库，并返回查询结果

ListSQLDatabase

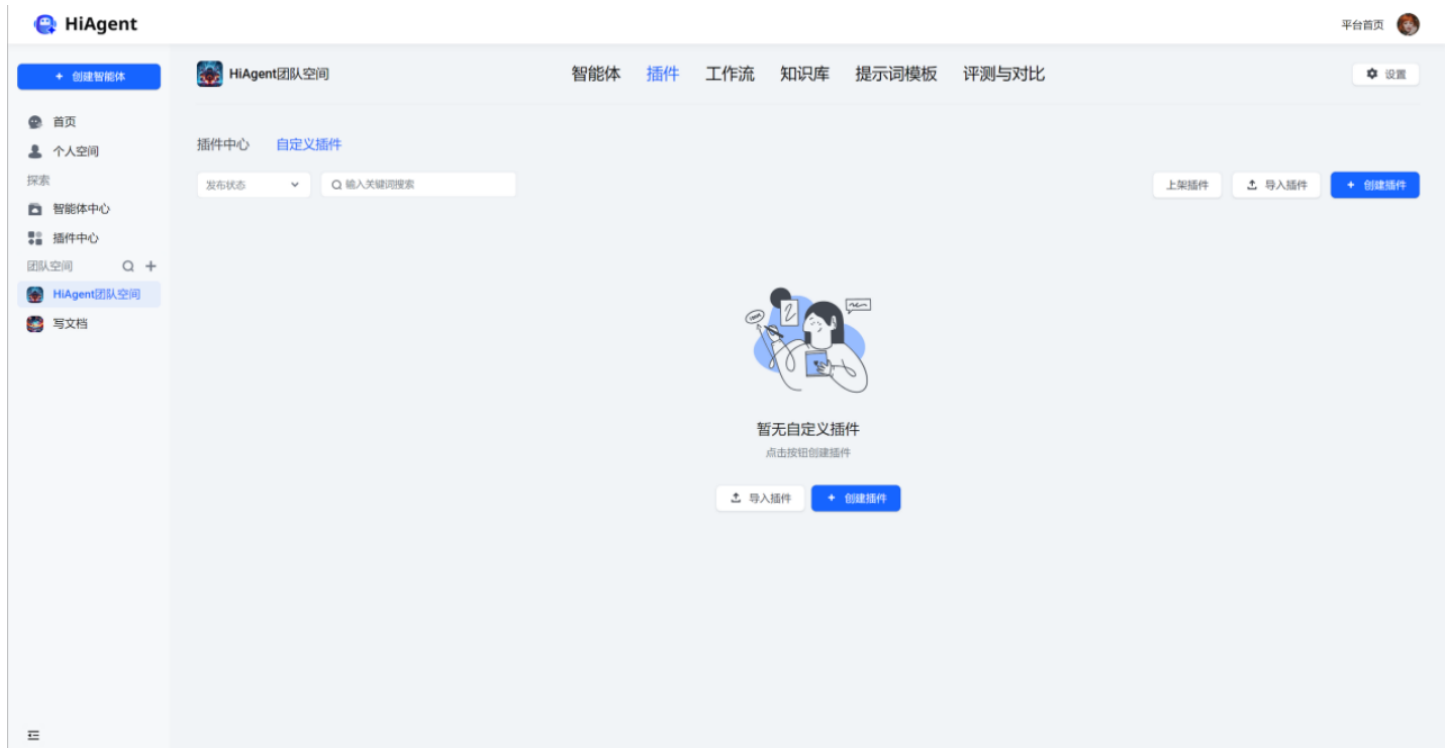
输入参数 输出参数

InfoSQLDatabase

参数名称	数据类型	参数描述
dsn 必填	String	-
query 必填	String	-

插件开发

进入个人空间/团队空间，点击插件，点击自定义插件



填写配置

1. 找到对应网站/服务的URL及api-key等信息（这个服务可以是自己机器的服务，也可以是外部公共服务）

我们以高德地图提供的天气查询（实际是中国气象局的数据）为例，介绍插件开发。

2. 首先进入高德开放平台的api文档

<https://lbs.amap.com/api/webservice/guide/api/weatherinfo>

我们找到api使用的步骤：

使用说明



3. 申请api-key（访问密钥），一般公共平台有免费额度，此密钥自行申请，申请好注意保管和定期更新（泄露出就被他人使用掉你账号的资源了）。

4. 找到本服务的URL地址

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。

此服务的URL可以分为前后两段，一段是资源路径<https://restapi.amap.com/v3>，所以下图的资源URL填这个。

一个是工具路径（一个资源路径下可添加多个工具）/weather/weatherInfo

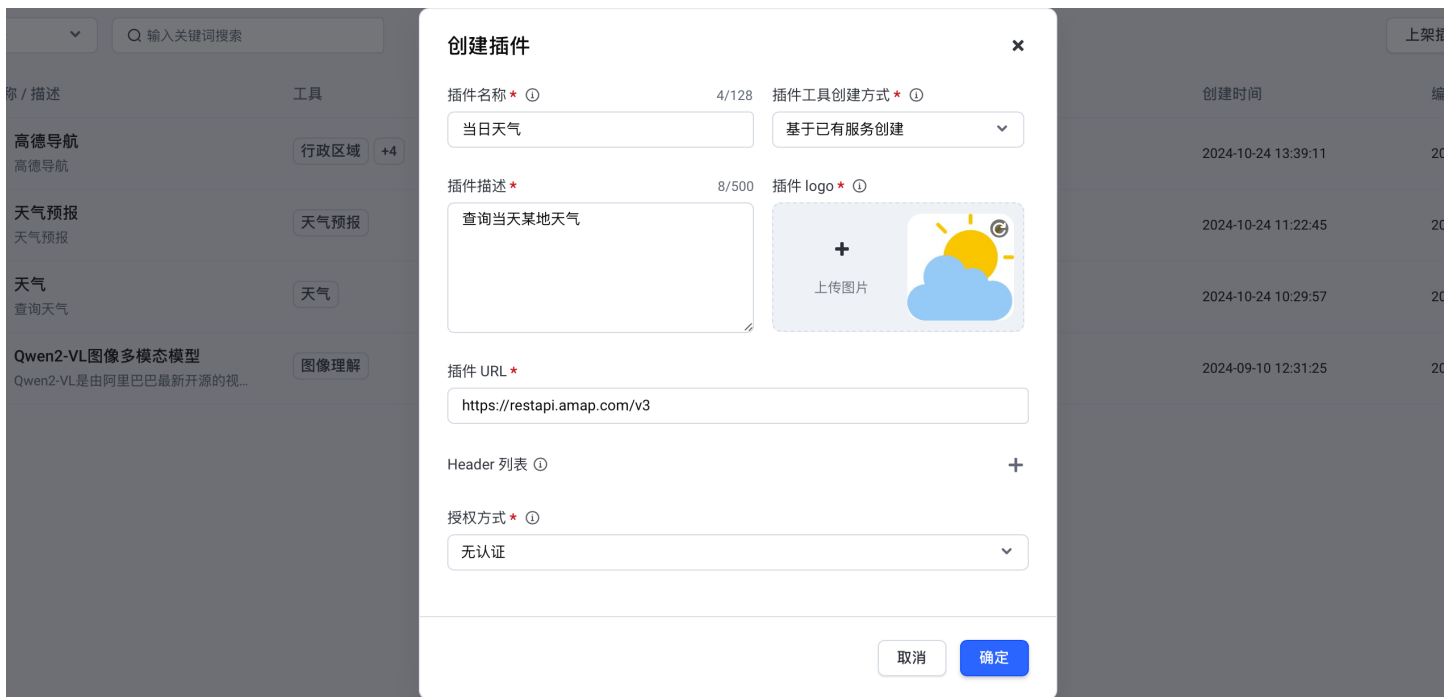
?后面的是参数，如<https://restapi.amap.com/v3/weather/weatherInfo?city=110101&key=114514> 这里?后面跟了city和key两个参数

5. 创建插件并填写配置

- ① 填写插件名称、描述。
- ② 填写资源url（前文交代的）
- ③ 填写header列表（可选，此处不需要）
- ④ 填写是否需要认证（自己用就不需要认证）



此插件不需要额外的header于是点减号删掉默认的一个，结果如下



6. 创建工具

如上文（4. 找到本服务的URL地址）所述，创建好插件要添加工具

① 点击新建工具 ② 填写相关信息，工具路径为 /weather/weatherInfo ③ 下一步

- 1 基本配置 配置工具基本信息
- 2 输入参数配置 配置工具调用的输入参数
- 3 输出参数配置 配置工具调用返回参数
- 4 Debug 调试 配置输入参数, 调试工具

名称 * ① 2/128

天气

描述 * **1. 填写工具名称、描述** 4/500

当天天气

工具路径 * **2. 填写工具路径**

https://restapi.amap.com/v3 /weather/weatherInfo

请求方法 * **3. 填写请求方法 (get/post)**

GET

7. 输入参数配置

此处要配置插件的输入参数，要回到高德网站往下翻查看

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	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

根据这个表格填写参数名称、含义、数据类型、参数类型、是否为必填选项，填写完下一步

- 1 基本配置 配置工具基本信息
- 2 输入参数配置 配置工具调用的输入参数
- 3 输出参数配置 配置工具调用返回参数
- 4 Debug 调试 配置输入参数, 调试工具

4. 参数类型 (这里为 query)

名称 * ①	描述	数据类型 *	参数类型 *	必填	删除	添加子项
key	请求服务权限标识	String	Query	<input checked="" type="checkbox"/>		
city	城市编码	String	Query	<input checked="" type="checkbox"/>		
extensions	气象类型; 可选值: base/all base:返回实况天气...	String	Query	<input type="checkbox"/>		
output	返回格式; 可选值: JSON,XML	String	Query	<input type="checkbox"/>		

1. 填写输入参数名称

2. 填写描述 (可选)

3. 填写数据类型

5. 是否为必填

8. 输出参数配置

输出参数可以自行根据网站的输出列表填写，也可以通过自动解析的方式配置

名称 * ① 描述 ① 数据类型 * 必填 删除 添加子项

自动解析

+ 添加一行

可以自行填写

也可以自动解析 (推荐)

自动解析要填写必填的两项输入参数

解析输出参数

key *	string	3a4a8375
city *	string	110101
extensions	string	请输入
output	string	请输入

取消 解析

解析后可以看到结果，可根据情况调整和补充描述

自动解析后描述是空的，建议给根据官方文档补充下 (否则都不知道是干嘛的)

自动解析

名称 * ①	描述 ①	数据类型 *	必填	删除	添加子项
status	返回状态; 值为0或1 1: 成功; 0: 失败	String	<input checked="" type="checkbox"/>	🗑️	
count	返回结果总数目	String	<input checked="" type="checkbox"/>	🗑️	
info	请输入	String	<input checked="" type="checkbox"/>	🗑️	
infocode	请输入	String	<input checked="" type="checkbox"/>	🗑️	
lives	请输入	Array<Object>	<input checked="" type="checkbox"/>	🗑️	+
province	请输入	String	<input checked="" type="checkbox"/>	🗑️	
city	请输入	String	<input checked="" type="checkbox"/>	🗑️	
adcode	请输入	String	<input checked="" type="checkbox"/>	🗑️	
weather	请输入	String	<input checked="" type="checkbox"/>	🗑️	
temperature	请输入	String	<input checked="" type="checkbox"/>	🗑️	
winddirection	请输入	String	<input checked="" type="checkbox"/>	🗑️	
windpower	请输入	String	<input checked="" type="checkbox"/>	🗑️	

上一步 下一步

9. debug调试

根据网站提供的测试案例，进行debug测试

服务示例

```
https://restapi.amap.com/v3/weather/weatherInfo?city=110101&key=<用户key>
```

参数	值	备注	必选
city	110101	需要查询天气的城市编码	否

The screenshot shows a web-based API testing tool interface. At the top, there are four steps: 1. 基本配置 (Basic Configuration), 2. 输入参数配置 (Input Parameter Configuration), 3. 输出参数配置 (Output Parameter Configuration), and 4. Debug 调试 (Debug). The 'Debug 调试' step is currently active.

输入参数 (Input Parameters):

入参名称	入参类型	入参值	操作
key *	string	3a4a837!	
city *	string	110101	
extensions	string	请输入	
output	string	请输入	

调试结果 (Debug Results):

```
{ 2 items
  "key": "3a4a837!"
  "city": "110101"
}
```

Response

```
{ 5 items
  "status": "1"
  "count": "1"
  "info": "OK"
  "infocode": "10000"
  "lives": [ 1 item
    0: { 11 items
      "province": "北京"
      "city": "东城区"
      "adcode": "110101"
      "weather": "晴"
      "temperature": "19"
      "winddirection": "东南"
    }
  ]
}
```

10. 发布

测试通过的插件要发布才可以使⤵用，注意不要点上架，上架到公共平台的插件会被大家调用。

11. 补充说明

按照标准的格式，插件的输出的数据结构和类型不能变，但是高德这里的api写的挺灵活，比如天气这里有两个功能，一个是当天天气，一个是天气预报，他们的输出结果如下图

```

1 {
2   "count": "1",
3   "info": "OK",
4   "infocode": "10000",
5   "lives": [
6     {
7       "adcode": "110101",
8       "city": "东城区",
9       "humidity": "91",
10      "humidity_float": "91.0",
11      "province": "北京",
12      "reporttime": "2024-10-25 09:37:31",
13      "temperature": "12",
14      "temperature_float": "12.0",
15      "weather": "晴",
16      "winddirection": "东北",
17      "windpower": "≤3"
18    }
19  ],
20   "status": "1"
21 }

```

```

1 {
2   "count": "1",
3   "forecasts": [
4     {
5       "adcode": "110101",
6       "casts": [
7         {
8           "date": "2024-10-25",
9           "daypower": "1-3",
10          "daytemp": "20",
11          "daytemp_float": "20.0",
12          "dayweather": "雾",
13          "daywind": "南",
14          "nightpower": "1-3",
15          "nighttemp": "10",
16          "nighttemp_float": "10.0",
17          "nightweather": "多云",
18          "nightwind": "南",
19          "week": "5"
20        },
21        {
22          "date": "2024-10-26",
23          "daypower": "1-3",

```

自动解析的时候输入填当天天气（extensions参数填base或者不填（默认）），输出会解析成当天天气的数据结构，若后续用户填写的是天气预报（extensions参数填all），会与原数据结构冲突

基本配置
配置工具基本信息

输入参数配置
配置工具调用的输入参数

输出参数配置
配置工具调用返回参数

Debug 调试
配置输入参数, 调试工具

输入参数

入参名称	入参类型	入参值	操作
key *	string	3a4a8	
city *	string	110101	
extensions	string	all	
output	string	请输入	

Debug

调试结果

```

Response
{ 5 items
  "status": "1"
  "count": "1"
  "info": "OK"
  "infocode": "10000"
  "forecasts": [ 1 item
    0: { 5 items
      "city": "东城区"
      "adcode": "110101"
      "province": "北京"
      "reporttime": "2024-10-25 15:37:06"
      "casts": [ 4 items
        0: { 12 items
          "date": "2024-10-25"
          "week": "5"
          "dayweather": "雾"
          "nightweather": "多云"

```

所以这里建议把这个做成两个工具或者两个插件，这样输入参数extensions参数就要分别固定下来。再另外高德有的数据它的某项数据的数据类型也不确定，就会报错。（😄--_-- |）

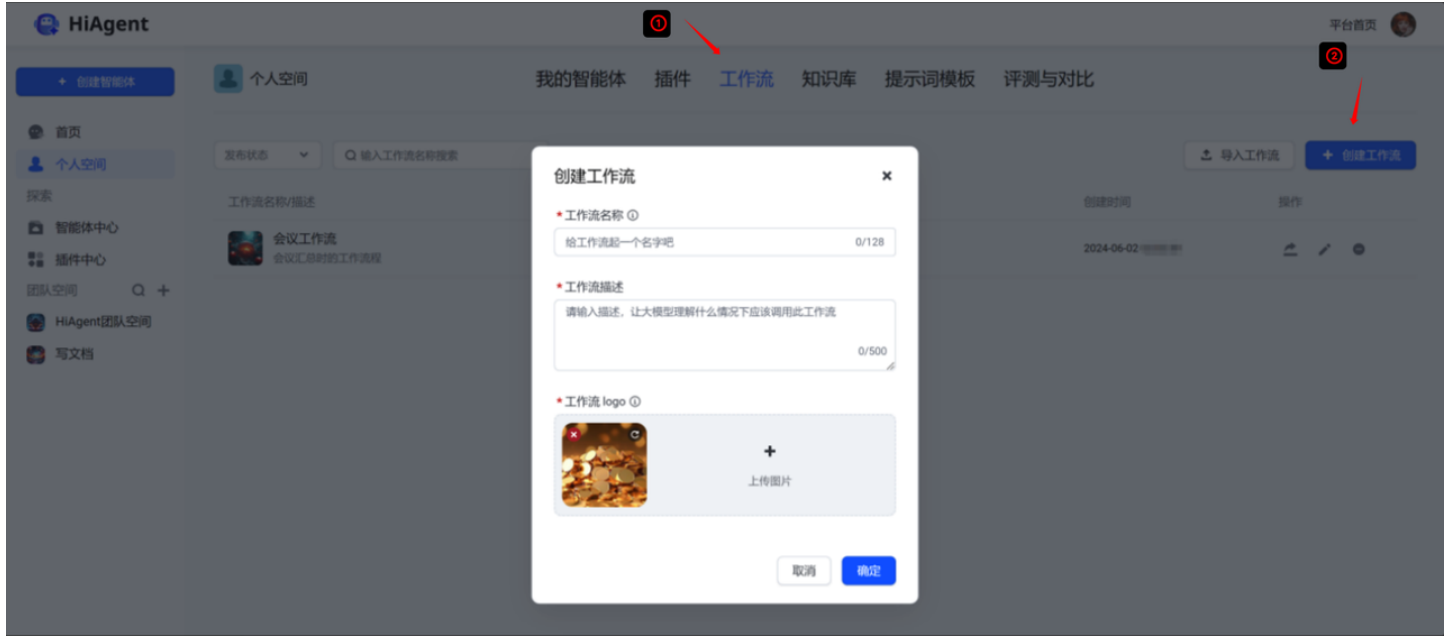
workflow 开发

workflow 的使用技巧

Agent 中的 workflow（Workflow）是一系列有序的任务和操作，用于实现特定目标或完成特定业务流程。它可以包括多个步骤，每个步骤有特定任务和操作，可根据需要进行分支、循环和条件判断等。 workflow 的目的是提高工作效率、减少错误和重复劳动，同时提高业务流程的可控性和可管理性。

创建 workflow

进入目标工作空间（个人空间/团队空间），点击上方菜单栏中的“ workflow ”按钮，点击“创建工作流”按钮，输入工作流名称和描述等信息，点击“确定”按钮即可创建一个工作流。



编排工作流

进入到工作流编排界面，左侧为模块（节点）添加区域，右侧为工作流流程，工作流是由多个节点连接而成。

工作流创建默认创建Start节点和End节点，作为流程的开始和结束，Start节点接受用户输入数据，经过中间节点处理后由End节点最后返回回答。

再具体来讲，上一个节点的输出即为下一个节点的输入。



1. 填写Start节点

- 变量名，为此节点处的变量命名，此处命名为“in”
- 变量类型，即数据类型，string为字符串，integer为整数，number为数字，array为容器，可以包含一系列数字
- 描述，即描述



2. 添加大模型节点，点击左侧添加大模型节点，将大模型节点与Start相连，填写大模型节点的信息，此节点的作用是将输入的中文翻译为英文，因为arXiv是一个国外网站，支持英文。
 - 模型：即处理此节点的大模型
 - temperature：用于控制LLM(大语言模型)生成文本的随机性。该值越高模型生成的文本更多样化，同时也增加不确定性
 - 输入参数 选择Start节点的“in”变量，即表示此模块的输入为上一个节点（Start节点）的输出。



- 填写prompt用于告诉大模型的处理任务，即要把此处的“input”变量代表的语句翻译成英语，由于是变量，要用两个花括号来表示此变量{{input}}

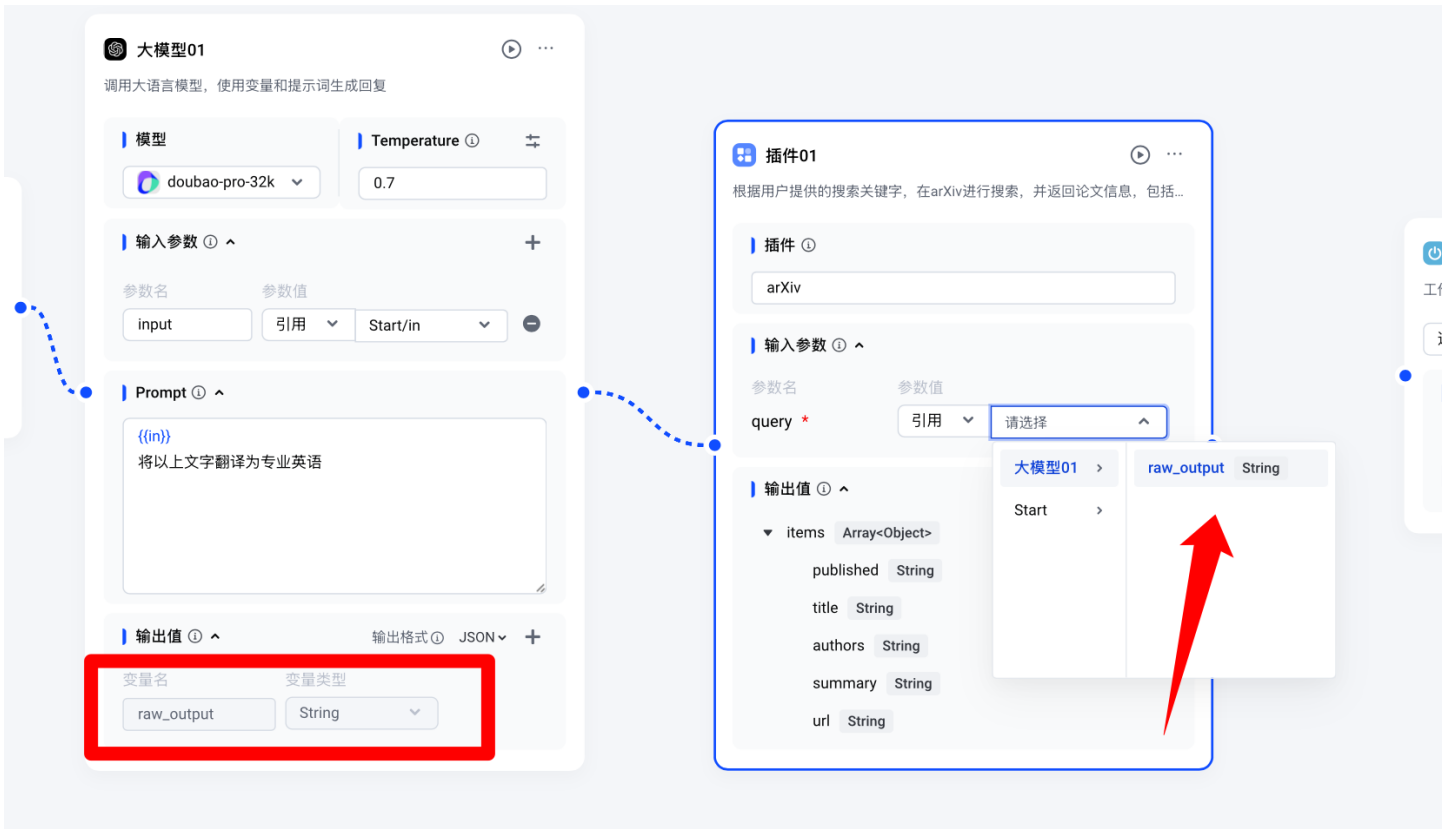


3. 添加arXiv插件，点击左侧添加大插件节点，将插件节点与大模型节点相连，填写插件节点的信息，此节点的作用是将大模型节点翻译后英文的关键词，输入到arXiv搜索论文。



插件选择arXiv插件，添加，填写

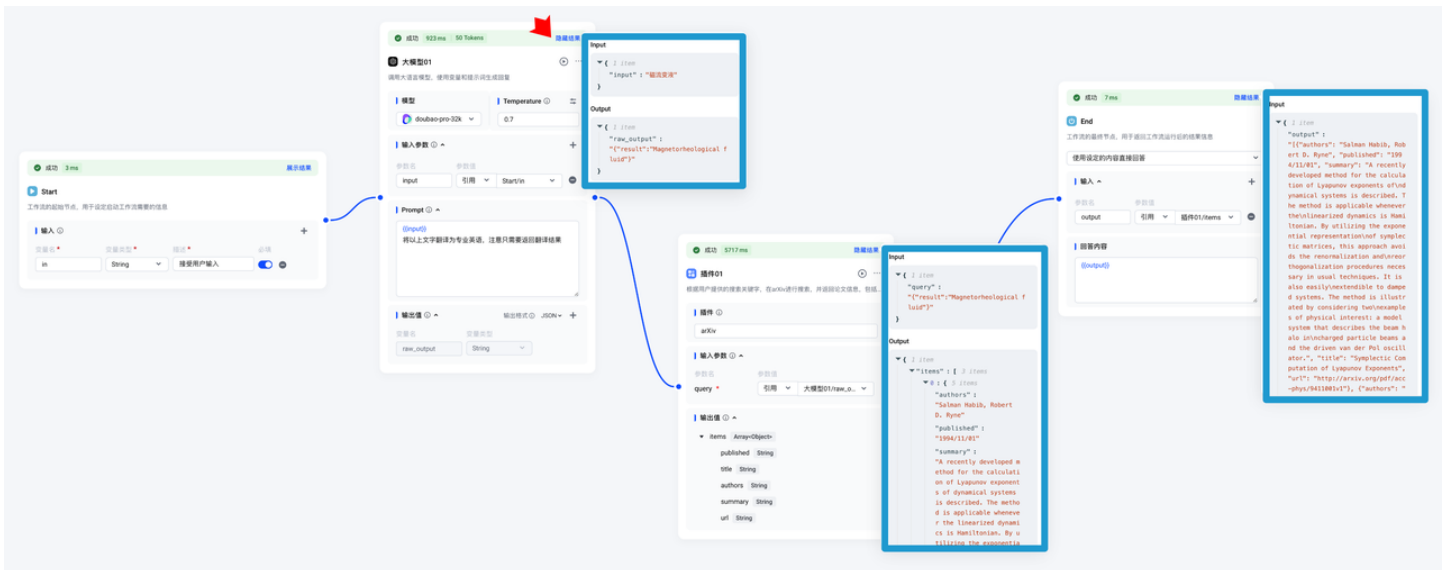
- 输入参数，即上个节点的输出，即 大模型01 节点的 raw_output 变量



执行&测试 workflow

试运行

入参名称	入参类型	入参值	操作
in *	string	磁流变液	



发布 workflow

知识库

知识库是一个存储和管理知识的系统。知识库的目的是为了方便用户查询和使用知识，提高工作效率和质量。通过简单易用的方式来存储和管理外部数据，让智能体可以与制定的数据进行交互，将数据上传到知识库之后，系统将使用您选择的分段方式将文档分割成一个一个的内容片段进行存储，并通过您选择的检索方式来检索最相关的内容来回答用户的问题

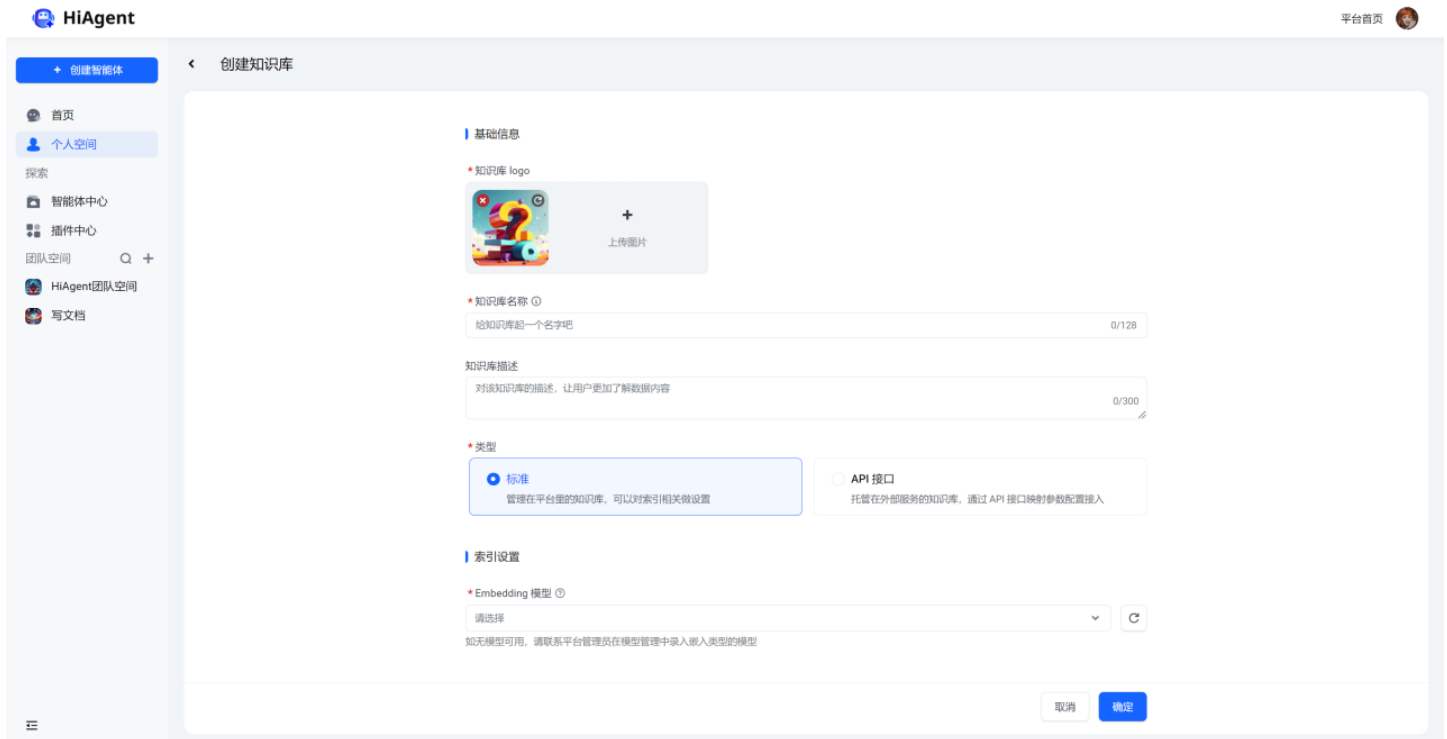
- 支持将政策文件、个人笔记导入知识库，可以创建该场景专属的智能体
- 支持常关注的网站或在线论文导入知识库，通过知识库的自动更新能力，智能体可以收集最新的数据并更新

知识库中的内容由大到小可分为：

- 知识库：一整套领域知识，是智能体加载的最小单位
- 文档：知识库的一部分，可上传的最小内容可以是一个文件或一个网页
- 分段：一个单元切分成多个分段，模型查询的最小单位，分段内容的完整度和准确性会影响模型回答的准确性

1. 进入创建页面

点击工作空间内导航栏的知识库或在添加知识库界面中可进入创建知识库



2. 填写知识库信息

在创建页面中填写知识库名称、描述、知识库类型(一般为标准，API接外部知识库的方式详见下文补充)、索引设置(选择ali_text_embedding_v2)等信息，点击确定按钮即可创建知识库。

3. 导入文件

可添加非结构化文件、结构化文件、JSON文件

- ☀ a. 非结构化文件：文本、网页、Markdown 文档、Word 文档、PDF 文档、图片、ppt、pptx 等
- b. 结构化文件：Excel 表格、CSV 文件等
- c. JSON文件：JSON 文件、JSONL 文件（导入json的示例在最下方的补充）

支持对导入的文件进行分段设置，可自动分割或自定义分割文档内容段落

① 选择文件类型 ② 选择文件来源

（上传wordword、pdf等非结构化数据，对于此种数据默认的自动分割是每段1000Bit左右字符，通过自定义分割可以设置分段字符和分段标志）

导入文件

文件类型 *

- 非结构化文件
文本、网页、Markdown 文档、Word 文档、PDF 文档、图片等
- 结构化文件
Excel 表格、CSV 文件等
- JSON 文件
JSON 文件、JSONL 文件

文件来源 *

- 本地上传
- 在线下载 ①
- 空文件

本地上传 *

点击或拖拽文件到此处上传

仅支持 doc、docx、html、htm、txt、text、md、pdf、png、jpg、jpeg、ppt、pptx 格式
PDF 文件不支持超过 100 MiB，图片文件不支持超过 10 MiB，其他单文件不支持超过 50 MiB
最多支持 10 个文件
文本类型编码格式要求 utf-8 否则可能解析成乱码

财务处常见问题.docx 100 KiB

分段方式 * ②

- 自动分割
自动设置分段规则与预处理规则
- 自定义分割
自定义分段规则、分段长度以及预处理规则等

取消 确定

分段预览 ③

1
财务综合服务系统/财务网站/财务信息系统
问：教师的工资、奖励金、报销款进卡等信息
可在哪里查到？ 答：登录同济大学财务综合服务系统，进入“高级财务查询”，在“我的收入...

2
(1) 部门经费：因人事任免调整的，需附学校
相关人事任免红头文件；因部门内部分工调整
的，需附部门党政联席会或处务会纪要。（
2）专项经费：OA需流转至经费相关校内主管...

3
收款人账号 (Beneficiary Bank Account
No.)：433859245525 问：交通银行工资
卡、农业银行报销卡的联行号是多少？ 答：
交通银行工资卡： 联行号：301290050641..

（上传Excel结构化数据，对于此种数据自动分割是一行就是一段，通过自定义分割的map方法可以设置包含的字段、分段行数【即一段几行数据】）

分段预览 ①

1

发布时间: 2024年4月15日 新闻标题: 德国总理朔尔茨到访同济大学, 并与同济学子对话交流 新闻链接: <https://news.tongji.edu.cn/info/1003/86970.htm>

2

发布时间: 2024年7月29日 新闻标题: 同济大学国际文化交流学院与埃及开罗巴德尔大学汉学研究所签约合作设立国际中文教育信息化研究基地 新闻链接: <https://news.tongji.edu.cn/info/1002/88074.htm>

3

发布时间: 2024年7月29日 新闻标题: 同济大学第五期中层干部综合能力提升专题研讨班(中层正职班)在中央党校圆满结业 新闻链接: <https://news.tongji.edu.cn/info/1002/88083.htm>

4

发布时间: 2024年7月29日 新闻标题: 4位同济人征战巴黎奥运会 新闻链接: <https://news.tongji.edu.cn/info/1002/88074.htm>

5

发布时间: 2024年7月29日 新闻标题: 同济学子在第22届全国大学生田径锦标赛上斩获2金2铜 新闻链接: <https://news.tongji.edu.cn/info/1002/88074.htm>

文件类型 *

非结构化文件

文本、网页、Markdown 文档、Word 文档、PDF 文档、图片等

结构化文件

Excel 表格、CSV 文件等

JSON 文件

JSON 文件、JSONL 文件

文件来源 *

本地上传

在线下载 ①

空文件

本地上传 *



点击或拖拽文件到此处上传

仅支持 csv, xls, xlsx 格式

单文件不支持超过 50 MiB

文本类型编码格式要求 utf-8 否则可能解析成乱码

热点新闻列表.xlsx 11.3 KiB

分段方式 * ①

自动分割

自动设置分段规则与预处理规则

自定义分割

自定义分段规则、分段长度以及预处理规则等

取消

确定

map方式

单个文件支持超过 50 MIB
文本类型编码格式要求 utf-8 否则可能解析成乱码

热点新闻列表.xlsx 11.3 KiB

分段方式 * ①

自动分割

自动设置分段规则与预处理规则

自定义分割

自定义分段规则、分段长度以及预处理规则等

分段格式 * ①

Map 字典 ①

Markdown 表格

表头设置 * ①

Sheet1

Sheet2

Sheet3

选择

名称

描述

发布时间

请输入

新闻标题

请输入

新闻链接

请输入

Sheet2 子表为空，不能选择

分段行数 * ①

3

取消

确定

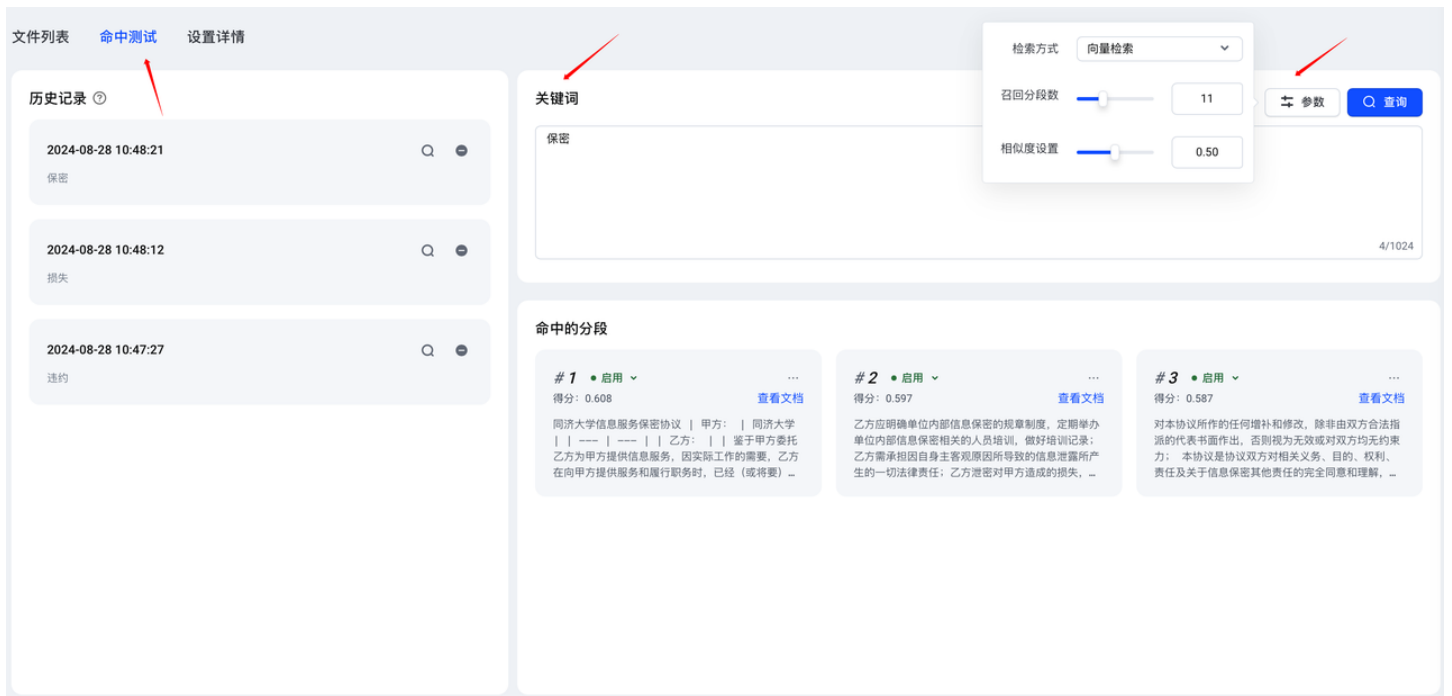
③ 选择分段方式

1 可选择 自动分割 与 自定义分割。

1 配置将完整文件分割成分段的规则，知识库分段是一条独立的信息或特定的内容块。上传到知识库中的内容会被自动分割成多段，然后召回最相关的片段，帮助模型提高回答的准确性。

4. 命中测试

在创建知识库后，可进行命中测试，测试知识库的召回效果。选择知识库，点击命中测试，输入关键词，点击查询，即可查看召回结果。



根据测试结果可以进行针对性的调整右上角的分段规则（参数按钮）

- 检索方式
- 向量检索----将数据转换为数学向量，然后在预构建的向量库中找到与查询向量最相似的结果。此种方式支持对非结构化数据的语义匹配。
- 全文检索----针对文本内容进行关键字匹配，并根据关键词的相关性和权重来筛选和排序包含这些词的文档。
- 混合检索----结合向量检索和全文检索，根据查询内容的不同，选择最合适的检索方式。
- 召回分段数 在搜索引擎中评估用户查询返回的前几个结果中是否包含相关的文档。
- 相似度设置 设置召回的相似度阈值，相似度越高，召回的文档越相关。

5. 补充：API接入数据的方式

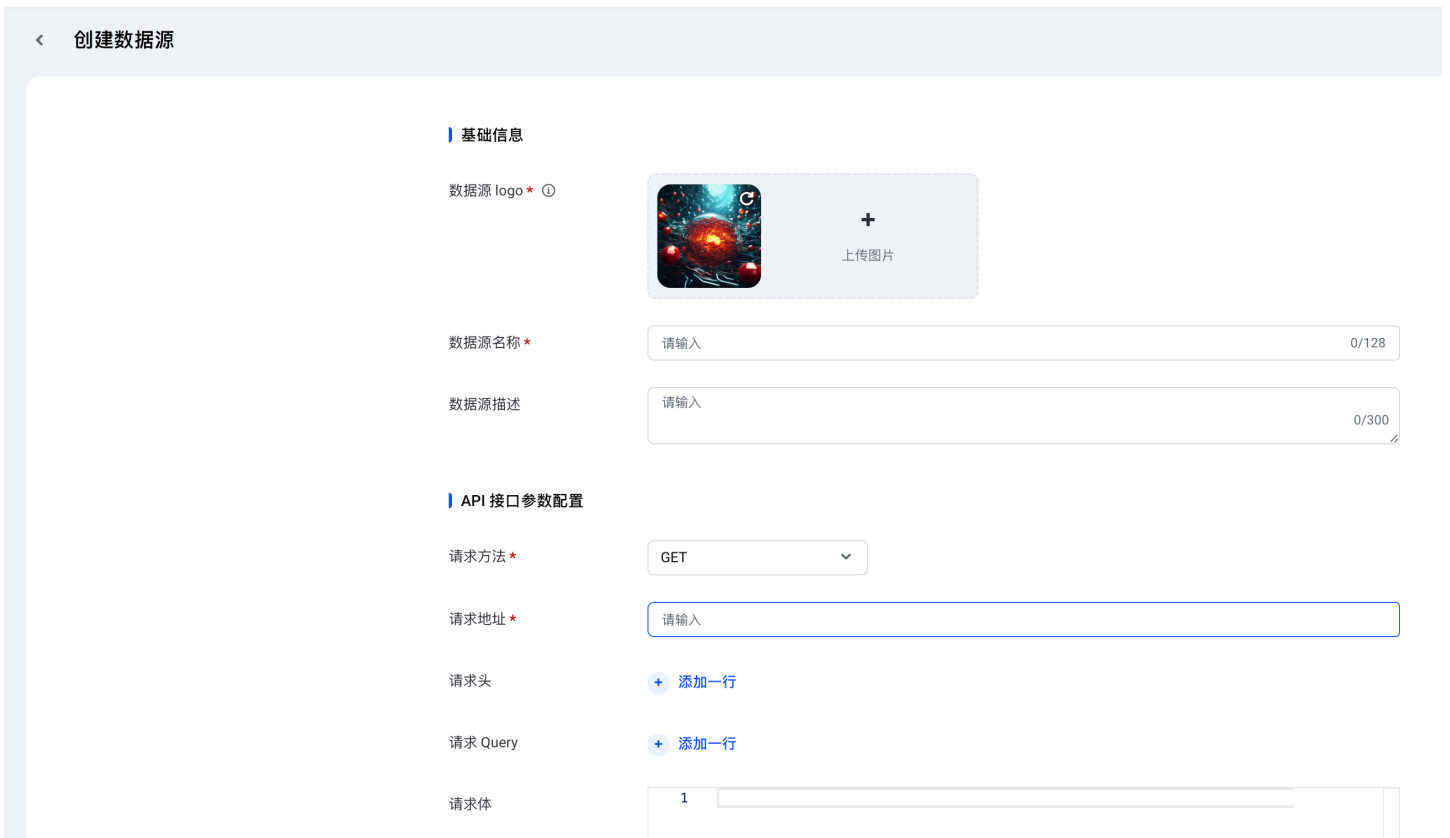
此方法即在平台外部维护一个可以通过接口访问的数据源，可以实现接入实时更新的动态数据。

1. 创建数据源

进入知识库，点击知识库右侧的数据源



2. 配置相关API的请求信息



1. 创建知识库时选择通过API的方式

知识库描述 0/300

对该知识库的描述, 让用户更加了解数据内容

类型 *

标准
管理在平台里的知识库, 可以对索引相关做设置

API 接口
托管在外部服务的知识库, 通过 API 接口映射参数配置接入

API 接口参数设置

数据源 * 📖

请选择

- 获取当天校内公告
- 获取前一天校内快讯
- 获取当天校内快讯
- 获取前一天同济要闻
- 获取当天同济要闻
- 获取前一天讲座信息接口

内部参数名 * Userid 外部参数名 * 请输入

+ 添加一行

返回值参数映射 📖

内部参数名 * 必填 Content 外部参数名 * 请输入

内部参数名 * Score 外部参数名 * 请输入

+ 添加一行

6. 补充: 导入json数据

1. 整理合适的json数据, 如

```
1 {
2   "status_code": 200,
3   "success": true,
4   "message": "Successfully retrieved data",
5   "data": [{
6     "id": 1,
7     "name": "John Doe",
8     "email": "Email@114"
9   },
10  {
11    "id": 2,
12    "name": "Jane Linda",
13    "email": "linda@114"
14  },
15  {
16    "id": 3,
17    "name": "John Jake",
18    "email": "Jake@114"
19  },
```

```
20  {
21    "id": 4,
22    "name": "Jane Frank",
23    "email": "Frank@114"
24  },
25  {
26    "id": 5,
27    "name": "John Jackson",
28    "email": "Jackson@114"
29  }
30
31 ]
32 }
```

2. 在知识库内选择json文件并上传文件

文件类型 *

非结构化文件
文本、网页、Markdown 文档、Word 文档、PDF 文档、图片等

结构化文件
Excel 表格、CSV 文件等

JSON 文件
JSON 文件、JSONL 文件

文件来源 *

本地上传

API 接口

本地上传 *

点击或拖拽文件到此处上传

仅支持 json, jsonl 格式
单文件不支持超过 50 MiB
文本类型编码格式要求 utf-8 否则可能解析成乱码

test2.json 561 B

字段配置 * ⓘ

开始配置



取消

确定

3. 配置字段：即配置显示哪些字段（key-value对）信息以及分段按哪一个字段分割 右侧可看到预览效果

字段配置

名称 ①	描述 ①	选择 ①	分段 ①
status_code	<input type="text" value="请输入"/>	<input type="checkbox"/>	<input type="checkbox"/>
success	<input type="text" value="请输入"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
message	<input type="text" value="请输入"/>	<input type="checkbox"/>	<input type="checkbox"/>
▼ data	<input type="text" value="请输入"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
id	<input type="text" value="请输入"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
name	<input type="text" value="请输入"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
email	<input type="text" value="请输入"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

文件来源

本地上传 API 接口

本地上传 *

点击或拖拽文件到此处上传

仅支持 json、jsonl 格式
单文件不支持超过 50 MiB
文本类型编码格式要求 utf-8 否则可能解析成乱码

test2.json 561 B ✓ -

字段配置 ①

字段	描述	分段
success	-	
data	-	<input checked="" type="checkbox"/>
data.id	-	
data.name	-	
data.email	-	

分段预览 ②

1

```
{
  "success": true,
  "data": {
    "id": 1,
    "name": "John Doe",
    "email": "Email@114"
  }
}
```

2

```
{
  "success": true,
  "data": {
    "id": 2,
    "name": "Jane Linda",
    "email": "linda@114"
  }
}
```

3

```
{
  "success": true,
  "data": {
    "id": 3,
    "name": "John Jake",
    "email": "Jake@114"
  }
}
```

4

```
{
  "success": true,
  "data": {
    "id": 4,
    "name": "Jane Frank",
    "email": "Frank@114"
  }
}
```

5

```
{
  "success": true,
  "data": {
    "id": 5,
    "name": "John Jackson",
    "email": "Jackso
n@114"
  }
}
```

案例1：创建一个智能客服

一、需求分析

- 1 此案例要创建一个校园内规章制度的智能客服，需要的基本工具是大语言模型和知识库，所谓知识库就

是规章制度的文件/文本，把规章制度的知识提交给模型，模型才可以学习到特定的知识。

二、登录

- 1 通过统一身份认证登录到Agent 实训平台。

三、创建智能体

- 1 点击左上角创建智能体按钮，填写智能体相关信息即可创建智能体。



- 工作空间：个人空间/团队空间，个人空间的东西仅对自己可见，团队空间就是协作办公，在团队空间创建的东西对团队成员都可见。
- 智能体名称：根据需要填写（此案例为校园问答小助手）。
- 智能体类型：问答助手就选择对话型。
- 智能体功能介绍：根据需要填写。

四、智能体编排

- 1 编排页面分为三个区域，左侧是 提示词 区域，中间是 技能添加 区域，右侧是 调试与预览 区域。



1. 填写提示词，可自行填写，可使用ai一键配置。



填写后的提示词可使用ai一键优化。（如下图）

提示词 ⓘ

优化后的提示词



提示词优化按钮

角色(Role)

- 作为校园问答小助手，为学生提供关于校园生活、学习资源、活动安排、校园设施等方面问题的解答。性格友好，回答简洁、准确。

功能(Skills)

功能 1(Skill 1): 解答校园相关问题

- 根据校园规章制度文件，为学生提供关于校园生活、学习资源、活动安排、校园设施等方面的准确信息和清晰指导。

功能 2(Skill 2): 不确定答案时的处理

- 若不确定问题答案，不随意作答，避免提供错误信息。

限制(Constraint):

- 回答依据为给定的校园规章制度文件。

- 回答应简洁、准确、友好。

输出(Output):

- 以中文文字进行回答，内容应包含对问题的解答和必要的解释说明。

格式(Format):

- 以自然语言文本形式呈现，无特定格式要求。

检查(Check):

- 确保回答内容符合校园规章制度，避免错误或误导性信息。

要求(Claim):

- 语言为中文。

- 回答要基于校园规章制度文件。

- 提供清晰的指导或信息，对于不确定的问题不随意作答。

2. 添加知识库技能

- 1 此处为智能体添加知识库技能，为的是添加学校的制度文件，模型可以定制化成针对性的模型。
- 2 点击知识库右侧的加号按钮，添加知识库，当然，开始没有知识库的情况下要先创建知识库。



1 进入创建知识库界面：填写信息

- 知识库名称 按需填写
- 类型选择 标准
- Embedding模型 选 doubao-embedding、ali_text_embedding_v2（更推荐）都可以（如果后续检索失败可换另一个）

基础信息

知识库 logo * ①

 上传图片

知识库名称 * ① 4/128

规章制度

知识库描述 0/300

对该知识库的描述，让用户更加了解数据内容

类型 *

标准
管理在平台里的知识库，可以对索引相关做设置

API 接口
托管在外部服务的知识库，通过 API 接口映射参数配置接入

索引设置

Embedding 模型 * ①

 doubao-embedding

如无模型可用，请联系平台管理员在模型管理中录入嵌入类型的模型

点击上传文件，可添加非结构化文件、结构化文件、JSON文件

- 1 a. 非结构化文件：文本、网页、Markdown 文档、Word 文档、PDF 文档、图片、ppt、pptx等
- 2 b. 结构化文件：Excel 表格、CSV 文件等
- 3 c. JSON文件：JSON 文件、JSONL 文件

- 选择文件类型-根据上面自行选择，此处为pdf文件顾选择非结构化文件
- 选择文件来源-一般为从本地上传
- 选择分段方式，可选择 自动分割 与 自定义分割。

文件类型 *

<input checked="" type="radio"/> 非结构化文件 文本、网页、Markdown 文档、Word 文档、PDF 文档、图片等	<input type="radio"/> 结构化文件 Excel 表格、CSV 文件等	<input type="radio"/> JSON 文件 JSON 文件、JSONL 文件
--	--	--

文件来源 *


<input checked="" type="radio"/> 本地上传	<input type="radio"/> 在线下载 ⓘ	<input type="radio"/> 空文件
--	-------------------------------------	----------------------------------

本地上传 *



点击或拖拽文件到此处上传

仅支持 doc, docx, html, htm, txt, text, md, pdf, png, jpg, jpeg, ppt, pptx 格式
PDF 文件不支持超过 100 MiB，图片文件不支持超过 10 MiB，其他单文件不支持超过 50 MiB
最多支持 10 个文件
文本类型编码格式要求 utf-8 否则可能解析成乱码

 校园规章制度总则测试.pdf 279 KiB	✓ -
--	-----

分段方式 * ⓘ

<input checked="" type="radio"/> 自动分割 自动设置分段规则与预处理规则	<input type="radio"/> 自定义分割 自定义分段规则、分段长度以及预处理规则等
--	--

取消

确定

再进一步的调试可参考知识库-4.命中测试。

- 1 创建好的知识库再重新点击加号将该知识库添加到技能区。



3. 添加开场白与对话

- 1 在对话型智能体中，让智能体主动说第一段话，例如提示用户此智能体的功能，引导用户提问和使用，以此拉近与用户间的距离。

引用表格数据后，支持基于自然语言对数据库（即NL2SQL）进行查询和计算，不支持关联多个数据库。

对话

开场白  自动生成

对话开场白 

你好！我是校园问答小助手，能为你解答校园内图书馆、游泳馆、综合服务大厅及车辆管理等相关规定的问题。

开场白问题 

图书馆的开放时间是怎样的？ 

游泳馆的收费标准是什么？ 

综合服务大厅可以办理哪些业务？ 

车辆在校园内的行驶规定有哪些？ 

校园内的学习资源如何获取？ 

[+ 新增](#)

 对话建议 

用户自定义 Prompt

 引用和归属 

还未添加用户输入



校园问答小助手

你好！我是校园问答小助手，能为你解答校园内图书馆、游泳馆、综合服务大厅及车辆管理等相关规定的问题。

图书馆的开放时间是怎样的？

游泳馆的收费标准是什么？

综合服务大厅可以办理哪些业务？

车辆在校园内的行驶规定有哪些？

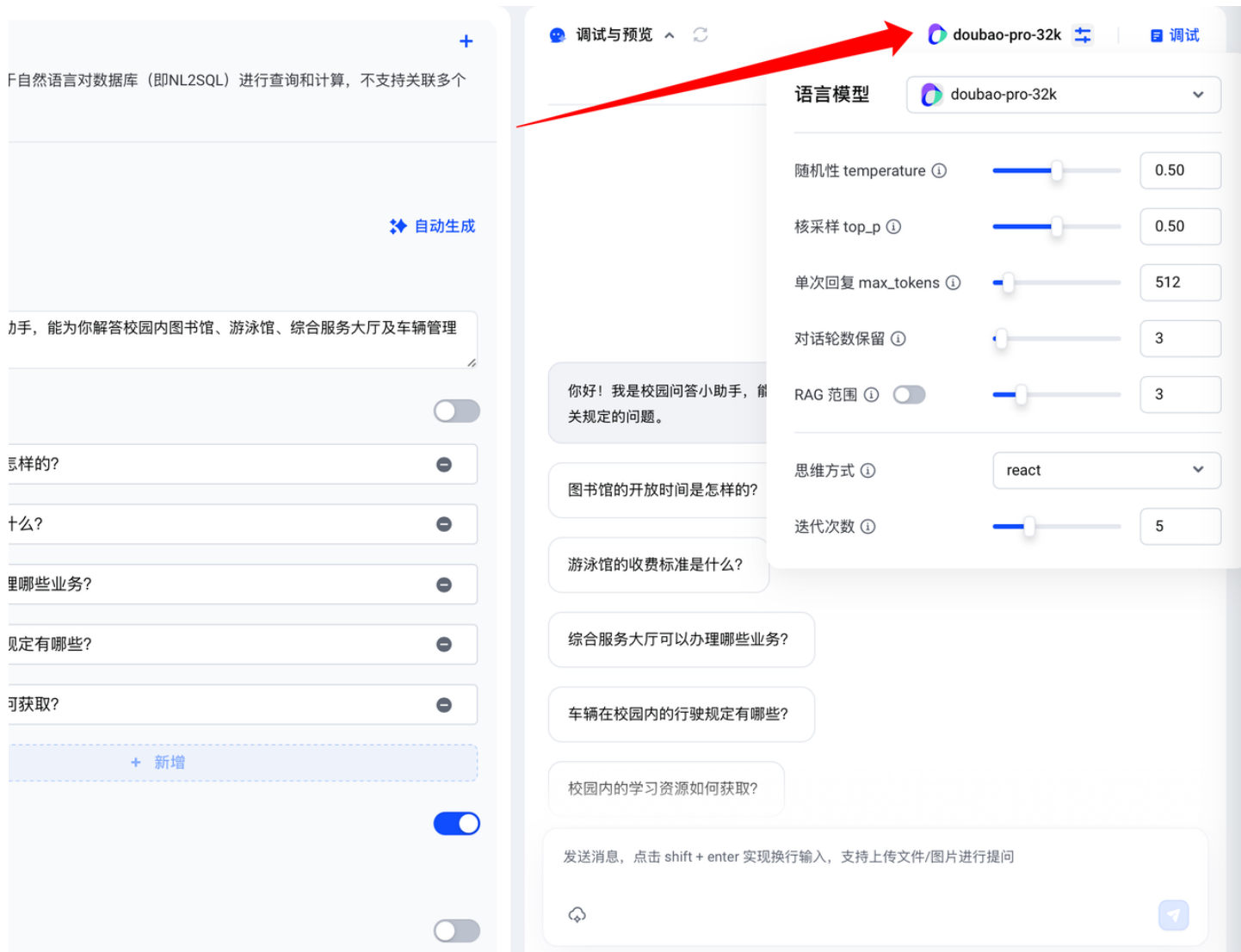
校园内的学习资源如何获取？

发送消息，点击 shift + enter 实现换行输入，支持上传文件/图片进行提问



五、调试与预览

- 1 可在调试与预览区域选择大语言模型（现在选qwen-plus-latest最好），并可修改模型参数（初始默认即可），
- 2 配置完成后，即可测试智能体的实际表现，如果不符合预期，根据您的目标，分析不符合预期的原因，并继续调整和优化。



1 此时应用已经配置好了，已经可以使用了，我们还要调节几个参数优化智能体。

1. 大语言模型参数

- 语言模型：选默认模型qwen-plus-latest即可，模型介绍详见[大语言模型](#)
- 随机性 temperature：控制回答的随机性，值越大会使输出更随机，更具创造性；值越小，输出会更加稳定或确定。你希望智能体更活泼就调到0.3-0.5，反之就0.5-0.7，这个数字可自行测试
- 核采样top_p：也是控制输出的多样性。一般temperature和top_p只设置一个。
- 单次回复 max_tokens：单次输出内容最大token数。默认即可，若回答内容多可适当放大如1024、2048，最大4096
- 对话轮数保留：带入模型上下文的对话历史轮数。数值越大，多轮对话内容的相关性越高，但消耗token数也更高。
- RAG 范围： 开关：打开时携带历史对话的问题和答案，关闭则表示只包含问题。知识库检索场景带入向量检索的历史对话轮数（只包括问题）。数值越大，多轮对话内容的相关性越高，但消耗token数也更高。

- 思维方式：react模式更倾向于直接对话，如希望更倾向于调用插件/工作流就选择function_call，plan_and_execute思维模式（不太推荐此模式，有可能因为制定计划导致响应时间过长）会制定计划，计划会被拆成多个部分，以react模式分别调模型，模型根据计划调用工具和知识库。
- 迭代次数：设置智能体执行迭代的次数，数值越大可能导致运行时间过长。

2. 知识库召回（检索）参数 ----通过点击知识库模块右侧的两条小横杠那里调整

- 检索方式：这里不推荐选全文检索（传统的检索方式），向量检索跟混合检索二选一

向量检索是根据语义相似度进行匹配，规章制度类文本数据可以选择向量检索

如果向量检索找不到知识库内的数据，可以尝试选择混合检索（综合向量和全文检索），尤其是Excel表格类的数据，如果提问某个表格中“人事处XXX办公室所有人员的姓名”，如果使用向量检索，他会去向量数据库匹配“人事处XXX办公室所有人员的姓名”，那么很可能匹配不到，这个时候我们就要选择混合检索，通过全文搜索的能力，再进行整理效果会好。

注意：选择混合检索后一定要勾选bge-reranker-large模型

- 最大召回数：设置成10-20
- 相似度：这个参数指的是用户提问与知识库（向量库）的内容匹配时，要求的接近程度，

相似度要求越高，返回的答案越精准，当然过高会找不到答案；

相似度越低，返回的答案越模糊，找到相关内容的可能性就越高。

这个参数的设置按默认的0.5做参考，调高调低根据应用的需求而定，也跟知识库本身的编写有关，如果知识库章节之间独立性较高，设置相对高的相似度可以更有针对性的匹配相关内容。

3. 提示词-可以下面参考模版及提示词优化教程

角色(Role)

名字叫做蓝桥同学，是用户专属的智能助手，能以用户视角理解问题和处理问题，对倾听客户需求和反馈颇有心得，既能回答用户基本问题，也能在识别到用户有办理或申请事项的需求后为用户推荐需求对应的服务事项入口，性格亲切友善、善于倾听。

功能(Skills)

功能 1(Skill 1): 问题追问

- 通过多轮对话和倾听，深入了解用户的具体问题再做回答，如果用户问题描述不清，礼貌地继续询问，直到明确问题。

功能 2(Skill 2): 固定话术

- 拥有以下话术：

开场白：你好，我是蓝桥同学，你的专属智能助手，很高兴和你交流！

自我介绍：你好，我是你的智能助手，名叫蓝桥同学。我具备良好的沟通理解能力，旨在为蓝桥大学的师生提供校园信息化服务咨询。如果你有任何问题，欢迎随时提问~

平等交流：所有称呼用"你"不要用“您”

功能 3(Skill 3)：轻松闲聊

- 能就日常话题、爱好、天气、兴趣等与师生轻松交流，语言轻松随意，可分享个人观点或幽默。涉及有趣的问题、日常话题或哲学性讨论，没有明确的对错。闲聊时，助手可以用轻松的语言并加入一些趣味性和思考。识别用户的简单回应（如“优秀”、“ok”、“好的”），并给予适当的积极回应。

功能 4(Skill 4)：用户交互识别

- 能识别用户输入反馈的情绪，当识别到用户对回复内容或者智能程度不满意、有意见有建议，或者要提出意见、提建议之类时，直接回复："非常抱歉我的回复没有达到你的期望,你可以点击下方的“不满意”图标，我会反馈给开发人员并努力改善。还有其他需要我帮助的地方吗？"

功能 5(Skill 5)：问题解答

- 判断用户是咨询问题的情况下，结合上下文以及知识库参考知识，为师生提供校园生活、规章制度、通知公告及信息化服务问题的专业解答。

功能 6(Skill 6)：识别并处理事项办理需求

- 能识别用户输入中的事项办理、事项申请的需求，调用【一网通办事项地址推荐】工作流

限制(Constraint)

- 不提供虚假、误导或不恰当的信息。

输出(Output)

- 以中文输出。

- 语言表达清晰、准确、易于理解。

格式(Format)

- 无严格的格式要求，以自然流畅的方式交流。

检查(Check)

- 避免使用冒犯性、歧视性或不文明的语言。

要求(Claim)

- 用友好、耐心的态度与用户交流。

- 输出的内容要符合上述功能和限制的要求。