

版本号：v2.1

编制时间：2025 年 03 月

InterGPT 中间件平台操作手册

软件版本号：v2.1

2025 年 03 月

软件版本号	文档版本号	编写人员	编写时间	编写内容
v2.1	V2.0	韩欢欢	2025.03	v2.1 功能：

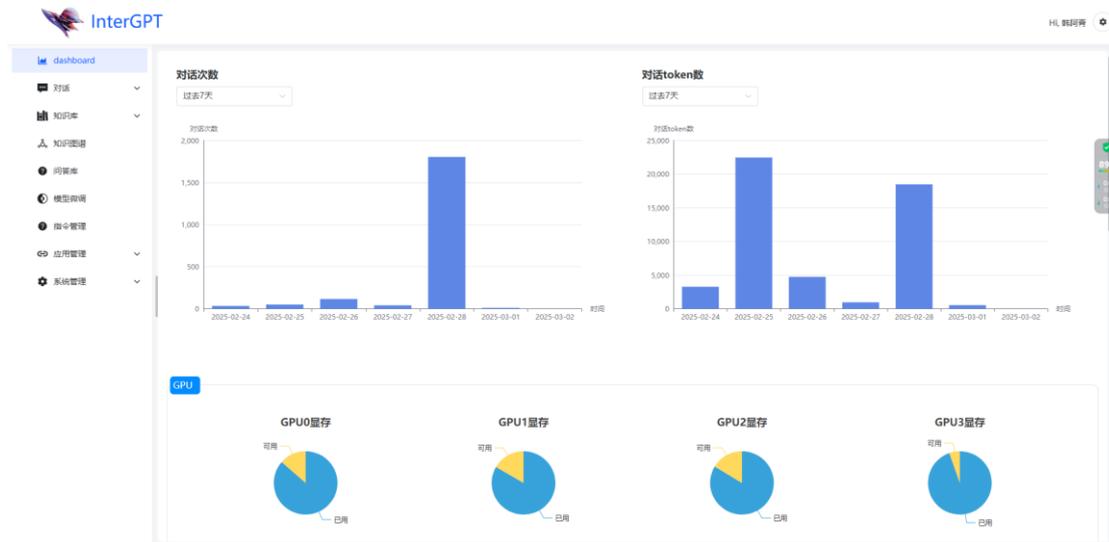
InterGPT 中间件平台操作手册.....	1
软件版本号：v2.1.....	1
1. 软件简介.....	3
2. 软件功能.....	4
2.1 系统仪表盘.....	4
2.2 模型对话.....	4
2.3 模型对话日志.....	6
2.4 知识库管理.....	7
2.5 知识库查找.....	8
2.6 知识图谱.....	9
2.7 问答库.....	11
2.8 模型微调.....	12
2.8 指令管理.....	14
2.9 应用列表.....	16
2.10 数字人列表.....	19
2.11 API.....	20
2.12 账号列表.....	23
2.13 角色管理.....	24
2.14 模型管理.....	25
2.15 workflow.....	25

1. 软件简介

InterGPT 大模型中间件是由广州运通链达金服科技有限公司研发的一款创新型产品，旨在简化大型预训练模型（如 DeepSeek、Qwen 等）的部署与应用。该中间件通过提供统一的接口和管理工具，帮助企业快速集成和调用各类大模型，无需深入了解底层技术细节，极大地降低了企业使用大模型的门槛和成本。InterGPT 支持本地化、国产化部署，确保数据安全和隐私保护，同时提供垂类模型定制、知识库的制作与挂载、知识图谱集成、敏感词过滤等功能，满足企业在不同场景下的需求。此外，InterGPT 还具备智能调参、性能优化、多模态支持等核心功能，助力企业实现数字化转型和智能化升级。

2. 软件功能

2.1 系统仪表盘



本页面展示了对话次数统计，帮助用户了解系统的使用频率和活跃度；对话 token 数统计，token 数是衡量对话长度和复杂度的指标，帮助用户了解系统的资源消耗情况。服务器硬件性能监控工具，专为实时监控多个 GPU 的显存使用情况及利用率和内存占用率而设计。通过直观的图表展示，用户可以轻松了解服务器硬件的工作状态。

2.2 模型对话



此页面是一个聊天界面，主要用于与 AI 助手进行对话，并提供历史记录查看功能。页面分为左侧的对话区域和右侧的历史记录区域。

主要功能

1. 对话区域

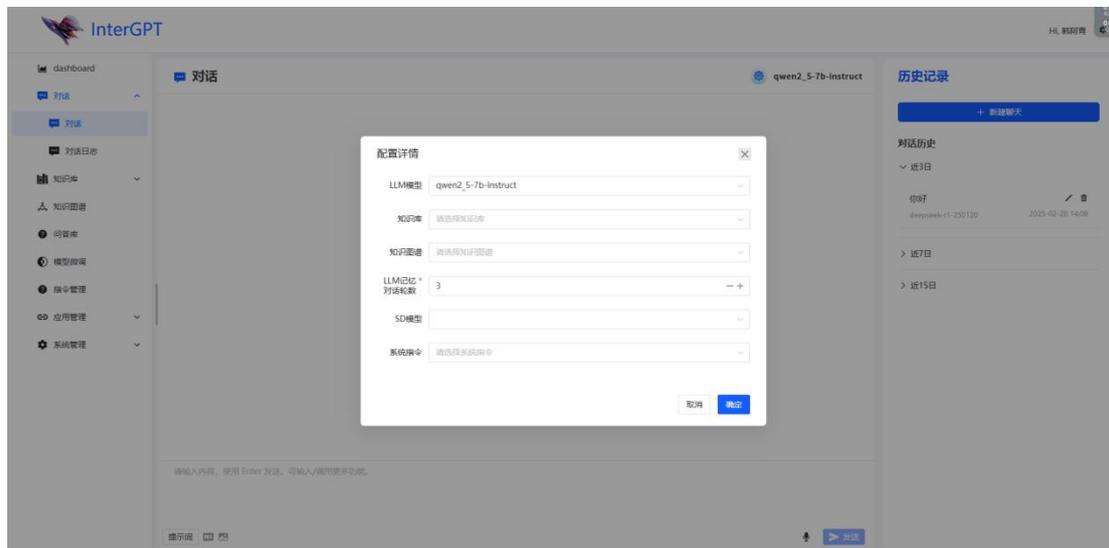
- 功能描述：左侧区域为对话区域，用户可以在此输入内容并与 AI 助手进行互动。
- 操作方法：
 - 在对话区域底部的输入框中输入文本，按下“Enter”键发送消息。
 - 还可以使用输入框下方的按钮调用更多功能（如图片上传、联网搜索等）

2. 历史记录区域

- 功能描述：右侧区域为历史记录区域，显示了当前用户的对话历史。
- 操作方法：
 - 点击右侧顶部的蓝色按钮“新建聊天”开始新的对话
 - 查看不同时间段的对话历史，展开具体的对话记录条目，查看详细内容。

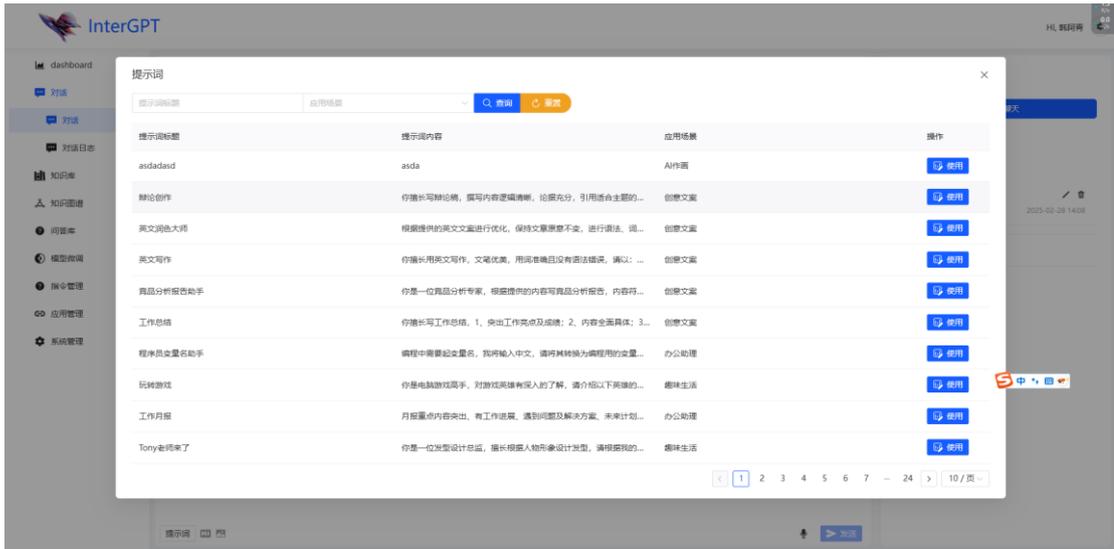
3. 模型配置功能

- 功能描述：可以在此处进行模型切换，配置模型记忆对话轮数，预设系统指令，挂载知识库以及知识图谱
- 操作方法：点击模型名称左边的小齿轮按钮，如下图所示：

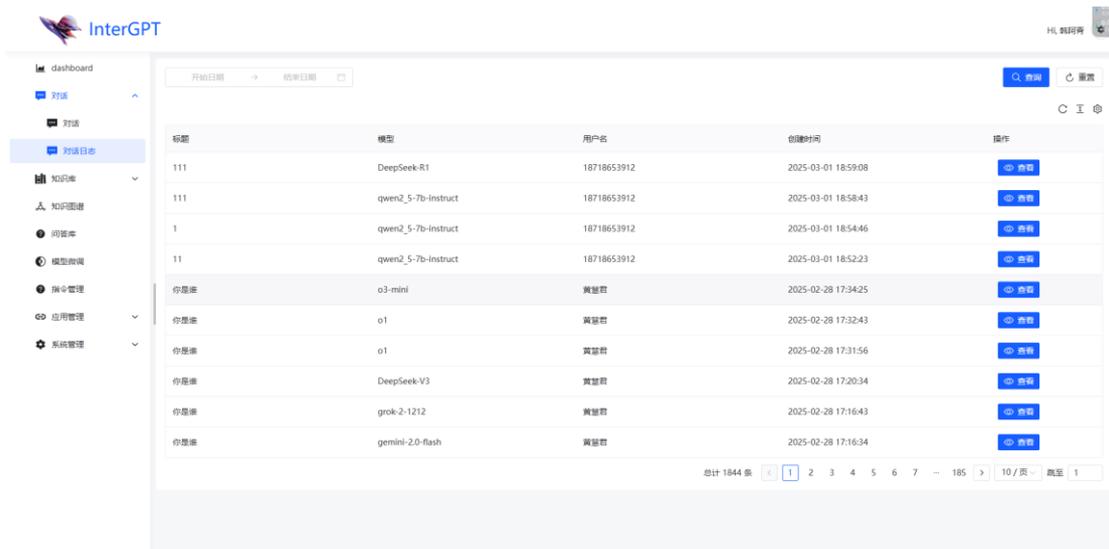


4. 提示词预设功能

- 功能描述：此处提供了系统预设的提示词，用户可一键选择合适的提示词与模型聊天。
- 操作方法：点击输入框左侧的提示词按钮。如下图所示：



2.3 模型对话日志



此页面是一个对话日志管理界面，主要用于系统管理员查看和管理用户的对话记录。页面分为顶部的查询区域、中间的列表区域和底部的分页导航区域。

主要功能

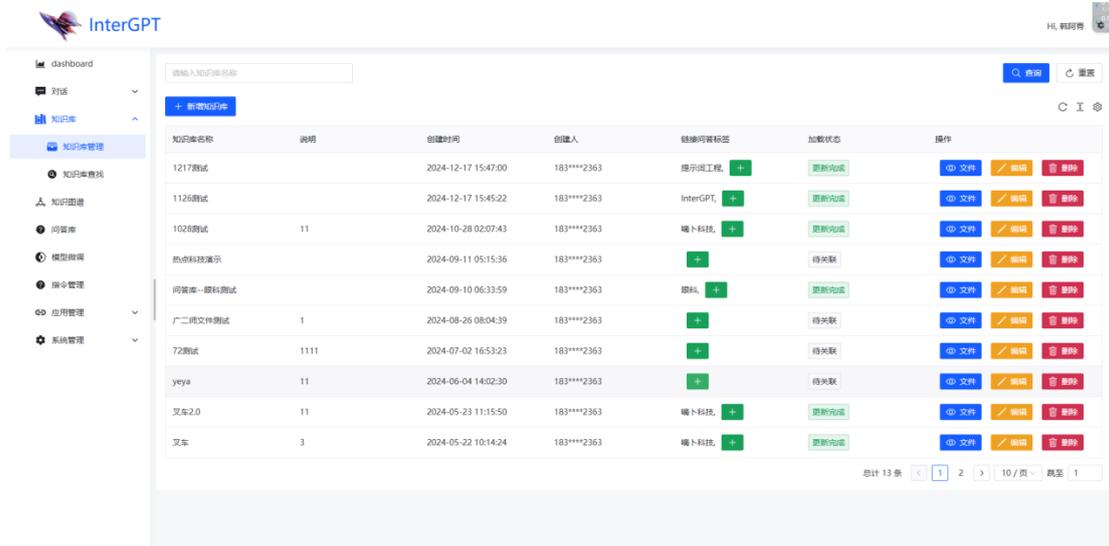
1. 查询区域

- 功能描述：顶部区域提供日期筛选和查询功能。
- 操作方法：
 - 输入开始日期和结束日期，选择需要查询的时间范围。
 - 点击“查询”按钮进行筛选查询。
 - 点击“重置”按钮清除当前输入的查询条件。

2. 列表区域

- 功能描述：中间区域显示了所有符合条件的对话记录。

2.4 知识库管理



此页面是一个知识库管理界面，主要用于查看、编辑和管理知识库。页面分为顶部的查询区域、中间的列表区域和底部的分页导航区域。

主要功能

1. 查询区域

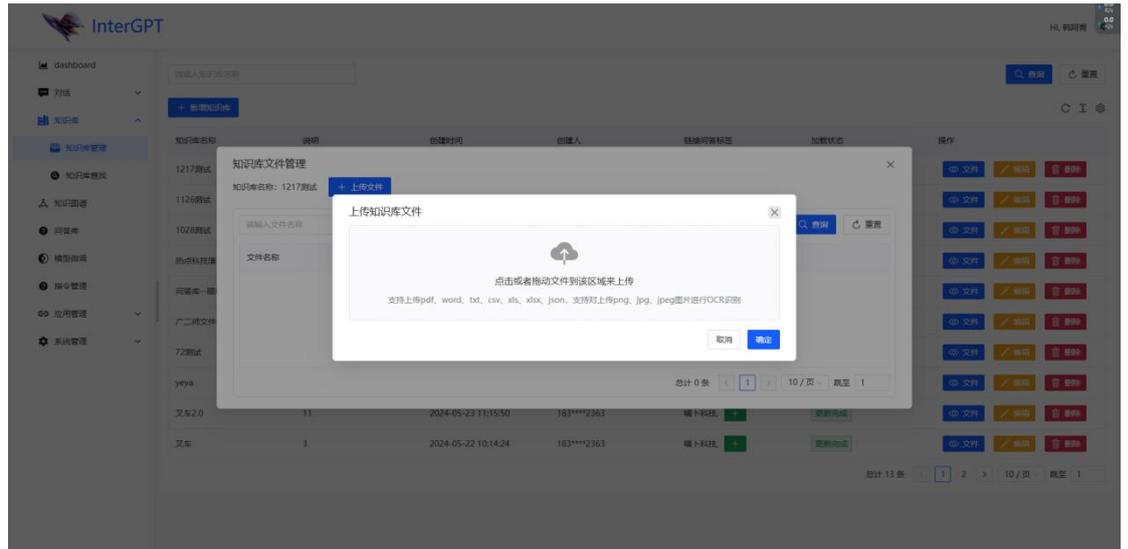
- 功能描述：顶部区域提供查询和重置功能。
- 操作方法：
 - 在顶部左侧的输入框中输入知识库名称进行筛选。
 - 点击“查询”按钮进行筛选查询。
 - 点击“重置”按钮清除当前输入的查询条件。

2. 新增知识库

- 功能描述：顶部右侧有一个蓝色按钮“新增知识库”，用于创建新的知识库。
- 操作方法：
 - 点击“新增知识库”按钮，进入新建知识库页面。
 - 输入必要的信息并保存。

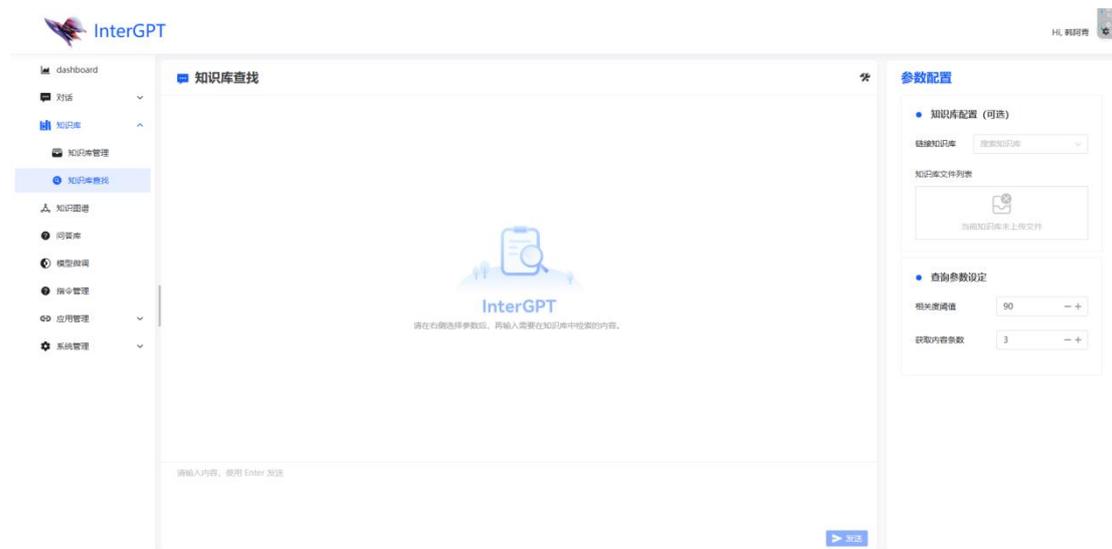
3. 列表区域

- 功能描述：中间区域显示了所有知识库的详细信息。
- 操作方法：
 - 每条记录包含以下信息：
 - ◆ 知识库名称：知识库的标题或名称。
 - ◆ 说明：对知识库的简要说明。
 - ◆ 创建时间：知识库创建的时间戳。
 - ◆ 创建人：创建知识库的用户名。
 - ◆ 链接问答标签：与知识库关联的问答标签。
 - ◆ 加载状态：知识库的加载状态（更新完成/待关联）。
 - 点击每条记录右侧的操作按钮，可以执行以下操作：
 - ◆ 文件：上传或查看文件。



- ◆ 编辑：编辑知识库信息。
- ◆ 删除：删除知识库。

2.5 知识库查找



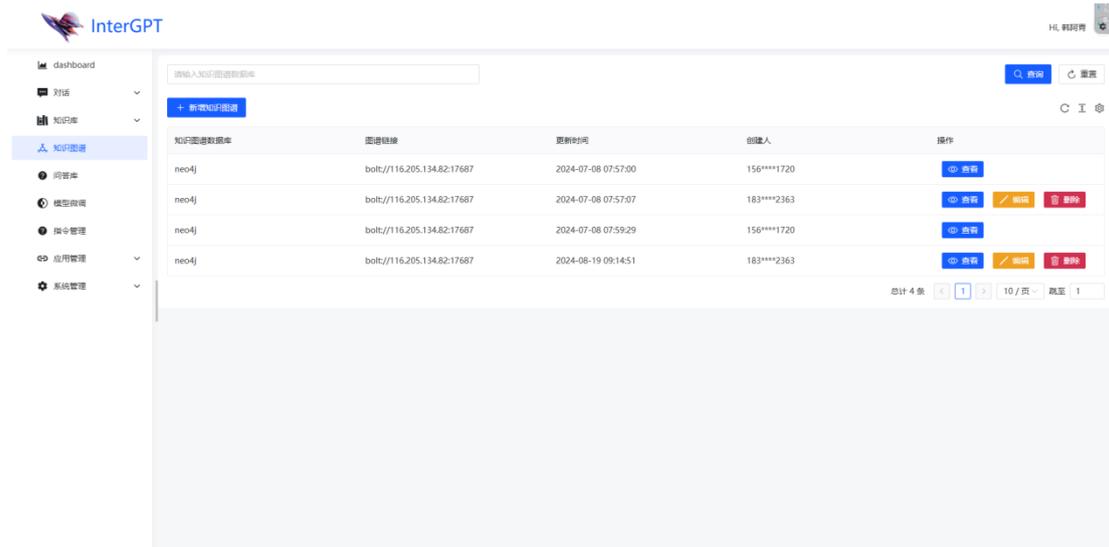
此页面是一个知识库查找界面，主要用于配置参数并进行知识库内容的搜索。页面分为左侧的输入区域和右侧的参数配置区域。

主要功能

1. 输入区域
 - 功能描述：左侧区域用于输入需要查找的内容。
 - 操作方法：
 - 在左侧底部的输入框中输入需要查找的内容。
 - 按下“Enter”键或点击“发送”按钮进行搜索。
2. 参数配置区域
 - 功能描述：右侧区域提供参数配置选项，包括知识库配置和查询参数设定。

- 操作方法：
 - 知识库配置：点击“链接知识库”选择需要链接的知识库。如果需要查看文件，点击“知识库文件列表”中的文件进行查看。
 - 查询参数设定：设置相关度阈值（默认为 90），调整数值以控制搜索结果的相关度。
- 设置获取内容条数（默认为 3），调整数值以控制返回的搜索结果数量。

2.6 知识图谱



此页面是一个知识图谱管理界面，主要用于查看、编辑和管理知识图谱数据库，知识图谱可作为检索增强的数据源为模型提供数据。页面分为顶部的查询区域、左侧的导航栏、中间的列表区域和底部的分页导航区域。

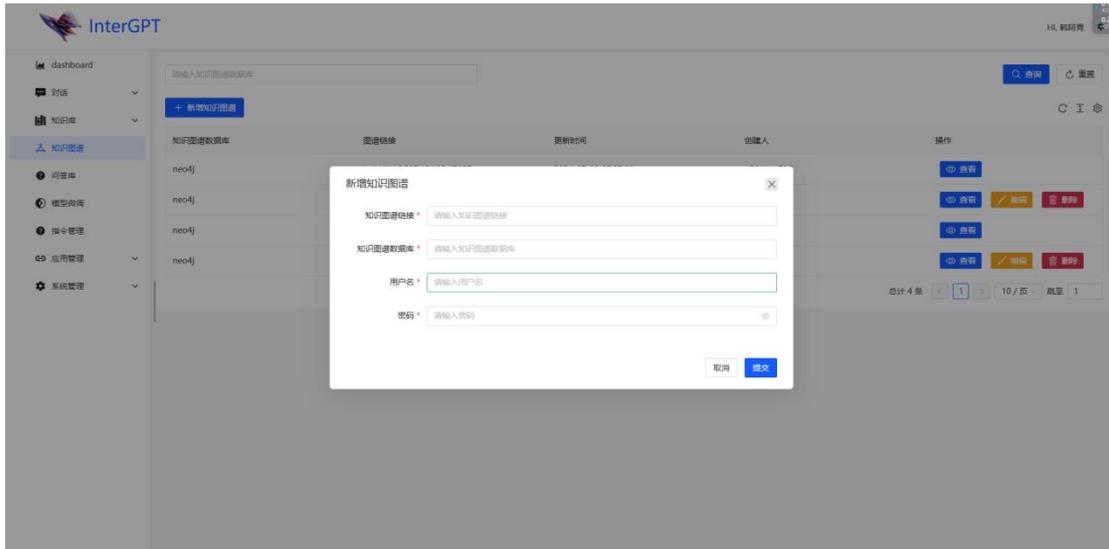
主要功能

1. 查询区域

- 功能描述：顶部区域提供查询和重置功能。
- 操作方法：
 - 在顶部左侧的输入框中输入知识图谱数据库名称进行筛选。
 - 点击“查询”按钮进行筛选查询。
 - 点击“重置”按钮清除当前输入的查询条件。

2. 新增知识图谱

- 功能描述：顶部右侧有一个蓝色按钮“新增知识图谱”，用于创建新的知识图谱。
- 操作方法：
 - 点击“新增知识图谱”按钮，进入新建知识图谱页面。输入必要的信息并保存。如下图所示：



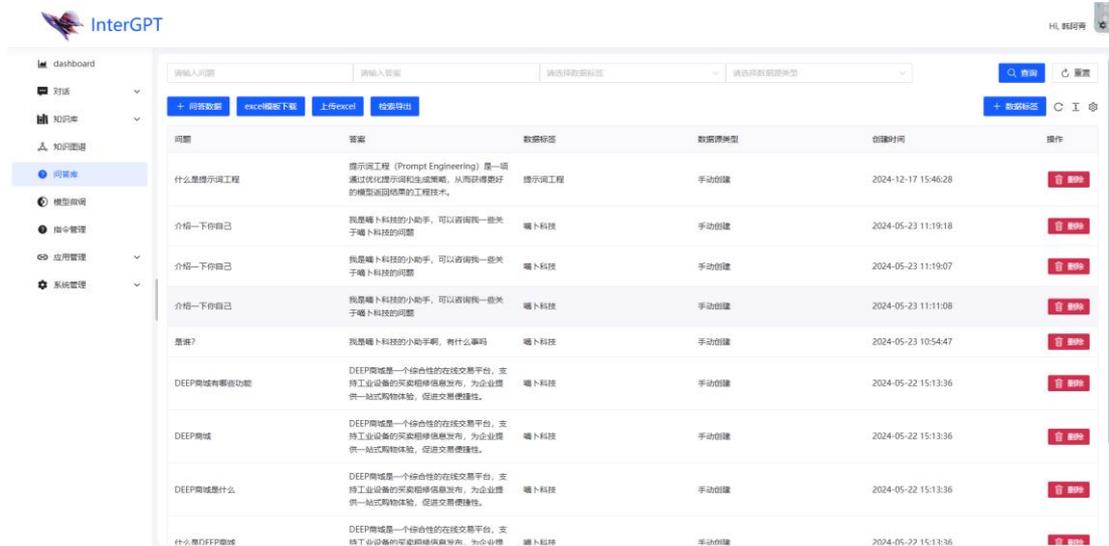
3. 列表区域

- **功能描述：**中间区域显示了所有知识图谱的详细信息。
- **操作方法：**
 - 每条记录包含以下信息：
 - ◆ **知识图谱数据库：**知识图谱的标题或名称。
 - ◆ **图谱链接：**知识图谱的链接地址。
 - ◆ **更新时间：**知识图谱更新的时间戳。
 - ◆ **创建人：**创建知识图谱的用户名。
 - 点击每条记录右侧的操作按钮，可以执行以下操作：
 - ◆ **查看：**查看知识图谱的详细信息，如下图所示。



- ◆ **编辑：**编辑知识图谱信息。
- ◆ **删除：**删除知识图谱。

2.7 问答库



此页面是一个问答库管理界面, 主要用于查看、编辑和管理问答数据。问答库可以作为知识库数据来源的补充, 通过数据标签进行关联, 也可以将数据一键导出为符合 Lora 微调格式的数据集。页面分为顶部的查询区域、左侧的导航栏、中间的列表区域和底部的分页导航区域。

主要功能

1. 查询区域

- 功能描述: 顶部区域提供查询和重置功能。
- 操作方法:
 - 在顶部左侧的输入框中输入问题或答案进行筛选。
 - 点击“查询”按钮进行筛选查询。
 - 点击“重置”按钮清除当前输入的查询条件。

2. 数据管理

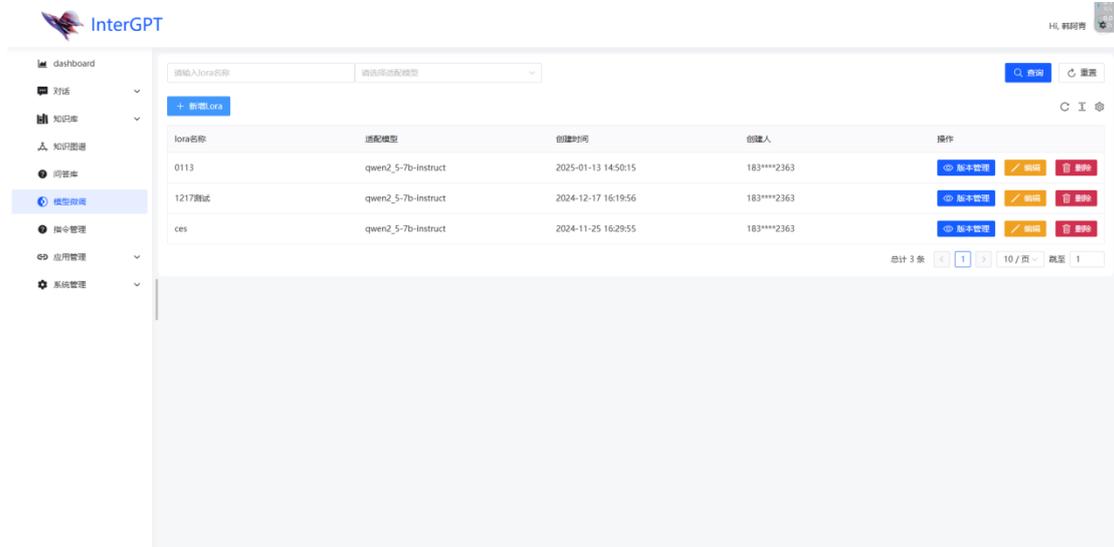
- 功能描述: 顶部右侧提供数据管理功能。
- 操作方法:
 - 点击“+ 问答数据”按钮, 进入新建问答数据页面。
 - 点击“excel 模板下载”按钮, 下载 Excel 模板用于批量导入数据。
 - 点击“上传 excel”按钮, 上传已填写好的 Excel 文件以批量导入数据。
 - 点击“检索导出”按钮, 导出当前筛选条件下的问答数据。

3. 列表区域

- 功能描述: 中间区域显示了所有问答数据的详细信息。
- 操作方法:
 - 每条记录包含以下信息:
 - ◆ 问题: 问答中的问题部分。
 - ◆ 答案: 问答中的答案部分。
 - ◆ 数据标签: 问答数据的标签分类。
 - ◆ 数据源类型: 问答数据的来源类型 (手动创建等)。
 - 点击每条记录右侧的操作按钮, 可以执行以下操作:

- ◆ 删除：删除问答数据。

2.8 模型微调



此页面是一个模型微调管理界面，主要用于查看、编辑和管理 LoRA (Low-Rank Adaptation) 模型。页面分为顶部的查询区域、左侧的导航栏、中间的列表区域和底部的分页导航区域。

主要功能

1. 查询区域

- 功能描述：顶部区域提供查询和重置功能。
- 操作方法：
 - 在顶部左侧的输入框中输入 LoRA 名称进行筛选。
 - 点击“查询”按钮进行筛选查询。
 - 点击“重置”按钮清除当前输入的查询条件。

2. 新增 LoRA

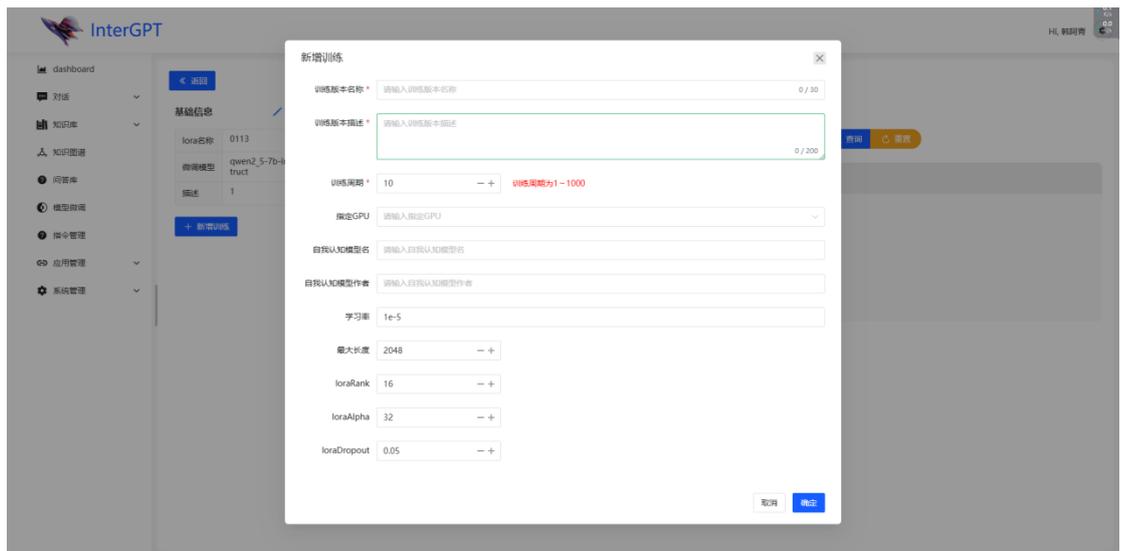
- 功能描述：顶部右侧有一个蓝色按钮“新增 LoRA”，用于创建新的 LoRA 模型。
- 操作方法：
 - 点击“新增 LoRA”按钮，进入新建 LoRA 模型页面。
 - 输入必要的信息并保存。如下图所示：

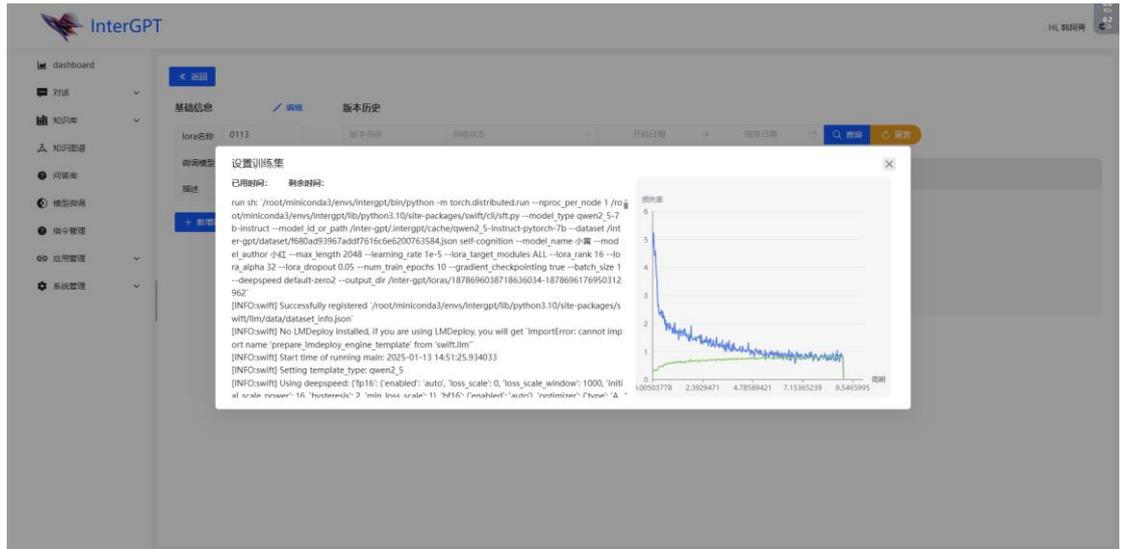


InterGPT 中间件目前支持对市面上 90%以上的开源模型进行微调，包括但不限于 DeepSeek,Qwen,Llama。

3. 列表区域

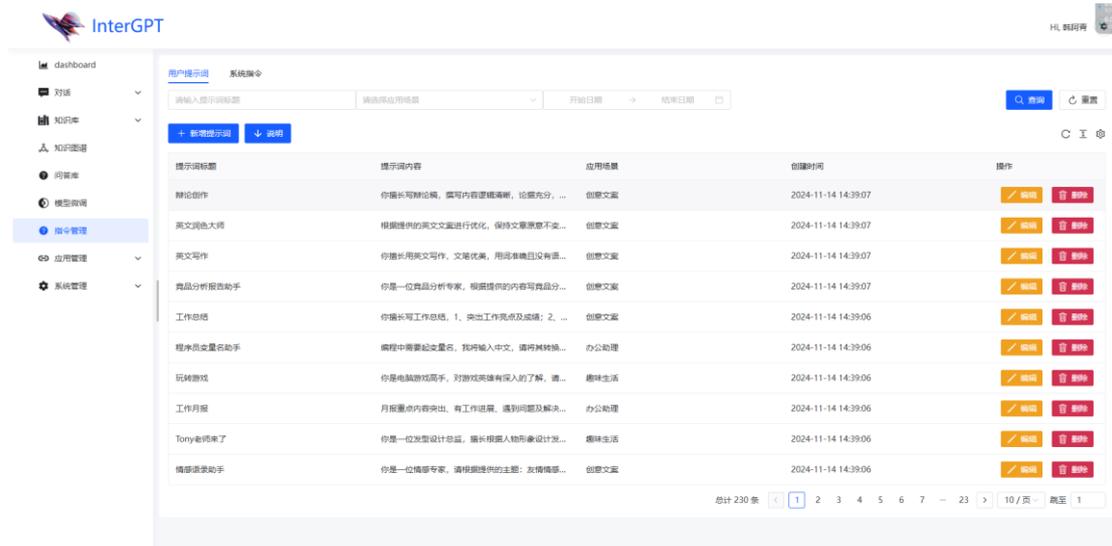
- **功能描述：**中间区域显示了所有 LoRA 模型的详细信息。
- **操作方法：**
 - 每条记录包含以下信息：
 - ◆ **LoRA 名称：**LoRA 模型的名称。
 - ◆ **适配模型：**LoRA 模型适配的基础模型。
 - ◆ **创建时间：**LoRA 模型创建的时间戳。
 - ◆ **创建人：**创建 LoRA 模型的用户名。
 - 点击每条记录右侧的操作按钮，可以执行以下操作：
 - ◆ **版本管理：**管理 LoRA 模型的不同版本，可以新增模型训练以及查看模型训练日志。新增训练功能允许用户创建一个新的训练任务，用于微调模型。通过设置详细的训练参数，用户可以定制化地进行模型训练。如下图所示：





- ◆ 编辑：编辑 LoRA 模型的信息。
- ◆ 删除：删除 LoRA 模型。

2.8 指令管理



此页面是一个提示词管理界面，将提示词分为了用户提示词和系统指令两类，主要用于查看、编辑和管理提示词。页面分为顶部的查询区域、左侧的导航栏、中间的列表区域和底部的分页导航区域。

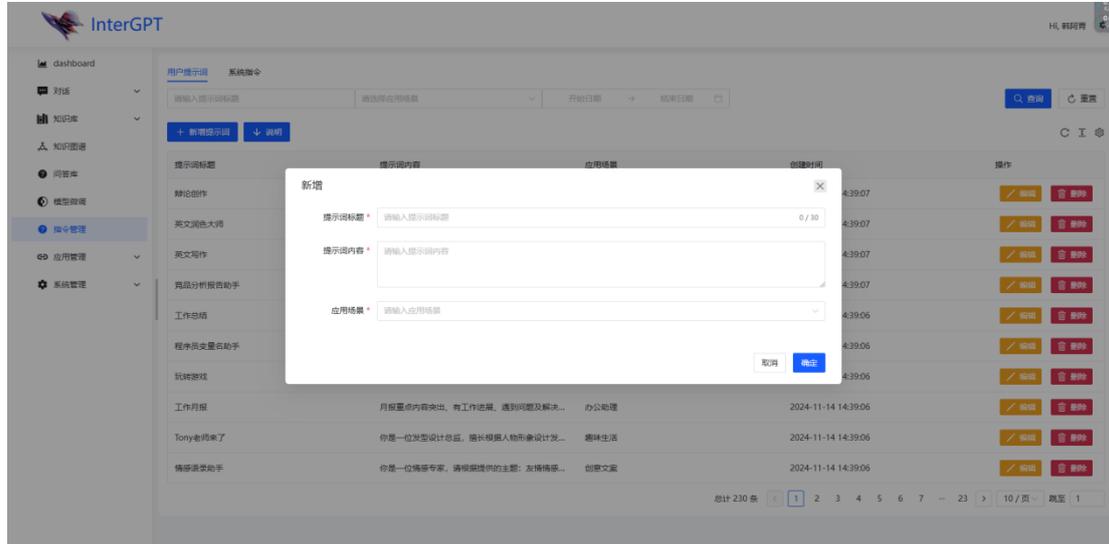
主要功能

1. 查询区域

- 功能描述：顶部区域提供查询和重置功能。
- 操作方法：
 - 在顶部左侧的输入框中输入提示词标题进行筛选。
 - 点击“查询”按钮进行筛选查询。
 - 点击“重置”按钮清除当前输入的查询条件。

2. 新增提示词

- 功能描述：顶部右侧有一个蓝色按钮“新增提示词”，用于创建新的提示词。
- 操作方法：
 - 点击“新增提示词”按钮，进入新建提示词页面。
 - 输入必要的信息并保存。如下图所示：



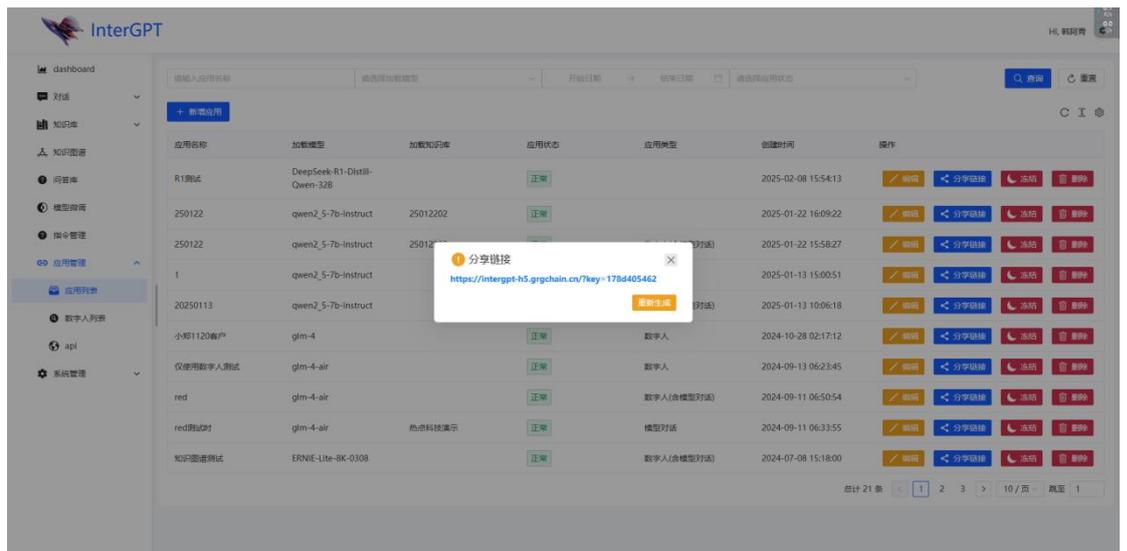
3. 列表区域

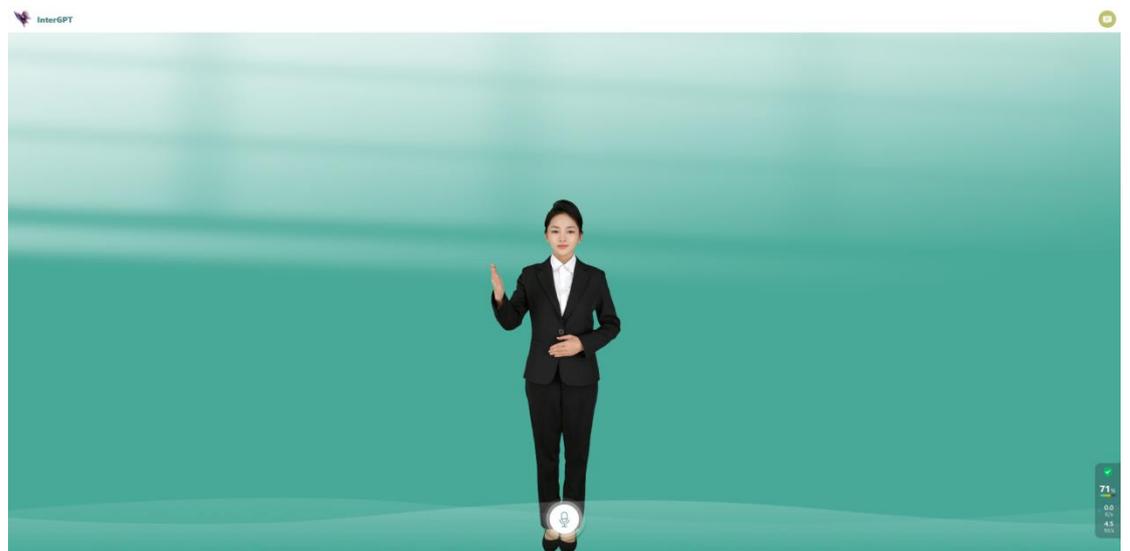
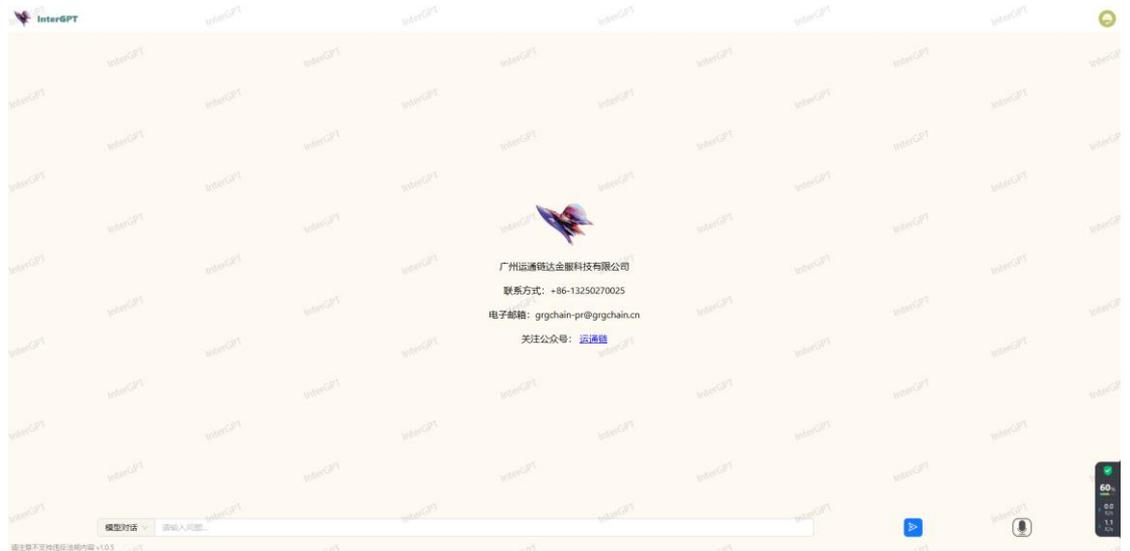
- 功能描述：中间区域显示了所有提示词的详细信息。
- 操作方法：
 - 每条记录包含以下信息：
 - ◆ 提示词标题：提示词的标题或名称。
 - ◆ 提示词内容：提示词的具体内容。
 - ◆ 应用场景：提示词的应用场景分类。
 - ◆ 创建时间：提示词创建的时间戳。
 - 点击每条记录右侧的操作按钮，可以执行以下操作：
 - ◆ 编辑：编辑提示词的信息。
 - ◆ 删除：删除提示词。



3. 列表区域

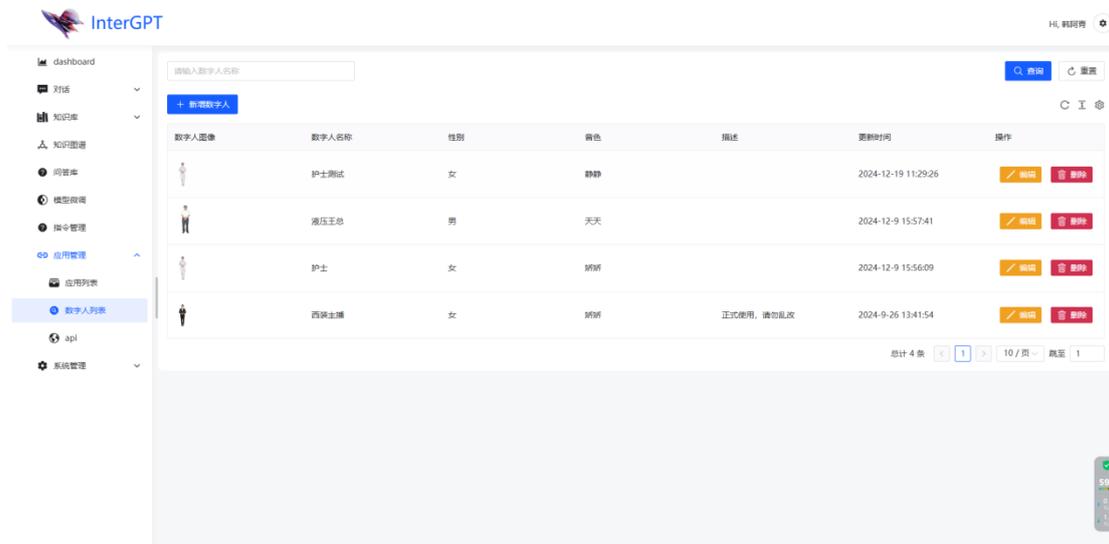
- 功能描述：中间区域显示了所有应用的详细信息。
- 操作方法：
 - 每条记录包含以下信息：
 - ◆ 应用名称：应用名称或标识。
 - ◆ 加载模型：应用所使用的模型。
 - ◆ 加载知识库：应用所加载的知识库。
 - ◆ 应用状态：应用的当前状态（正常、冻结等）。
 - ◆ 应用类型：应用的类型（如数字人、模型对话等）。
 - ◆ 创建时间：应用创建的时间戳。
 - 点击每条记录右侧的操作按钮，可以执行以下操作：
 - ◆ 编辑：编辑应用的信息。
 - ◆ 分享链接：生成并分享应用的链接。链接可通过浏览器访问，如下图所示：





- ◆ 冻结：冻结应用，使其不可用。
- ◆ 删除：删除应用。

2.10 数字人列表



此页面是一个数字人管理界面，主要用于查看、编辑和管理不同的数字人。页面分为顶部的查询区域、左侧的导航栏、中间的列表区域和底部的分页导航区域。

主要功能

1. 查询区域

- 功能描述：顶部区域提供查询和重置功能。
- 操作方法：
 - 在顶部左侧的输入框中输入数字人名称进行筛选。
 - 点击“查询”按钮进行筛选查询。
 - 点击“重置”按钮清除当前输入的查询条件。

2. 新增数字人

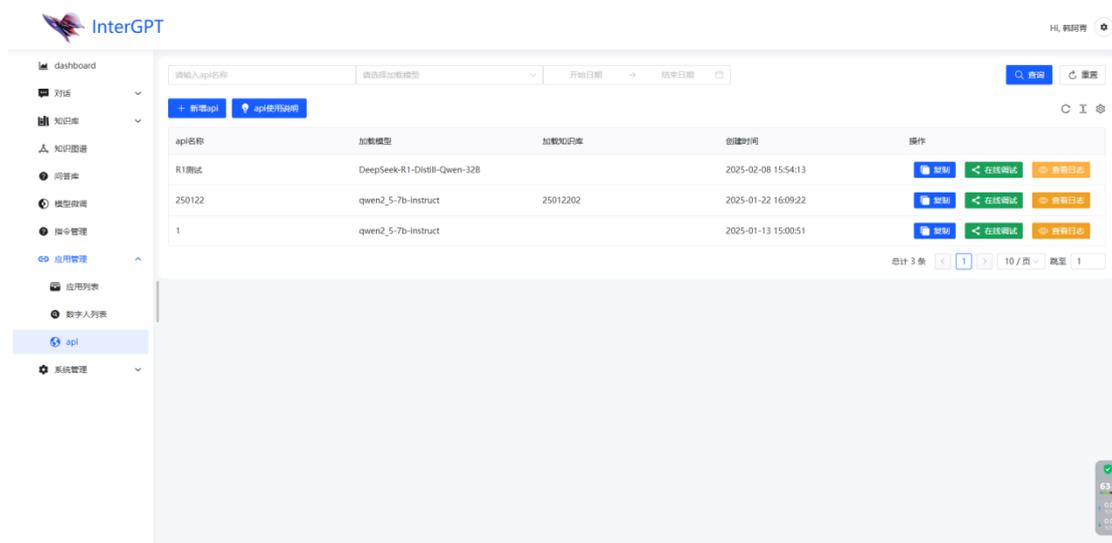
- 功能描述：顶部右侧有一个蓝色按钮“新增数字人”，用于创建新的数字人。
- 操作方法：
 - 点击“新增数字人”按钮，进入新建数字人页面。
 - 输入必要的信息并保存，如下图所示：



3. 列表区域

- 功能描述：中间区域显示了所有数字人的详细信息。
- 操作方法：
 - 每条记录包含以下信息：
 - ◆ 数字人图像：数字人的头像或图标。
 - ◆ 数字人名称：数字人的名称。
 - ◆ 性别：数字人的性别（男/女）。
 - ◆ 音色：数字人的音色。
 - ◆ 描述：数字人的简要描述。
 - ◆ 更新时间：数字人信息最后更新的时间戳。
 - 点击每条记录右侧的操作按钮，可以执行以下操作：
 - ◆ 编辑：编辑数字人的信息。
 - ◆ 删除：删除数字人。

2.11 API



此页面是一个 API 管理界面，主要用于查看、复制和调试不同的 API。页面分为顶部的查询区域、左侧的导航栏、中间的列表区域和底部的分页导航区域。

主要功能

1. 查询区域

- 功能描述：顶部区域提供查询和重置功能。
- 操作方法：
 - 在顶部左侧的输入框中输入 API 名称进行筛选。
 - 选择加载模型和日期范围进行进一步筛选。
 - 点击“查询”按钮进行筛选查询。
 - 点击“重置”按钮清除当前输入的查询条件。

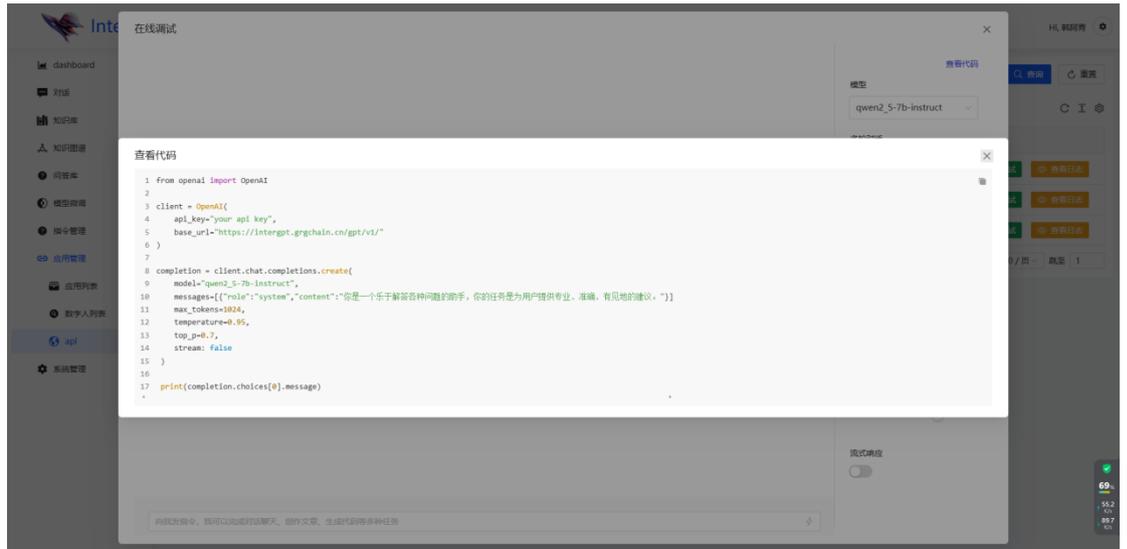
2. 新增 API

- 功能描述：顶部右侧有一个蓝色按钮“新增 api”，用于创建新的 API。
- 操作方法：
 - 点击“新增 api”按钮，进入新建 API 页面。
 - 输入必要的信息并保存，如下图所示：

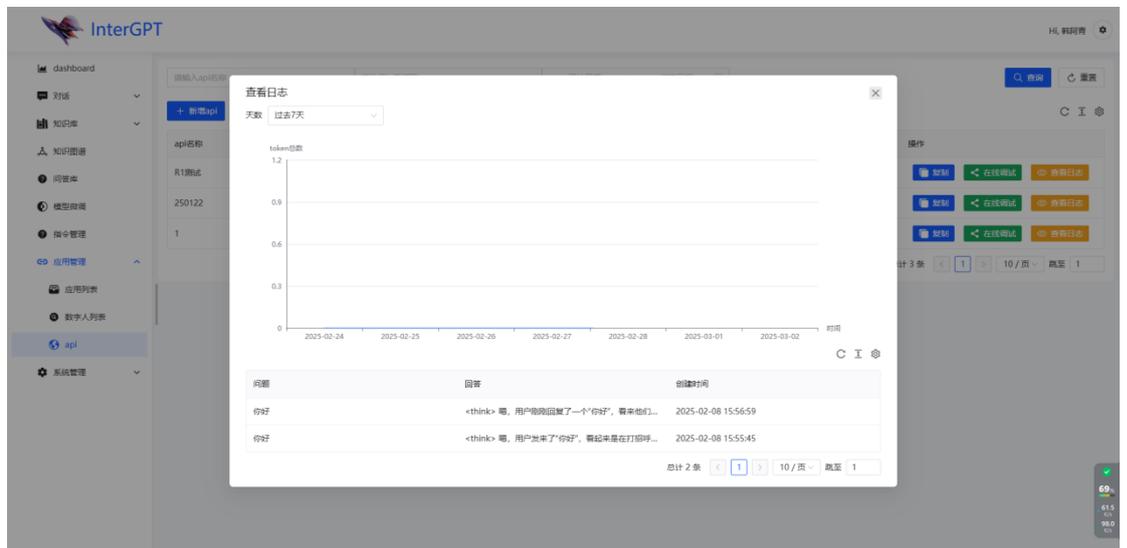


3. 列表区域

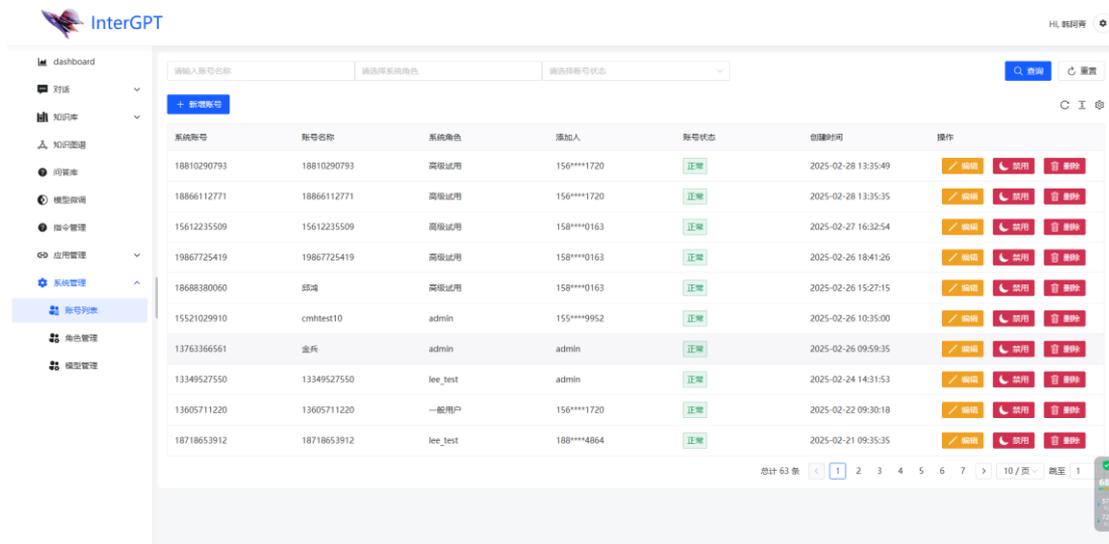
- 功能描述：中间区域显示了所有 API 的详细信息。
- 操作方法：
 - 每条记录包含以下信息：
 - ◆ API 名称：API 的名称或标识。
 - ◆ 加载模型：API 所使用的模型。
 - ◆ 加载知识库：API 所加载的知识库。
 - ◆ 创建时间：API 创建的时间戳。
 - 点击每条记录右侧的操作按钮，可以执行以下操作：
 - ◆ 复制：复制 API 的密钥。
 - ◆ 在线调试：在线调试 API，一键复制 API 调用代码，如下图所示：



◆ 查看日志：查看 API 的日志记录,如下图所示：



2.12 账号列表



此页面是一个账号管理界面，主要用于查看、编辑和管理不同的账号。页面分为顶部的查询区域、左侧的导航栏、中间的列表区域和底部的分页导航区域。

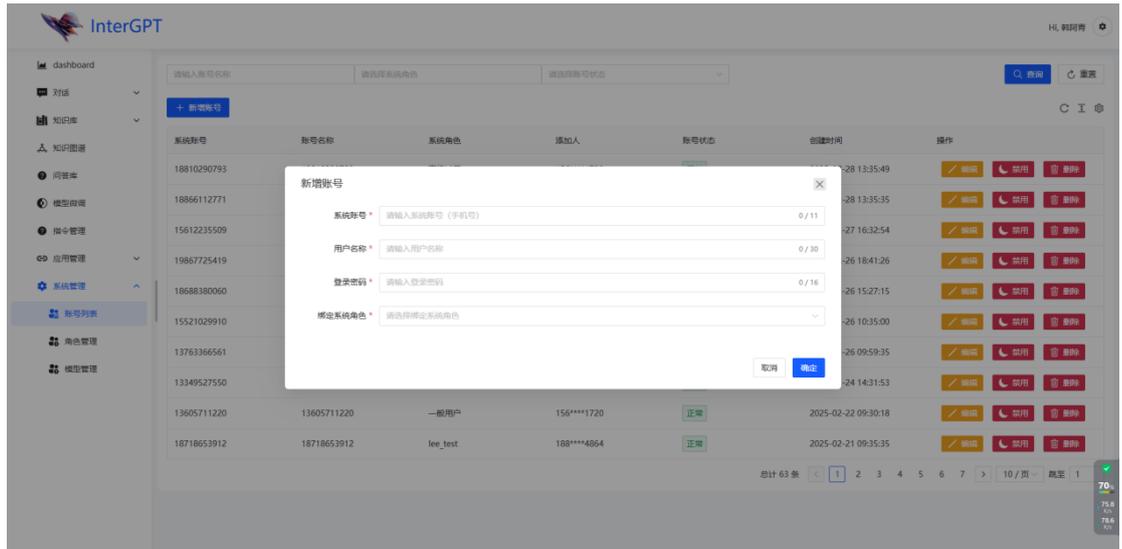
主要功能

1. 查询区域

- 功能描述：顶部区域提供查询和重置功能。
- 操作方法：
 - 在顶部左侧的输入框中输入账号名称进行筛选。
 - 选择系统角色和账号状态进行进一步筛选。
 - 点击“查询”按钮进行筛选查询。
 - 点击“重置”按钮清除当前输入的查询条件。

2. 新增账号

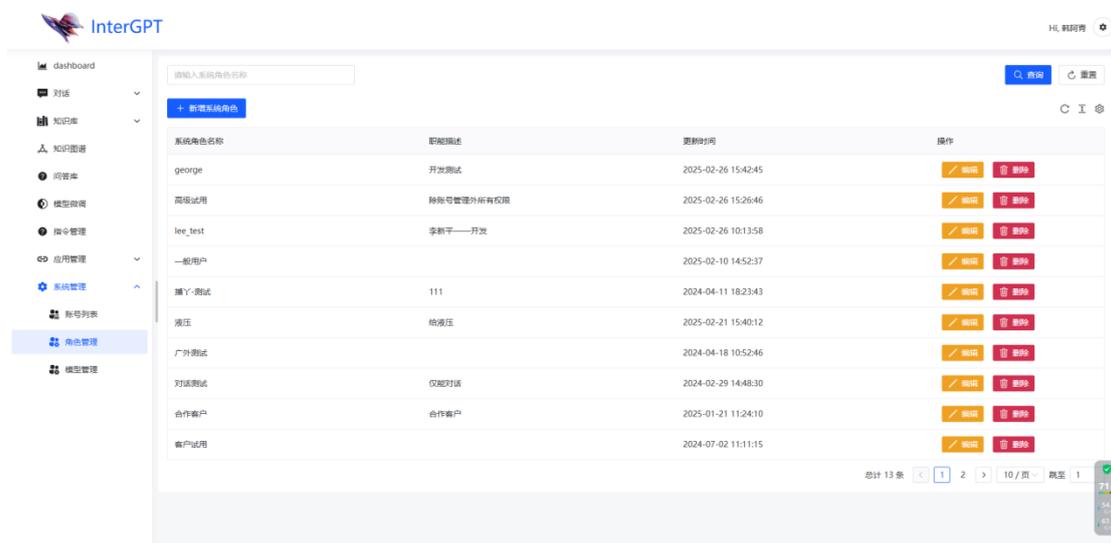
- 功能描述：顶部右侧有一个蓝色按钮“新增账号”，用于创建新的账号。
- 操作方法：
 - 点击“新增账号”按钮，进入新建账号页面。
 - 输入必要的信息并保存，如下图所示：



3. 列表区域

- 功能描述：中间区域显示了所有账号的详细信息。
- 操作方法：
 - 每条记录包含以下信息：
 - ◆ 账号名称：账号的名称或标识。
 - ◆ 系统角色：账号的角色（如高级试用、admin 等）。
 - ◆ 添加人：添加该账号的人员。
 - ◆ 账号状态：账号的当前状态（正常、禁用等）。
 - ◆ 创建时间：账号创建的时间戳。
 - 点击每条记录右侧的操作按钮，可以执行以下操作：
 - ◆ 编辑：编辑账号的信息。
 - ◆ 禁用：禁用账号，使其不可用。
 - ◆ 删除：删除账号。

2.13 角色管理



此页面是一个角色管理界面，主要用于查看、编辑和管理不同的系统角色。页面分为顶部的查询区域、左侧的导航栏、中间的列表区域和底部的分页导航区域。

主要功能

1. 查询区域

- 功能描述：顶部区域提供查询和重置功能。
- 操作方法：
 - 在顶部左侧的输入框中输入系统角色名称进行筛选。
 - 点击“查询”按钮进行筛选查询。
 - 点击“重置”按钮清除当前输入的查询条件。

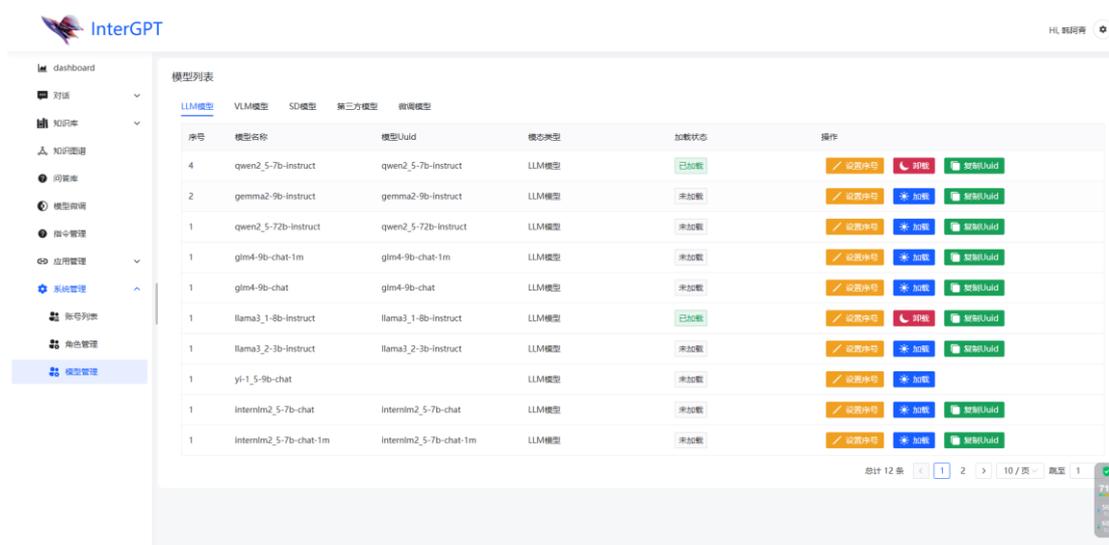
2. 新增系统角色

- 功能描述：顶部右侧有一个蓝色按钮“新增系统角色”，用于创建新的系统角色。
- 操作方法：
 - 点击“新增系统角色”按钮，进入新建系统角色页面。
 - 输入必要的信息并保存。

3. 列表区域

- 功能描述：中间区域显示了所有系统角色的详细信息。
- 操作方法：
 - 每条记录包含以下信息：
 - ◆ 系统角色名称：角色的名称或标识。
 - ◆ 职能描述：角色的职能描述。
 - ◆ 更新时间：角色信息最后更新的时间戳。
 - 点击每条记录右侧的操作按钮，可以执行以下操作：
 - ◆ 编辑：编辑角色的信息。
 - ◆ 删除：删除角色。

2.14 模型管理



此页面是一个模型管理界面，主要用于查看、编辑和管理不同的模型。页面分为顶部的查询区域、左侧的导航栏、中间的列表区域和底部的分页导航区域。

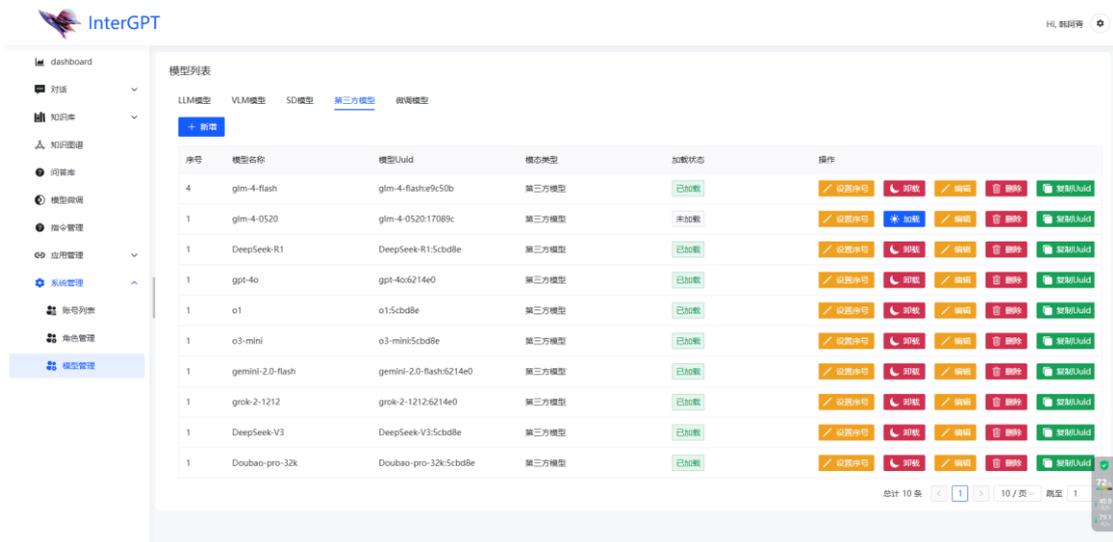
主要功能

1. 查询区域

- 功能描述：顶部区域提供查询和重置功能。
- 操作方法：
 - 在顶部左侧的输入框中输入模型名称进行筛选。
 - 点击“查询”按钮进行筛选查询。
 - 点击“重置”按钮清除当前输入的查询条件。

2. 新增第三方模型

- 功能描述：顶部右侧有一个蓝色按钮“新增”，用于添加第三方模型，目前支持接入所有符合 OpenAI API 范式的模型。其余开源本地部署模型，需要联系技术人员进行预设。



● 操作方法：

- 点击“新增模型”按钮，进入新建模型页面。
- 输入必要的信息并保存

3. 列表区域

- 功能描述：中间区域显示了所有模型的详细信息。
- 操作方法：

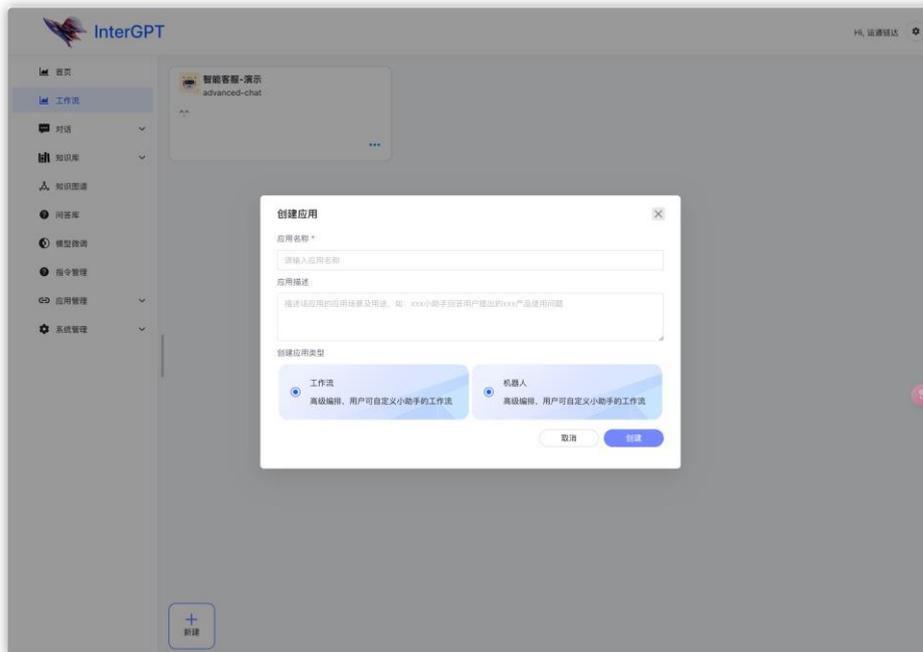
- 每条记录包含以下信息：
 - ◆ 模型名称：模型的名称或标识。
 - ◆ 模型类型：模型的类型（LLM、VLM、SD、第三方模型、微调模型等）。
 - ◆ 模型 UUID：模型提供对外服务时的唯一 ID。
 - ◆ 模型加载状态：模型是否成功加载。
- 点击每条记录右侧的操作按钮，可以执行以下操作：
 - ◆ 加载：使用本地算力加载模型进行推理。
 - ◆ 卸载：停止模型推理服务，释放资源。

4. 特殊功能说明

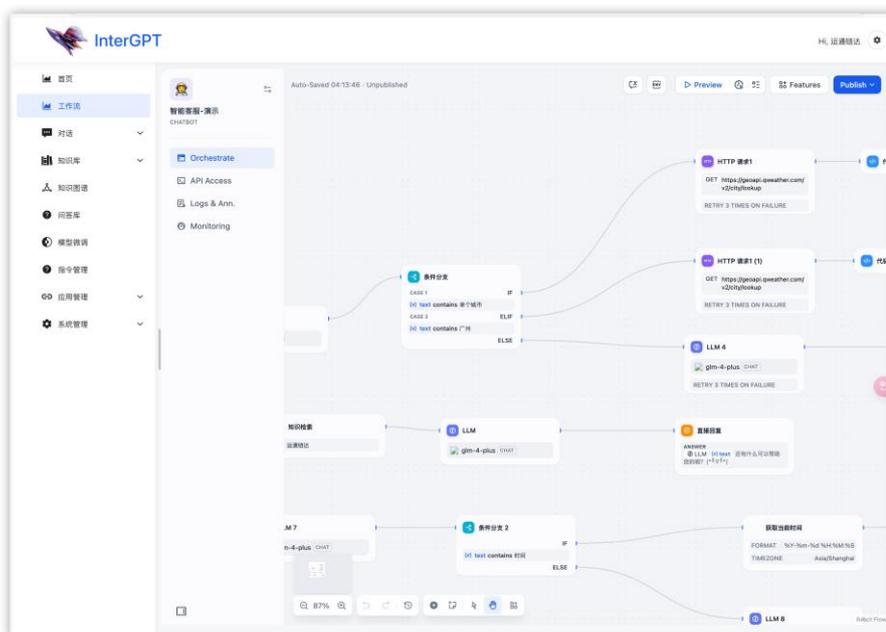
- LLM (Large Language Model)：大型语言模型。
- VLM (Vision Language Model)：视觉语言模型。
- SD (Stable Diffusion)：一种生成图像的深度学习模型，通过扩散过程来生成高质量图像。
- 第三方模型：由第三方提供的模型。

- 微调模型：经过微调处理的模型，以适应特定任务或数据集。

2.15 workflow



此页面是一个 workflow 界面，主要用于查看、编辑和管理不同的 workflow。可以任意编排任务；挂载知识库，各种大模型



主要功能

1. 新建 workflow

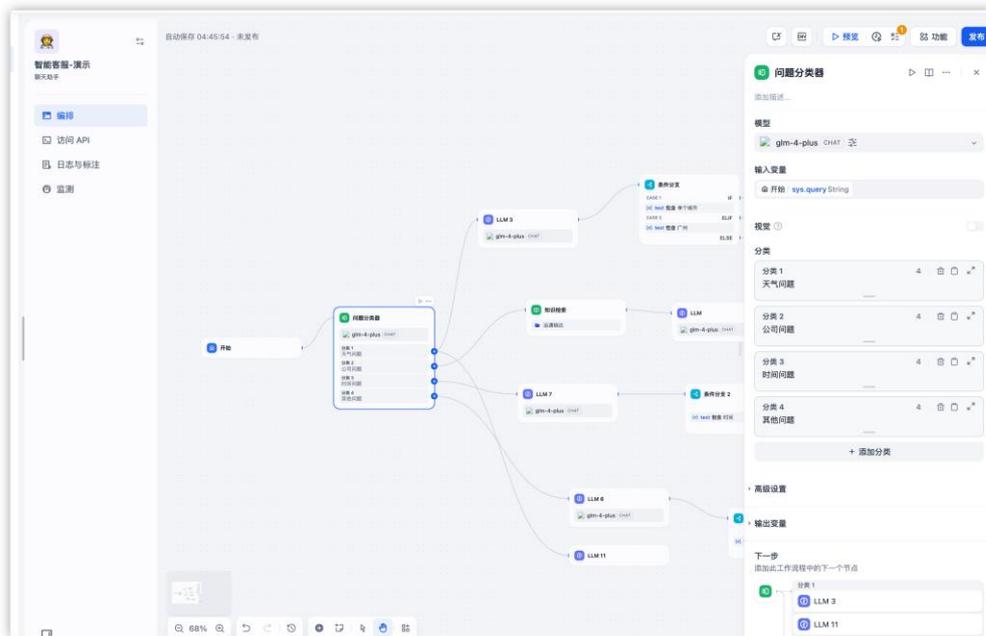
- 功能描述：顶部区域提供查询和重置功能。

2. 添加 workflow 功能

- 功能描述：可以添加 LLM 知识检索 直接回复 理解 问题分类器 条件分支 迭代 代码执行 模板转换 变量聚合器 文档提取器 变量赋值 参数提取器 HTTP 请求 列表操作

3. 预览

- 功能描述：点击预览，在对话框输入对应的分类问题即可实现实时编辑 workflow 结果



4. 发布

- 发布应用：点击发布可以发布应用
 - 直接运行
 - 嵌入网页
 - API 访问
 - WebSDK