



**基于RAGFlow的推理问答功能使用指南**

**编者：3023244322蒋茜，3023244328马佳一**

**资料提供：3023244327邵玺冉，3023244338张婉毓**

**（姓名不分先后，按首字母排序）**

**2025年 6 月 30 日**

目 录

[Ollama安装](#_Toc29496_WPSOffice_Level1) [1](#_Toc29496_WPSOffice_Level1)

[Docker Desktop安装 2](#_Toc5928_WPSOffice_Level1)

[RAGFlow安装 3](#_Toc5394_WPSOffice_Level1)

[模型选择与配置](#_Toc20631_WPSOffice_Level1) 4

[知识图谱构建与生成](#_Toc26352_WPSOffice_Level1) 6

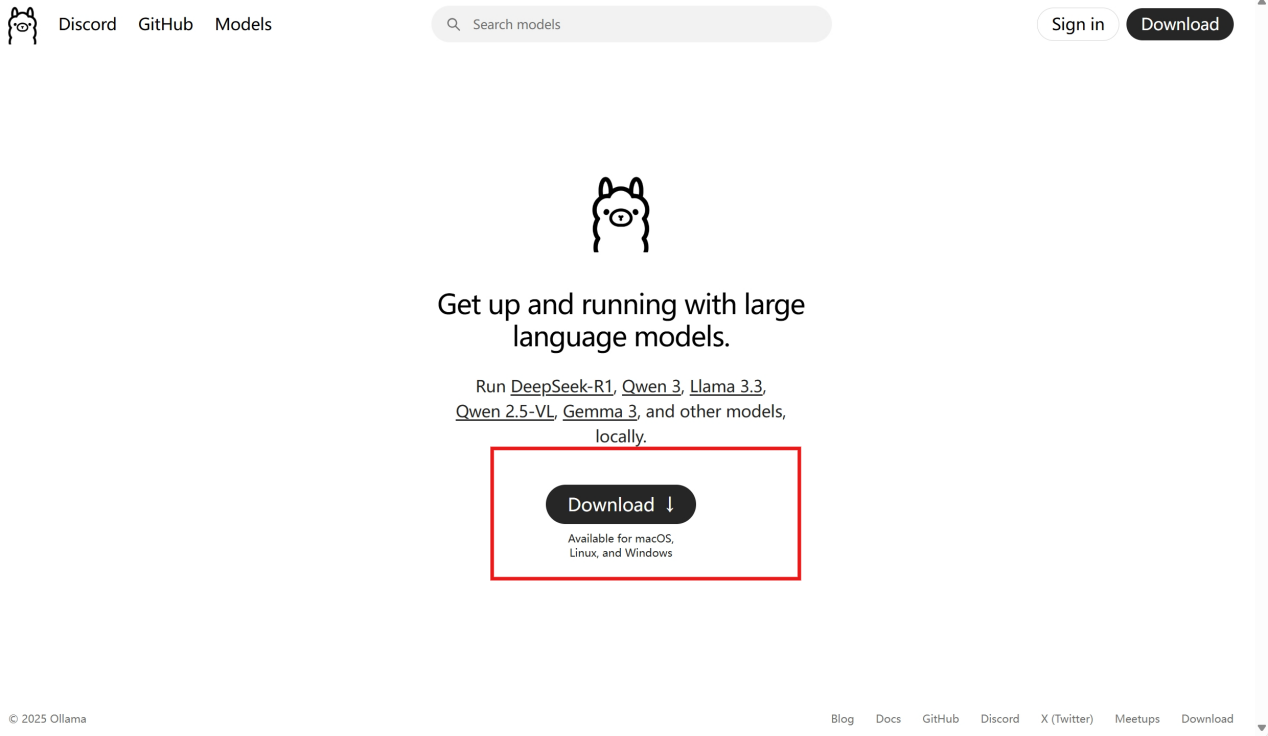
[基于RAG的检索与聊天实现](#_Toc20251_WPSOffice_Level1) 8

[小结](#_Toc26352_WPSOffice_Level1) 13

1. Ollama安装
   1. 官网下载

登录 <https://ollama.com/>

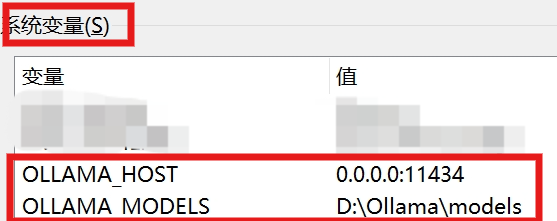
选择相应版本下载



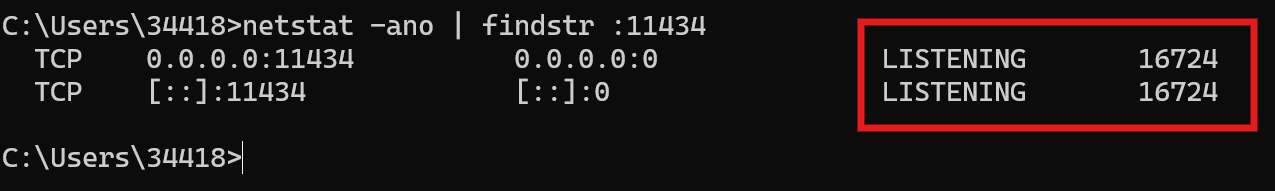
* 1. 环境配置

注意model的配置路径选择自己空闲的盘区，我这里放在了D盘

OLLAMA\_HOST 必须选择0.0.0.0，但后面的侦听端口可以选择自己没有被占用的端口，我这里选择了11434。

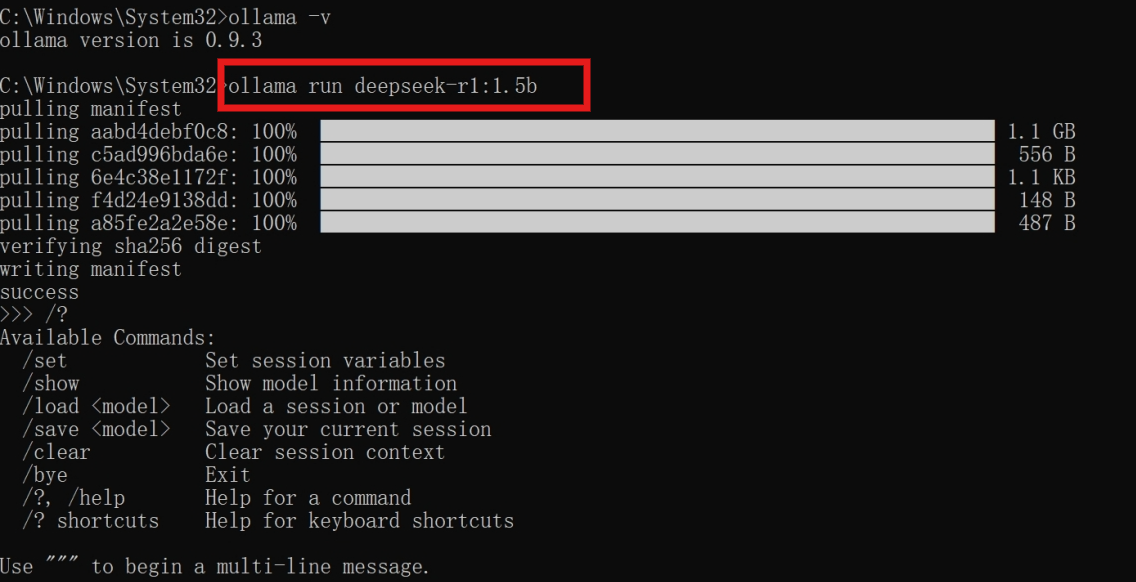


如果不知道端口有没有被占用，可以输入以下指令，下图显示端口11434已经被PID=16724的进程占用。

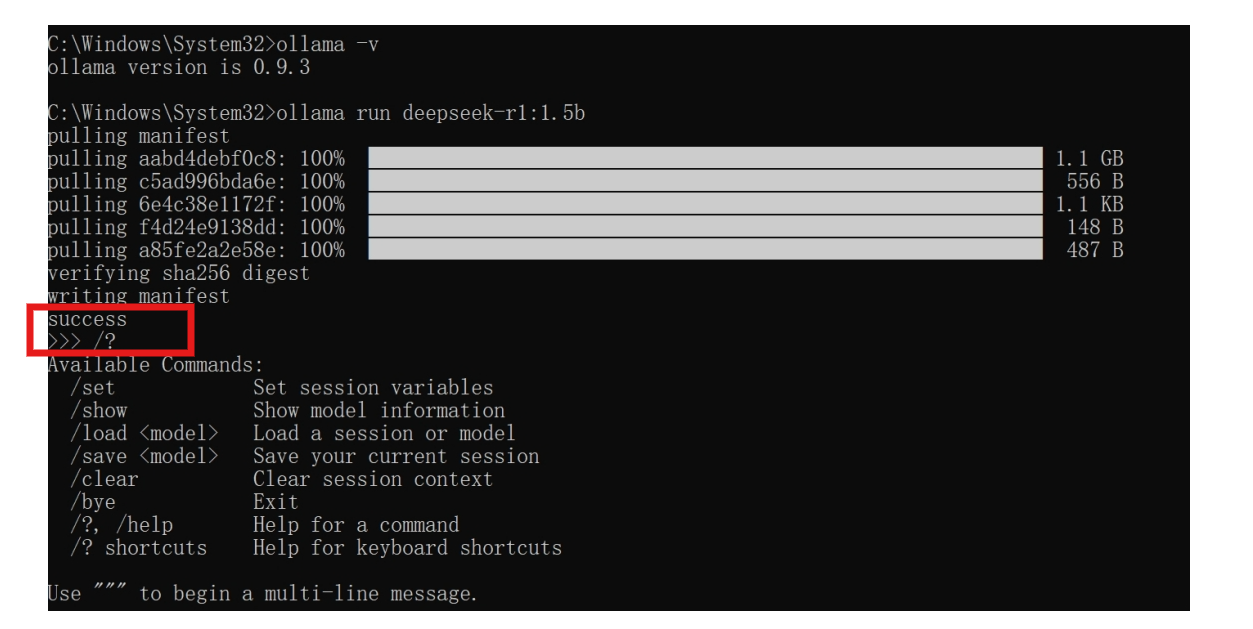


* 1. 模型下载

运行以下指令进行下载



出现以下结果说明下载完成



1. Docker Desktop安装

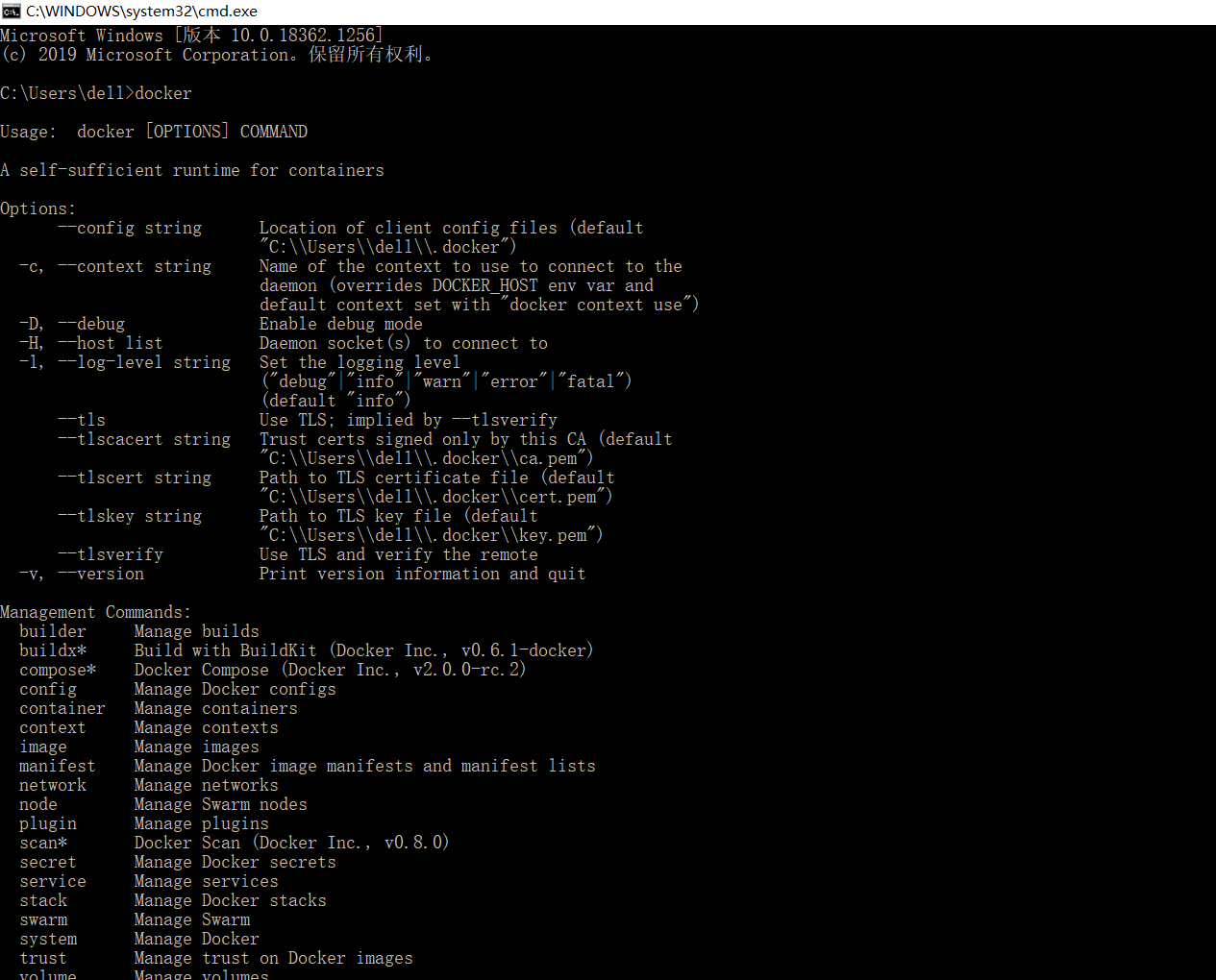
2.1 WSL安装

在cmd中可直接通过指令安装

2.2 Docker Desktop安装

登录官网安装：<https://docs.docker.com/get-docker/>

安装好docker后，在命令行输入 docker ，可以看到docker的一些命令使用方法，这证明docker已经安装成功了



1. RAGFlow安装

3.1 代码包下载

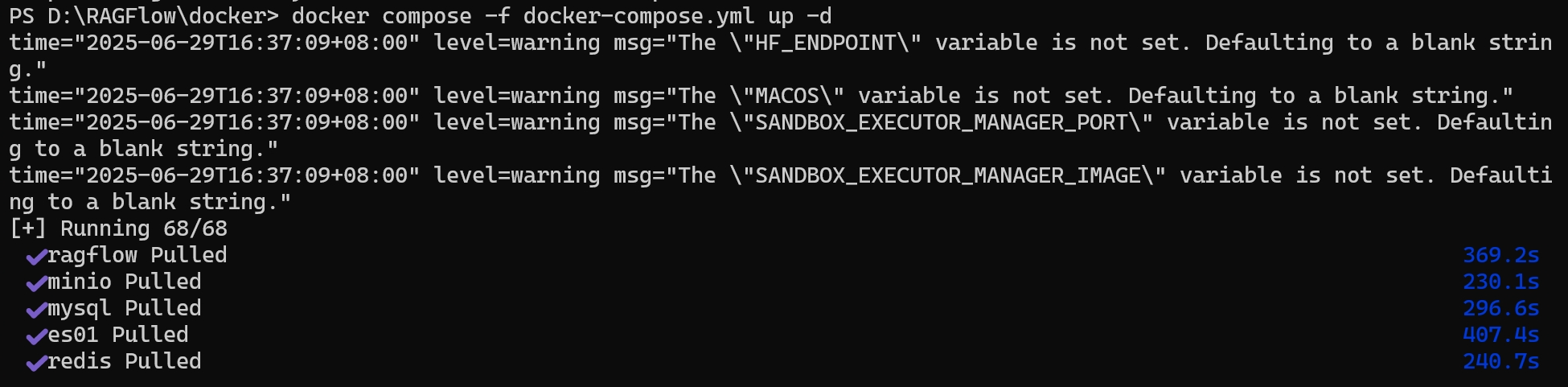
登录<https://github.com/infiniflow/RAGFlow>

在cmd中git clone <https://github.com/infiniflow/RAGFlow.git>

3.2 安装

docker compose -f docker-compose.yml up -d 安装 RAGFlow

出现√则安装正确

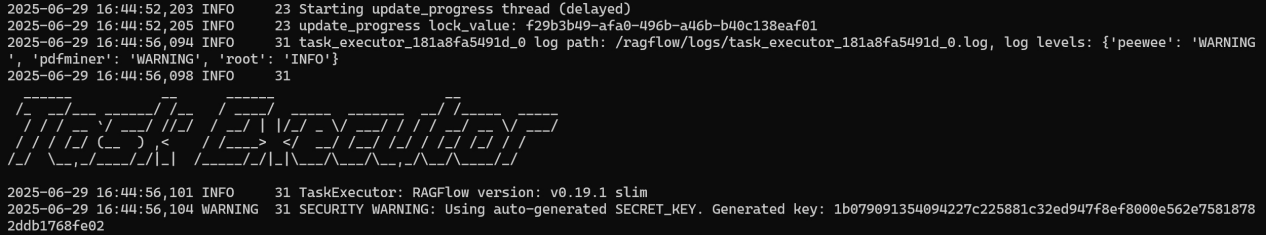




可以使用 docker ps 看一下容器列表，找到 RAGFlow 容器 id



docker logs -f（容器 id）查看容器运行情况，出现图像说明成功



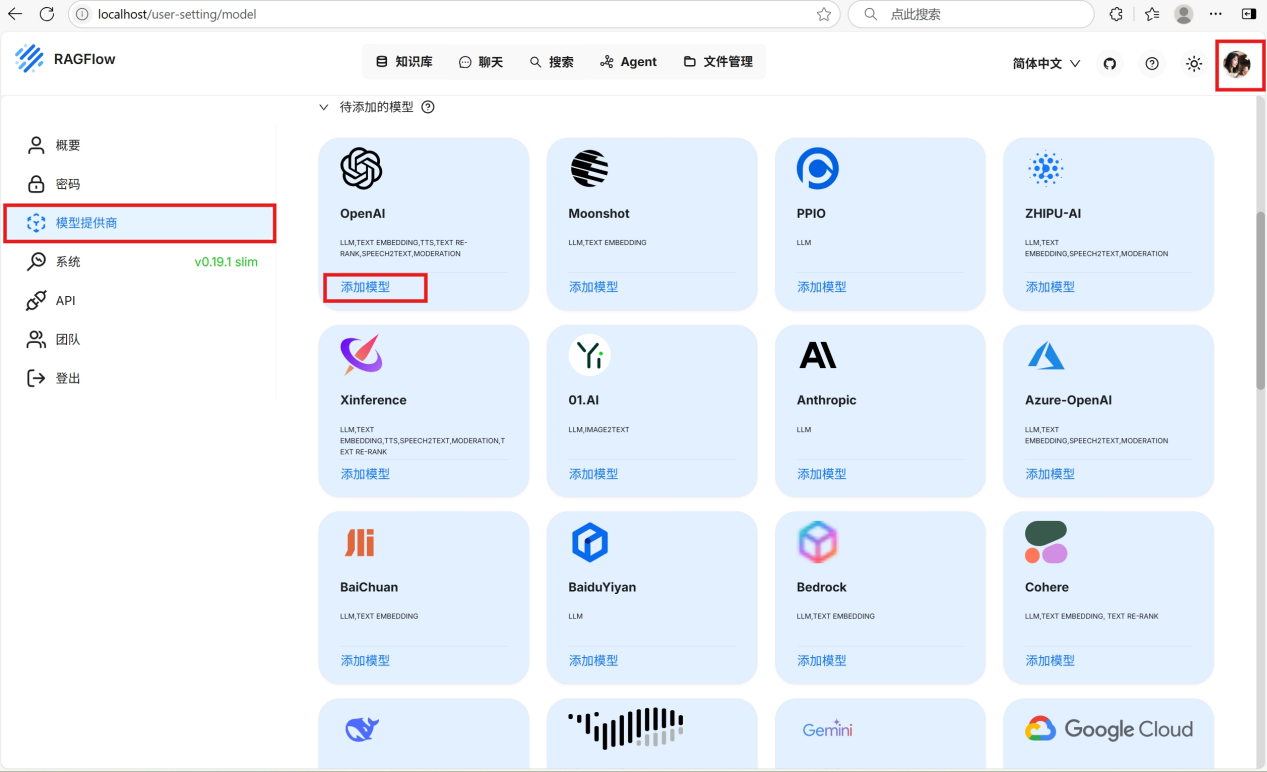
1. 模型选择与配置

4.1 登录RAGFlow

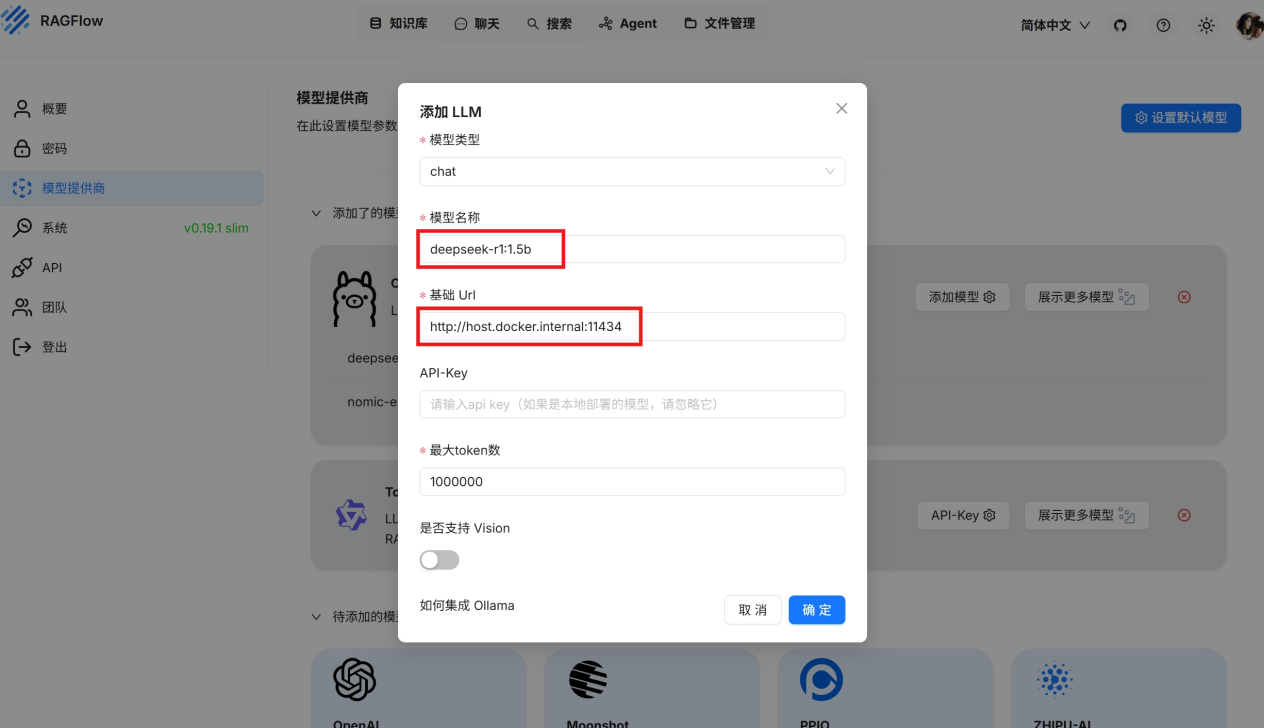
在上一步基础上，访问localhost，进入web页面

4.2 添加本地模型

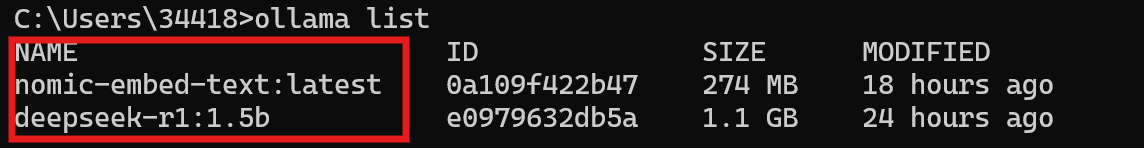
点击头像-“模型提供商”-“添加模型”



如果是添加本地模型，如刚下载的Ollama。



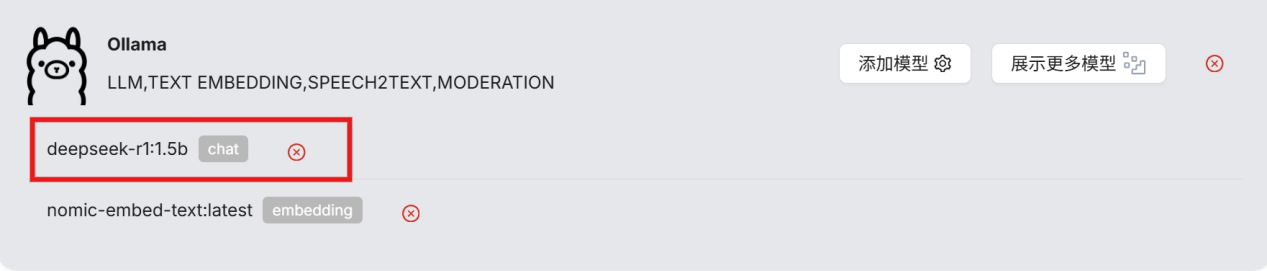
模型名称可以查询



基础Url用<http://host.docker.internal:11434>

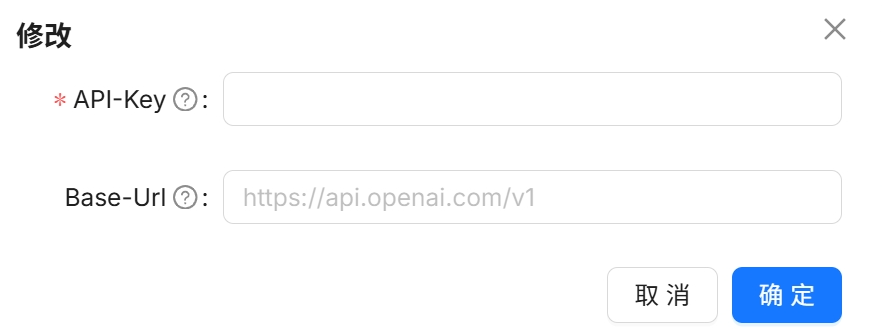
根据之前配的环境变量，11434为ollama侦听的端口。

在展示更多模型下出现模型名称即为添加成功



4.3 添加API

只要填写模型API即可



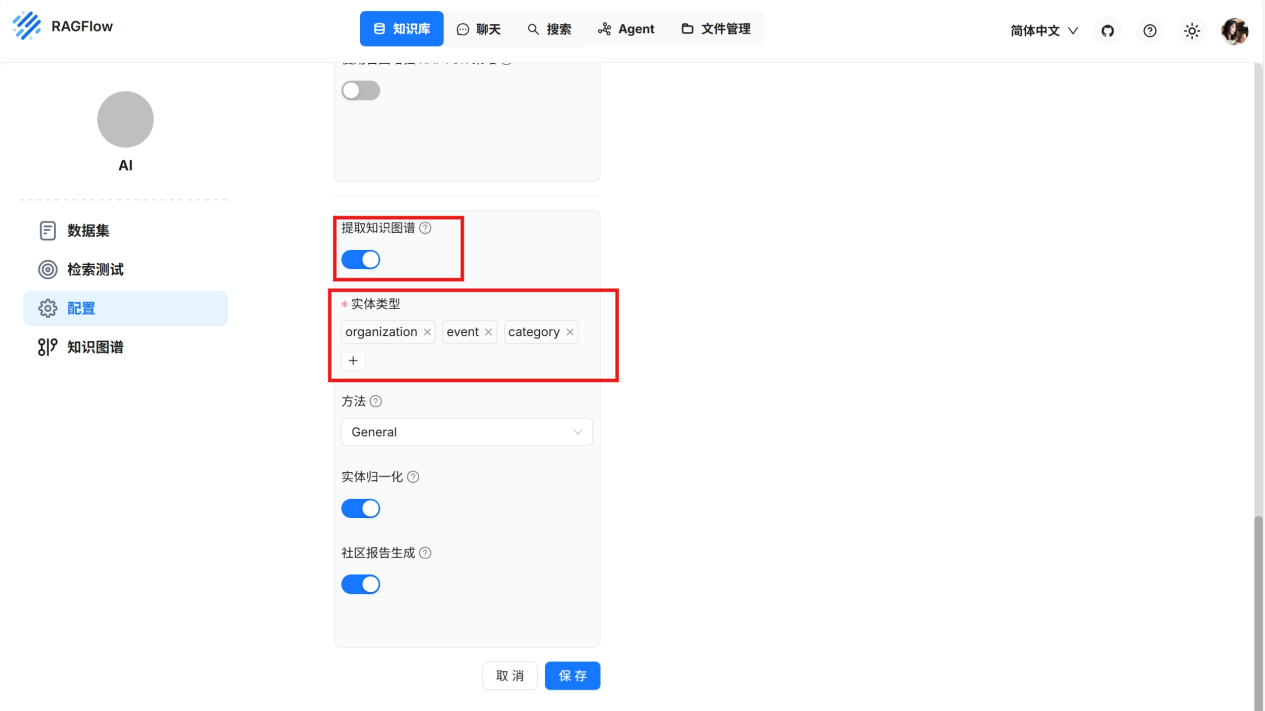
1. 知识图谱构建与生成

5.1 知识图谱属性设置

在知识库配置里，选择PDF解析器和嵌入模型等参数

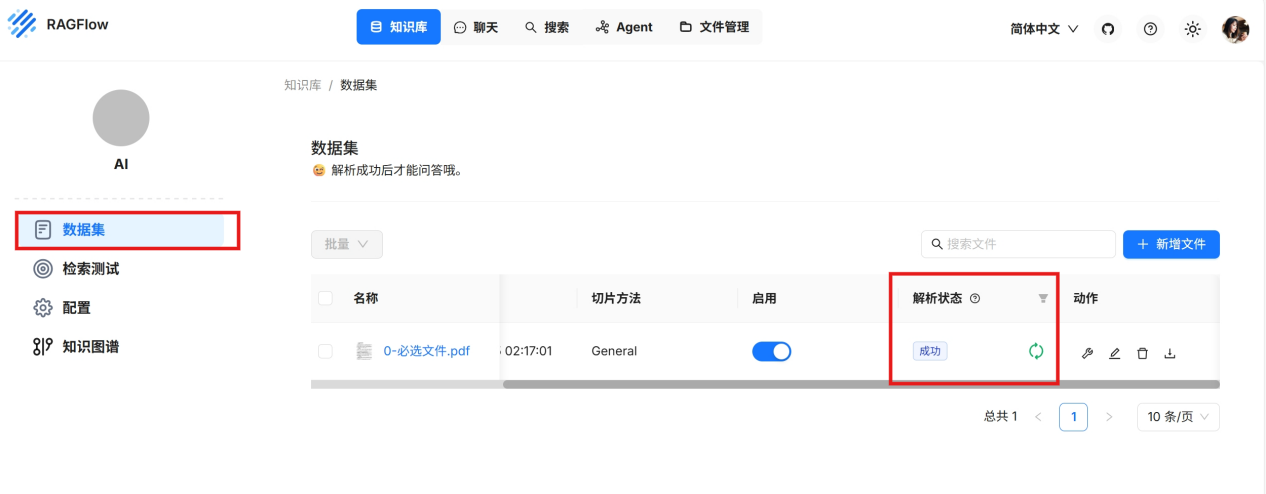


还可以根据内容，定义实体类型，便于抽取

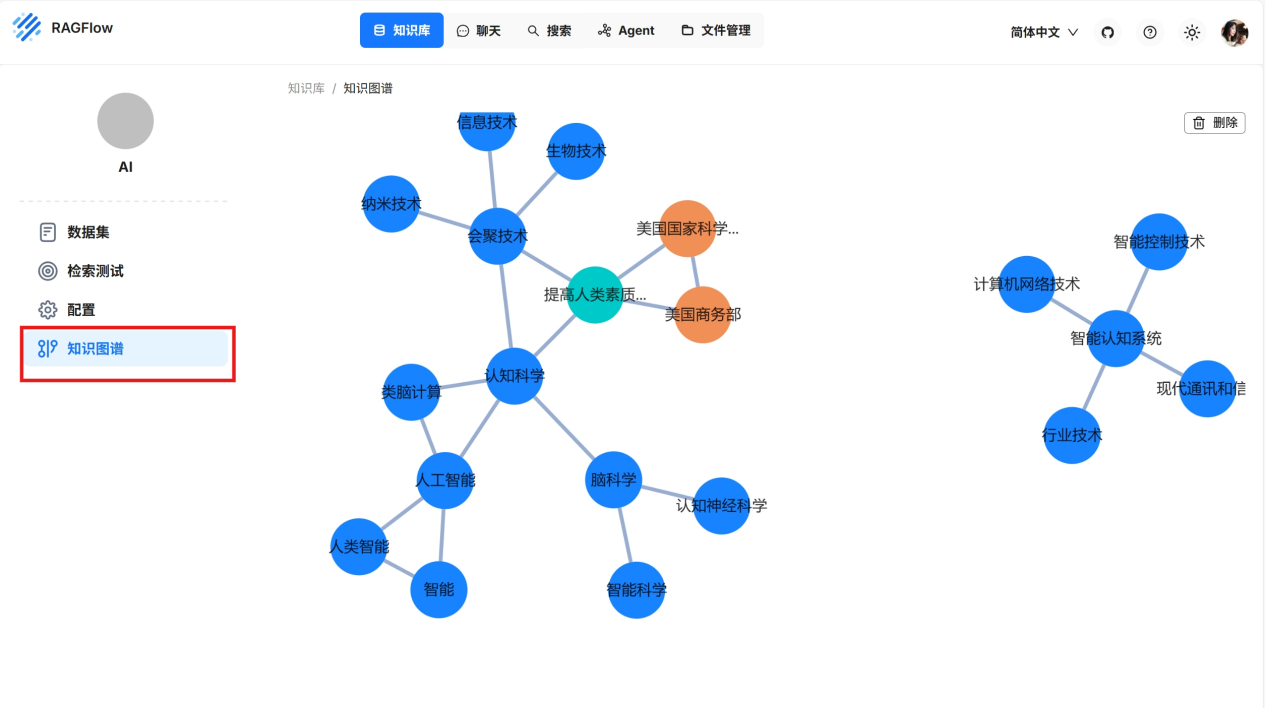


5.2 图谱构建

导入数据集，开始解析，解析一栏“成功”，则代表图谱构建完成



点开知识图谱，可以看见已经绘制好的图谱



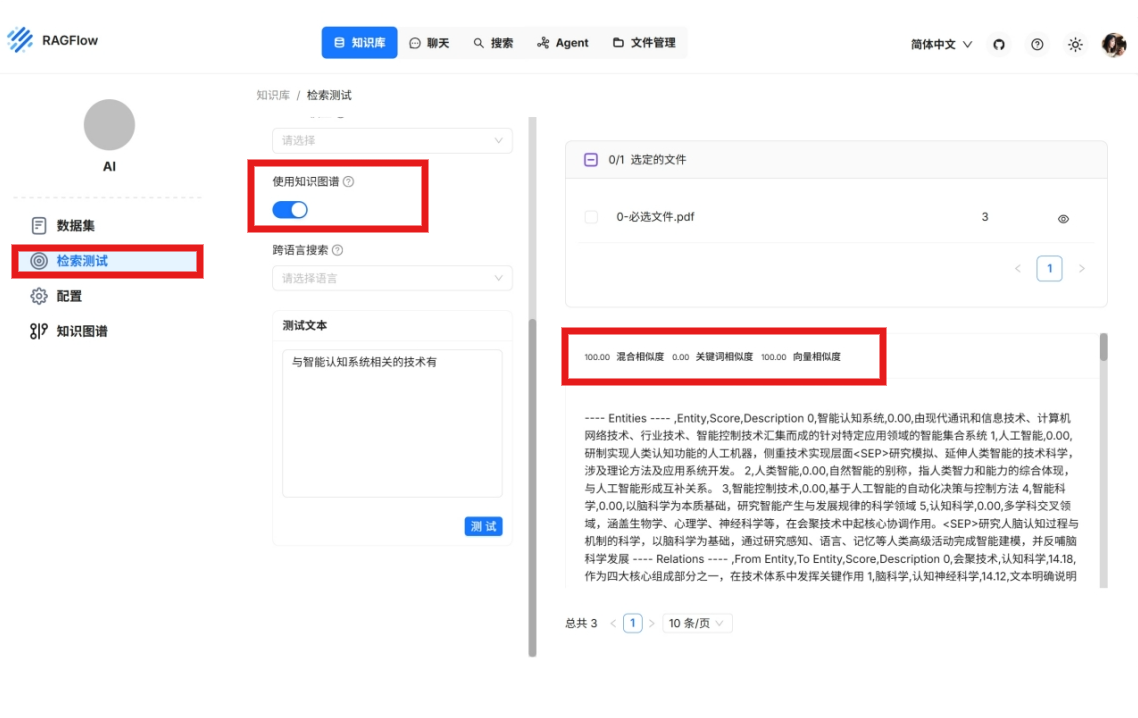
1. 基于RAG的检索与聊天实现

6.1 检索

先测试未使用知识图谱的检索



再测试使用知识图谱的检索



经过对比，我们可以发现，使用知识图谱的检索不仅内容更加丰富，而且混合相似度更高了，说明，检索到的内容准确率更高了。具体对比如下：

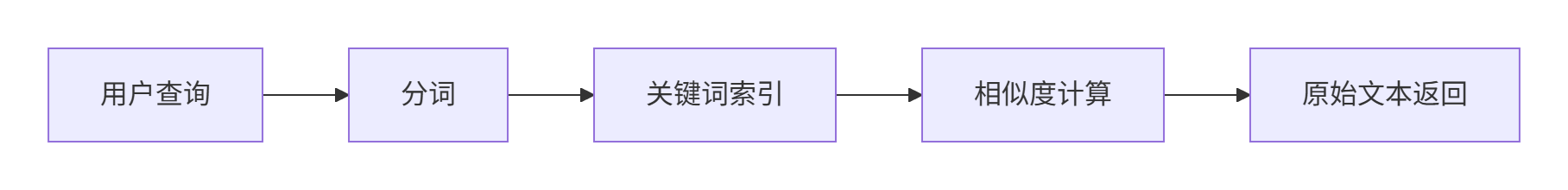
### 对比分析：知识图谱在RAGFlow检索中的作用差异

1、效果对比：

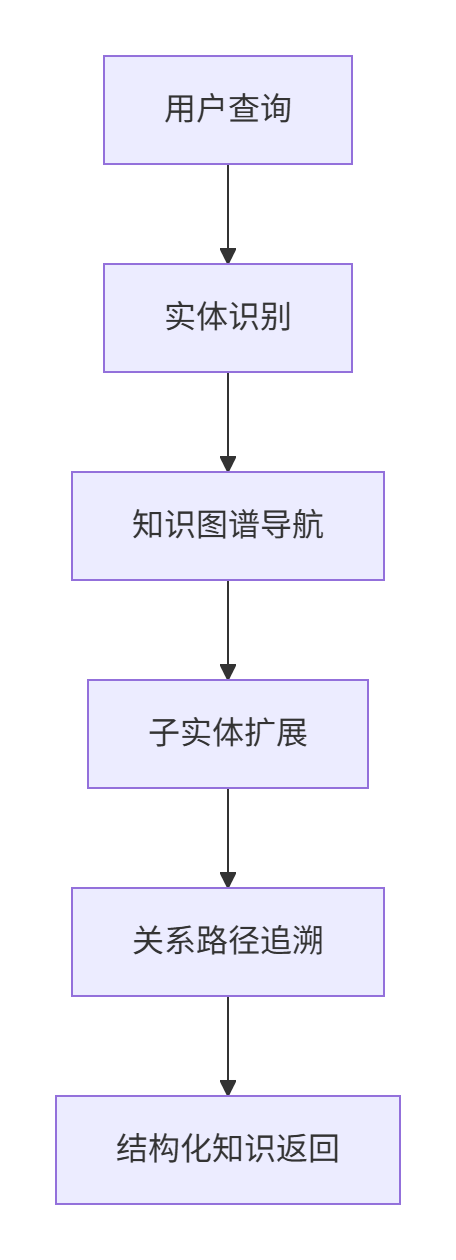


1. 技术对比：

未使用知识图谱时



使用知识图谱时：



1. 对比小结

通过两图对比可见：

（1）未启用知识图谱时，检索结果仅返回模糊匹配的文本片段（如"智能是知识与智力的总和"等），需人工解析技术关联，混合相似度最高仅85.99；​​

（2）启用知识图谱后，系统自动构建结构化认知网络：精准定位"智能认知系统"核心实体（0.00分完全匹配），智能关联5项关键技术（人工智能/认知科学等），揭示学科深层关联："会聚技术→认知科学"(14.18)、"脑科学→认知神经科学"(14.12)三大相似度指标均达100.00完美匹配

总结来说：知识图谱将碎片化文本检索升级为认知网络解析，使技术体系脉络清晰可辨，相似度计算维度从表层语义扩展到实体关系网络，精准度提升16%以上（85.99→100），为技术决策提供量化支持。

6.2 聊天

我们先设置聊天模型和一些参数

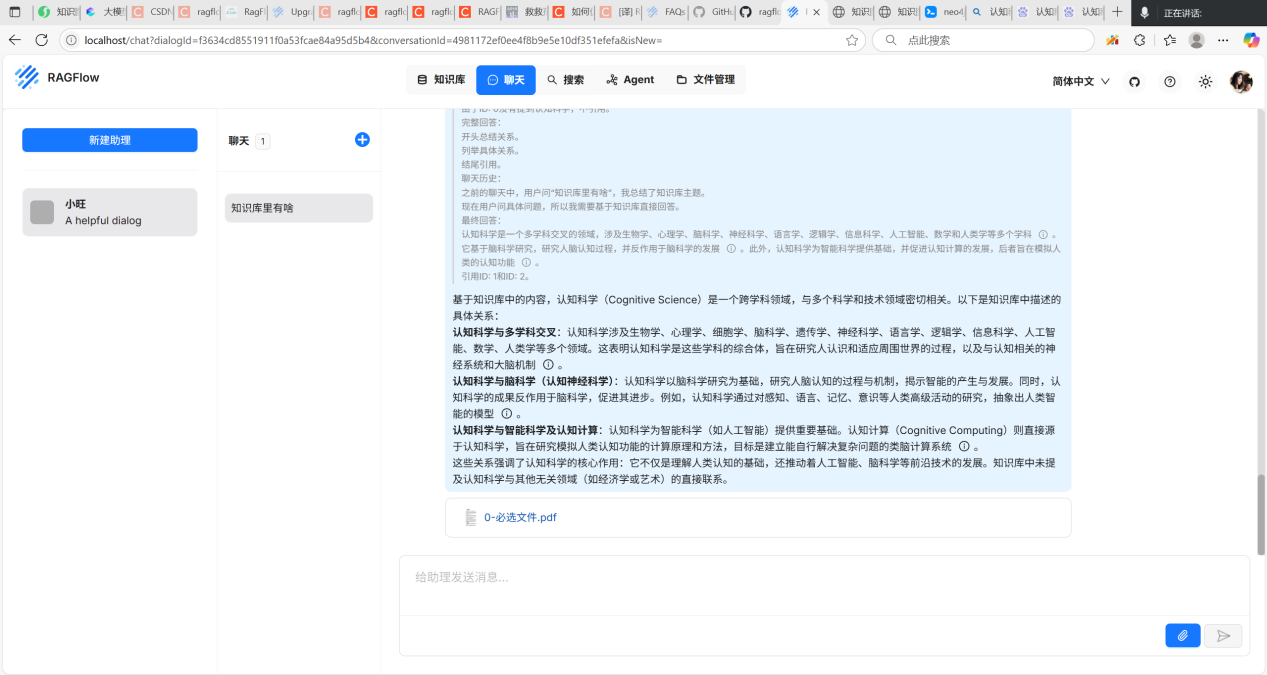
我们可以选择聊天的知识库以及引文，关键词分析等参数。



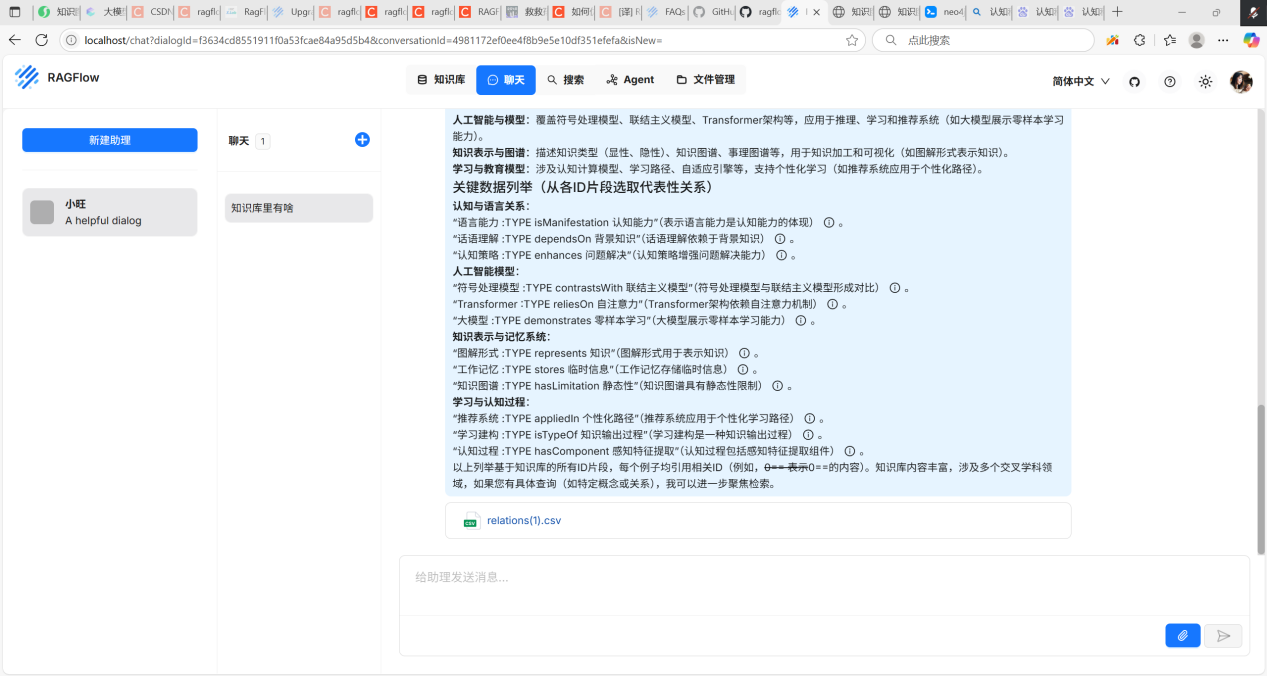
我们还可以选择模型的参数



先测试未使用知识图谱的聊天



再测试使用知识图谱的聊天



我们可以看出，有知识图谱的聊天回答的内容更加丰富，来源都有引用，可信度增加。具体对比如下：

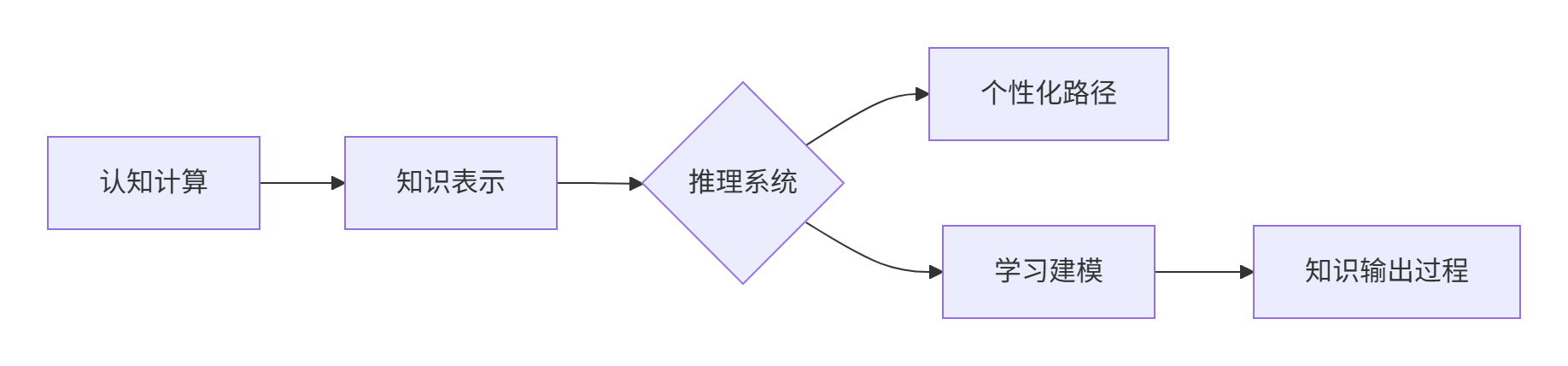
1. 认知关联强度差异对比：



1. 技术对比：

未使用KG：仅笼统提到"认知计算来源于认知科学"，未说明具体实现路径

使用KG：



1. 对比小结：

通过两图对比可见：​​

未启用知识图谱时（图1）的聊天结果呈现连续的描述性文本（如"认知科学涉及生物学、心理学..."），虽然涵盖多学科交叉知识，但存在三重局限：信息组织扁平化：纯文本堆砌认知科学相关学科名称，缺乏层级关联。机制解释薄弱：仅提及"认知计算来源于认知科学"等表面结论，无技术实现路径。认知深度不足：各概念呈离散态，未揭示学科间本质关联。

​​

启用知识图谱后（图2）实现三重突破：结构化认知网络：语法规解 → dependsOn → 背景知识（揭示依赖机制）知识图谱 → hasLimitation → 静态性（暴露技术边界）。跨领域深度关联：Transformer → reliesOn → 自注意力机制（技术原理溯源）工作记忆 → stores → 临时信息（认知功能具象化）。应用导向增强：推理系统 → applied → 个性化路径（明确应用场景）学习建模 → isProcess → 知识输出（展示实现流程）。

1. 小结

RAGFlow是一个很好的关系抽取和知识图谱构建的方法。

他有深度文档解析能力，多模态支持。他能解析PDF、Word、Excel、PPT、扫描件、图像、音视频等50+格式，保留原始排版（表格/公式/标题结构）。他也有OCR智能增强。内置OCR引擎精准识别扫描件、图片中的文字，支持多语言混排文档。他还可以逻辑结构重建。自动还原文档目录层级、表格关联性，确保知识抽取不丢失语义关系。