# AGENT-LLM

## 数据库大作业展示

胡瑞康,莫子昊,陈俊帆

# 大纲

# 1. 团队成员

- 胡瑞康:整体项目结构设计,知识库 RAG 搜索,模仿 OpenAI API 实现,前端修改
- 莫子昊,陈俊帆：表格增删改查实现,后台测试,知识上传功能实现

# 2. 技术概览

# 2.1 数据库

使用安装了 DataVec 向量引擎插件的 openGauss 数据库

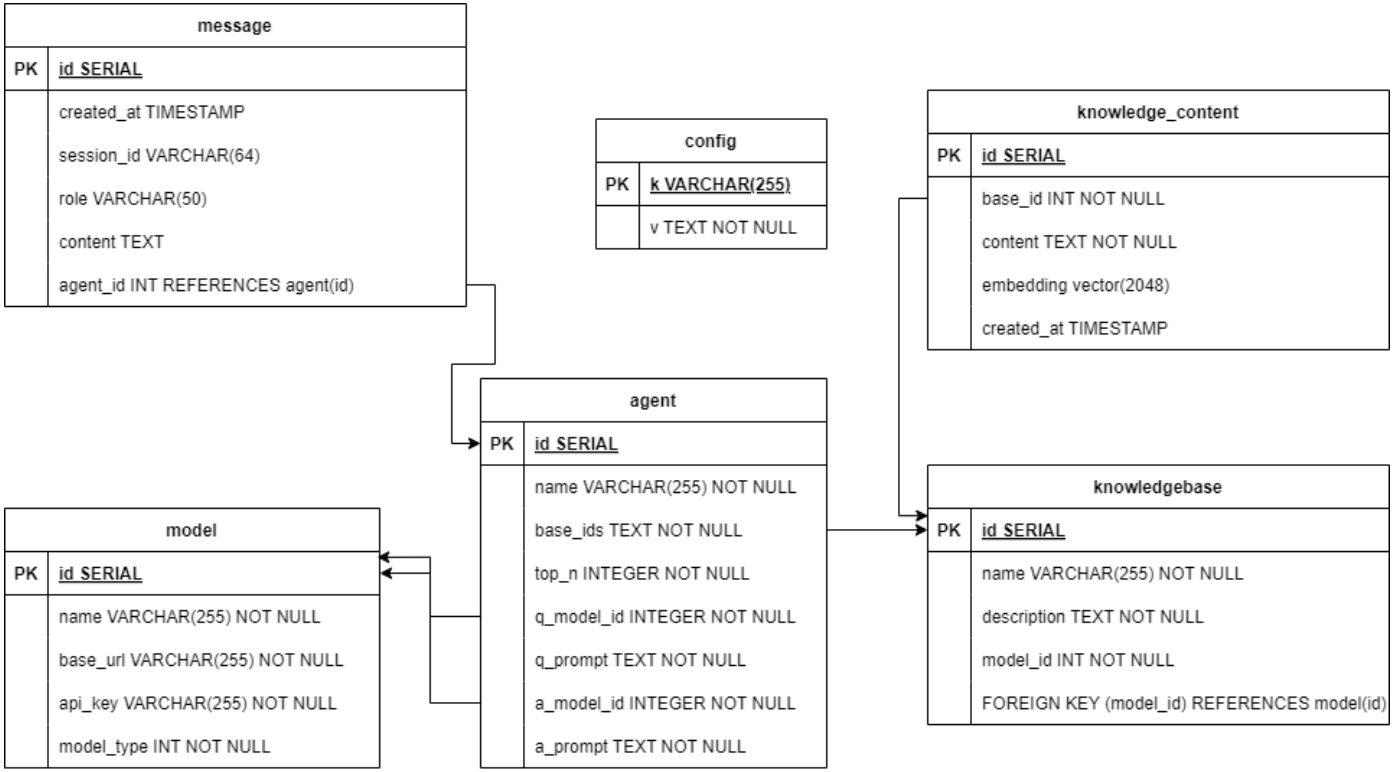


Figure 1: 数据库 ER 图

## 2.2 后端

- 基于 Python fastapi 开发,使用非阻塞式协程
- 用 psycopg3 协程链接数据库
- 封装统一 Model 接口对接数据表,减少重复 sql 撰写
- 通过 AsyncOpenAI 库链接兼容 OpenAI 接口的国产大模型平台 DeepSeek,智谱
- 根据 OpenAI 接口规范,模拟出了他们的 chat 接口,方便任何支持对接 OpenAI 的项目使用我们 API

## 2.3 前端

- 后台使用 LayuiAdmin 配合 Jinja2 模板引擎
- 前台使用 Github 上的 chatgpt-web 做了一定修改对接后端 API,传递对话 id

# 3. 效果展示

## 3.1 前台

各位同学可以使用校园网访问 db.dorm.skyw.cc 预览网页

目前导入了网络中心的一些 QA 做知识库,可以问该类型问题测试,比如校园网,电子邮箱等问题.

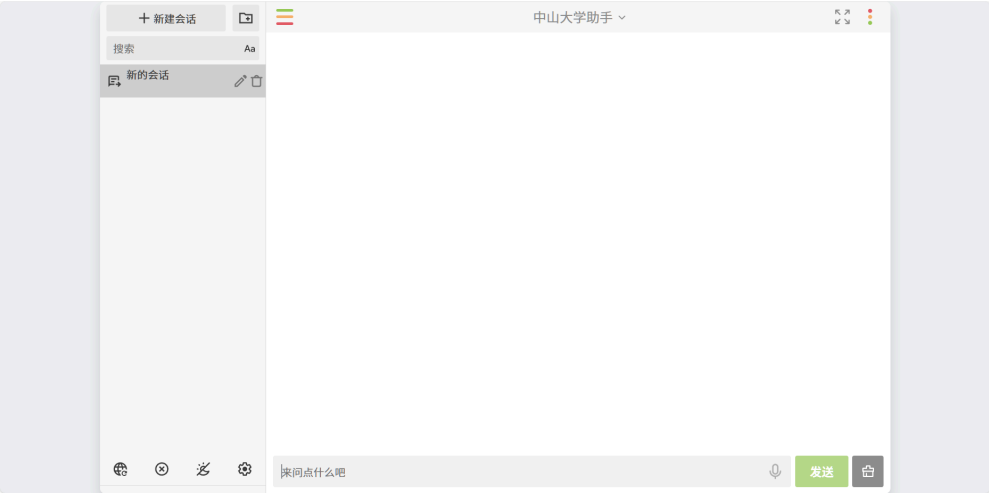


Figure 2: 前台

# 3.2 后台

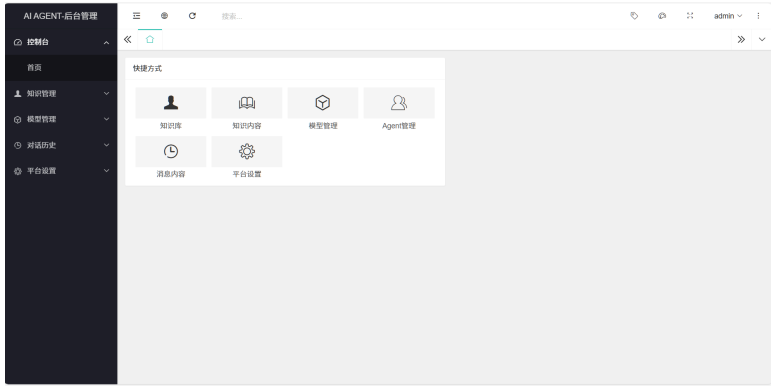


Figure 3: 后台首页

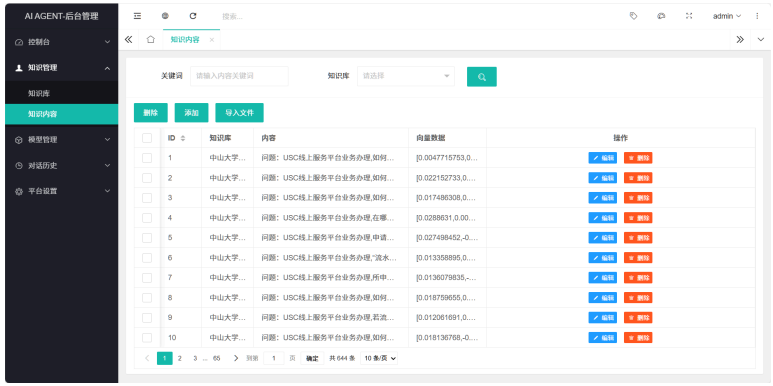


Figure 5: 知识内容

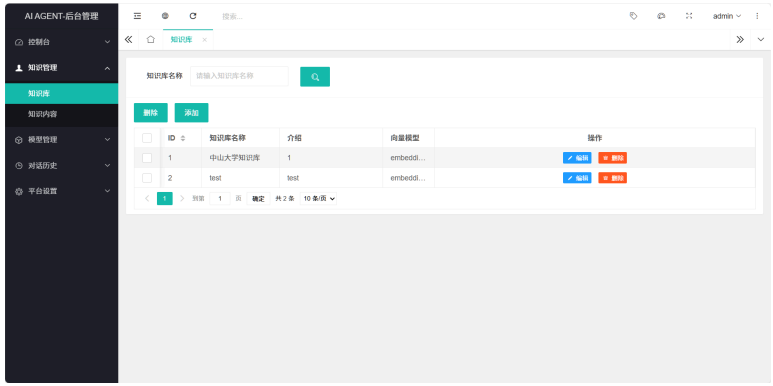


Figure 4: 知识库

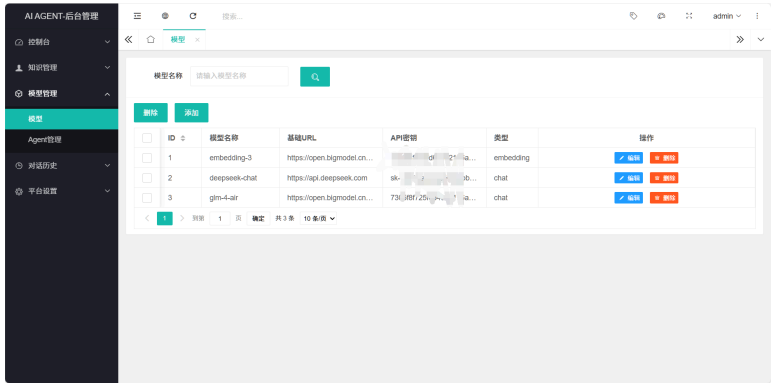


Figure 6: 模型管理

## 3.2 后台

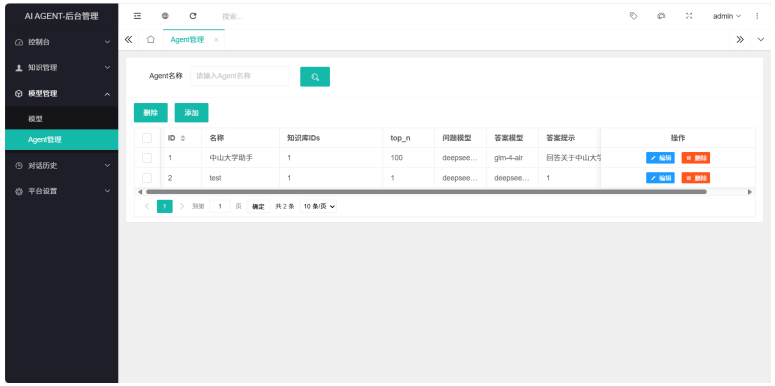


Figure 7: 智能体管理

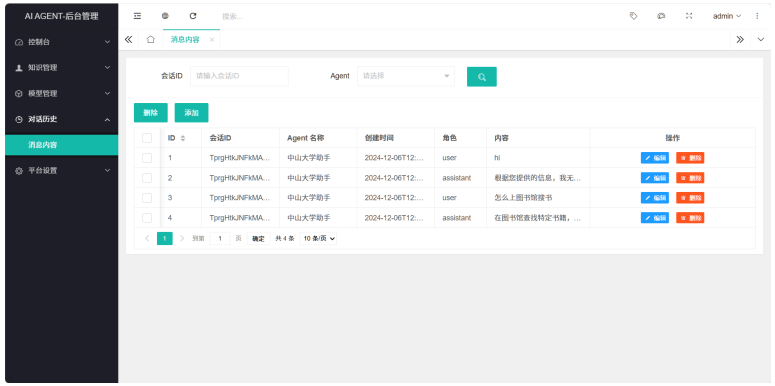


Figure 8: 消息内容

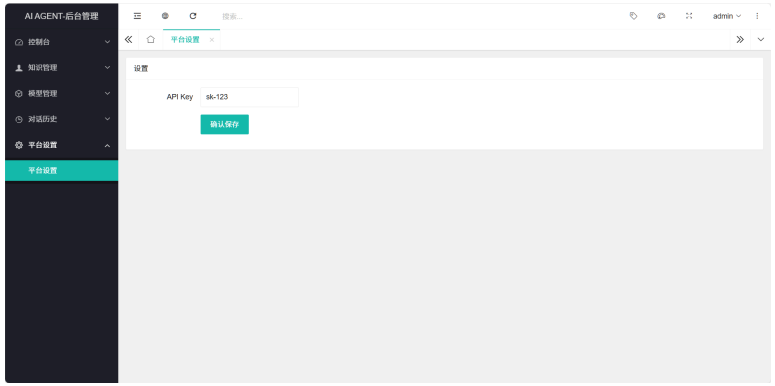


Figure 9: 平台设置

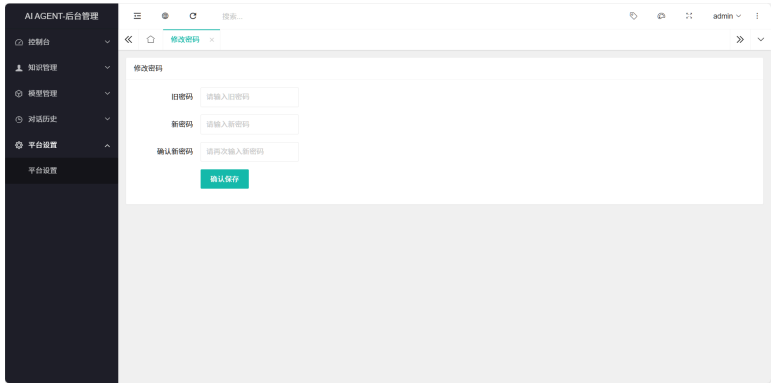


Figure 10: 密码设置

# 3.3 知识库 RAG 搜索

用户问题会先被大模型扩展为更多问题，之后调用接口 Embedding，然后把向量在数据库中匹配相似的 TopN 个结果

由于 FastAPI 特性，链接数据库，调用大模型接口均使用协程 使得我们的搜索速度非常快


Figure 11: RAG 搜索

# 3.4 兼容 Openai API

兼容 Openai 格式可以方便被各种第三方应用调用，第三方开发者无需对已有功能做修改

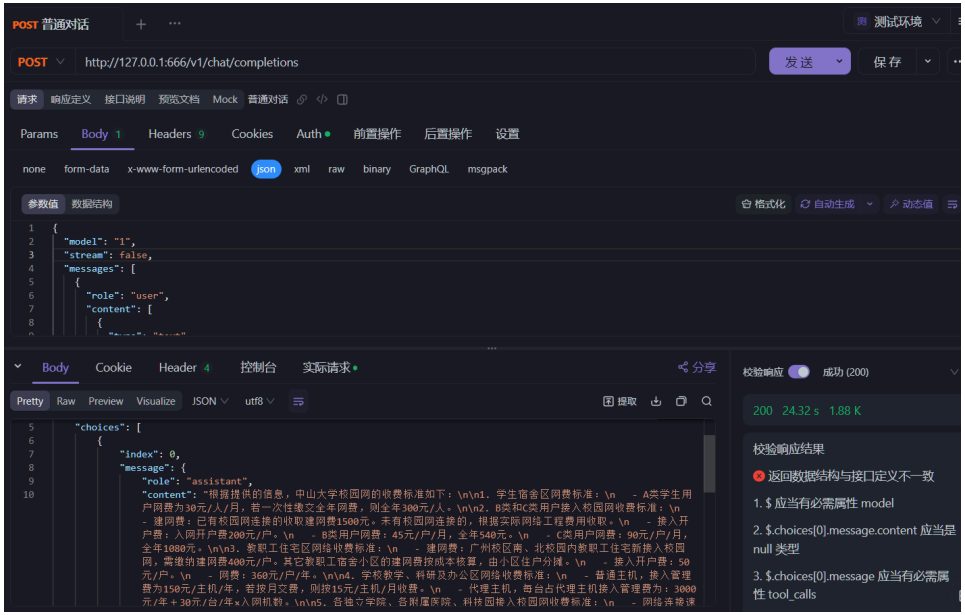


Figure 12: API 回答

# 3.4 兼容 Openai API

SSE（Server-Sent Events）是一种 Web 技术，它允许服务器实时向客户端推送数据。

相比于传统的轮询和长轮询机制，SSE 提供了一种更高效且实时的数据推送方式。

用于给前端实现 API 对话。

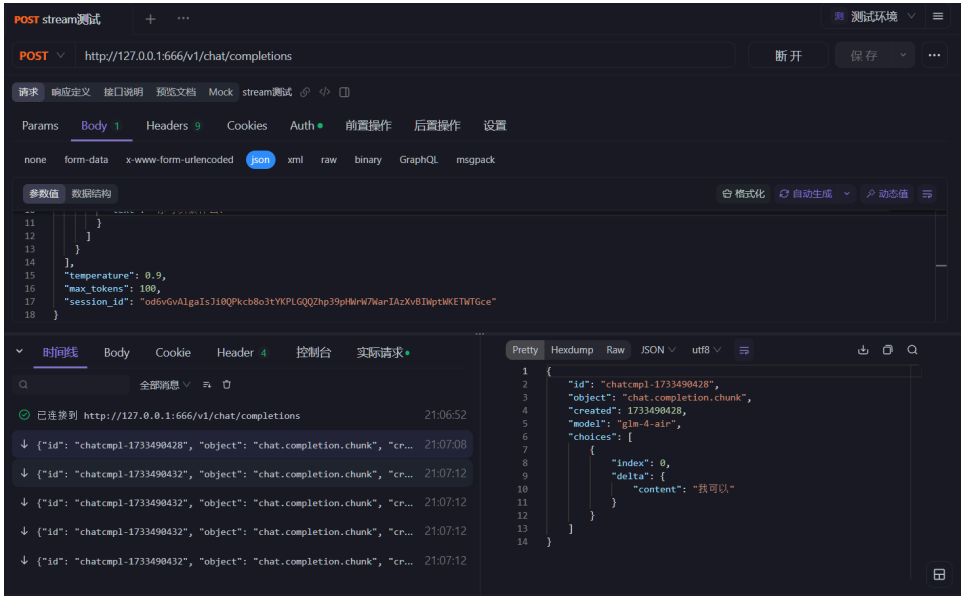


Figure 13: API-SSE 支持