个人知识库助手

|  |  |
| --- | --- |
| 项 | 说明 |
| 名称 | 个人知识库助手 |
| 行业 | 人工智能教育 |
| 技术 | 1. API接口；  2. API调试技术；  3. 文本向量模型：Embedding-2；  4. 编码格式：JSON；  5. 网页请求：POST；  6、API调用方式：同步、异步、SSE 三种调用方式  7. 编程语言：Python；  8. 运行操作系统：Ubuntu |
| 其他 | 代码的调试经验 |

# 一． 项目摘要

国内各大模型公司都推出了自己的模型ChatGLM、文心一言等，为了降低用户的使用成本和门槛，建立了自己的MAAS平台，通过直接调用API的方式来使用大模型，把大模型应该于自己的产品当中，降低了成本，又免除了部署维护的麻烦；本项目以智谱开放平台为例，来介绍大模型MAAS平台的使用，使用文本向量模型，将输入的文本信息进行向量化表示，以便于结合向量数据库为大模型提供外部知识库，提高大模型推理的准确性。

* 使用场景：高职、本科计算机专业、人工智能专业、自动化专业、智能制造专业等。
* 用户人群：高职、本科老师、学生、科研人员，工厂实践基地等
* 产品内容覆盖：计算机技术、人工智能、大模型等
* 就业方向：人工智能系统集成与运维、人工智能数据服务、人工智能应用开发工程师等。

# 二． 项目需求

个人知识库助手是一款致力于帮助用户高效管理知识和信息的工具。它能够从多种来源收集信息，并保存到个人知识库中。同时，提供便捷的文档编辑功能，支持创建和编辑笔记、清单等。为了便于检索和管理，用户可以对知识和信息进行分类和添加标签。此外，个人知识库助手还具备强大的搜索功能，支持关键词、标签等多种检索方式，让用户能够快速找到所需信息。

个人知识库助手还支持多人协作，允许用户邀请他人共同编辑和管理知识库，促进知识的共享和传播。为了确保数据在多个设备上保持最新，并提供数据备份功能，我们还实现了云同步功能。此外，为了降低成本和部署维护难度，我们使用大模型MAAS平台进行开发。结合文本向量模型，将输入的文本信息进行向量化表示，以便于结合向量数据库为大模型提供外部知识库，提高大模型推理的准确性。

个人知识库助手后期可以开发友好的界面，操作简便，易于上手。还可以提供丰富的个性化设置，满足用户个性化需求。在确保数据安全的同时，后面也可以私有化部署保护用户隐私。通过提供优质服务，我们希望吸引用户付费订阅或购买增值服务，实现盈利。

本项目把底层代码调试通过，能够通过API实现个人知识库助手的基本功能，学员可以根据自身开发不同的前端界面，来丰富个人知识库助手的相关功能。

# 三． 项目功能

## 1. API KEY

请不要与他人共享您您的 API Keys，避免将其暴露在浏览器和其他客户端代码中。为了保护您帐户的安全,我们还可能会自动更换我们发现已公开泄露的密钥信息。 新版机制中平台颁发的 API Key 同时包含 “用户标识 id” 和 “签名密钥 secret”，即格式为 {id}.{secret} 。



## 2. SDK 安装说明

安装 Python SDK

Python SDK 地址：[https://github.com/zhipuai/zhipuai-sdk-python-v4](https://github.com/zhipuai/zhipuai-sdk-python-v4" \t "https://open.bigmodel.cn/dev/_blank)

首先请通过如下方式进行安装 SDK 包：

pip install zhipuai

我们升级了最新的模型 GLM-4、GLM-3-Turbo，支持了System Prompt、FunctionCall、Retrieval、Web\_Search等新功能。 使用以上新功能需升级最新版本的 Python SDK。如您已安装老版本SDK，请您更新到最新版SDK。

pip install --upgrade zhipuai

## 3. SDK 用户鉴权

所有 API 使用 API Key 进行身份验证。在开放平台[API Keys 页面](https://open.bigmodel.cn/usercenter/apikeys" \t "https://open.bigmodel.cn/dev/_blank)查找您将在请求中使用的 API Key。

**Python SDK 创建 Client**

我们已经将接口鉴权封装到 SDK，您只需按照 SDK 调用示例填写 API Key 即可，示例如下

from zhipuai import ZhipuAI

client = ZhipuAI(api\_key="") # 请填写您自己的APIKey

response = client.chat.completions.create(

model="glm-4", # 填写需要调用的模型名称

messages=[

{"role": "user", "content": "你好！你叫什么名字"},

],

stream=True,

)

for chunk in response:

print(chunk.choices[0].delta)

## 3. SDK 向量模型

模型编码：embedding-2

创建向量模型同步请求

接口请求参数



接口响应参数



**Python 调用示例**

from zhipuai import ZhipuAI

client = ZhipuAI(api\_key="your api key")

response = client.embeddings.create(

model="embedding-2", #填写需要调用的模型名称

input="你好",

)

响应示例

{

"model": "embedding-2",

"data": [

{

"embedding": [(一共1024个元素)

-0.02675454691052437,

0.019060475751757622,

......

-0.005519774276763201,

0.014949671924114227

],

"index": 0,

"object": "embedding"

}

],

"object": "list",

"usage": {

"completion\_tokens": 0,

"prompt\_tokens": 4,

"total\_tokens": 4

}

}

## 4. 前端网页的开发

搭建Web前端的交互页面，实现Web界面同API的交互。

# 四．项目使用的技术与学习建议大纲

|  |  |
| --- | --- |
| 技术名称 | 推荐学习资料 |
| 1. API技术 | 《API设计的模式》 |
| 2. 向量模型 | 《深入浅出Embedding:原理解析与应用实践》 |
| 3. 编码格式：JSON | 《[JSON必知必会](http://www.baidu.com/link?url=aPLgzmANhDOlBR-HYTpRhTq4_ufz_PGv5wLHduEnBYdVLqAaIA6qw-3r_yCUD8olneJX5KQkRjjCCyS8_XlZba" \t "https://www.baidu.com/_blank)》 |
| 4. 编程语言：Python | 《人工智能编程语言Python基础》 |
| 5. 运行操作系统：Ubuntu | 《Linux操作系统基础》 |