

# CS PhD 申请助手

---

## 1. 项目背景:

在进行海外PhD申请时，繁琐的申请流程让人望而却步。AI的出现，使得申请者在信息获取、整合方面的压力有了释放的空间。然而，由于申请任务的复杂性，实现通用的申请助手无疑需要大量的专用数据和智能的匹配算法，现有的大模型无法满足该需求。使用RAG技术结合语言大模型，可以在回答申请者的问题时实现知识查询和逻辑整合，有机会帮助申请者完成选校、邮件编写、文书写作等任务。

## 2. 项目阐述:

本项目采用RAG方法，结合yuan2.0大模型进行计算机专业指定方向的选校辅助、导师推荐。

## 3. 产品功能:

方向筛选：用户可以在网页端提问，系统为其匹配研究方向。

选校建议：用户选定指定研究方向后，系统为其推荐学校。

导师推荐：用户在网页端提问感兴趣的学校，系统会输出匹配的导师信息，并给出详细推荐理由

## 4. 技术方案:

### 4.1. 整体架构

大语言模型（LLM）：用于生成详细的推荐理由和回答用户问题。

嵌入模型（Embedding Model）：用于将文本转换为向量表示，支持相似性查询。

向量库索引（Vector Store Index）：用于存储和查询分段后的文本向量，支持基于相似度的推荐。

### 4.2. 模块解析

TextSplitter 类

功能：将长文本分割成较小的段落，以便后续的向量化和相似性计算。

主要参数：

max\_chunk\_size: 每个文本块的最大长度。

overlap: 相邻块之间的重叠部分，用于避免在分段时丢失信息。

LLM 类

功能：加载预训练的大语言模型（如 LLaMA），根据上下文生成文本。

模型加载：使用 transformers 库中的 AutoModelForCausalLM 类加载模型。

生成方法：

通过提供背景信息（上下文）和用户问题，生成回答。

使用采样策略生成新的文本，支持的参数包括 temperature 和 top\_p。

## EmbeddingModel 类

功能：将文本转换为向量表示，用于相似性计算。

模型使用：采用有道训练的bce-embedding-base-v1向量模型，具有中英文跨语种能力，且有RAG优化，适配真实业务场景。

模型加载：使用 transformers 库中的 AutoModel 类加载模型。

嵌入计算：通过 get\_embeddings 方法，将文本批量转换为嵌入表示，并进行归一化。

## VectorStoreIndex 类

功能：存储文档的嵌入表示，并根据用户查询找到最相似的文本块。

文本切分：在初始化时，将文档切分为多个文本块，并计算这些块的嵌入表示。

相似性查询：通过 query 方法，根据输入问题找到最相似的文本块，返回排名靠前的结果。

## 4.3. 页面布局与交互

### Streamlit 页面布局

页面设置：使用 st.set\_page\_config 设置页面标题和布局。

侧边栏：用户可以在侧边栏中选择研究方向。

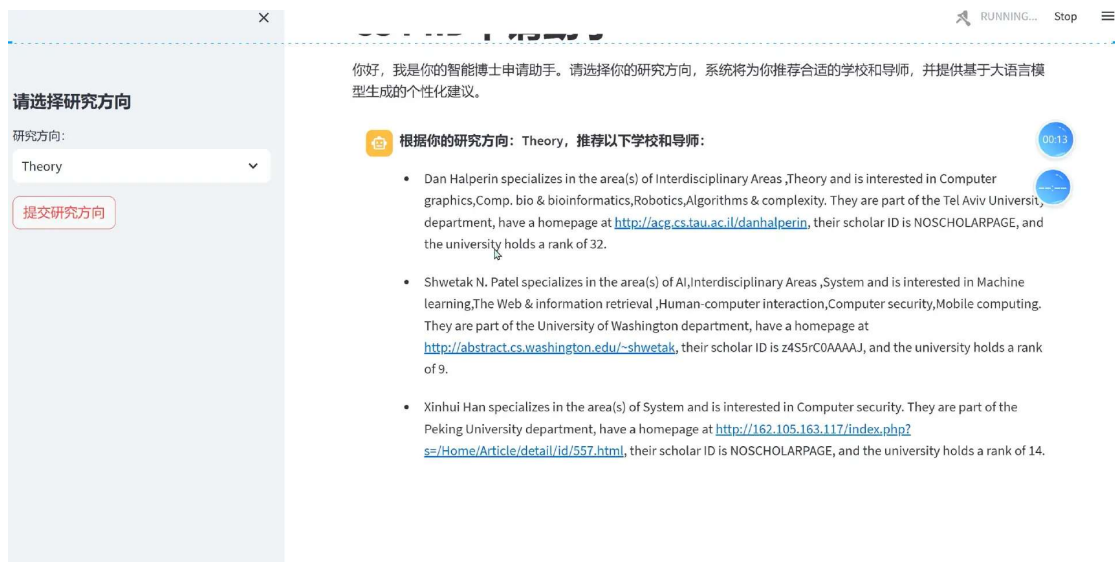
主页面：显示推荐的学校和导师，并提供生成详细建议的功能。

### 交互流程

选择研究方向：用户在侧边栏选择研究方向后，系统会查询相似的文档块，并显示推荐内容。

输入问题：用户可以输入具体问题，系统根据上下文生成回答。

## 5. 运行效果：



## 6. 迭代计划：

进行内存优化，提高系统稳定性。（9/1）

进行语料扩充，智能判断匹配信息个数。（9/8）

增加网络搜索和网页分析模块，根据用户需求进行网络信息获取。（9/13）

## 7. 商业模式：

暂无商用计划。

## 8. 团队介绍

团队名：Mustbecrazy

姓名	性别	学校	学历	职责
梁凯晴	女	中国科学技术大学	研究生	项目策划，代码部署
贺维易	女	Michigan State University	研究生	代码编写
许霆	男	Umass boston	研究生	数据处理，前端编写