**项目建议书：交大科研牵线人**

**——基于大语言模型的学生-导师配对助手**

上海交大 纪佳翼

**一、项目背景与问题描述**

**1.1 项目针对的问题**

本项目旨在解决上海交通大学本科生在寻找合适的科研导师和方向时所面临的问题。

**1.2 社会背景**

交大新生素质优秀，是未来科研的潜在人才，但是很多同学并不了解自己本专业的研究方向，交大有哪些老师，各自的研究方向如何。虽然官网上有上述这些信息，但信息量庞大，很难查找。交大老师也很希望优秀的同学加入自己的课题组，但寻找学生方式有限，据观察大多数老师采用给自己任课的教学班的优秀学生发邮件私聊，但所找到的学生不一定适合本课题组。

**1.3 学术背景**

自ChatGPT公布以来，大语言模型性能逐步涌现，大模型基座的泛化能力很强，但是缺乏在特定领域的知识储备，一旦问到细节的问题无法发挥其智能优势。而现实生活中需要人工智能解决的问题大多不是其现有的吟诗写作等创作能力，而是希望其成为具有专业知识的智能助手，专业领域的大模型具有较强的落地应用可能性，大模型的知识注入是一片未曾涉足的蓝海，需要各个领域的人共同打造适用于不同领域的大模型，才能让其能力充分落地。

**二、解决方案**

**2.1 数据集构建**

在数据集方面，通过从学校各学院官网教师介绍（含研究兴趣，论文发表情况），学校对于各老师的报道（如某老师某方向论文发表），不同老师近期发表的论文摘要中提取信息，整理成结构化的数据，采用合适的方法将知识向量化，构建知识库。同时，还可从学校教务处官网获取各专业课课程名称及任课教师，构建课程-教师对应的知识体系。





用于构建知识库的信息示例1：学院官网教师简介



用于构建知识库的信息示例2：教务处公布的各专业课任课教师



用于构建知识库的信息示例3：交大官网公布的交大教授最新科研发现

**2.2 大模型基座**

在大模型基座方面，采用上海人工智能实验室开源的**InternLM-20B**(或7B）[1]模型作为基座，利用其开源的RAG工具将整理好的知识库注入大模型，这样同学通过与大模型交互，可以最快捷的找到适合自己的导师。例如：告诉其自己比较感兴趣的内容，模型可自动推荐交大相关方向的老师，便于同学们进一步与教授沟通，或者同学告诉大模型自己取得高分的专业课，模型自动对应相关方向的老师，介绍其研究兴趣。

**三、项目意义**

当前同学找课题组只能一个一个翻阅学校官网或prp，大创等网站，效率低下。这个大模型可以解决同学希望迈出科研的第一步但却苦于无法找寻到合适的老师的问题，使得大模型成为学生找导师的媒介，通过与模型交互便能获取导师履历，研究兴趣，研究方向摘要等信息，既减轻了学生的压力，也减少了老师找合适的学生的工作量。未来可以设法对模型进行实时更新，或拓展更丰富的知识，如交大课程学习助手，交大校史讲解员等，如果能够持之以恒可能创造一个服务交大师生的全方位智能体。

**四、方案基础与外部支持**

本项目在撰写建议书时得到了**上海人工智能实验室**的支持，如能成功立项，可以获得他们的**全链路大模型开发工具，系统的大模型入门课程，A100级别的算力支持**。项目需要的工具和算力都已经有了很好的保证。

**五、参考文献**

[1] InternLM Team. InternLM: A Multilingual Language Model with Progressively Enhanced Capabilities [EB/OL]. [2023-03-10]. https://github.com/InternLM/InternLM-techreport.