

# 大学生创新训练项目计划申请书

项目编号\_\_\_\_\_

项目名称 基于 LLM 的国际中文学习融媒体系统

项目负责人 李晓鹏 联系电话 18560741975

所在学院 信息与控制工程学院

学 号 202011050028 专业班级 软件工程 201

指导教师 房斐斐 冯志国

申请日期 2023 年 6 月 1 日

起止年月 2023. 6. 1-2024. 6. 1

青岛理工大学制

## 填写说明

1. 本申请书所列各项内容均须实事求是，认真填写，表达明确严谨，简明扼要。
2. 申请人可以是个人，也可为创新团队，首页只填负责人。“项目编号”一栏不填。
3. 本申请书为大 16 开本(A4)，左侧装订成册。可网上下载、自行复印或加页，但格式、内容、大小均须与原件一致。
4. 负责人所在学院认真审核，经初评和答辩，签署意见后，将申请书(一式两份)报送项目管理办公室。

一、 基本情况

项目名称	基于 LLM 的国际中文学习融媒体系统						
项目级别							
项目类型	创新训练项目						
所属学科	学科一级门：0812 计算机科学与技术 学科二级类：081203 计算机应用技术						
申请金额	10000 元	起止年月		2023 年 6 月 至 2024 年 6 月			
负责人	李晓鹏	性别	男	民族	汉族	出生年月	2001 年 10 月
学号	202011050028	联系电话	宅：手机:18560741975				
指导教师	房斐斐	联系电话	宅：手机:18669882758				
	冯志国		宅：手机:13953219144				
项目简介	<p>国际中文作为一种独特的语言形态，致力于通过中文教育弥合跨文化沟通的鸿沟，跨越语言和文化的障碍。相较于传统汉语言，国际中文更加注重培养跨文化交流能力和实际应用技能。为达成此目标,国际中文学习融媒体系统将汉语知识与中华优秀传统文化相互融合，以提升学习者的兴趣度和参与感。此外，该系统将整合图文、视频等多种表现形式,采用多模态教学策略。这一综合性学习系统将为全球中文学习者提供坚实支持，并在推动国际中文教育研究方面发挥着关键作用。</p> <p>系统采用大型语言模型(Large Language Model, LLM)对海量中文语料进行训练，旨在构建全知型中文学习 AI 助手，以便在融媒体环境中为学习者提供实时解答。同时，通过整合多个语音模型，精确捕捉学习者的口语发音，快速分析语义并生成恰当回应，实现多场景口语互动。项目所采用的算法可将中文学习者的发音与标准发音对比，精确纠正发音。此外，系统还能运用学习者的对话数据进行自学习分析,智能推荐针对性的学习内容，匹配难度适当的话题和中华优秀传统文化篇章,以提高中文学习者的学习效果。</p> <p>汉语水平考试(HSK)作为一项国际标准化考试，旨在评估非中文母语者在生活、学习和工作场景中运用汉语进行交际的能力，涵盖 HSK 一级至六级。本研究项目构建的国际中文学习融媒体系统，未来将依托于 HSK 词汇和中华优秀传统文化典籍，建立丰富的中华语料库。针对每个话题场景，系统将为学习者提供相应的对话任务。在互动过程中，系统未来将自动纠正学习者对话的语法错误。完成对话后，系统会根据学习者回答的话题相关性、语义连贯性等因素进行评分，并给出针对性的改进建议。</p> <p>融媒体(Media Convergence)，是一种将传统媒体、新兴媒体及互联网技术整合的媒体发展模式。本研究项目充分利用各类媒介的独特优势，以多模态的形式呈现国际中文知识与中华优秀传统文化，使中文学习过程更加直观、易于理解。</p> <p>本研究项目将提供从语料库构建、模型训练到创新应用的完整解决方案，旨在继承中华传统优秀文化,弘扬时代精神,为构建新时代中国特色社会主义的国际中文学习体系做出贡献。</p>						

负责人曾经参与科研的情况		2022-2023 学年在美国大学生数学建模竞赛获得国际级二等奖； 2022-2023 学年在全国大学生数学建模竞赛获得省级一等奖； 2020-2021 学年在青岛理工大学获互联网+创新创业大赛院级一等奖；			
指导教师承担科研课题情况		1. 国家科技支撑计划项目子课题“社区互动服务综合整合关键技术研发和海量数据分析子系统研发”，2012. 1-2014. 12； 2. 山东省教育厅科技项目，基于主动学习的大规模图文结合垃圾邮件过滤技术研究，已结题，国内领先； 3. 山东省博士基金，网络海量动态文本流内容发展演化挖掘及可视化平台，已结题，国际先进； 4. 青岛市科技计划基础研究项目：服务公共安全的网络舆情发展及演化智能挖掘技术研究，2012. 1-2014. 9； 5. 青岛市科技计划基础研究项目：网格服务组合、调度及访问机制的安全可信性研究； 6. 多项国创、省创项目的指导老师； 7. 多次指导学生参加蓝桥杯软件设计大赛、挑战杯软件设计大赛等并多次获奖；			
指导教师对本项目支持情况		指导教师对于系统设计与实现提供技术指导, 提供所需数据, 团队合作等软实力培养。			
项目组主要成员	姓名	学号	专业班级	所在学院	项目中的分工
	李晓鹏	202011050028	软件 201	信息与控制工程学院	融媒体系统的总体设计 研究项目网站平台搭建 LLM 模型教学助手训练
	李路轩	202111050228	信息 212	信息与控制工程学院	系统功能模块软件测试 融媒体平台的运行维护 中华语料库的整合优化
	朱骊安	202111010143	中美土木 211	土木工程学院	融媒体系统的需求分析 学习者记录的数据挖掘 多模态教学的内容整合 项目创新点的文档撰写
	周一诺	202201090234	英语 221	人文与外国语学院	HSK 教材大纲数据分析 中英口语对话模型训练 语音语法纠错功能实现 跨文化交流的教学设计
	陈鑫雨	202011020352	广告 211	人文与外国语学院	智能化推荐算法的设计 传统文化的多模态呈现 融媒体传播的模式创新 全球学习者的社群管理

二、 立项依据

(一) 研究目的

伴随着我国国际地位的日益提升,中外文化交流活动也变得日益频繁,中华民族优秀传统文化逐步赢得了国际友人的认可和喜欢。在“汉语热”的科学视角下,国际中文学习的传播力度得到了空前的强化。

近年来,教师、教材、教学法的“三教”问题仍然是国际中文教育和汉语国际推广的基本问题。不仅如此,新冠疫情后,国际中文的教学模式也有了革命性的改变——从线下转变为线上,在“三教”方面提出了新的需求:

教师方面,不但需要教师精通汉语教学的技巧,还需要教师熟悉互联网的运作原理才能做好线上教学工作。

教材方面,线上的教学资源不能局限于书面形式,还需要结合图文、视频等融媒体信息,打造多模态的汉语教材资源。

教学法方面,需要适应线上教学的特征,设计出新的教学思路 and 教学体系,既要注重实时在线互动,也不能忽略教学内容在云端的存续性。

针对“三教”问题,该项目通过大语言模型开发的中文学习 AI 助手解决“教师”问题;通过融媒体技术的多模态教学,解决“教材”问题;通过以上技术的探索,结合语料库中增加中华优秀传统文化典籍以及实时更新内容,取 AI 助手的 24 小时在线之特长,最终解决“教学法”问题。

同时,通过本项目系统的设计和实现,能够培养项目成员的综合实践能力、创新能力以及分析问题解决问题的能力,为今后的学习研究打下坚实基础。

(二) 研究内容

国际中文学习融媒体系统的主要功能是为国际中文学习者打造一个以 HSK 体系为衡量标准的,兼顾跨文化沟通能力的智能化融媒体。

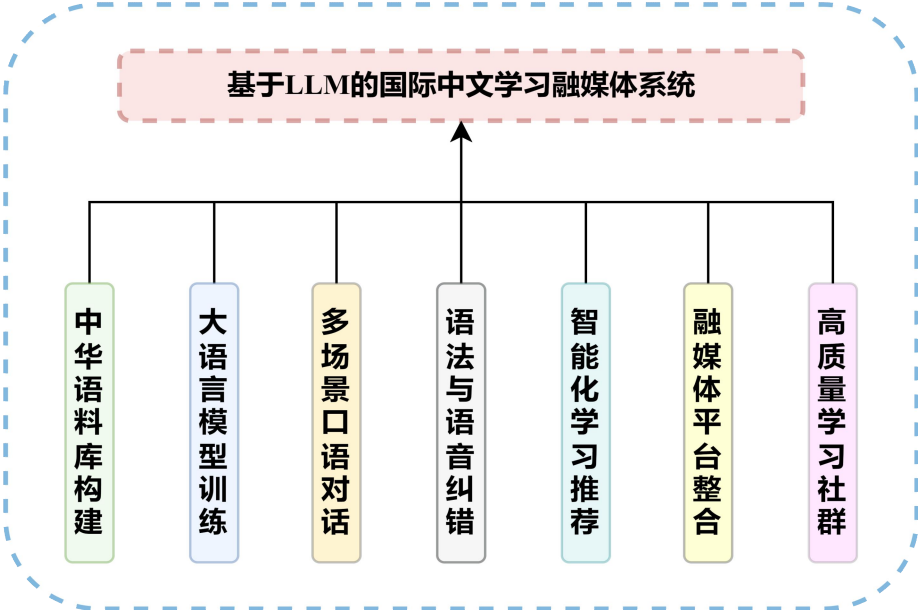


图 1 系统功能模块图

受新冠疫情影响,国际中文学习逐步由线下实体课堂转变为线上的虚拟课堂,由于时间和空间的特殊性,师生之间几乎只在上课时进行交流和互动,课后缺乏互动,学习者时常遇到疑问而无从解答。此外,线上教学还存在文化交流薄弱,缺乏口语训练,教材枯燥乏味,课堂形式单一等问题。

## 1、中华语料库构建

针对文化交流薄弱的问题,本项目未来将运用海量的中华优秀传统文化典籍构建多模态语料库,语料库的技术框架包含基础层、中间层和交互层。

基础层任务是语料元素采集整理,即多模态数据的采集、整理和归档,具体包括典籍中图像的收集整理、音视频等的文本故事编纂、多模态数据元素对齐方案设计与实施。

中间层是算法模型搭建和业务逻辑层设计,涉及多模态语料关联分析与呈现算法,多模态语料检索算法,多模态语料融合下的情感分析、跨模态翻译、特殊元素定位检索、系列故事关联分析算法,以及算法之间融合和关联的逻辑层设计。

交互层是学习者界面层,服务于学习者检索和动态交互,促进检、研、用、学一体化虚拟社区形成。多模态语料库通过数据处理、存储管理、可视化交互分析等技术,实现模态关联查询、可视化关联分析及多模态融合下具体任务信息的发掘等功能。

在上述中国传统文化多模态语料库的基础上,再收集和整理大量中文教育相关的文本资料,包括课本内容、教师讲义、习题解析等。对收集到的文本进行预处理,如去除特殊符号、分词、词性标注等,以便后续模型训练。

## 2、大语言模型训练

针对课后缺乏互动的问题,本项目未来将采用基于 Transformer 的大型语言模型 GPT-3.5,利用迁移学习中的 PLMF 方法(Pretrained Language Model Fine-tuning),中华语料库作为数据基础,设计并实现一款网页版的中文教学 AI 助手。通过高效生成中文词汇释义、例句、学习提示等功能,AI 助手未来将实现 24 小时不间断的高质量答疑服务。

## 3、多场景口语对话

针对缺乏口语训练的问题,我们未来将基于大型语言模型训练适用于不同场景的 AI 对话助手。该助手将既能够进行特定场景的文字对话,也能模拟特定话题的口语对话。本项目将基于多模态语料库,通过自学习技术训练多场景的专项练习助手。学习者通过点击话题卡片的方式可选择不同专业的 AI 进行具体场景的口语练习。

### (1) 自学习技术实现细节

我们未来将融合自然语言处理和深度学习算法来实现自学习功能。具体而言,我们的模型包含以下几个关键步骤:

评级模型训练:我们采用大量人工标注的口语训练数据,涵盖各个难度等级和话题。通过监督学习和迁移学习技术,模型能够在训练过程中不断提高预测的准确性。

中文能力评估:模型首先对学习者的输入进行分析,包括词汇、语法和语义理解。通过比较学习者输入与各难度级别的标准答案,模型能够评估学习者的口语能力。

### (2) 分场景对话模型训练

为了模拟现实对话,我们未来将针对不同生活场景进行分类训练模型。这些场景包括但不限于:

旅行出行:如机票预订、酒店入住、观光旅游等。

学习工作:如课堂讨论、求职面试、项目汇报等。

对于每个话题场景, AI 都会为学习者提供相应的对话任务。当完成对话后, AI 会根据学习者回答的话题相关性、语义连贯性等因素进行打分, 并给出有针对性的改进建议。

### **(3) 分难度对话模型训练**

为了满足不同程度学习者的需求, 我们将口语训练分为基础话题、进阶话题和高级话题三个难度等级。模型通过上述自学习技术判断学习者的口语水平, 并匹配难度适当的口语训练。分级依据主要参考 HSK 大纲和往年学生的得分情况。每个等级的具体内容和细节如下:

**基础话题:** 针对 HSK 1-3 级学习者, 主要涵盖日常生活中常见的场景和话题, 如问候、购物、餐饮等。训练内容注重基本词汇、语法和发音的掌握。

**进阶话题:** 针对 HSK 4-5 级学习者, 包括一些较复杂的场景和话题, 如旅行、工作、学习等。训练内容强调词汇量的提升、语言表达的准确性和流利性。

**高级话题:** 针对 HSK 6 级及以上学习者, 涉及抽象主题、社会议题和文化交流等。训练内容关注深入的思考、逻辑推理和论证能力, 以及地道的表达方式。

## **4、语法与语音纠错**

### **(1) 语法纠错**

语法纠错涉及到对中文学习者输入的文本进行分析和修正。我们未来将采用以下的技术路径和方法实现此功能:

**依存句法分析:** 通过对句子进行依存句法分析, 我们可以识别句子中的主谓宾等语法成分, 从而判断句子是否符合语法规则。

**基于规则纠错:** 我们未来将根据汉语语法规则, 设计一套规则库。通过比对学习者输入与规则库, 我们可以发现并纠正句子中的语法错误。

**统计语言模型:** 利用中华语料库训练出的统计语言模型, 我们可以评估句子的语言规律性。模型会根据概率值对句子进行修正, 使其更符合实际语言使用习惯。采用基于神经网络的深度学习方法, Transformer、BERT 等, 结合大量标注过的语料, 训练高性能的语法纠错模型。

**多种模型融合:** 为了提高纠错的准确性和鲁棒性, 我们未来会融合不同类型的语法纠错模型, 研发一款综合性能更优的纠错模型。

### **(2) 语音纠错**

针对语音纠错, 我们未来将采用以下技术路线实现此功能:

**语音识别:** 我们利用自动语音识别(ASR)技术将学习者的语音输入转换为文本, 以便后续的数据处理。

**韵律分析:** 我们对学习者语音的韵律特征(如音高、音强、语速等)进行分析, 以评估其发音的准确性和自然度。

**基于规则纠音:** 我们将根据汉语发音规则和音位学知识, 设计一套发音纠音规则库。通过比对学习者发音与规则库, 我们可以发现并纠正发音中的错误。

**多种模型融合:** 为了提高纠音的准确性和鲁棒性, 我们未来会融合不同类型的语音纠错模型, 研发一款综合性能更优的纠音模型。

## **5、智能化学习推荐**

针对教材枯燥乏味的问题, 本项目未来将依托于 HSK 词汇与中华优秀传统文化典籍, 构建丰富的中华语料库。系统将运用学习者的对话数据进行自学习分析, 智能推荐有针对性的学习内容, 匹配难度适当的话题和中华优秀传统文化篇章, 以提高中文学习者的学习效果。

## 6、融媒体平台整合

针对课堂形式单一的问题,本项目未来将依托 YouTube、TikTok 等短视频与直播平台,建设面向海外中文学习者的多终端在线教学工具——HSKhelper。平台整合视频、语音、图文等多模态的教学资源,为国际中文学习者提供丰富、趣味的学习体验。通过定期举办 HSK 趣味闯关答题、口语风采展示、汉字作文展示等活动,增强学习者的学习动力。

## 7、高质量学习社群

本项目未来将构建中文学习社群,通过实时互动与在线讨论,促进学习者间相互交流与经验分享。社群将集结来自不同国家的学习者,旨在推进跨文化交流和中华文化传播。将采用机器审核和人工管理相结合,利用徽章、打卡、排行榜等激励手段,定期举办专题活动与汉语学习经验分享,营造积极的学习氛围。确保及时过滤不良信息,维护社群环境。

### （三） 国、内外研究现状和发展动态

#### 1、国内研究现状

2023 年 3 月 30 日,北京大学的刘晓雨副教授主讲的系列课程 Chinese for Beginners 虚拟现实(VR)体验版在 Coursera 平台全球首发。学习者通过 VR 耳机或桌面访问,能够如同身临其境地在中国生活场景中进行沉浸式体验,在获得实时反馈的同时,既能练习听力和口语技能,又能了解中国的社会文化。

北京大学的此课程做到了将国际中文教育与虚拟现实(VR)结合,打造了多模态的汉语教材资源。但是无法实时在线互动,也无法依据每一位学习者的特点及需求,提供个性化的学习策略和内容推荐。与本项目的智能汉语教学 AI 助手相比,缺乏灵活性。

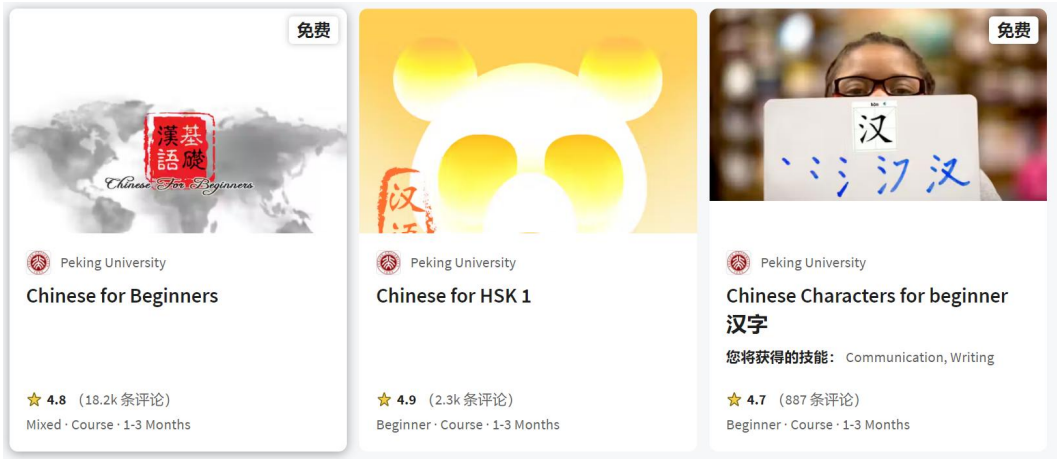


图 2 北京大学对外汉语教育学院的国际中文系列课程

#### 2、国外研究现状

美国密歇根州立大学研发了一种用于自动评估学生的汉语写作技能的系统,该系统能够分析学生的作文内容并提出个性化的反馈和建议。

但相较于本项目中采用的 GPT 技术,该系统的智能存在一定的局限性。此外,该系统的应用范围仅限于写作技能的评估,缺乏通用性。相比之下,本项目基于 LLM 的国际中文学习融媒体系统具有更强的多功能性和多模态教学能力,可为学生提供更全面的辅助。



### 3、发展动态

近年来,汉语学习人数不断增加,全球180多个国家和地区开展了中文教育,81个国家将中文纳入国民教育体系,国外正在学习中文的人数超过3000万。汉语教学朝多方面、多层次、多领域发展,人才培养多样化、专业化,国际中文教育将从单纯的汉语教学转变为多元化的教学模式、朝内涵式高质量方向发展。

2023年3月22日,全球教育智库HolonIQ发布了一项关于中文学习市场的报告,预计到2027年,全球中文学习市场体量将翻一番,以每年12.1%的复合年增长率,从74亿美元增长到131亿美元。目前中文已成为现在全球增长最快的语言学习市场,而中文学习需求也正在快速增长,尤其是来自约6,000万总人口的华裔群体中的K-12学生。

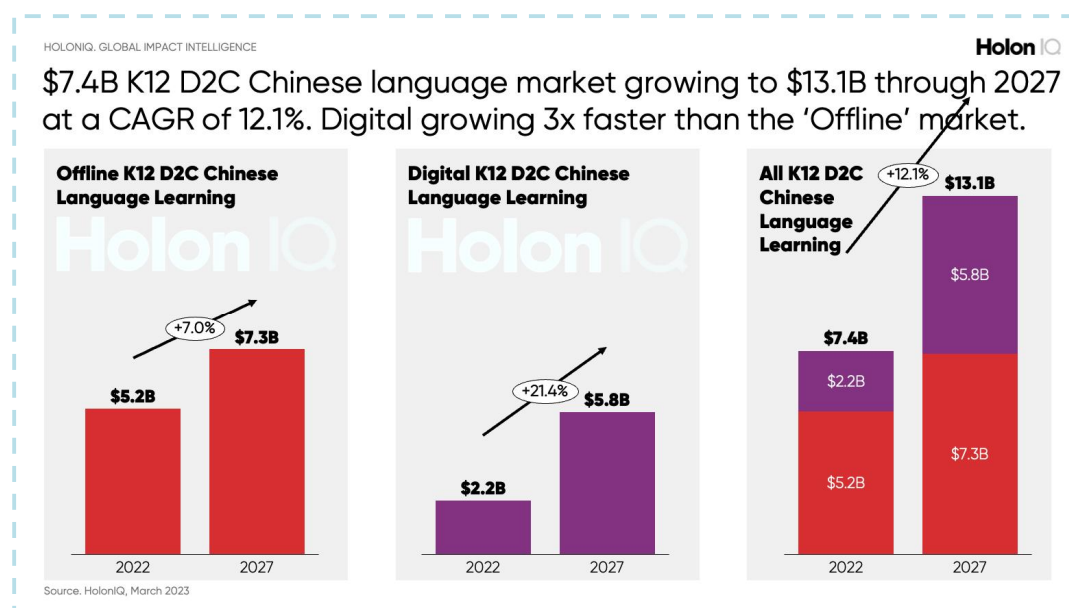


图3 HolonIQ关于中文学习市场的报告

自新冠疫情爆发以来,人工智能技术与教育领域的融合日趋加深,尤其在外语教学与英语学习领域,深度学习算法使得个性化学习与推荐成为可能,融媒体技术使得教材知识变得多模态化,受到广泛关注与运用。

国际中文教育领域也开始尝试将人工智能技术与融媒体技术相结合,但大语言模型在国际中文教育中的应用研究还比较浅显,拥有较大的发展空间。

#### (四) 创新点与项目特色

##### 特色与创新点一：基于大语言模型的智能化问答助手

本研究项目以GPT-3.5大语言模型和融媒体技术为基础,推动国际中文教育智能化和综合化发展。运用自然语言处理、深度学习等技术,实现智能化交互式的学习,学习者能够自主选择学习内容和方式。项目重点关注跨文化交流能力的培养,旨在继承中华优秀传统文化,弘扬时代精神,为构建新时代中国特色社会主义的国际中文学习体系做出贡献。

##### 特色与创新点二：基于融媒体平台的多模态教材呈现

本项目致力于构建基于融媒体的多模态教材呈现,采用图文、视频等多媒体形式,为学习者提供丰富的学习资源。项目将建立一个高质量的学习社群,通过实时的互动,促进学习者相互学习。该社群将汇聚来自不同国家和地区的学习者,有利于跨文化交流,经验分享,提高学习者的学习效果。

### 特色与创新点三：基于多语音模型的分场景口语对话

本项目融合多种语音模型,包括语音识别(ASR)、语音合成(TTS)以及端到端语音对话系统等,以实现多场景的口语对话。学习者可以通过与智能语音助手进行实时互动,提高口语表达能力,同时加深对语言和文化的理解。

### 特色与创新点四：基于自学习算法的智能化推荐策略

本项目通过对学习者的对话数据进行自学习分析,能够为学习者提供最符合其兴趣和需求的学习内容,实现高度智能化的推荐。基于自学习算法的推荐策略能够自适应地调整模型参数,以实现最佳的推荐效果,这有助于系统持续优化推荐算法,提高学习者的满意度。采用注意力机制的 LLM 模型,是技术创新和实用性的结合。

## (五) 技术路线、拟解决的问题及预期成果

### 1、技术路线

本项目未来将基于 GPT-3.5 等大语言模型,使用中华语料库训练,采用 pgvector 存储嵌入向量,提高模型的语言理解和生成能力,以支持智能化的汉语教学。

未来的大语言模型训练过程将主要包括以下步骤:

#### (1) 数据处理

未来将对大量无标签的中文文本进行预处理,包括分词、词性标注、去除特殊符号等。将处理后的文本转换为模型输入所需的 Token ID 序列。

#### (2) 构建计算图

未来会通过实例化 GPT-3.5 模型,并为其指定超参数(如隐藏层大小、注意力头数等)。然后将 Token ID 序列输入模型,得到预测输出。

#### (3) 训练策略

将采用梯度累积(Gradient Accumulation)和混合精度训练(Mixed Precision Training),以降低显存消耗并提高训练速度。使用学习率预热(Learning Rate Warmup)或余弦退火(Cosine Annealing)策略调整学习率,促进模型收敛。

#### (4) 优化器选择

未来将使用 AdamW 优化器进行权重更新。AdamW 是一种自适应学习率的优化器,结合权重衰减(Weight Decay)技术,可以有效地防止过拟合。

#### (5) 损失函数

未来将采用交叉熵损失(Cross-Entropy Loss)计算预测输出与真实标签之间的误差,并通过反向传播(Backpropagation)更新模型参数。

对于 AI 交互口语功能,我们未来将设计三个模块库。采用跨平台的 PortAudio 库实现音频录制和播放功能,为后续音频处理提供基础;GPT 和 ASR 技术库,实现智能聊天和音频转文字服务,保障语音助手的语音识别能力;基于 TensorFlow 实现的 TensorFlowTTS 库,完成文本到语音的高质量转换,模拟真人对话的效果。

### 2、拟解决的问题

针对课后缺乏互动的问题,本项目未来将设计并实现一款网页版的中文教学 AI 助手,通过高效生成中文词汇释义、学习提示等功能,实现 24 小时不间断的高质量答疑服务。

针对缺乏口语训练的问题,本项目未来将融合多种语音识别模型,实时分析对话语义并生成高质量回复,实现 AI 口语的纠正、练习、教学;

针对课堂形式单一的问题,本项目未来将通过 YouTube 等融媒体平台发布图文和视频教程,实现多模态教学,并通过引入动画和传统文化故事来增加学习的趣味性;

针对教材枯燥乏味的问题,本项目未来将融合 HSK 词汇与中华文化典籍构建语料库。并对语料进行分难度、分场景的数据标注,实现多模态内容的智能化推荐;

### 3、预期成果

- (1)研发出满足项目需求的国际中文学习融媒体系统；
- (2)训练出高精度的智能化推荐模型和口语对话模型；
- (3)培养成员的创新、跨文化交流和团队协作的能力；
- (4)申请软件著作权；
- (5)发表 1 至 2 篇相关领域论文。

## （六）项目研究进度安排

### 2023 年 6 月—2023 年 7 月：系统可行性分析及相关技术准备

综合设计方案进一步分析“国际中文学习融媒体系统”的可行性；掌握实现该系统所需技术与技能，主要包括网页前后端开发、Python 数据处理和挖掘等技术、Tensorflow 深度学习框架、LLM 模型训练技术。

### 2023 年 8 月—2023 年 10 月：系统需求分析及中国文化多模态语料库构建

对系统的功能需求进一步细化；获取构建中国传统文化多模态语料库所需的数据，前期主要以 word、excel 等格式保存，设计对此类文档的自动化数据清洗算法，设计自动爬取中华优秀传统文化典籍内容的爬虫工具。

### 2023 年 11 月—2024 年 2 月：国际中文学习融媒体系统的设计与实现

实现“国际中文学习融媒体系统”，具体包括大语言模型的训练与软件开发环境的搭建，基于深度学习的语法和语音模型与智能化推荐算法的设计、实现与训练，系统软件架构设计，软件编码和测试，以及阶段评测。

### 2024 年 3 月—2024 年 4 月：系统整合、优化与最终测试

该阶段对上一阶段初步实现的“国际中文学习融媒体系统”的各个子模块进行最后的整合、优化和测试。

### 2024 年 5 月—2024 年 6 月：项目总结

整理编写相关文档，发布完整软件，准备结题答辩。

## （七）已有基础

### 1、与本项目有关的研究积累和已取得的成绩

参加本次创新训练项目的成员，分别来自不同专业领域，实现了大数据、人工智能、广告学、国际中文教育的学科交叉，未来将打造基于 LLM 的国际中文学习融媒体系统。我们掌握了各自专业的基础知识，熟悉了开发一个软件应有的步骤，以及需要准备的各种信息，并已在科研竞赛中取得了许多优异的成绩。

本项目组各位成员积极进取，乐于创新，敢于直面难题，并认真思考解决方法，在学习方面，成绩优良，均积极参加学校组织的活动，具备了较高的理论知识水平，并有专业知识丰富的老师指导我们，掌握本研究项目所需的技术和方法。

指导老师 1 的科研方向为自然语言处理和数据挖掘，有丰富的 NLP 领域的学术经验，可以指导学生迅速提高自然语言处理方面的能力。

指导老师 2 的科研方向为语言学和翻译传播学，有丰富的国际中文教学研究经验和融媒体传播项目经验，可以指导学生正确充分地将国际中文学习与融媒体结合。

综上所述，项目成员不仅有着扎实的专业基础知识，同时也具备优秀的学习能力，对所开展的项目都持有积极的兴趣和创新的想想法，并针对项目进行了初步的实践工作，能够在项目成员不懈的努力和老师的指导下圆满完成项目。

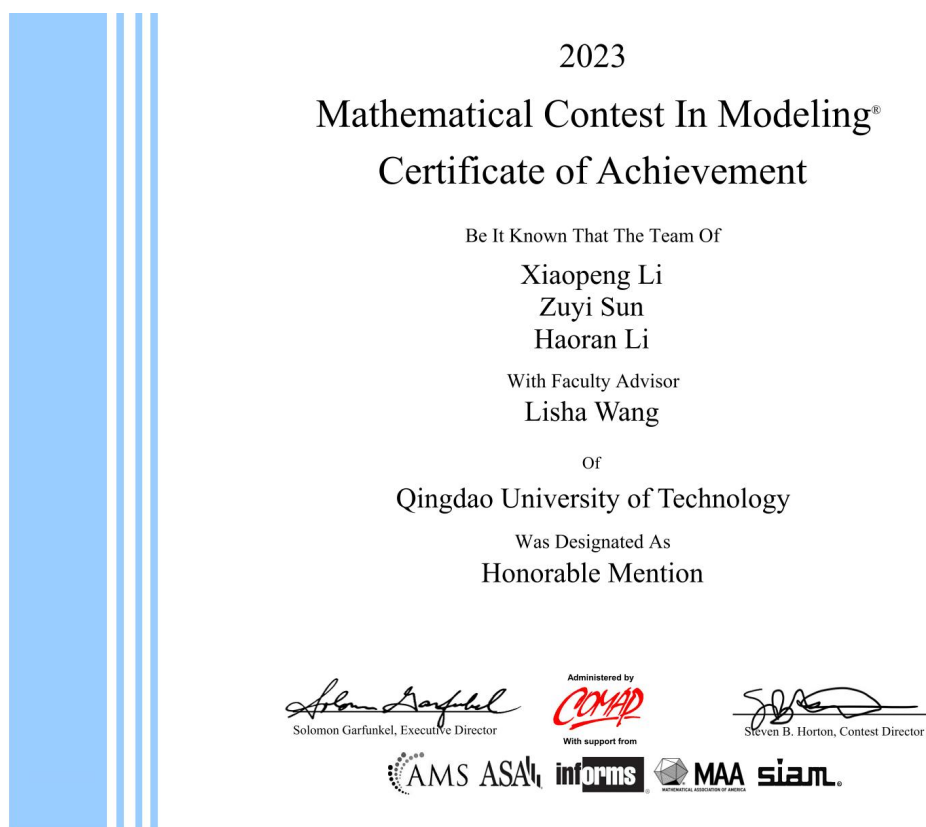


图 4 美国大学生数学建模竞赛 MCM Honorable mention



图 5 全国大学生数学建模竞赛省级一等奖

证明:

域名 **hskhelper.top** 已由 **Xiao Peng Li** 注册，并已在国际顶级域名数据库中记录。

域名(Domain Name) :	hskhelper.top
域名注册人(Registrant,中文) :	李晓鹏
域名注册人(Registrant,English) :	Xiao Peng Li
注册时间(Registration Date) :	2023-04-20 18:29:40
到期时间(Expiration Date) :	2033-04-20 18:29:40

图 6 国际顶级域名 HSKhelper. top

## 2、已具备的条件, 尚缺少条件及解决方法

已初步完成基于 LLM 的中文学习助手 HSKhelper 的网页前后端设计。

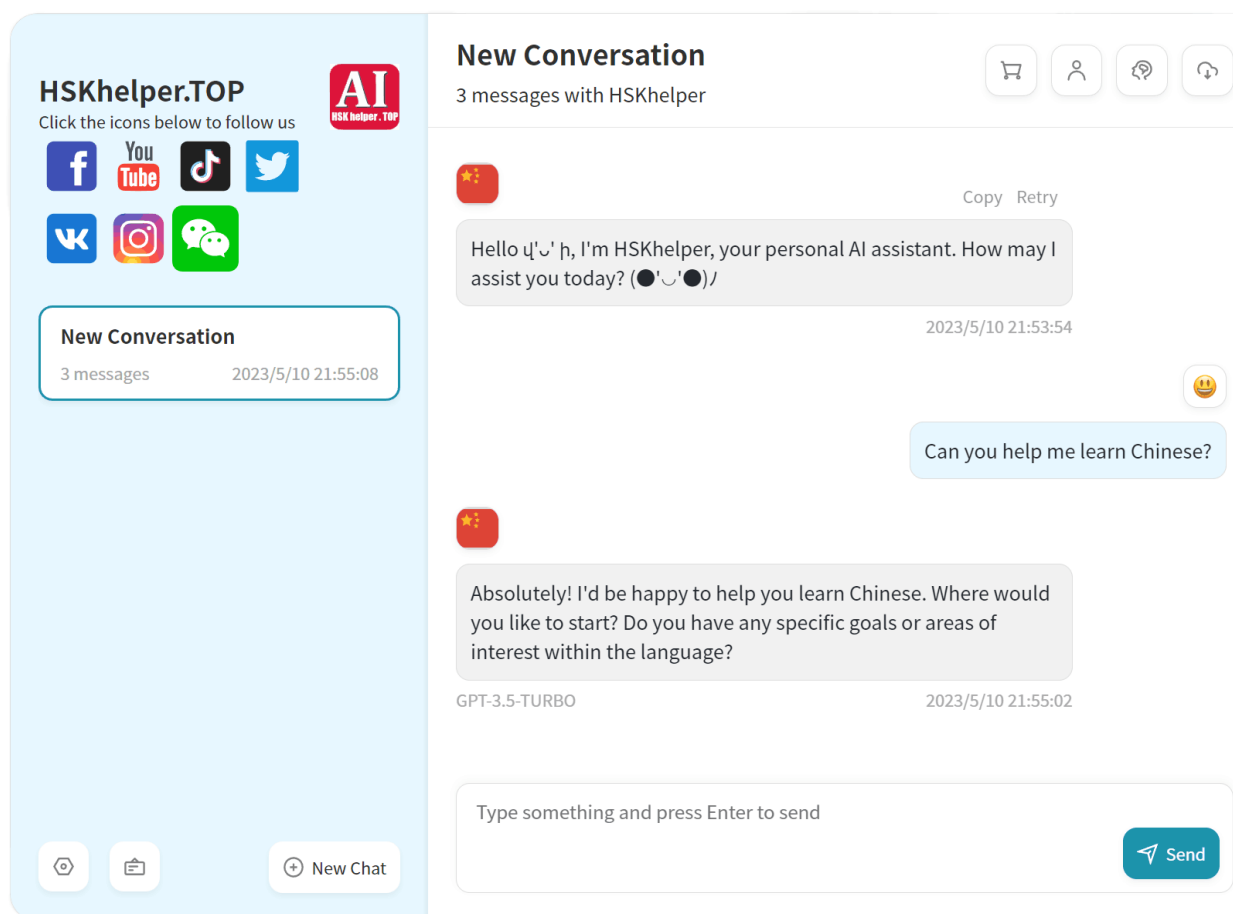


图 7 HSKhelper 前端网页设计

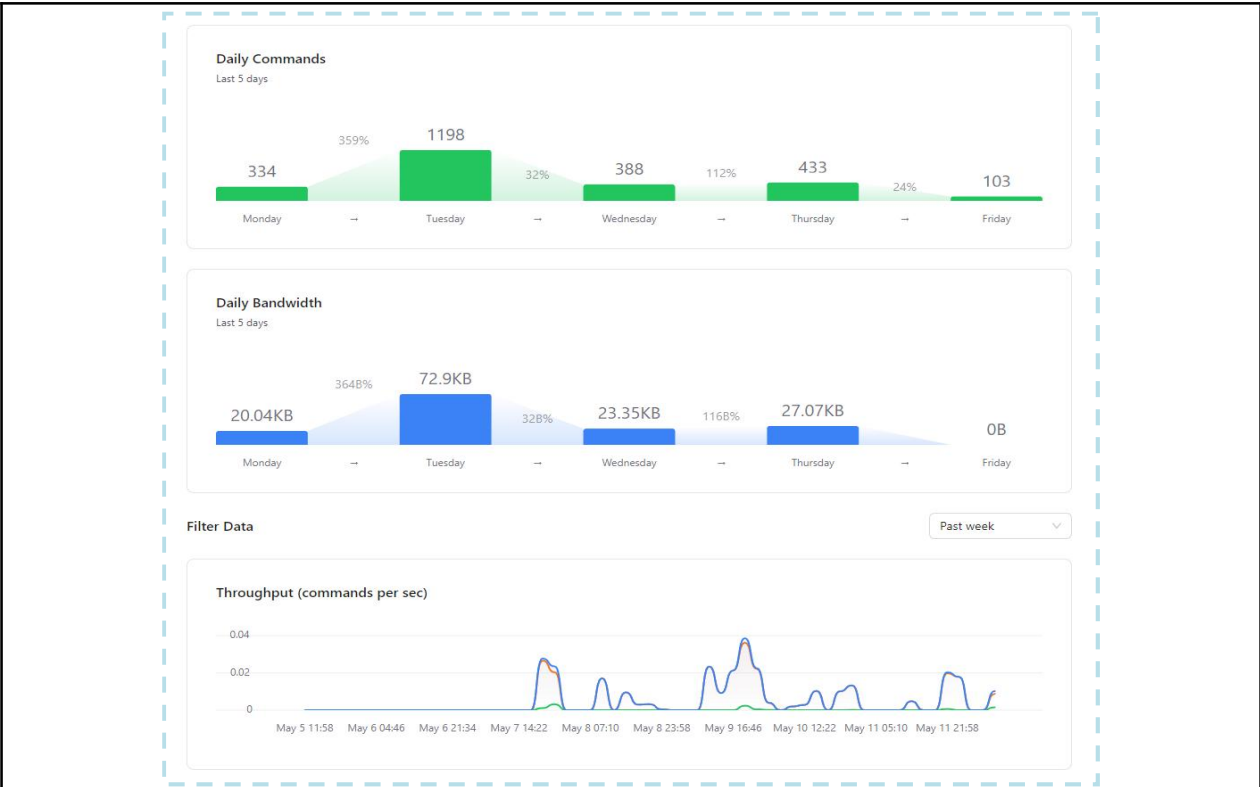


图 8 HSKhelper 云数据库的可视化面板

尚缺少的条件：  
未购买阿里云服务器，Cloudbase 内容审核。

### 三、 经费预算

开支科目	预算经费(元)	主要用途	阶段下达经费计划(元)	
			前半阶段	后半阶段
预算经费总额	10000	网站运行和教学成本	5200	4800
1. 业务费	2500	国际中文教师证考试	1500	1000
(1)计算、分析、测试费	1000	内容审核服务费	600	400
(2)能源动力费	0		0	0
(3)会议、差旅费	1000	会议费	1000	0
(4)文献检索费	0		0	0
(5)论文出版费	3000	论文版面费	0	3000
2. 仪器设备购置费	0		0	0
3. 实验装置试制费	1500	购买云服务器	1500	0
4. 材料费	1000	教学书籍费用	600	400
学校批准经费				



#### 四、 指导教师意见

该项目具有较高的研究价值和应用价值,有广阔的开拓空间,不仅有助于项目组学生自身组织、策划能力的培养,其建设成果将对本院学生软件的开发能力的培养有极大的帮助。建成后可以为国际中文的学习者助一臂之力。具有一定的应用价值。

综上所述,该项目有良好的发展前景,拟同意该项目的申报。

导师(签章): 房斐斐 2023年5月26日

2023年5月26日

#### 五、 院系意见

同意

学院(盖章):



2023年5月28日

#### 六、 学校意见

同意立项

学校(盖章):



2023年6月7日