

# 项目概要介绍

## 一、前言

近年来，AIGC 技术蓬勃发展，带来全新的生产力革新并正在影响教育生态。如何正确使用 AIGC 技术助力教学是一个亟待解决的问题。通慧智教使用精调 LLM、构建高质量数据集、打破传统交互方式并采用先进算法助力个性化，解决当下教育资源同质化严重、教学过程缺少互动、个性化学习内容少以及学习效果反馈弱的问题。

## 二、创意描述

通慧智教利用多维数据构建学习者画像，并使用精调 LLM 生成个性化学习报告，以及辅助用户理解题目和生成相似例题。数据来源包括分布式爬虫、IDP 和开源数据集，经过清洗得到高质量的数据。通过虚拟教学助理-小慧和 VR 课堂，结合 STT、LLM、TTS 技术和开发工具，提供互动与反馈。我们采用精调方法对本地 LLM 进行优化，并根据 Rouge 得分评估 LLM 性能。

## 三、功能简介

通慧智教可分为两个层面：

### 1. 网页端：

主要有以下五个功能模块：

虚拟教学助理：使用自然语言交互的方式控制平台全部功能。

相似题目生成：使用本地精调大语言模型对用户提供的题目进行相似题目的生成。

学习报告生成：使用多维度构建的学习者画像和商用大语言模型生成用户学习报告，助力用户学习把控和科学教育。

题目在线评估：使用商用大语言模型对用户作答题目进行在线评估，增强学习反馈，提高学习效果。

智能题目推荐：使用商用大语言模型结合高效混合过滤算法智能进行题目的推荐，助力用户攻克学习薄弱点，提高学习效率。

### 2. VR 端：

将网页端实现的全部功能迁移至 VR 中，并且虚拟教学助理置于身旁，协助用户学习。

## 四、特色综述

## 4.1 大语言模型数据：

### 4.1.1 数据获取方式多样

数据来源包括使用 scrapy-redis 进行的分布式爬虫获取，以及多源智能文档处理和协作平台，采集方式多样化，能够有效整合和处理大量信息。

### 4.1.2 数据预处理高效精确

数据预处理通过 LDA 主题分析进行内容分类，并构建多样化的 FAQ 问答对，以提高处理效率和精确性。

## 4.2 个性化：

### 4.2.1 多维用户画像构建

通过分析学生的学习行为、成绩历史和兴趣偏好等多维数据，构建详细的学习者画像，进一步利用大型模型技术生成针对性的个性化学习报告。

### 4.2.2 混合过滤算法推荐例题

结合混合过滤算法，系统能够推荐与学生薄弱领域相关的精选例题，旨在提高学习效率和成效。

### 4.2.3 大模型辅助理解题目与答案评估

通过深度理解题目要求和答案内容，模型能够提供准确的解释和评估，帮助学生更好地理解问题本质、掌握解题方法，并提供针对性的反馈和建议。

## 4.3 交互方式：

### 4.3.1 虚拟教学助理

通过整合 Unity、Blender 和 VRoid 技术构建了虚拟教学助理“小慧”，该助理融合了语音转文本（STT）、大语言模型（LLM）、文本转语音（TTS）等先进技术，创新性地实现了自然语言的交互式教学平台。

4.3.2 VR 虚拟现实

VR 虚拟现实技术提供了沉浸式学习体验，通过全方位互动增强学习效果和参与感。

4.3.3 本地 GPT-SoVITS 大模型

部署在本地的 GPT-SoVITS 大模型能够实现快速的语音合成，以实现更自然、个性化的交互体验。

4.4 训练大模型：

4.4.1 本地大模型

使用了多种大模型在不同精调方式下（FT 、GaLore 、LoRA、Qlora）训练，完成相似例题生成这一功能。

4.4.2 商用大模型

训练精调后的商业大模型能够评估作业、制定学习计划、进行对话交互、控制虚拟人物及网站路由，实现教育领域的多方面应用。

五、开发工具与技术

前端	Vue3、WebGL
算法	LDA、DACF、STT、TTS
后端	Django、Ninja
数据库	MySQL、Redis
建模工具	VRoid、Blender、Unity
虚拟设备	Unity VR、Steam VR
大模型	百度智能云、Mistral-7B

六、应用对象

网页版课堂、虚拟 VR 课堂使用对象支持学生，老师，系统管理员等多种角色。

七、应用环境

GPU	NVIDIA GeForce GTX 1650 Ti 及以上
CPU	i5-13600KF、R5 7600X 及以上
数据库版本	MySQL5.7、Redis
JDK 版本	1.8
PC 端	Win10 及以上

八、结语

通慧智教巧妙地融合了大语言模型和教育,通过精细的 AI 算法和个性化学习方式,有效提升了教学效果和学习体验。借助 AIGC 技术实现教学个性化、互动性和反馈的全面提升,为教育培训注入新的活力和创新。