AI智能·学习搭子

解决方案文档



**发包方：数字马力**

**承接方：声像科技**

**日期：2025年10月**

**项目背景**

* **前言**

随着人工智能技术与教育领域的深度融合，学习场景正经历深刻变革。在信息爆炸时代，知识更新速度极快，学习者面临海量信息筛选与高效吸收的难题。传统“千人一面”的学习模式难以匹配个体学习节奏与偏好，学习过程中存在的孤独感、缺乏即时反馈与个性化指导等问题，也严重阻碍了学习者的进步。在此形势下，亟需借助人工智能技术，为学习者提供更贴合需求的学习支持。

* **创意描述**

随着人工智能技术的快速发展，教育领域对智能化教学工具的需求日益增长，而传统语音合成（TTS）服务在教育场景中则难以满足个性化教学和多元化场景的需求。同时，两会期间，关于“换脸拟声”技术的提案也对信息安全提出了严峻挑战。在此背景下，声像科技团队以“守正创新”为核心理念，推出“基于AI语音合成的教学声音处理软件”，旨在通过多模态内容分析、实时数字人交互和深度伪造检测技术，构建起一套安全、智能、高效的教学声音处理解决方案。融合语音、图像、视频的多模态感知能力，为教育行业提供从内容生成到安全鉴伪的全产业链服务，助力教学方式的数字化转型。

本项目立足传统语音合成，提出三大拓展创新方向：

1. **多模态内容分析与生成：**通过音视频内容感知系统，从语音、图片、视频中自动提取语义信息，生成符合教学场景的合成内容，解决传统TTS依赖人工输入文字的痛点；
2. **数字人实时推流与交互：**项目构建数字人与拟音系统，将语音TTS服务升级为音视频数字人形式；支持自然语言控制情感、口音、节奏，并通过WebRTC实现实时推流，打造沉浸式教学体验；
3. **换脸拟声检测与伦理保障：**结合自主研发的多模态声纹特征提取与筛选神经网络、离散余弦引导与动态特征选择等技术，鉴别合成音视频的真伪，确保教学合成内容的安全合规且可控，响应国家对AI伦理的监管要求。

* **功能简介**

项目包含三大核心系统：

1. **谛听·音视频内容感知系统**
2. **语音情感事件转换**

支持语音转写、语言识别、情感识别和事件检测功能

可借由语音分析学生课堂情绪状态及教师授课风格

1. **图像视频内容分析**

采用视觉大模型与混合注意力特征融合技术

解读图像、视频中的内容表征，更可识别文档类图像的篡改痕迹

1. **摩耶·数字人与拟音TTS系统**
2. **自然语言可控的语音合成**

支持零样本语音克隆、多语言语种支持

以自然语言进行情感、口音、节奏、停顿、语调的精细控制

1. **实时数字人音视频推流**

借由WebRTC技术，实现端到端的实时推流

高分辨率唇形同步框架，实现在消费级显卡上的实时渲染能力

1. **白泽·深度伪造内容鉴别系统**
2. **合成语音检测**

通过多模态声纹提取网络与知识蒸馏技术

有效捕捉合成伪造音频的细微特征，应用于伪造合成声音的鉴别

1. **人脸伪造检测**

基于离散余弦变换与动态特征选择框架

对操纵线索进行解耦度量学习，有效识别监考检测中的伪造换脸

* **特色综述**
* **多模态内容融合与生成技术**

整合语音、图像、视频的感知与生成能力，突破单一模态的局限性

增广内容的获取形式，达到备课授课过程的降本增效

* **自然语言可控的零样本情感语音克隆**

低延迟流式响应，支持零样本、多语言支持的语音克隆

结合文本基座大模型，控制合成语音的情感、口音、节奏、停顿和语调

* **实时流式数字人音视频推流**

在消费级显卡上的实时端到端多并发调用，支持多语言音频驱动

高保真面部唇形驱动模块，具有序列化批处理加速能力

* **合成伪造音视频鉴别能力**

基于混合专家系统，对使用深度伪造及合成音频、视频进行鉴别

应对在线监考等身份验证需要，同时保障合成内容的安全合规

* **开发工具与技术**
* **编程语言与框架：**

Python（编程语言，基于3.10.16版本）

PyTorch（人工智能框架，基于v2.3.0+cu118版本）

FastAPI（服务器与算力协同）

QFluentWidgets（前端界面，基于PyQt6绑定）

WebRTC（信息流式传输）

* **采用的模型与算法：**

[SenseVoice](https://github.com/FunAudioLLM/SenseVoice)（语音内容识别）

[Qwen2.5-VL](https://github.com/QwenLM/Qwen2.5-VL)（图像、视频内容识别）

[CosyVoice](https://github.com/FunAudioLLM/CosyVoice)、[GPT-SoVits](https://github.com/RVC-Boss/GPT-SoVITS/)（语音合成）

[MuseTalk](https://github.com/TMElyralab/MuseTalk)（音频唇形同步）

* **项目知识产权：**

3项发明专利（人脸伪造检测、图像篡改识别、跨模态篡改检测）

1项软件著作权（实时消息通讯）

* **应用对象**
* **教育机构**：

中小学、高校、职业院校，用于在线教育教学、校园心理驿站等

* **教育科技企业：**

在线教育平台（如猿辅导）、智能硬件厂商（学习机、机器人）等

* **公共服务部门：**

教育监管部门（在线监考身份验证）、展览馆（有声资源生成与互动）等

* **应用环境**
* **服务端部署环境**

Windows/Linux操作系统

带宽≥4Mbps

安装有英伟达品牌显卡，显存≥16GB

内存≥16GB

储存≥40GB

* **客户端应用环境**

具备现代浏览器环境（Edge、Chrome、Safari等）

带宽≥4Mbps

内存≥2GB

* **结语**

本项目通过技术创新与场景深耕，重新定义AI语音合成在教育领域的应用边界。从多模态内容生成到数字人交互，再到深度伪造检测，构建了覆盖教学全周期的智能化上下游解决方案。未来，团队还将继续以“守正创新”为主要导向，为教育等诸多行业的数字化转型提供安全、可靠的技术支撑与保障！

以下为项目解决方案正文部分

**目录**

[**项目背景概要** 1](#_Toc195526137)

[**教育行业背景** 1](#_Toc195526138)

[**社会背景** 1](#_Toc195526139)

[**项目目标** 2](#_Toc195526140)

[**团队调研** 3](#_Toc195526141)

[**问题分析与解决方案** 4](#_Toc195526142)

[**国内外研究现状** 4](#_Toc195526143)

[**用户角色与需求分析** 5](#_Toc195526144)

[**项目解决方案** 7](#_Toc195526145)

[**系统功能与实现方案** 8](#_Toc195526146)

[**项目技术路线基础：数据增广** 8](#_Toc195526147)

[**谛听·音视频内容感知系统** 10](#_Toc195526148)

[**摩耶·数字人与拟音TTS系统** 12](#_Toc195526149)

[**白泽·深度伪造内容鉴别系统** 15](#_Toc195526150)

[**项目管理与可行性分析** 19](#_Toc195526151)

[**项目人员结构** 19](#_Toc195526152)

[**可行性分析** 19](#_Toc195526153)

[**软件设计与实现** 22](#_Toc195526154)

[**软件体系结构与设计规范** 22](#_Toc195526155)

[**软件前端设计** 22](#_Toc195526156)

[**软件后端架构** 23](#_Toc195526157)

[**市场分析与商业模式** 24](#_Toc195526158)

[**项目市场分析** 24](#_Toc195526159)

[**项目商业模式** 25](#_Toc195526160)

[**参考文献** 26](#_Toc195526161)

[**项目知识产权** i](#_Toc195526162)

[**国内发明专利（3项）** i](#_Toc195526163)

[**软件著作权（1项）** iv](#_Toc195526164)

**解决方案须详细标注每部分的负责人，并阐述参赛团队内各选手在其中的角色及作用。应该至少包括以下四部分内容：**

**S2A：目标与服务模型(描述赛题的价值与解决思路)；**

**针对赛题文档和企业文档进行分析**

**引出四个解决方案的重点，结合调研实际背景**

* **数字人伴学系统**

**数字办**

* **多智能体-知识库协作系统**

**温州中学**

* **多学科客制化评估系统**

**学院、OJ平台等**

* **游戏化学习平台及评估系统**

**对应东语学院需求**

**S2B：组织管理与业务分析方案(主要由项目经理负责)；**

**甘特图显示成员写作和时间分配**

**Github显示代码写作**

**JetBrains-Code-with-me显示代码写作**

**针对提出的四个方案的业务成本、开发分析（体现经济效益成本性、加上应用环境配置分析）**

**展示项目调研合作（联通、网信办、公安厅、云栖大会照片等）**

**S2C：技术路线及实现方案(主要由技术经理负责)；**

* **数字人伴学系统**
  + **采用阿里巴巴通义实验室**[**OpenAvatarChat**](https://github.com/HumanAIGC-Engineering/OpenAvatarChat)**项目提供的数字人工作流**
  + **于校内数字办平台进行系统部署**
  + **取得相关软件著作权授权：《终身学伴-数字虚拟人合成平台》**
* **多智能体-知识库协作系统**
  + **采用蚂蚁集团**[**agentUniverse**](https://github.com/agentuniverse-ai/agentUniverse)**项目提供的多智能体框架**
  + **与一所省内重点高中、一所市属小学展开合作试用**
  + **于校内数字办平台进行系统部署试用**
* **多学科客制化评估系统**
  + **采用蚂蚁集团**[**OpenSumi**](https://github.com/opensumi/codeblitz)**项目提供的IDE基础框架**
  + **与一所省内重点高中展开合作试用**
  + **于计算机系平台进行系统部署试用**
* **游戏化学习平台及评估系统**
  + **采用蚂蚁集团**[**Ant Design**](https://4x-ant-design.antgroup.com/index-cn)**项目提供的前端设计框架**
  + **与一所市属小学展开合作试用**
  + **于校内日语系平台进行系统部署**

**S2D：成本模型及可行性分析(主要由客户关系经理负责)。**