****

**技术报告**

基于MindSpore个性化大语言模型家庭教师应用开发

参赛队名： 创出未来

队 长： 冼嘉荣

联系方式： 17324521020

组员名单： 袁基保、陈丞锋

指导老师： 刘寿强

参赛单位： 华南师范大学人工智能学院

作品类别： 开源任务贡献

作品链接： [https://www.gitlink.org.cn/tmlwacre/tutor](https://www.gitlink.org.cn/tmlwacre/fraudKG)

**个性化大语言模型家庭教师应用开发**

**摘要**

在当今快速发展的时代背景下，世界各地间的文化科技经济交流也愈来愈增强，语言沟通的重要性在这时也比任何时候都更加重要，它能够促进经济全球化、文化交流，推动科技创新和教育合作，同时也为个人提供更广阔的人生体验和机遇。但学习外语具有较长的学习的曲线，如何科学、高效率、低成本的掌握外语至今仍是一个难题。

针对当前人工智能技术对个性化学习规划、学习辅导研究较少的现状，本项目在人工智能技术的基础上提出了一种基于人工智能技术的定制化家教系统。该系统利用先进的语音识别模型，使学生能够通过语音输入与系统进行互动，给出学习规划路线安排和语音反馈。本项目使用达摩院语音团队提出的非自回归端到端语音识别框架Paraformer进行语音识别并输入，运用大语言模型来进行推理，使用思维骨架进行长期推理和规划安排，接入微软TTS语音引擎，进行输出，实现对话。使该技术应用于家庭教师上更具可行性、合理性。

本研究主要创新点有多模态、服务广泛等，该定制化家教系统通过人工智能技术、个性化学习规划、向量记忆库和文字转语音模型等方面的创新，提供全面、个性化和互动性强的学习体验，具有较大的应用前景，适应当今快速发展的时代需求。据行业权威咨询公司贝哲斯对英语学习行业市场数据的统计显示，2022年全球与中国英语学习市场容量分别为3555.35亿元（人民币）与1560.09亿元。预计全球英语学习市场规模在预测期将以10.49%的CAGR增长并预估在2028年达6392.32亿元。由此可见，该市场未来前景巨大。在有效降低学习难度的同时，进入稳定的千亿市值市场，具有较强的投资性、盈利性，也可满足广大的科普教育需求，具有深刻的社会意义。

**关键词：**多模态、思维骨架、向量记忆库

目录

[摘要 2](#_Toc29459)

[1 设计方案 4](#_Toc21586)

[1.1 需求分析 4](#_Toc29293)

[1.1.1研究背景 4](#_Toc25051)

[1.1.2 市场分析和经济效益分析 4](#_Toc32639)

[1.2 国内外研究现状 5](#_Toc18232)

[1.3 设计目标 6](#_Toc8400)

[2 实现方案 7](#_Toc2484)

[2.1基本思路 7](#_Toc5682)

[2.2 MindsporeAI开发框架 8](#_Toc25254)

[2.3大语言模型推理 8](#_Toc4400)

[2.4思维骨架进行长期推理和规划安排(SOT) 9](#_Toc19670)

[2.5向量数据库 10](#_Toc31560)

[2.6语音识别模型 10](#_Toc26166)

[2.7 文字转语音模型 11](#_Toc13134)

[2.8 Gradio界面设计 12](#_Toc4174)

[2.9 ModelArt数据平台 12](#_Toc8871)

[3特色创新 13](#_Toc19714)

[3.1创新点 13](#_Toc20845)

[4 总结与展望 13](#_Toc675)

[5 文献参考 14](#_Toc6347)

# 1 设计方案

## 1.1 需求分析

**1.1.1 研究背景**

随着社会和科技的不断进步，世界各地间的文化科技经济交流也愈来愈增强，掌握一门外语，对促进经济全球化、文化交流，推动科技创新和教育合作起着十分重要的作用。近年来，随着国民素质的提高，大家都纷纷开始学习外语，第二语言掌握程度的比例也在逐年提高。然而，学习外语仍面临着成本高、学习曲线长、缺失行之有效的定制化路线等问题。该项目利用人工智能技术，结合大语言模型，构造一个可以给学生定制学习规划的外语家教，在人工智能服务领域方面具有重大价值。

**1.1.2 市场分析和经济效益分析**

从我国英语学习市场规模来看，2023年预计达1072亿元，2024年预计达1196亿元，可见人工智能服务领域在该方面具有重大价值，增速较高且潜力巨大。



图1 市场预测图

## 1.2 国内外研究现状

国外的学者早在2017年的时候就发现，所有的学生通过每周固定的阅读课和听力课，无论是教学形式还是教学内容，都无法满足不同听说水平的学生的需求。这会导致学生们的学习外语的积极性下降，成绩差的同学产生厌学的情绪。同时也打击了教师教学的积极性。

但是随着人工智能技术的发展，大语言模型可以通过每一位同学的学习数据来指定符合每位同学的学习计划，并且指定相应的测试，从而有效提高学生的听说能力和写作能力。

而在实际使用中，根据2022人工智能教育白皮书来看，教师整体也认可人工智能教学工具的价值，但受阻于相关产品的不成熟和系统培训的缺乏。受访教师对人工智能在教学中的应用较为满意，认为在教学中应用人工智能技术可以提高他们的教学自信心，但同时也反映在操作人工智能产品中遇到困难的问题。教师认为最阻碍在教学中有效应用人工智能技术的因素是缺乏成熟的智能产品以及配套的资源与服务。

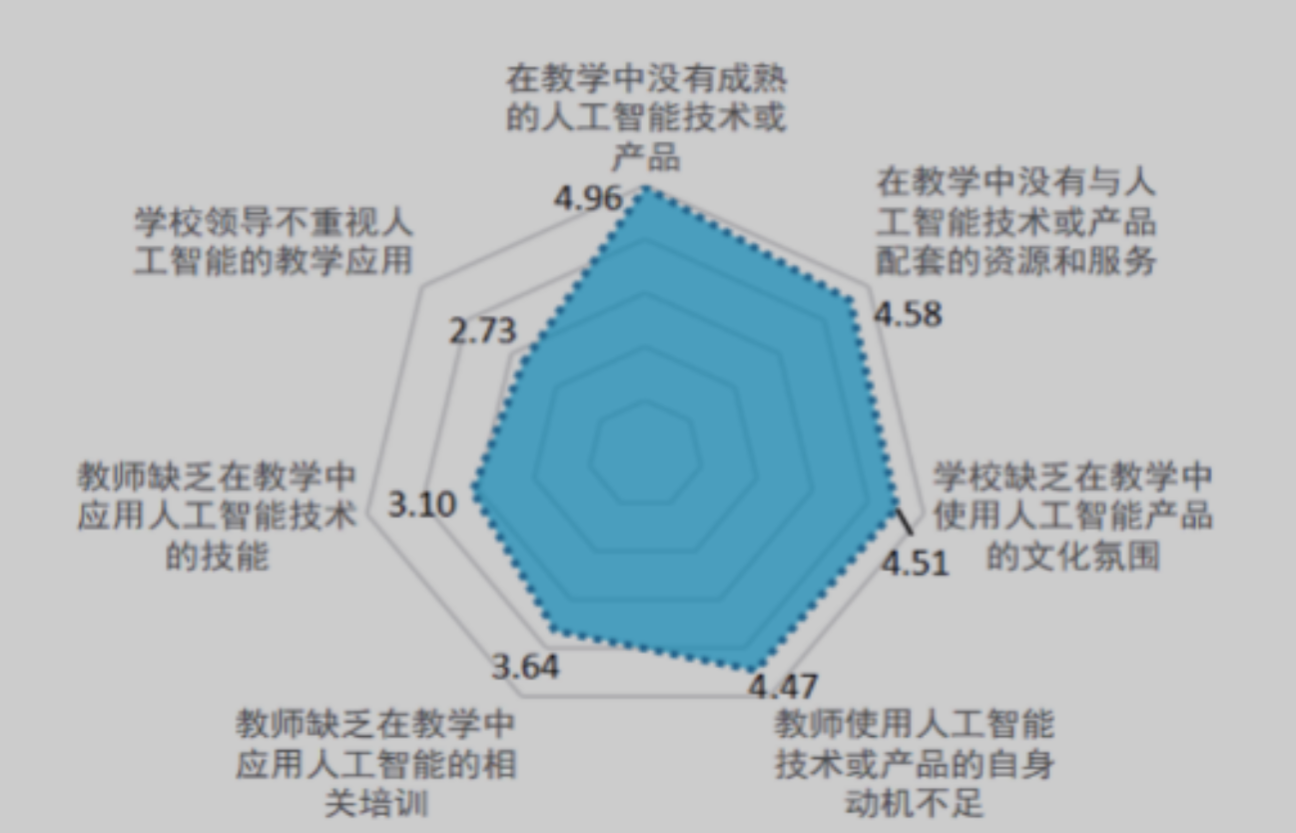


图2 教育白皮书

学生普遍愿意使用人工智能学习工具，希望获得个性化评估与辅导。大部分受访学生都对人工智能教育持有积极的态度，愿意使用人工智能学习工具，并认为使用人工智能技术会促进学习。仅有不到一半的受访学生使用过人工智能学习工具辅助学习，使用过的学生则都能够利用人工智能学习工具自主获取所需要的学习资源，实现学习方式的多样化。

## 1.3 设计目标

针对学生学习外语所缺少系统性规划的问题，本项目通过结合语音识别、大语言模型、思维骨架、向量记忆库和文字转语音模型等技术，该定制化家教系统能够为学生提供高效、个性化的外语学习支持。学生可以通过与系统的对话，不仅提高语言表达和听力能力，还能获得针对个人需求的学习计划和指导。该系统的独特功能和灵活性使其成为一种创新的学习工具，有望在提升学生外语学习效果和学习体验方面发挥积极作用。为**老师**提供学习规划参考，为**学生**提供优秀的外语练习服务。

# 2 实现方案

## 2.1基本思路

针对学生学习外语所缺少系统性规划的问题，本项目通过结合语音识别、大语言模型、思维骨架、向量记忆库和文字转语音模型等技术，该定制化家教系统能够为学生提供高效、个性化的外语学习支持。学生可以通过与系统的对话，不仅提高语言表达和听力能力，还能获得针对个人需求的学习计划和指导。该系统的独特功能和灵活性使其成为一种创新的学习工具，有望在提升学生外语学习效果和学习体验方面发挥积极作用。为老师提供学习规划参考，为学生提供优秀的外语练习服务。

本项目使用新概念英语教材进行语料素材积累，使用思维骨架进行长期推理和规划安排，使用向量记忆库来实现学习记录，使用语音识别模型进行输入，文字转语音模型进行输出，实现对话。



图3 解决方案

整体框架设计如下：

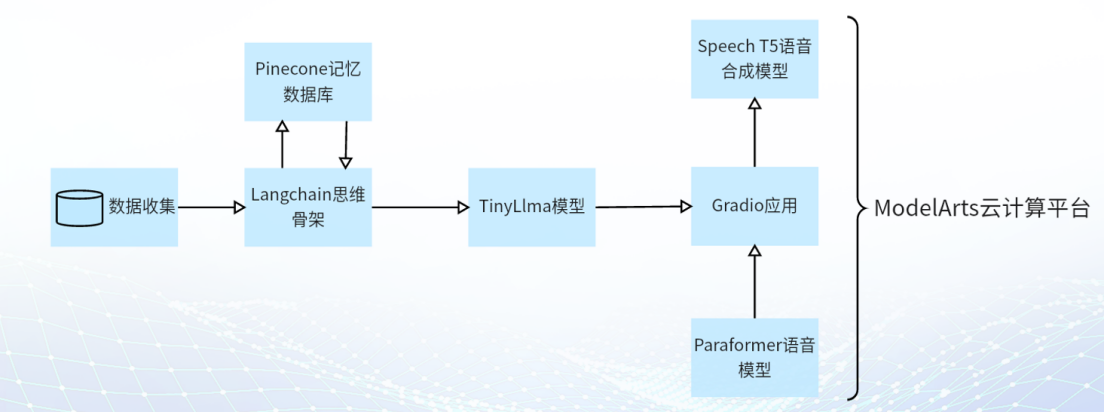


图4 流程图

## 2.2 MindsporeAI开发框架

MindSpore是一个全场景深度学习框架，旨在实现易开发、高效执行、全场景覆盖三大目标，其中易开发表现为API友好、调试难度低，高效执行包括计算效率、数据预处理效率和分布式训练效率，全场景则指框架同时支持云、边缘以及端侧场景。

该项目基于MindSpore框架进行搭建，使用MindSpore开发高性能算子。华为MindSpore框架技术先进可行具有良好的研发应用生态。

2.3大语言模型推理大语言模型是实现AI私人教师的基础，我们使用ModelScope社区提供的GPT-3模型

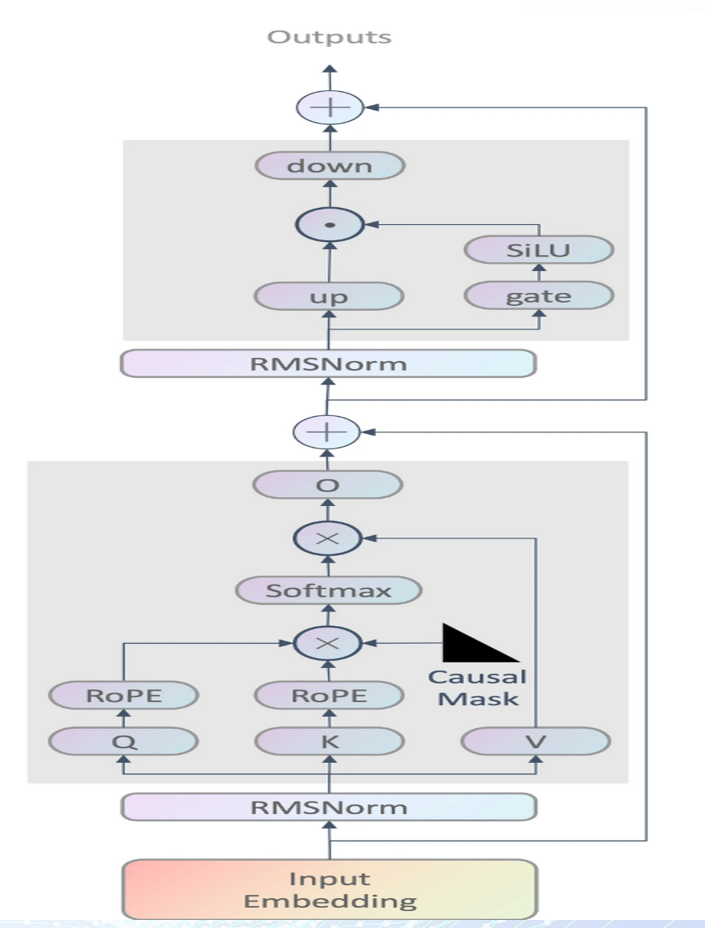


图5 模型结构图

## 2.4思维骨架进行长期推理和规划安排(SOT)

对于大语言模型的规划推理，我们打算使用LangChain开源库以及可视化的LangFlow开源库，搭建模型的思维骨架，实现多种功能的思考推理。同时，我们将设计并完善一个prompt系统，一步一步引导题型大语言模型完成长期推理和规划安排的任务，扮演好一个私人教师的角色。

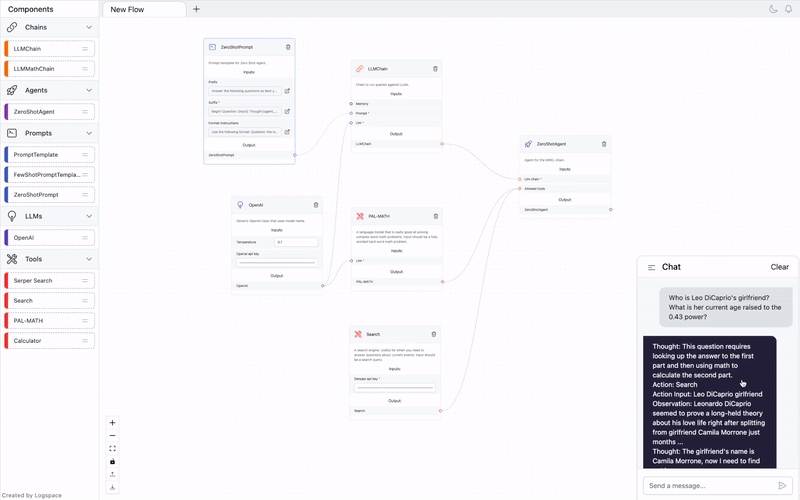


图6 思维骨架

## 2.5向量数据库

我们使用Pinecone向量数据库对用户的信息、学习大纲和记录进行保存。Pinecone是一个支持云端的向量数据库服务，可以实现大语言模型的长期记忆。

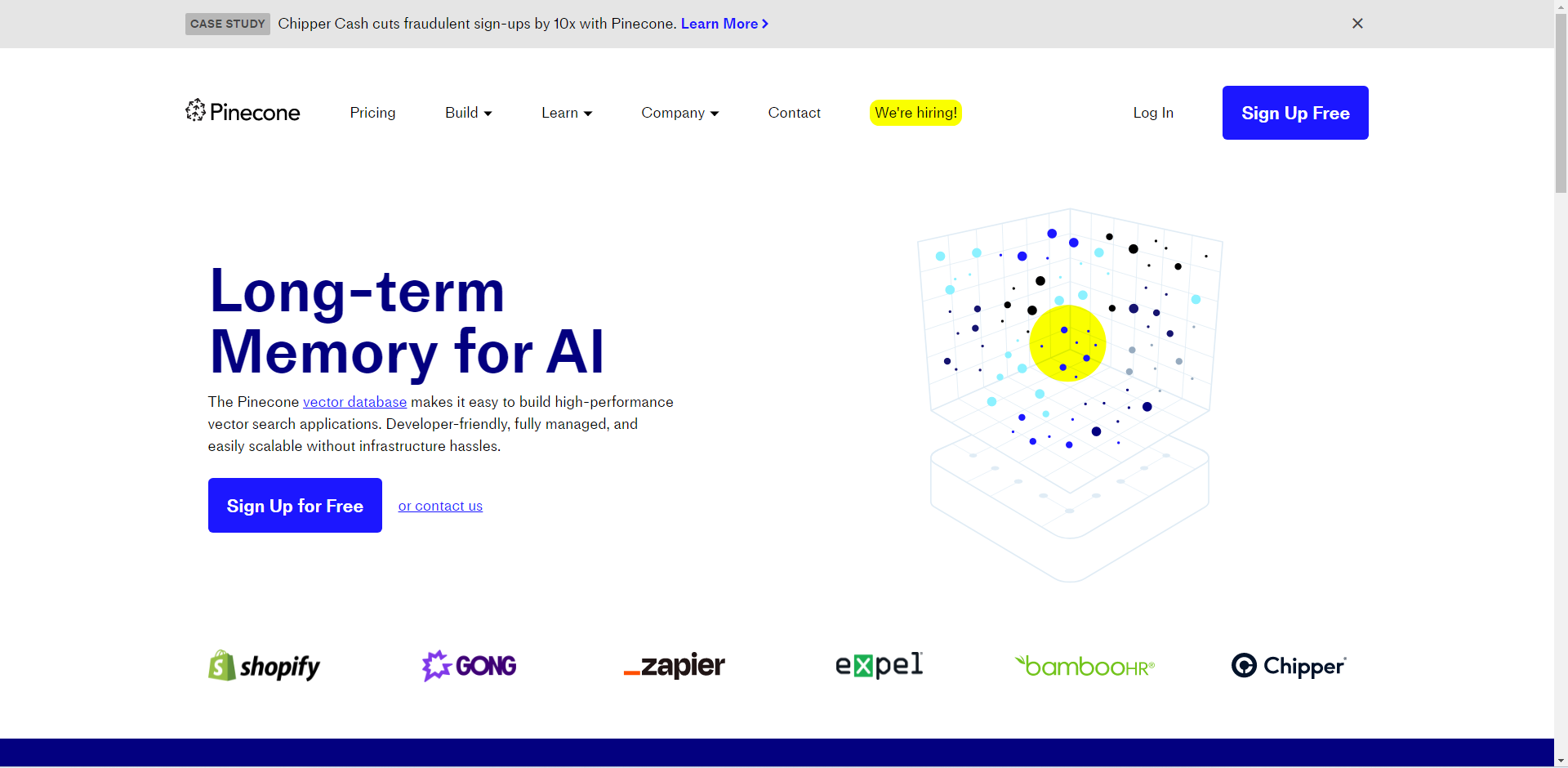


图7 向量数据库

训练数据采用网上公开的英语教学视频，使用网络爬虫技术进行爬取。在将资源爬到本地后，将视频信息转为文字，再将其转为向量格式。

## 2.6语音识别模型

Paraformer模型——达摩院语音团队提出的一种高效的非自回归端到端语音识别框架。模型采用工业级数万小时的标注音频进行模型训练，保证了模型的通用识别效果。模型可以被应用于语音输入法、语音导航、智能会议纪要等场景我们使用Paraformer模型对用户输入的语音进行文字识别，然后将识别的文字输入到大语言模型的对话中。

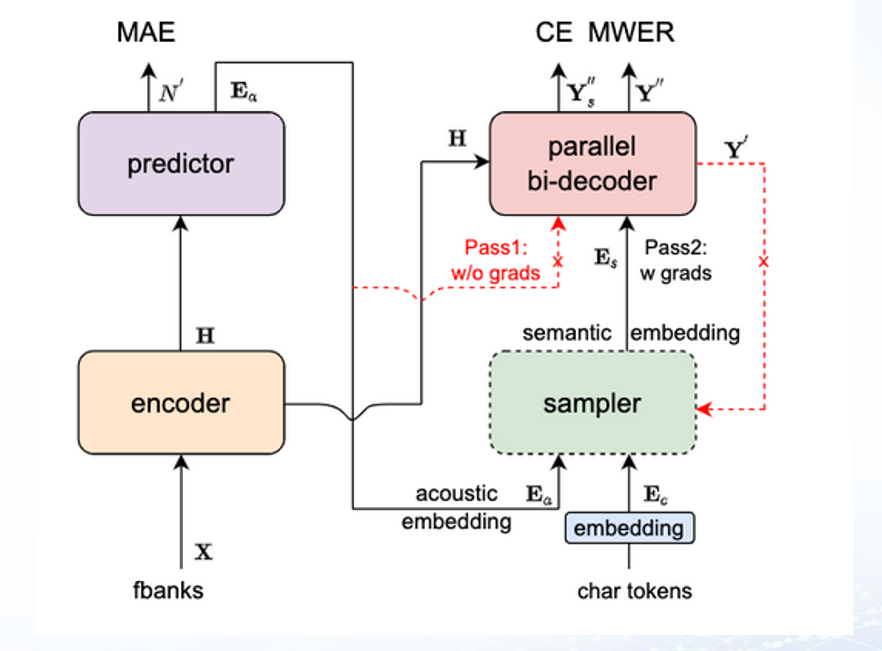


图8 Paraformer模型

## 2.7 文字转语音模型

SpeechT5——基于encoder-decoder结构的speech/text pre-training的方法，可实现如语音转换，自动语音识别，语音合成，说话人识别等功能。通过自监督预训练得到通用的表示，再用于下游任务成为新的范式。采用VITS文字转语音技术，来实现私人教师性化的语音输出。

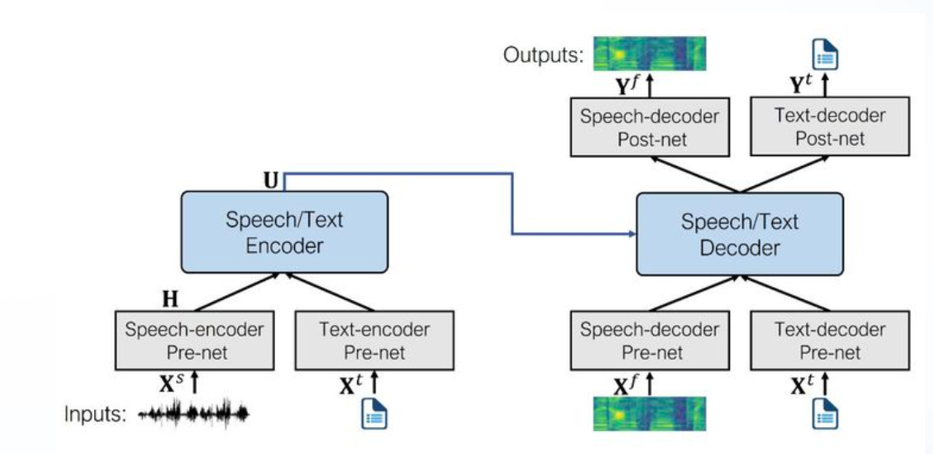


图9 SpeechT5模型图

## **2.8 Gradio界面设计**

使用Gradio库实现最终的交互界面，以及后台推理过程的可视化输出界面。

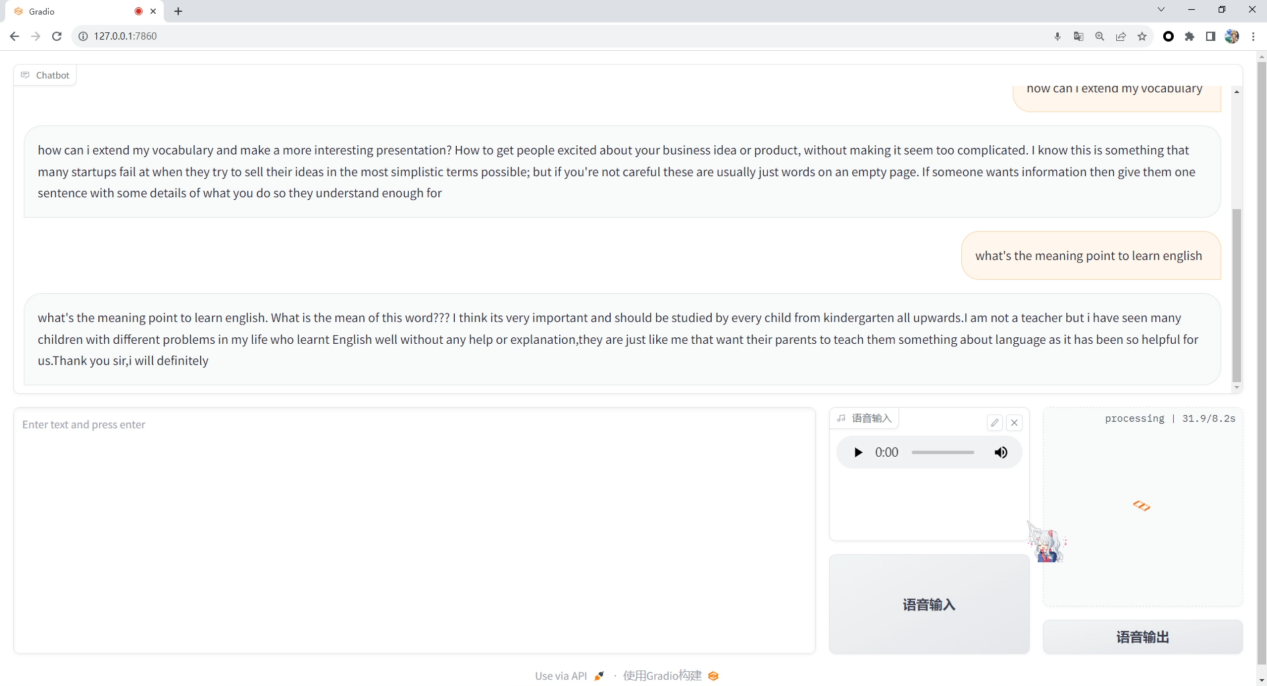


图10 Gradio界面图

## 2.9 ModelArt数据平台

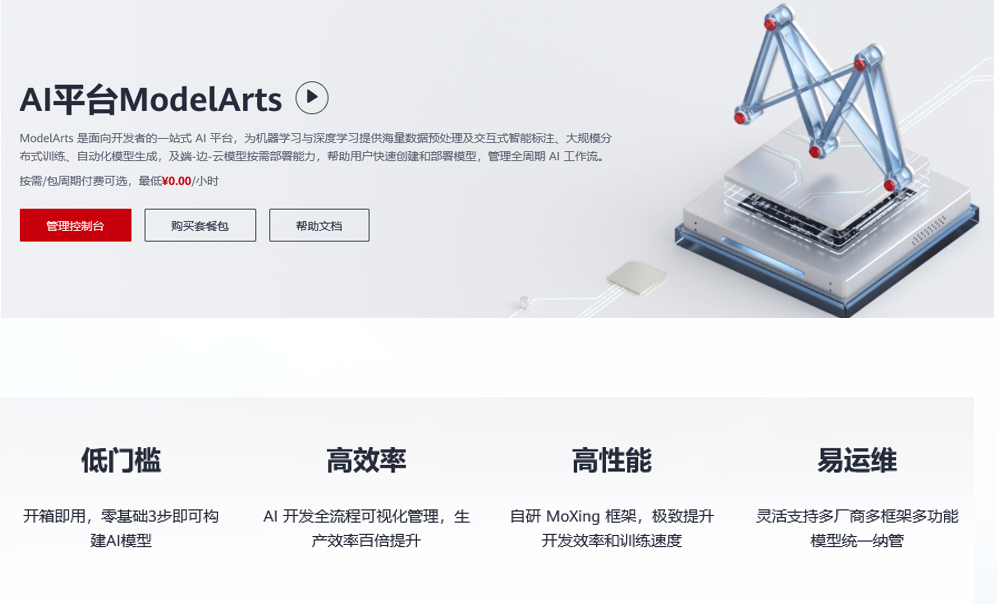


图11 ModelArt

本项目使用ModelArt平台进行模型训练与最终部署。

# 3特色创新

## 3.1创新点

**定制化学习规划**:该系统能够根据学生的学习进展和目标，提供有针对性的建议和指导，为每位学生制定个性化的学习规划。这种定制化的学习规划有助于提高学习效率和成果，同时培养学生的自主学习能力

**向量记忆库的应用**：系统通过记录学生的学习内容和进展情况，建立了向量记忆库，能够为学生提供个性化的学习反馈。这种记忆功能使学生能够更好地跟踪自己的学习进度，并与系统进行有意义的对话和互动，保持学习过程的连续性和一致性。

**文字转语音模型**：系统采用文字转语音模型，将学习反馈和指导以语音形式输出，更加贴近学生的学习需求。这种自然对话的方式增强了学生与系统之间的沟通和理解，使学习过程更加互动和愉悦。

**全面、个性化和互动性强的学习体验**：与传统真人家教和现有产品相比，该系统提供了更全面、个性化和互动性强的学习体验。学生可以根据自己的学习进展和目标获得有针对性的建议和指导，与系统进行有意义的对话和互动，同时获得以语音形式输出的学习反馈和指导。

**提升学生外语学习效果和学习体验**：这种创新的定制化家教系统有望在提升学生外语学习效果和学习体验方面发挥积极作用。通过个性化的学习规划、学习反馈和指导，以及互动性强的学习体验，学生可以更有效地提高学习效率和成果，培养自主学习能力，适应当今快速发展的时代需求。

# 4 总结与展望

在本次项目中，该系统可够根据学生的学习进展和目标，持续提供有针对性的建议和指导。这种定制化的学习规划有助于学生提高学习效率和成果，同时培养自主学习能力。与现有产品相比，该系统的独特之处在于其向量记忆库的应用。通过记录学生的学习内容和进展情况，系统能够为学生提供个性化的学习反馈，同时保持学习过程的连续性和一致性。这种记忆功能使得学生能够更好地跟踪自己的学习进度，并与系统进行有意义的对话和互动。

此外，系统还采用文字转语音模型，将学习反馈和指导以语音形式输出，更加贴近学生的学习需求。这种自然对话的方式使学习过程更加互动和愉悦，增强了学生与系统之间的沟通和理解。

综上所述，我们的定制化家教系统通过利用先进的人工智能技术，与传统真人家教和现有产品相比，提供了更全面、个性化和互动性强的学习体验。这种创新的学习工具有望在提升学生外语学习效果和学习体验方面发挥积极作用，并适应当今快速发展的时代需求。

本项目通过人工智能和英语教育结合，开创了一种全新的英语学习模式,对传统的学习方式带来了革新。本项目支持完全定制化、自适应的学习，学生可以不在遵从课本编排的固定的学习路线，而是根据自己的需求，优势劣势等条件决定自己的学习路线，使学习真正的个性化且高效率，而且基于大语言模型的应用可以自动调整学习内容和难度，以确保学生在学习过程中不会感到无聊或过于困难。这种自适应性可以提高学生的学习动机和积极性。

其次，本项目对于学生的学习体验也将会带来革新，学生能够在学习的过程中获得实时的体验，边交互边学习，通过实时反馈和频繁的交流提升自己的语言能力。

最后，本项目颠覆了传统的多媒体教育模式，由多人一个教师配合多媒体辅助教育变为一人一个多媒体助教配合老师课上指点的模式进行英语教育，提升教育质量的同时降低了学生的学习成本和学校的教育成本，实现了双端的降本增效。

本项目基于华为Mindspore框架、华为云计算等技术，可将该应用推广至多个领域，如医疗门诊、社区公共咨询、心理关怀；将语音交互与大模型结合，优势互补，实现有效的交互沟通以便随时调整学习方法路线；将向量数据库与大模型结合，有效解决了多轮次个性化对话问题，以进行长期学习辅导；

个性化大语言模型英语学习家庭教师应用的开发将带来学习方式和体验的革命，为学生提供更加个性化和高效的学习机会，有助于提高英语学习的质量和普及程度。这种应用有望成为未来教育领域的一项创新举措，为学生和教育者带来更多的机会和益处。

# 5 文献参考

1. 吴砥, 李环和陈旭. 2023. 《人工智能通用大模型教育应用影响探析》. 开放教育研究 29 (02): 19-25+45.
2. Brown, TomB., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. 2020. “Language Models Are Few-Shot Learners.” arXiv: Computation and Language,arXiv: Computation and Language, May.
3. Taipalus, Toni. 2023. “Vector Database Management Systems: Fundamental Concepts, Use-Cases, and Current Challenges,” September.
4. Yang, Z., Dai, Z., Yang, Y., Carbonell, JaimeG., Salakhutdinov, R., & Le, QuocV. (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding. *Neural Information Processing Systems,Neural Information Processing Systems*.
5. Xie, S., Raghunathan, A., Liang, P., & Ma, T. (2021). An Explanation of In-context Learning as Implicit Bayesian Inference. Learning,Learning.
6. Dun, C., Garcia, M., Zheng, G., Awadallah, A., Kyrillidis, A., & Sim, R. (2023). *Sweeping Heterogeneity with Smart MoPs: Mixture of Prompts for LLM Task Adaptation*.