AI智能·学习搭子

技术架构文档



**发包方：数字马力**

**承接方：声像科技**

**日期：2025年10月**

**技术架构图、核心代码示例、关键技术选型说明，重点体现蚂蚁集团技术在系统中的应用与融合方案。**

**数字人伴学系统**

本系统以阿里巴巴通义实验室开源的OpenAvatarChat为基座，构建了一套端到端的实时数字人对话工作流。该工作流采用模块化设计，将语音识别（ASR）、大语言模型（LLM）、语音合成（TTS）和数字人驱动（Avatar）四大核心引擎解耦，便于独立优化与替换。

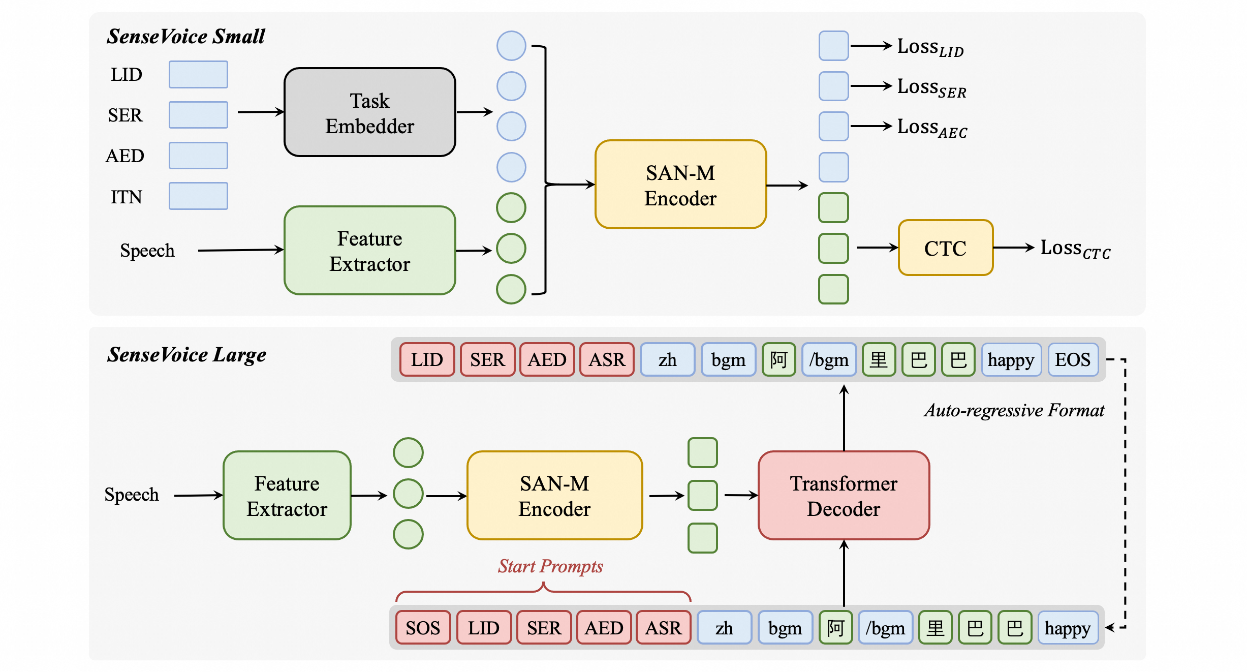
**图示

AI 生成的内容可能不正确。**

在本地部署测试环境中（Ultra7-265K CPU + Nvidia RTX 5080 GPU），我们选用了以下技术栈，并在局域网HTTPS（由OpenSSL提供SSL证书）环境下进行推流，实现了约3秒的平均端到端延迟。以下是对四大功能引擎的选型介绍与优化方案：

**ASR语音识别引擎：SenseVoice**

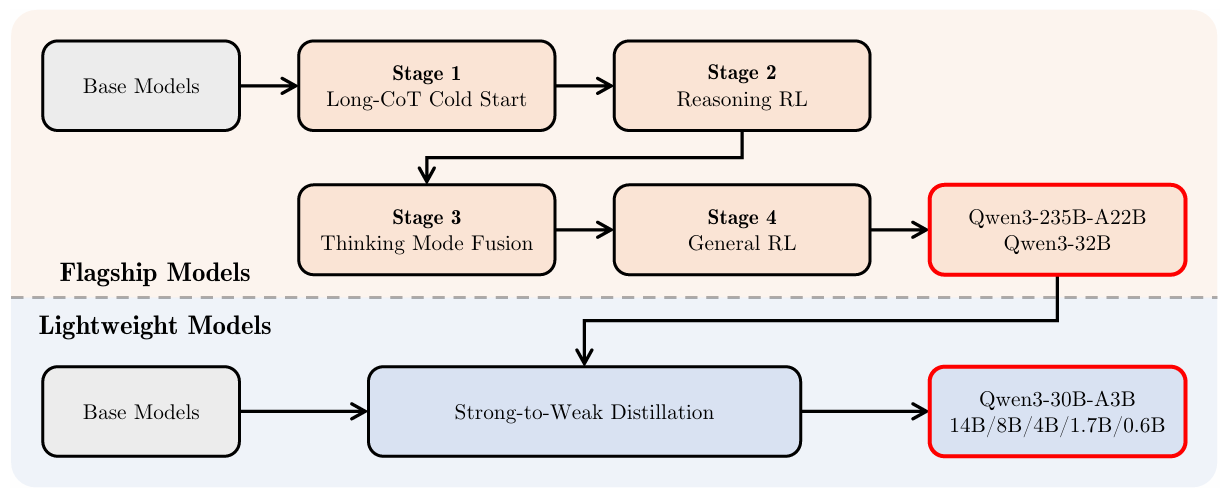
SenseVoice是阿里达摩院推出的工业级语音识别基础模型，不仅支持高精度的自动语音识别（ASR），还集成了语种识别（LID）、情感识别（SER）等多模态能力，能为后续的个性化交互提供丰富的上下文信息。



为降低交互延迟，我们对SenseVoice进行了伪流式处理改造。通过将用户的连续语音流按固定时间窗口（chunk）进行分块，并采用截断注意力机制（truncated attention）对每个语音块进行独立推理 。这种方式在保证识别准确率的同时，显著减少了用户说完话后到系统开始响应的等待时间，提升了对话的流畅感。

**LLM大语言模型引擎：Qwen3-4B + AgentUniverse**

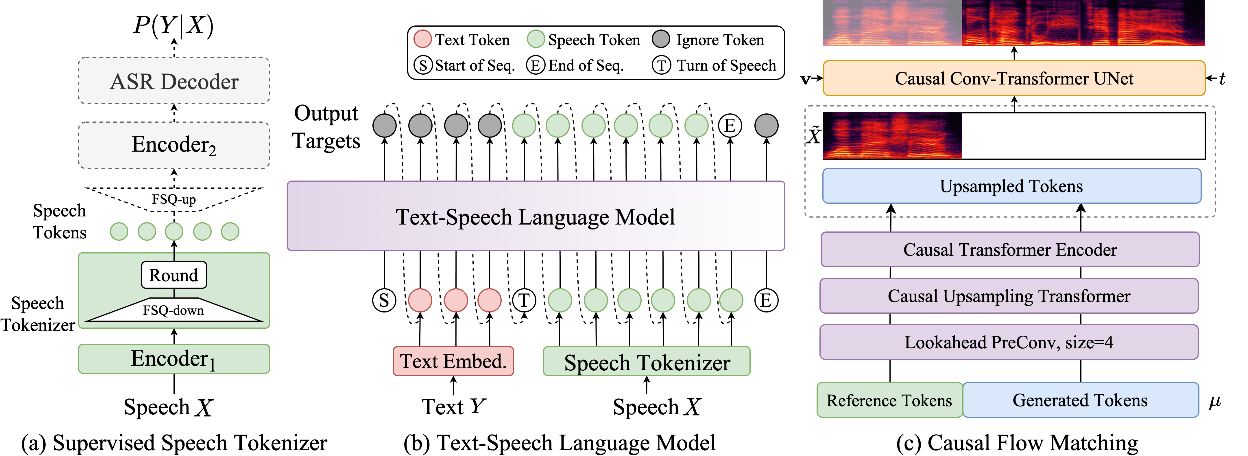
Qwen3-4B是阿里巴巴最新一代的大语言模型，以其卓越的参数效率和强大的多语言、工具调用能力著称。其性能可与上一代更大参数量的模型相媲美，非常适合在本地有限算力下部署。



我们将Qwen3-4B深度集成到AgentUniverse多智能体框架中。数字人不再是一个孤立的对话模型，而是作为一个“伴学智能体”（Companion Agent），能够与其他智能体（如“知识库检索智能体”、“学习规划智能体”）协同工作。例如，当学生提出一个复杂问题时，伴学智能体会调用知识库智能体检索最新资料，并将结果整合后以更易懂的方式回答，实现了知识的动态更新与深度整合。

**TTS语音合成引擎：CosyVoice 2.0**

CosyVoice 2.0在发音准确性、语音自然度和响应速度上均有显著提升，其发音错误率相比1.0版本降低了30%-50%。更重要的是，它支持通过自然语言指令控制语音的情感和韵律，这对于营造有温度的伴学体验至关重要。



我们利用其情感控制能力，根据LLM对对话上下文的情感分析结果，动态调整TTS输出的语调。例如，在学生回答正确时，数字人会用更欢快、鼓励的语气；在学生遇到困难时，则会切换到更温和、耐心的语气，实现真正的情感化陪伴。

**Avatar数字人驱动引擎：LAM**

LAM是一种创新的大型数字人模型，能够从单张图片快速生成超写实的3D高斯头像，并支持跨平台实时驱动与渲染。这极大地简化了数字人形象的创建流程，未来可支持学生上传自己的照片生成专属学习伙伴。

图表

AI 生成的内容可能不正确。

我们正在探索将ASR的情感识别结果（SER）与LAM的面部表情驱动相结合，使数字人的表情能与语音情感同步，打造更具沉浸感和真实感的交互体验。

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**

本系统是“AI智能·学习搭子”解决方案的知识中枢与智能引擎。它并非一个孤立的问答机器人，而是依托蚂蚁集团在复杂金融业务场景中淬炼出的AgentUniverse多智能体框架，构建了一个由多个领域专家智能体协同工作的动态知识网络。其核心价值在于，能够将静态的知识库转化为可推理、可协作、可进化的活知识，为数字人伴学系统和评估系统提供强大、精准、实时的智能支持。

图形用户界面, 网站

AI 生成的内容可能不正确。

AgentUniverse是一个面向复杂业务场景设计的多智能体协作框架，其核心是一个“模式工厂（pattern factory）”，允许开发者对多智能体协作模式进行开发和定制。我们基于此框架，设计了以下关键智能体及其协作流程：

**知识库检索智能体(Knowledge Retrieval Agent)：**

作为系统的“图书管理员”，负责监听来自数字人伴学系统或评估系统的查询请求。它能理解查询的语义，并在我们构建的多学科、多来源（教材、论文、开源项目、行业报告）的知识库中进行高效、精准的检索。

技术实现：采用向量数据库（如Milvus、FAISS）存储知识嵌入，并结合关键词索引，实现混合检索，确保召回率与准确率。

**内容生成与摘要智能体(Content Generation & Summarization Agent)：**

作为系统的“内容编辑”，接收检索智能体返回的原始资料。它利用大语言模型（如Qwen3）的能力，对冗长、专业的原始信息进行提炼、总结、改写，生成适合学生当前认知水平和学习阶段的、易于理解的答案或学习材料。

技术实现：该智能体被配置为遵循特定的“写作规范”，确保输出内容的准确性、简洁性和教育性；同时在应用接口返回数据时被严格规范为使用JSON格式进行返回，这也有效保证了智能体对各个子系统的有效兼容。

**学习规划智能体(Learning Planner Agent)：**

作为系统的“学业顾问”，它与“多学科客制化评估系统”深度联动。当评估系统诊断出学生的知识薄弱点后，该智能体会综合学生的学习目标（校园考试/就业技能）、历史学习数据和知识库内容，动态生成或调整个性化的学习路径和任务清单。

**反思与进化智能体(Reflection & Evolution Agent)：**

作为系统的“质量监督员”和“进化引擎”，它会定期分析所有智能体的交互日志和用户反馈。通过反思（Reflection）机制，它能发现知识库的缺失、答案的不准确或规划的不合理之处，并自动触发知识库的更新流程或向管理员发出优化建议，实现系统的持续自我进化。

图示

AI 生成的内容可能不正确。

**与蚂蚁集团生态的融合**

AgentUniverse框架是蚂蚁集团探索大模型技术在金融等严谨产业场景中的最新成果，已在真实业务中得到验证。这为我们系统的稳定性、可靠性和处理复杂任务的能力提供了坚实基础。

系统设计严格遵循企业文档中提出的MCP/A2A协议规范，确保每个智能体都是一个标准的、可插拔的服务单元。这不仅便于我们内部系统的模块化开发和维护，也为未来将本系统作为能力开放给蚂蚁集团内外的其他教育应用奠定了基础；所有智能体间的通信和数据交互，均在微服务架构下通过Service Mesh进行治理，并采用OAuth2/JWT等机制进行认证授权，确保整个知识协作流程的安全可控。

* + 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**

本系统是“AI智能·学习搭子”解决方案中实现精准诊断与效果闭环的关键一环。它并非一个通用的题库系统，而是针对不同学科（如计算机、化学）的独特知识结构和能力评估标准，提供高度客制化的测评与反馈服务。系统旨在解决“学得怎么样”和“哪里不会”的核心问题，为数字人伴学系统和学习规划提供数据驱动的决策依据，契合赛题中“一阶段（校园考试）”与“二阶段（就业技能）”的双场景需求。

我们以高校中计算机学科的实际教学为切入点，构建了首个客制化评估模块，并为其他学科的快速扩展奠定了技术基础；其采用蚂蚁集团开源的OpenSumi IDE框架作为核心载体。OpenSumi是一个高性能、高可定制的双端（Web/Electron）IDE开发框架，能够快速构建出云端或桌面端的集成开发环境。

图表

AI 生成的内容可能不正确。

基于OpenSumi提供的前端IDE框架，团队集成了隔离的代码执行沙箱环境。学生可以直接在浏览器中编写、运行和调试代码，无需担心环境配置问题；同时，团队也开发了与项目深度集成的代码自动评测系统。该引擎不仅能判断代码的正确性（通过单元测试），还能从代码风格、算法效率、内存占用等多个维度进行综合评分，提供远超“对/错”的精细化反馈**；**同时，结合大语言模型（LLM），对学生的错误代码进行智能分析，不仅能指出错误原因，还能生成针对性的修复建议和相关知识点链接，实现“以评促学”，让编程学习不止停留在“纸上谈兵”。

**跨学科扩展框架：**

面向其他综合学科，我们也搭建了相应的跨学科拓展框架。

**数学：**团队通过集成视觉大模型（Qwen-VL）实现数学公式识别和符号计算；学生可以手写或输入数学公式，系统能自动解析并评估其推导过程的逻辑与正确性。

**化学：**团队基于Three.js的前端3D界面构建分子结构的可视化模型；使得教师能够在实际授课中全面展现分子结构，实现可视化的教学过程。

**通用技术：**团队结合教学过程中的授课PPT、教科书PDF等多元信息，对Qwen3大模型进行微调（Fine-Turning），实现题目对应相关知识的溯源过程。

**与整体解决方案的深度协同**

**数据驱动学习规划：**评估系统的诊断结果（如“动态规划算法掌握薄弱”、“有机化学反应机理理解不清”）会实时同步给“多智能体-知识库协作系统”中的学习规划智能体，后者据此动态调整学生的学习路径，推送针对性的学习资料和练习题。

**赋能数字人伴学：**当学生在评估中受挫时，评估结果会触发数字人伴学系统的主动关怀。数字人会以鼓励的语气，结合评估报告，与学生一起分析错因，并引导其进入复习流程，实现情感激励与知识巩固的结合。

**衔接“二阶段”就业场景：**针对“就业导向”的学习目标，评估系统可直接对接数字马力产教融合项目中定义的行业技能图谱。例如，对“后端开发工程师”岗位，系统可模拟真实的编码面试题或项目任务进行评估，并生成一份对标企业用人标准的能力雷达图，让学生清晰了解自身与岗位要求的差距。

* + 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**

本系统是“AI智能·学习搭子”解决方案中面向入门级学习者（如小学生）和兴趣驱动型学习场景（如语言学习）的核心模块。其核心价值在于，通过将游戏化（Gamification）设计理念深度融入学习过程，有效解决低龄或初学者“学习动力不足、注意力难以集中、缺乏即时正向反馈”的痛点，让学习变得像游戏一样有趣、有挑战、有成就感，从而激发内在学习动机，为后续的进阶学习打下坚实基础。

**技术架构与游戏化设计**

我们以小学英语为切入点，构建了首个游戏化学习模块，并为其他语言学科（如日语）的快速扩展提供了范式。

**技术基座：**采用蚂蚁集团主导开发的Ant Design企业级UI设计系统作为前端开发框架。Ant Design不仅提供了丰富、美观、一致的UI组件，其设计价值观中也蕴含着“创造快乐工作”的理念 ，这与我们的游戏化目标高度契合。我们利用其响应式布局能力，确保平台在平板和电脑上均能提供流畅、沉浸的用户体验。

**核心游戏化机制：**

**角色与成长体系：**学生创建自己的虚拟角色（Avatar），通过完成学习任务（如单词拼写、听力练习、口语对话）获得经验值（XP）和金币，用于解锁新装扮、新技能及进入更高级别的学习关卡。

**即时反馈与奖励：**借鉴“百词斩”等成功产品的经验，系统在学生完成每个微小任务后，立即给予视觉（动画、徽章）和听觉（音效）的正向反馈，强化学习行为。

**挑战与任务系统：**将学习目标分解为一系列由易到难的“每日挑战”和“主题任务”，如“一周内掌握20个动物单词”。任务设计紧密结合教学大纲，确保游戏性与教育性的统一**。**

**智能评估与反馈：**

**嵌入式评估：**评估不再是独立的考试，而是完全融入游戏流程中。例如，在一个“餐厅点餐”的角色扮演游戏中，系统会通过语音识别（ASR）和自然语言理解（NLU）技术，实时评估学生的发音准确度和句型使用是否恰当。

**数据驱动的个性化：**平台会记录学生在游戏中的所有行为数据（如错误类型、反应时间、重试次数），并利用这些数据动态调整后续任务的难度，确保挑战始终与学生的能力相匹配，维持其“心流”状态 。

**与整体解决方案的深度协同**

**无缝衔接进阶学习：**当学生在游戏化平台中达到一定水平时（如完成小学英语全部课程），系统会自动推荐并引导其进入“多学科客制化评估系统”中的更高阶英语学习路径，实现从“兴趣入门”到“目标导向”的平滑过渡。

**赋能普惠教育：**本系统适合通过蚂蚁公益平台进行推广。我们可以将其打包为一个公益项目，例如“乡村儿童英语启蒙计划”，通过支付宝公益频道的“行为公益”（如用户捐步数）或直接捐赠，为资源匮乏地区的学校和学生提供免费的、高质量的游戏化英语学习资源，践行数字马力的普惠教育使命。

**融入教育机构合作：**与数字马力已合作的高校及中小学合作，将本平台作为其英语教学的官方辅助工具，教师可以在后台查看班级整体和学生个体的学习数据报告，实现精准教学。

* + 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