

AI 智能 · 学习搭子

产品方案文档



声像科技

Sheng Xiang Ke Ji

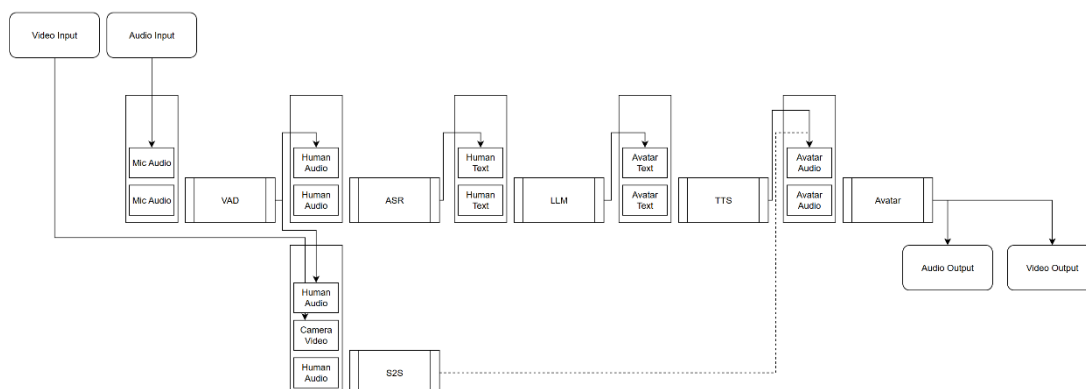
发包方：数字马力

承接方：声像科技

日期：2025 年 10 月

数字人伴学系统

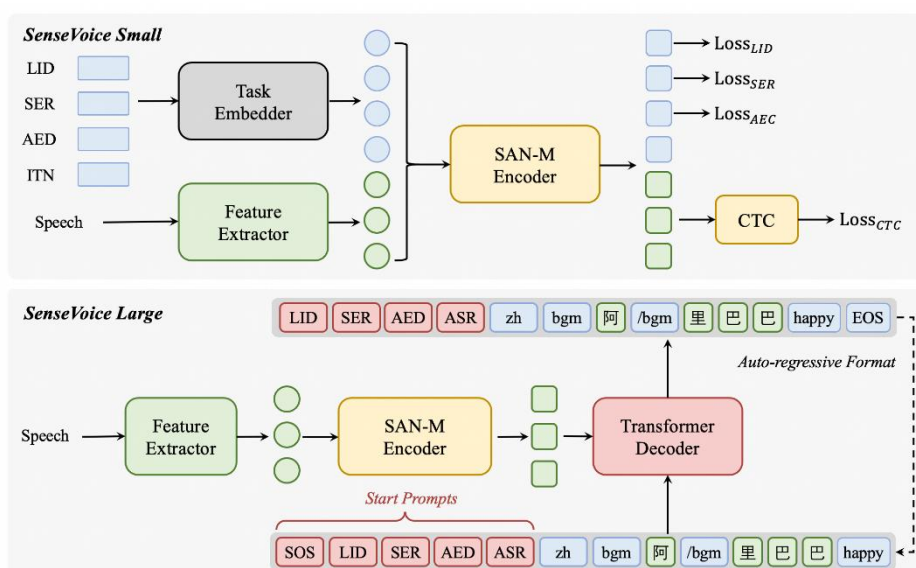
本系统以阿里巴巴通义实验室开源的 OpenAvatarChat 为基座，构建了一套端到端的实时数字人对话工作流。该工作流采用模块化设计，将语音识别（ASR）、大语言模型（LLM）、语音合成（TTS）和数字人驱动（Avatar）四大核心引擎解耦，便于独立优化和替换。



在本地部署测试环境中（Ultra7-265K CPU + Nvidia RTX 5080 GPU），我们选用以下技术栈，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和优化方案：

ASR 语音识别引擎：SenseVoice

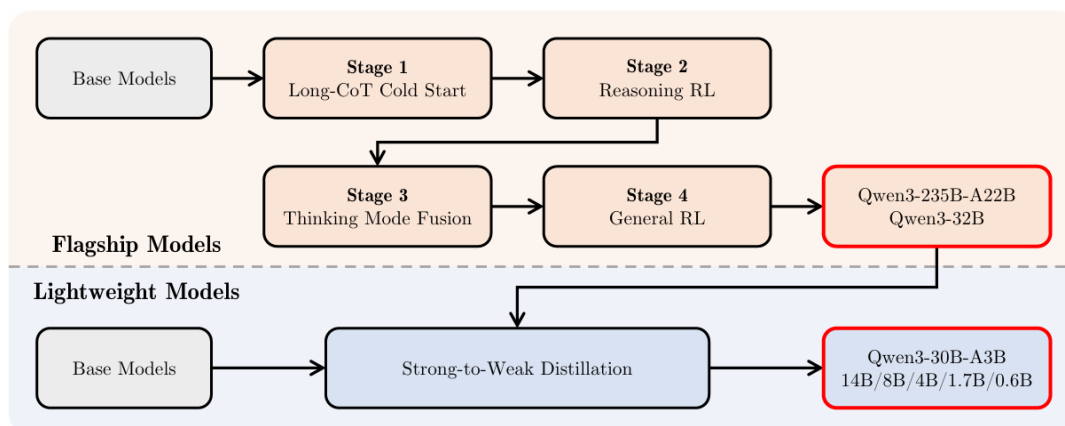
SenseVoice 是阿里达摩院推出的工业级语音识别基础模型，不仅支持高精度的自动语音识别（ASR），还集成了语种识别（LID）、情感识别（SER）等多模态能力，以便为后续的个性化交互提供丰富的上下文信息。



为了降低交互延迟，我们对 SenseVoice 进行伪流式处理改造。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，显著减少了从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

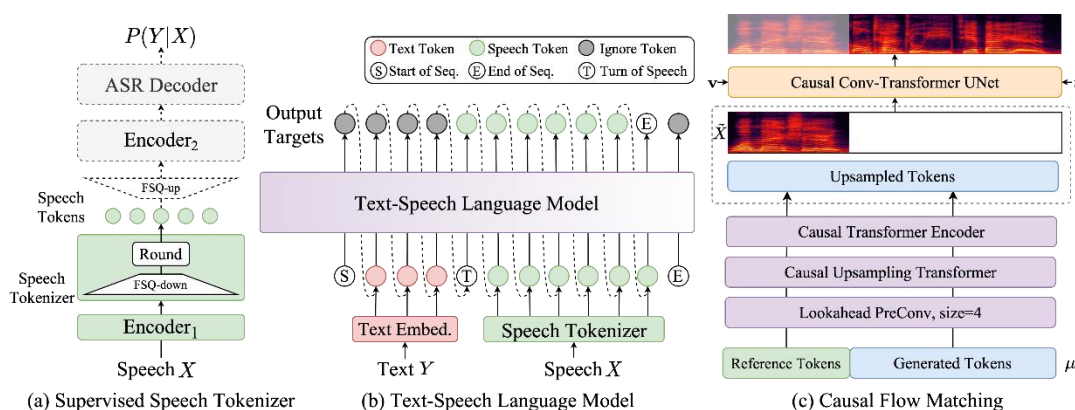
Qwen3-4B 是阿里巴巴最新一代的大语言模型，以其卓越的性能效率和强大的多语言、工具调用能力著称。其性能可与上一代更大参数量的模型相媲美，非常适合在本地有限算力下进行部署。



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去检索最新资料，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和深度整合。

TTS 语音合成引擎：CosyVoice 2.0

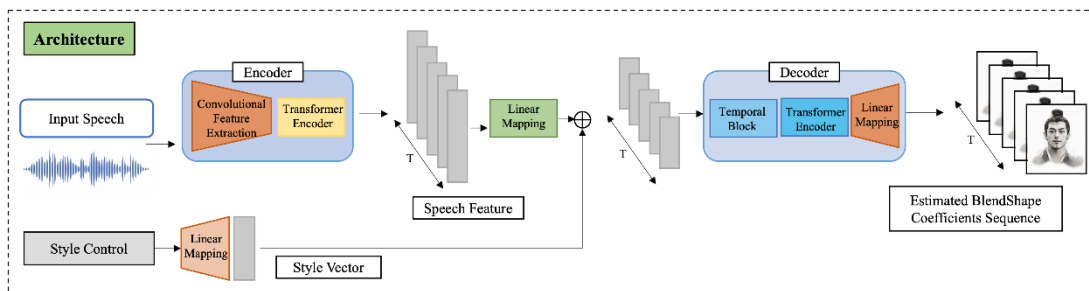
CosyVoice 2.0 在发音准确性、语音自然度和响应速度上均有显著提升，其发音错误率相较于 1.0 版本降低了 30%-50%。更重要的是，它支持通过自然语言指令控制语音的情感和韵律，这对于营造有温度的伴学体验至关重要。



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，能够根据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和渲染。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。



我们正在探索将 ASR 的情感识别结果（SER）和 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。

企事业单位的生态融合共建

团队就“数字人伴学系统”赴浙江联通开展了实地调研，重点围绕数字人技术在教育场景中的落地可行性、算力基础设施支持、以及应用数据的安全合规性等方面进行了深入交流。调研主要内容包括：

技术落地场景探讨：

在与浙江联通行业解决方案团队沟通后，了解到数字人教学场景在 K12 及高等教育中的实际应用案例，同时明确了其在智慧教室、远程教学、个性化辅导等场景中，对低延迟、高并发、多模态交互的数字人技术需求。



算力与网络基础设施支持：

浙江联通介绍了其在“算网融合”方面的布局，包括边缘计算节点、5G 专网和教育云平台，可为数字人系统提供稳定、低延迟的运行环境。团队据此验证了将数字人伴学系统进行本地部署与云边协同部署两种模式的可行性。

数据安全与合规建议：

针对学生隐私保护和教育数据安全，浙江联通强调了教育行业数据规范的重要性，建议在设计系统中强化身份认证、数据加密和访问审计机制，这与企业文档中提出的 OAuth2、JWT、HTTPS 等安全要求高度一致；同时，就数字人系统在实际场景中的应用规范，团队也与浙江联通展开了相关探讨。

团队系统将严格遵循企业文档中建议的安全规范，采用 OAuth2 和 JWT 进行用户认证授权，并通过 HTTPS 加密所有数据传输，确保学生隐私数据的安全；前端界面也将采用蚂蚁集团的 Ant Design 组件库进行开发，与蚂蚁系产品的 UI/UX 风格一致，确保团队产品良好的前端可交互性、降低用户学习成本。

除此之外，在数字人系统的应用安全层面上，声像科技团队以“守正创新”为核心理念，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、高效的数字人伴学系统。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。

声像科技团队同时也研发了一套数字人深度伪造鉴别系统，助力生态共建的相关企事业单位实现从系统到应用层面的全方位安全防护；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，可以鉴别合成音视频的真伪，确保数字人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。

团队就数字人应用系统申请了两项软件著作权并取得授权：

1. 软件名称：《面向深度伪造音视频的智能检测软件》

证书号：软著登字第 16024611 号

授权日期：2025 年 7 月 25 日

2. 软件名称：《终身学伴-数字虚拟人合成平台》

证书号：软著登字第 16545571 号

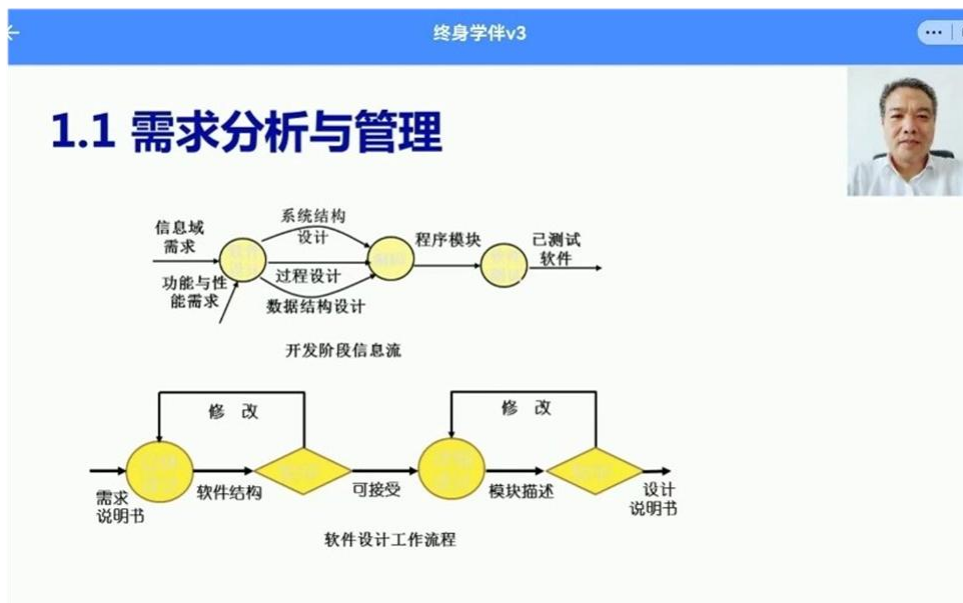
授权日期：2025 年 9 月 28 日



其中《面向深度伪造音视频的智能检测软件》主要面向数字人伴学系统的安全层面，为数字人的内容安全提供了相关保障与支持；《终身学伴-数字虚拟人合成平台》则面向数字人伴学系统的应用层面，搭建了一套高度模块化的数字人应用平台。

系统部署与应用：

声像科技团队所研发的数字人伴学系统在本校数字化办公室得到了实际部署和应用；以下是数字人伴学系统在实际产品中的应用画面：



产品迭代方向

V1.5：实现 ASR 情感识别（SER）和 Avatar 数字人驱动引擎中表情驱动的联动效果，增强数字人情感交互的真实性。

V2.0：打通与“多学科客制化评估系统”的数据接口。当评估系统发现学生知识薄弱点时，可自动触发数字人进行针对性的讲解和鼓励。

V3.0+：探索与“AI 豆计划”等蚂蚁公益项目的结合，为特定群体（如乡村学生）提供定制化的数字人伴学服务，践行普惠教育理念。

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和智能引擎。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、可进化的活知识库，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。



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，确保召回率和准确率。

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):

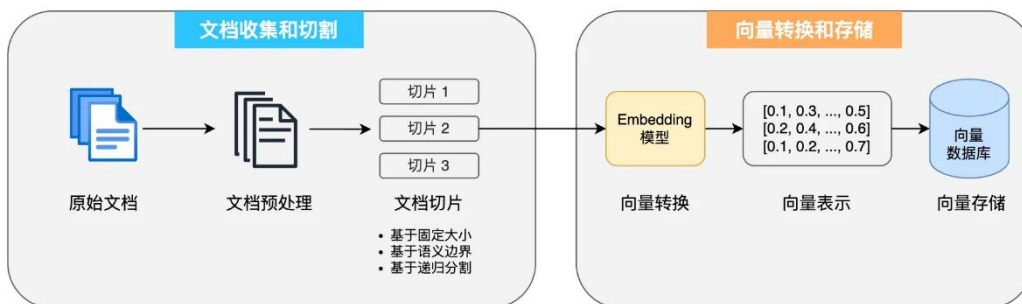
作为系统的“学业顾问”，它与“多学科客制化评估系统”深度联动。当评估系统诊断出学生的薄弱知识点后，该智能体会综合学生的学习目标（校园考试/就业技能）、历史学习数据和知识库内容，动态生成或调整个性化的学习路径和任务清单。

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

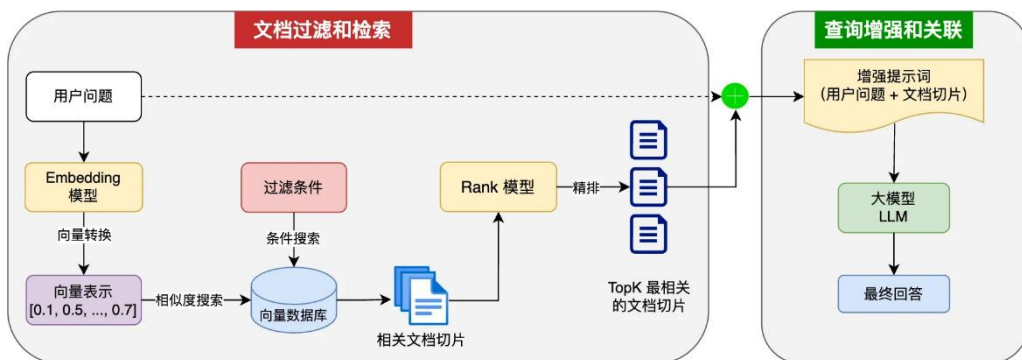
作为系统的“质量监督员”和“进化引擎”，它会定期分析所有智能体的交互日志和用户反馈。通过反思（Reflection）机制，它能发现知识库的缺失、答案的不准确或规划的不合理之处，并自动触发知识库的更新流程或向管理员发出优化建议，实现系统的持续自我进化，让系统“越用越懂你”。

RAG 检索增强生成工作流程

建立索引



检索生成



与蚂蚁集团生态的融合

AgentUniverse 框架来源于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。这不仅便于内部系统的模块化开发和维护，也让本系统得以沉淀为可复用的核心能力，为后续开放给蚂蚁集团内外其他教育应用打下基础；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。

技术落地场景探讨：

团队赴云栖大会调研了多个基于多智能体架构的行业解决方案，如智能客服、科研文献分析助手、企业知识管理平台等。这些案例为团队在“知识检索—内容生成—学习规划”智能体协作流程的设计上提供了宝贵参考；团队特别关注了相关企业在云栖大会展示的教育科技解决方案和产教融合实践成果，并确认了将多智能体系统和数字马力已有的“助教、助学、助就业”能力体系深度融合的可行性。



系统部署与应用：

声像科技团队所研发的多智能体-知识库协作系统在本校数字化办公室得到了实际部署和应用；以下是多智能体-知识库协作系统在实际产品中的应用画面：



产品迭代方向

V1.5：实现“反思与进化智能体”的初步功能，建立基于用户反馈的半自动知识库更新闭环，实现多智能体-知识库协作系统“越用越懂你”的研发目标。

V2.0：引入更多垂直领域的专家智能体（如“编程导师智能体”、“金融分析师智能体”），以更好地支撑“二阶段学习场景”（就业导向）的需求，直接对接数字马力在产教融合中积累的行业技能图谱。

V3.0+：探索将整个多智能体协作网络开放为一个教育智能体市场，允许高校教师或合作教育机构贡献自己的教学智能体，共同丰富“AI 智能·学习搭子”的知识生态，构建一套完整的、面向未来的 MCP 智能体协作系统。

通过构建这样一个基于 AgentUniverse 的多智能体-知识库协作系统，团队不仅解决了知识问答的准确性问题，更构建了一个能够主动规划、持续进化、紧密贴合“从校园到职场”双阶段学习目标的动态智能学习引擎，有力强化了“AI 智能·学习搭子”的核心竞争力。

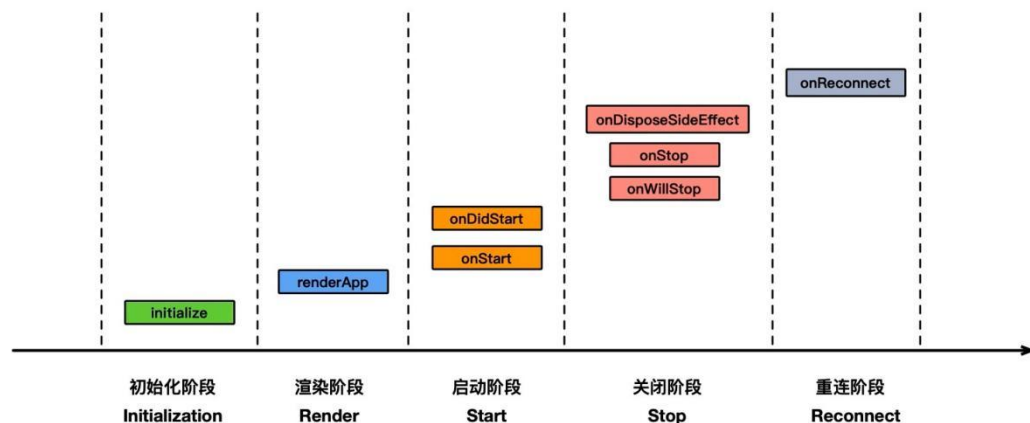
多学科客制化评估系统

本系统是“AI 智能·学习搭子”解决方案中实现精准诊断和效果闭环的关键一环。它并非一个通用的题库系统，而是针对不同理工科学科的独特知识结构和能力评估标准，提供高度客制化的测评和反馈服务。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，并为该系统在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。

生命周期贡献点

ClientAppContribution



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性和正确性。

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。

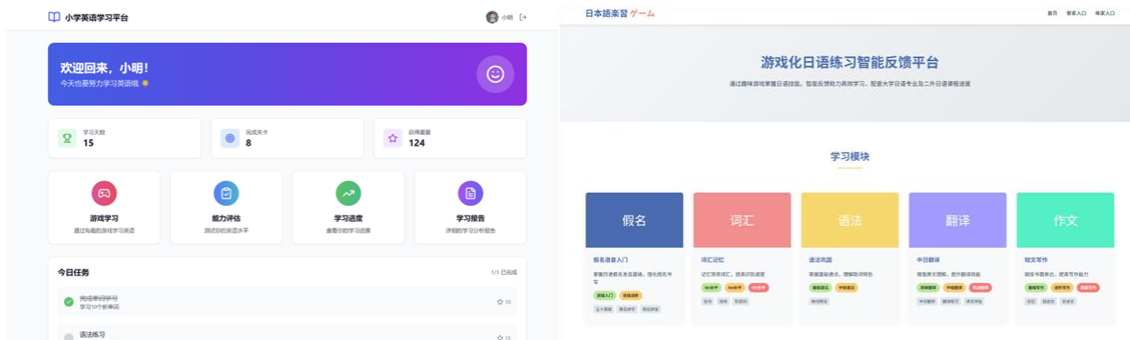
通用技术：团队结合教学过程中的授课 PPT、教科书 PDF 等多元信息，对 Qwen3 大模型进行微调（Fine-Tuning），实现题目对应相关知识的溯源过程。

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，并为其他语言类学科（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：团队借鉴了“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，激励并强化用户学习行为，形成正向激励的学习循环。

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”；任务设计紧密结合教学大纲，确保游戏性和教育性的统一，实现“以玩助学”。

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“单词句法对对碰”游戏中，玩家可以通过说出对应单词的词组匹配来消除对应的区块内容；在游戏过程中，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“玩中学、学中玩”正向激励的学习循环。

与整体解决方案的深度协同

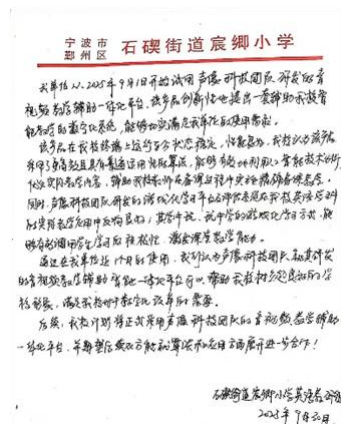
无缝衔接进阶学习：当学生在游戏化平台中达到一定水平时（如完成小学英语全部课程），系统会自动推荐并引导其进入“多学科客制化评估系统”中的更高阶英语学习路径，实现从“兴趣入门”到“目标导向”的平滑过渡。

赋能普惠教育：本系统适合通过蚂蚁公益平台进行推广。团队所研发的游戏化学习平台及评估系统可以被打包为一个公益项目，同“乡村儿童英语启蒙计划”等支付宝公益频道的“行为公益”（如用户捐步数）形成有机结合，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，实现以玩助学、精准教学。

系统部署与应用：

声像科技团队所研发的游戏化学习平台及评估系统在宁波市海曙区宸卿小学与本校东方语言与哲学学院得到了实际部署与应用；以下是游戏化学习平台及评估系统在实际产品中的应用画面和试用报告：



产品迭代方向

V1.5：增加并完善日语、法语等第二外语的游戏化学习模块，并完善 AI 语音评测功能，从语调、语法等角度提供更专业的发音指导。

V2.0：深化与“数字人伴学系统”的结合，营造“元宇宙”式的游戏场景；在游戏中，数字人可以扮演 NPC（非玩家角色）和学生进行互动对话，营造身临其境的语言学习场景，提供更生动的沉浸式体验。

V3.0+：探索与蚂蚁集团 IP（如蚂蚁庄园）的联动，推出联名主题活动和虚拟道具，利用蚂蚁集团强大的用户触达能力，进一步扩大平台影响力。

通过构建这样一个以游戏化机制为核心、并与蚂蚁生态深度协同的游戏化学习平台，我们为“AI 智能·学习搭子”注入了显著的入门吸引力和浓厚的情感温度，有效覆盖了从“激发兴趣”到“成就目标”的完整学习旅程。