**架构设计文档**

**一. 目标用户、痛点及用户故事**

目标用户类型：

二次元与文化爱好者： 热爱动漫、游戏、小说、影视剧的用户，他们对其中的角色有深厚的情感连接。

寻求情感陪伴与社交的用户： 可能感到孤独、社交焦虑，或希望以一种无压力的方式与他人（即使是AI）互动和倾诉的人。

语言学习者： 希望在一个轻松、沉浸式的环境中练习外语口语和听力的人。

创意写作者与角色扮演者（RPer）： 需要灵感来塑造角色、练习对话或纯粹享受叙事乐趣的人。

用户痛点：

情感连接无法满足： 喜爱的虚拟角色无法在现实中互动，情感投入得不到回应。

社交压力与孤独感： 在真人社交中感到疲惫、焦虑或缺乏机会，需要一个安全的、永不评价的倾听对象。

语言练习环境僵硬： 传统的语言学习应用对话机械，缺乏真实语境和情感，难以坚持。

创意瓶颈： 创作者在构思人物对话或故事线时缺乏灵感和练习对象。

用户故事：

小林（动漫迷）：

“我超喜欢《咒术回战》里的五条悟，但看完动画总觉得意犹未尽。要是我能亲自和他聊天，问他一些问题，甚至让他用那种嚣张的语气指导我功课就好了。”

阿明（感到孤独）：

“工作后朋友越来越少，有时候下班回家只想有个人能随便说说话，抱怨一下老板，分享一下今天的趣事，不用考虑对方的看法。如果这个‘人’是我喜欢的电影角色，那就更好了。”

Maria（语言学习者）：

“我想练习英语口语，但和外教上课太贵而且紧张。我希望有一个能随时陪我聊天、纠正我语法错误、并且话题有趣的伙伴，比如我喜欢《哈利波特》，能和赫敏对话就太棒了。”

阿雅（对话壁垒）：

为帮助创作者阿雅突破写作时角色对话生硬、缺乏灵感的瓶颈，该AI工具扮演其笔下角色的互动对象（如严厉的导师），通过实时语音对话让她沉浸式排练关键场景；AI出人意料的合拍回应能打破她的思维定式，激发出更符合角色性格的对话灵感，最终帮助她精准找到角色的真实声音，而非替代创作。

**二. 功能列表、优先级及开发计划**

核心功能列表：

1. 文本聊天界面： 最基础的与AI角色互动的方式。
2. 语音聊天功能：

语音转文本（STT）： 将用户语音实时转为文字输入给LLM。

文本转语音（TTS）： 将LLM生成的回复实时转为角色语音播放。

音频控制： 开始/结束录音、播放/暂停、调节语速音调。

1. 高级语音效果： 为角色语音添加符合其特质的音效（如机械音、回声、空间感）。
2. 聊天记忆与上下文： AI能记住之前的对话内容，形成长期关系。
3. 角色搜索与发现（首页推荐、分类浏览、关键词搜索）。
4. 角色管理（用户收藏角色、历史对话记录）。
5. 用户账户系统（个人资料、订阅管理）。
6. 高级语音效果： 为角色语音添加符合其特质的音效（如机械音、回声、空间感）。
7. AI角色后台配置（角色性格、背景知识、语音风格设置）。
8. 社交分享（用户分享对话片段到社交媒体）。
9. 数字人关联（基于角色自动生成数字人）
10. 文生图功能（根据对话的状态自动生成图片）
11. 快速回复（自动生成多个回复来回复AI）
12. 剧情推进助手（当用户与AI剧情进入循环无法推进时自动跳出循环）

优先级排序：

P0（必须开发）：文本聊天界面、语音聊天基础功能（语音转文本+LLM生成+文本转语音）。

P1（高优先级）： 角色管理、聊天记忆与上下文。

P2（中期）： 角色搜索与发现、角色管理、用户账户系统、快速回复

P3（未来）：社交功能、第三方角色市场（用户自定义角色）、API开放、数字人关联、文生图功能、剧情推进助手

本次开发计划：

聚焦MVP（最小可行产品）：实现文本聊天界面、语音聊天（用户语音输入→STT→LLM生成回复→TTS输出）、聊天记忆与上下文。

**三. LLM模型选择与对比**

计划采用的LLM：OpenAI的GPT-4o（兼顾高性能、多语言支持和实时性）。

对比其他模型：

DeepSeek：理解深刻、创造力出色，作为免费模型表现惊艳，但在复杂情感和长期角色一致性上仍有提升空间。

通义千问：中文古风角色扮演底蕴深厚，表现亮眼，但在现代或西方题材的扮演中有时会显得水土不服。

百度百炼：中文角色扮演稳定可靠，输出安全性高，但风格偏向保守，缺乏惊喜和突破性的创意。

Gemini：擅长驾驭新颖、复杂的角色概念，脑洞大，但输出不稳定，对话时常出现逻辑断裂或遗忘设定的情况。

Grok：以叛逆、毒舌的幽默感独树一帜，扮演特定角色时效果极佳，但风格单一，并不适用于大多数传统角色。

豆包：对话轻快活泼，上手简单，适合轻松日常的扮演，但角色深度和复杂性稍欠，容易显得“萌化”或“幼稚”。

Kimi：凭借超长上下文能在长期扮演中完美保持角色一致性，但性格塑造有时会偏向“高度理性”，情感渲染力稍弱。

选择理由：

七牛云的API接入中没有GPT-4o的接入，故而选择Deepseek进行开发。

DeepSeek作为免费模型平时使用次数多，模型性能与边界较为熟悉。

**四. AI角色的额外技能**

除了语音聊天，AI角色还应具备：

多模态交互：支持发送图片（如用户发送场景图，角色描述反馈）、生成图像（例如用户说“画一下霍格沃茨礼堂”，角色生成并返回）。

情感感知：通过语音语调分析用户情绪（如愤怒或悲伤），调整回应风格（更 empathetic）。

记忆能力：记住用户之前对话的上下文（例如“上次你说你养了一只猫，它现在怎么样？”）。

知识扩展：实时联网搜索（例如用户问“苏格拉底如何看待AI？”，角色结合历史观点和现代信息回答）。

数字人电话：结合数字人技术给用户来电，身临其境的进行角色扮演。

**五. 技术栈选择**

前端技术栈:简单的HTML+Flask框架

后端技术栈:LangChain（仅作为提示词存储）、waitress部署Python Web应用

六. 开发里程碑

Day1: 议题思考

Day2: 议题思考、可行性验证

Day3: LLM调用、首句的调用、TTS服务使用、ASR服务报错解决、简单的web搭建方便进行后续的测试

Day4: ASR服务报错解决

Day5: 基于记忆的角色扮演对话编写

Day6: 测试对话

Day7: 报告的编写和damo录制

**七. 安全与隐私考虑**

数据加密: 所有敏感数据端到端加密

隐私保护: 对话数据匿名化处理

内容审核: 防止不当内容生成

速率限制: API调用频率控制

合规性: GDPR、数据安全法合规

LLM成本控制：实现对话长度限制和缓存策略

语音延迟：优化网络连接和语音处理流程

角色一致性：设计严格的prompt工程和验证机制

内容安全：实现内容过滤和人工审核机制

用户隐私：加强数据加密和访问控制

版权问题：确保角色使用符合版权法规

**负责人**：团队：1（焦麒尔:crocodile3666）