

VirtualStream架构设计文档

版本：1.0 最后更新：2025-6-14

1. 引言

1.1 目的

定义虚拟主播系统的技术架构，指导开发团队实现多模态交互（语音、文本、虚拟形象）的实时流媒体平台。

1.2 范围

覆盖核心功能模块：

- 语音处理（STT/TTS）
- 大语言模型（LLM）交互
- 虚拟形象驱动
- 直播间管理
- 国际象棋游戏集成

2. 架构目标与约束

2.1 核心目标

维度	要求
实时性	语音输入→AI响应延迟
可扩展性	支持动态加载新功能模块（如未来新增小游戏）
容错性	单一模块崩溃不影响系统整体运行
多平台兼容	支持B站直播、本地客户端、网页端接入

2.2 技术约束

1. 实时性约束

场景	要求	应对措施
语音交互延迟	用户输入→AI响应 ≤1.5秒	优化STT/TTS云服务请求批处理
虚拟形象同步	表情变化延迟 ≤200ms	本地预加载动画资源
象棋游戏指令响应	走棋指令处理 ≤500ms	本地轻量级象棋引擎

2. 第三方服务约束

服务商	约束内容	应对方案
科大讯飞TTS	QPS限制、音频长度≤5分钟	请求队列+本地音频拼接
DeepSeek LLM	输入Token上限（如4096）	自动截断历史对话（signals.py控制）
B站直播API	WebSocket心跳间隔≤30秒	定时任务保活（room_manager.py）

3. 开发与部署约束

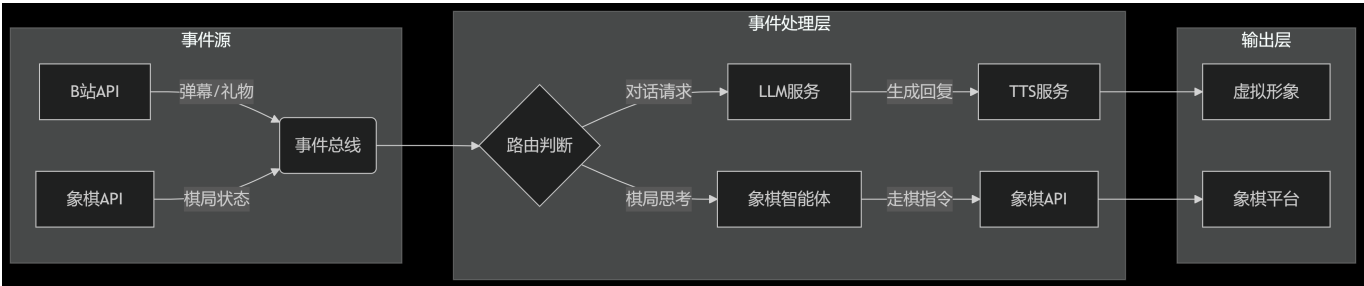
类别	约束内容	解决方案
Python版本	仅支持3.10+（因match语法）	开发环境强制校验（pyproject.toml）
操作系统	需兼容Windows（音频驱动差异）	抽象音频设备层（sounddevice封装）
网络环境	防火墙可能阻断WebSocket	提供HTTP轮询备选方案

4. 安全与隐私约束

风险点	约束要求	实现方式
用户语音数据	内存中驻留时间 ≤1分钟	定时清理音频缓冲区（stt.py）
API密钥存储	禁止硬编码在代码中	环境变量+Vault服务动态获取
内容过滤	实时屏蔽违规词（如政治敏感）	LLMState.py集成关键词黑名单

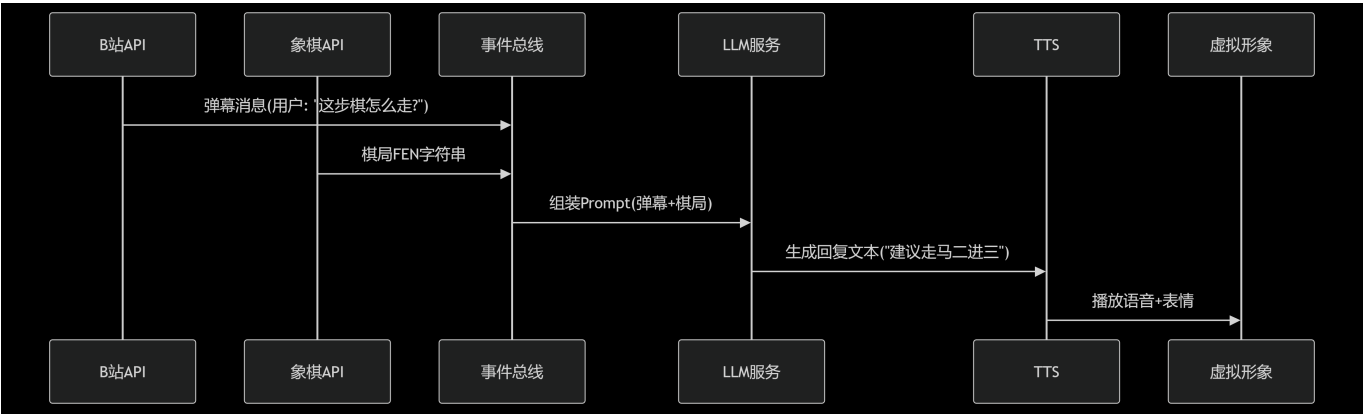
3. 架构视图

3.1 逻辑架构（事件驱动架构）



3.2 事件流程详细设计

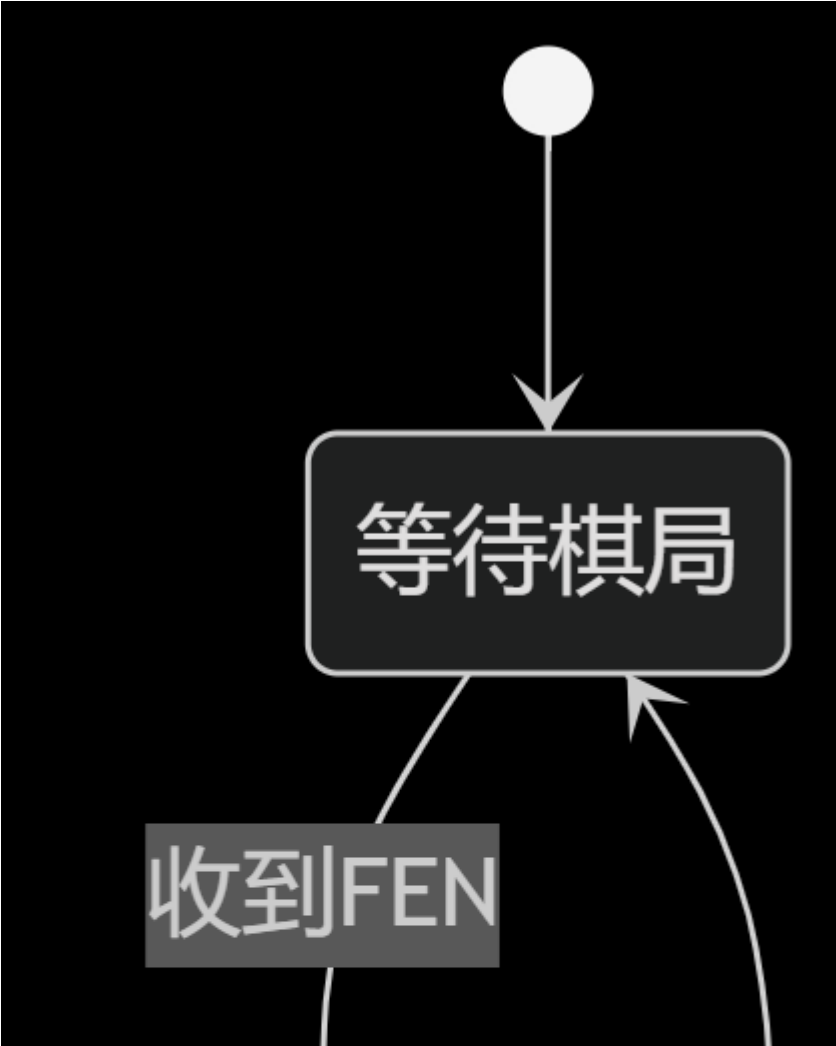
1. B站/象棋API → LLM 对话

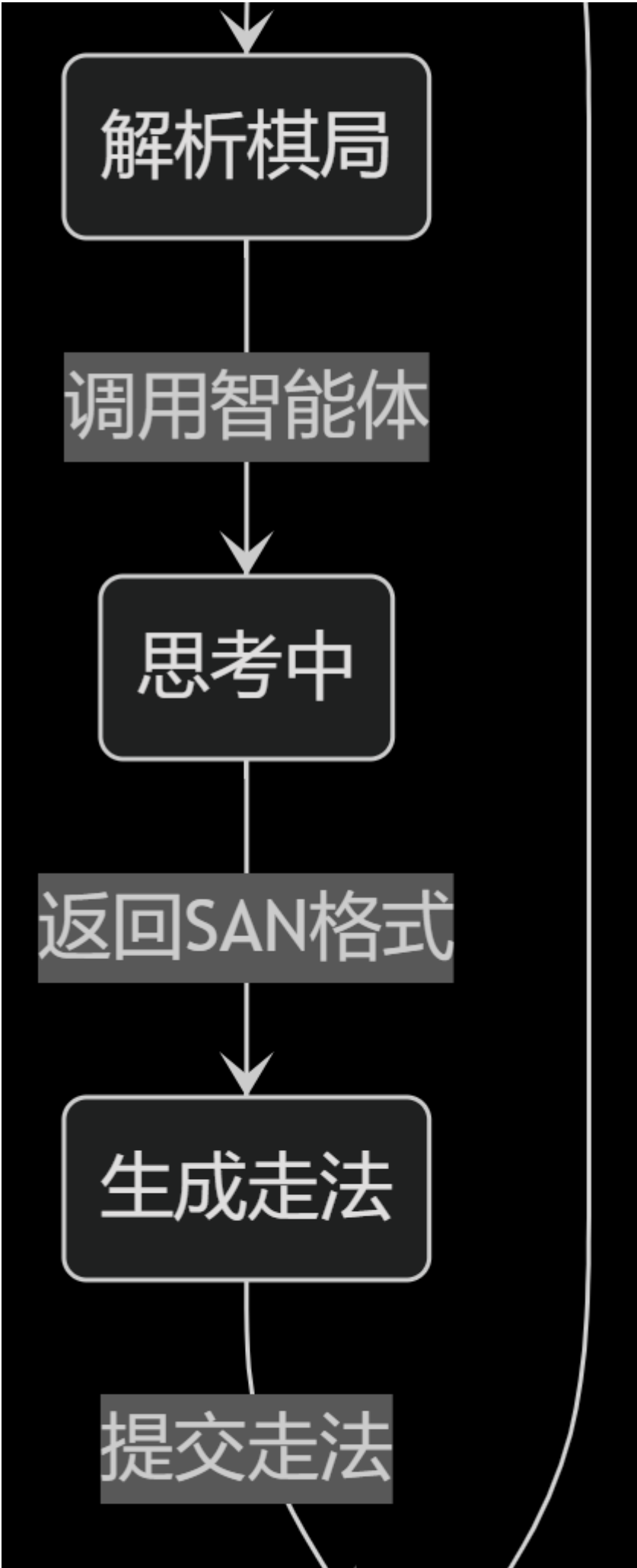


2. LLM回复 → TTS生成

```
# 在textLLMWrapper.py中的处理
async def generate_reply():
    reply = await llm_api.call(prompt)
    if "[chess]" in reply: # 检测象棋相关指令
        signals.chess_command_queue.put(reply)
    else:
        signals.tts_queue.put((reply, "neutral")) # (文本, 默认表情)
```

3. 象棋API → 智能体思考







4. 模块设计

4.1 核心模块职责

模块	职责	关键接口
signals.py	全局状态管理	@property 状态监听
textLLMWrapper.py	LLM请求封装	generate(text, prompts)
room_manager.py	直播间状态同步	join_room(room_id)
module.py	模块基类	run(), cleanup()

4.2 异常处理设计

- 语音服务故障：自动重试3次后降级为纯文本模式
- LLM超时：触发LLMState.cancel_next()并返回预设回复
- B站断连：room_manager.py启动指数退避重连机制

5. 数据设计

5.1 关键数据结构

```
# signals.py中的核心状态
class SystemSignals:
    _recent_messages: Deque[str] # 环形队列存储最近10条消息
    _tts_queue: Queue[Tuple[str, str]] # (文本, 语音参数)
    _AI_expres: Enum('neutral', 'happy', 'angry') # 表情状态机
```

5.2 数据流规范

数据流	协议	频率
用户语音→STT	PCM音频流	50ms/包
LLM回复→TTS	JSON{text, emotion}	异步触发
虚拟形象控制	WebSocket二进制协议	100-300ms/次

数据流	协议	频率
B站直播间信息与心跳控制	B站自带应用层协议	心跳包20s/次

6. 部署视图

6.1 资源需求

组件	CPU	内存	网络
主进程	4核	8GB	10Mbps↑
STT服务	2核	4GB	5Mbps↑
象棋引擎	1核	2GB	低延迟内网