- 技术架构
  - 前端层（Streamlit）
    - 交互组件
      - 输入组件：图片上传框（支持JPG/PNG/JPEG）、录音按钮、文本输入框（chat_input）
      - 配置组件：录音时长滑块（1-10秒）、发音人下拉选择框
      - 功能按钮：清空对话、结束项目
    - 展示组件
      - 对话展示：用户/助手聊天气泡（chat_message）
      - 媒体展示：图片预览、音频播放器（分段+合并）
      - 状态反馈：加载中Spinner、成功/错误提示框
    - 前端工具
      - Web Audio API：音频分段合并、暂停/继续播放
      - HTML嵌入：通过components.v1.html注入JS代码
  - 输入处理层
    - 图片输入处理
      - 步骤：用户上传→获取文件流→Base64编码→拼接data:image头
      - 状态管理：存储至session_state.uploaded_image_base64
      - 去重机制：通过「文件名+文件大小」生成唯一标识，避免重复上传
    - 语音输入处理
      - 录音模块：sounddevice录制→16000采样率→单声道→int16格式
      - 格式转换：生成WAV字节流→截取44字节后PCM数据（适配百度ASR）
      - 识别调用：百度ASR API（dev_pid=1537，中文普通话）→返回识别文本
  - 核心服务层
    - 大模型服务（智谱GLM）
      - GLM-4V（多模态）
        - 调用条件：存在上传图片时触发
        - 输入格式：text（含历史对话+用户问题）+ image_url（Base64编码）
        - 提示词设计：强制先识别品种→再解答问题→错误时主动指正
      - GLM-4（纯文本）
        - 调用条件：无上传图片时触发
        - 输入格式：整合历史对话上下文+当前用户问题
        - 核心要求：个性化回复，贴合宠物养护垂直领域
    - 语音合成服务（百度TTS）
      - 预处理：文本清洗（去除换行/特殊字符/多余空格）→避免合成报错
      - 分段处理：单段≤500字（百度API限制）→超长自动拆分
      - 合成参数：vol=5（音量）、spd=5（语速）、pit=5（音调）、aue=3（MP3格式）
      - 发音人选择：0=女声（默认）、1=男声、3=情感女声、4=情感男声
  - 数据存储层
    - 会话状态管理（streamlit.session_state）
      - chat_history：存储多轮对话（角色+内容）
      - uploaded_image_base64：当前上传图片的Base64编码
      - tts_audio_segments：语音合成后的MP3字节流列表
      - last_image_uploaded：最后一次上传图片的唯一标识
    - 临时存储：tempfile.SpooledTemporaryFile（录音生成的WAV临时缓存）