

# PR #26959 完整报告

sgl-project/sglang

[diffusion] add WebUI

合并时间: 2026-06-02 14:13

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/26959>

## 执行摘要

- 一句话: 新增 LingBot 实时 WebUI 前端, 支持实时视频流交互
- 推荐动作: 该 PR 展示了前端 WebUI 的良好架构设计, 特别是 decoder\_worker 使用 Web Worker 分离解码逻辑, 以及丰富的预设数据管理。值得前端和后端开发者阅读, 理解多模态生成场景下的实时视频流交互模式。

## 功能与动机

该 WebUI 从 LingBot 实时运行时 PR (#26954) 中独立拆分, 以便前端审查与运行时 / 模型变更解耦, 同时为 diffusion 提供交互式实时预览界面。

## 实现拆解

1. 在 python/sglang/multimodal\_gen/apps/realtime\_webui/ 下创建静态资源: index.html 定义页面布局和控件, styles.css 提供视觉样式, app.js 实现核心交互逻辑 (WebSocket 连接、帧渲染、camera 控制、预设切换、遥测统计), decoder\_worker.js 作为 Web Worker 负责解码多种帧格式 (原始 RGB、delta-gzip、WebP/JPEG)。
2. 修改 python/sglang/multimodal\_gen/runtime/entrypoints/http\_server.py, 添加 FastAPI CORSMiddleware (允许所有源、方法、头), 确保浏览器可跨域访问后端 WebSocket 和静态资源。
3. 修改 python/pyproject.toml, 将 multimodal\_gen/apps/realtime\_webui 目录打包进 wheel, 用户安装后可访问该 WebUI。
4. 添加单元测试 test/unit/realtime/test\_realtime\_webui.py, 通过扫描 app.js、index.html、styles.css 的内容验证预设数据、组件 ID、样式类等资产契约, 确保预设不包含摄像头脚本等。
5. 通过 9 次提交迭代优化: 添加更多预设 (Ziggy Stardust)、调整播放帧率至 25fps、修复播放节奏、bump 缓存版本、保留编码帧队列、默认关闭帧插帧、动态查询后端模型名等。

关键文件:

- python/sglang/multimodal\_gen/apps/realtime\_webui/app.js (模块 实时 WebUI; 类别 source; 类型 core-logic; 符号 \$, setStatus, addHistory, drawIdle) : WebUI 核心逻辑, 包含 WebSocket 连接、帧渲染、相机控制、预设切换、遥测统计等全部交互功能
- python/sglang/multimodal\_gen/apps/realtime\_webui/decoder\_worker.js (模块 实时 WebUI; 类别 source; 类型 core-logic; 符号 reset, rawFramesToRgbBuffers) : Web

Worker 解码器，负责将原始 RGB、delta-gzip、WebP/JPEG 帧转换为 RGBA 缓冲区

- `python/sclang/multimodal_gen/test/unit/realtime/test_realtime_webui.py` (模块 实时测试; 类别 `test`; 类型 `test-coverage`; 符号 `test_realtime_webui_presets_do_not_emit_camera_scripts`): 静态资产契约测试, 确保 WebUI 预设和行为符合预期 (不包含摄像头脚本等)
- `python/sclang/multimodal_gen/runtime/entrypoints/http_server.py` (模块 HTTP 服务; 类别 `source`; 类型 `dependency-wiring`): 添加 CORS 中间件以允许浏览器跨域访问 WebSocket 和静态资源
- `python/sclang/multimodal_gen/apps/realtime_webui/styles.css` (模块 实时 WebUI; 类别 `other`; 类型 `core-logic`): WebUI 样式定义, 提供布局、主题色和控件样式
- `python/sclang/multimodal_gen/apps/realtime_webui/index.html` (模块 实时 WebUI; 类别 `other`; 类型 `core-logic`): WebUI 入口 HTML, 定义页面结构和所有控件
- `python/sclang/multimodal_gen/apps/realtime_webui/README.md` (模块 文档; 类别 `docs`; 类型 `documentation`): WebUI 使用说明文档
- `python/pyproject.toml` (模块 构建配置; 类别 `config`; 类型 `configuration`): 将 WebUI 目录打包进 wheel, 确保用户安装后可访问

关键符号: `setStatus`, `resetStreamStats`, `rejectPendingDecodes`, `ensureDecoderWorker`, `resetDecoderState`, `rawFramesToRgbaBuffers`, `test_realtime_webui_presets_do_not_emit_camera_scripts`

## 关键源码片段

### `python/sclang/multimodal_gen/apps/realtime_webui/decoder_worker.js`

Web Worker 解码器, 负责将原始 RGB、delta-gzip、WebP/JPEG 帧转换为 RGBA 缓冲区

```
// 将原始帧数据转换为 RGBA 格式的 ArrayBuffer 数组
// header: 包含 width, height, channels, num_frames, bytes_per_frame
// payload: Uint8Array 原始帧数据
function rawFramesToRgbaBuffers(header, payload) {
  const width = Number(header.width);
  const height = Number(header.height);
  const channels = Number(header.channels);
  const count = Number(header.num_frames);
  const frameBytes = Number(header.bytes_per_frame);
  const pixels = width * height;
  const buffers = [];

  for (let f = 0; f < count; f++) {
    const offset = f * frameBytes;
    // 如果已经是 4 通道, 直接 slice 即可
    if (channels === 4) {
      buffers.push(payload.buffer.slice(
        payload.byteOffset + offset,
        payload.byteOffset + offset + frameBytes,
      ));
    }
  }
}
```

```

    continue;
}

// 非 4 通道需要填充 alpha 通道
const rgba = new Uint8ClampedArray(pixels * 4);
let src = offset;
let dst = 0;
for (let p = 0; p < pixels; p++) {
    rgba[dst++] = payload[src++];
    rgba[dst++] = payload[src++];
    rgba[dst++] = payload[src++];
    src += channels - 3;
    rgba[dst++] = 255; // alpha 通道设为 255
}
buffers.push(rgba.buffer);
}
return buffers;
}

```

## python/sglang/multimodal\_gen/test/unit/realtime/test\_realtime\_webui.py

静态资产契约测试，确保 WebUI 预设和行为符合预期（不包含摄像头脚本等）

```
# SPDX-License-Identifier: Apache-2.0
```

```
from pathlib import Path
```

```

def test_realtime_webui_presets_do_not_emit_camera_scripts():
    """验证 WebUI 预设数据不包含摄像头脚本等危险内容"""
    repo_root = Path(__file__).resolve().parents[6]
    app_js = (repo_root / "python/sglang/multimodal_gen/apps/realtime_webui/app.js").read_text()
    index_html = (repo_root / "python/sglang/multimodal_gen/apps/realtime_webui/index.html").
read_text()
    styles_css = (repo_root / "python/sglang/multimodal_gen/apps/realtime_webui/styles.css").
read_text()

    # 验证预设中没有 action 和重复动作，没有摄像头脚本
    assert "preset.actions" not in app_js
    assert "repeatActions" not in app_js
    assert 'id="eventFrames"' not in index_html # 无事件帧输入
    # 验证标准组件存在
    assert "ControlStateController" in app_js
    assert 'const DEFAULT_PREVIEW_OUTPUT_FORMAT = "webp";' in app_js
    assert 'id="transportFormat"' in index_html
    assert 'id="fps" type="number" value="25"' in index_html
    assert 'id="frameInterpolation" type="checkbox" />' in index_html
    assert 'id="serverUrl" value="ws://127.0.0.1:30000/v1/realtime_video/generate"' in index_
html
    assert '<option value="webp" selected>WebP preview</option>' in index_html

```

```
assert 'id="serverSendText"' in index_html
assert 'id="theoreticalFpsText"' in index_html
assert 'id="renderFps"' in index_html
assert 'id="stageRenderFps"' not in index_html # 没有 stage 级别渲染 FPS
assert "sglang-diffusion Realtime Studio" in index_html
assert "SGLD" not in index_html
# 验证预设顺序
assert app_js.index("Dragon Ride") < app_js.index("Dragon Dolly")
assert app_js.index("Ziggy Stardust") < app_js.index("Plastic Beach")
assert app_js.index("Dragon Dolly") < app_js.index("Kid A")
```

## 评论区精华

该 PR 无公开 review 讨论，作者 mickqian 通过 9 次 commit 独立推进前端优化。PR 描述中强调了从 #26954 拆分以便独立审查前端部分。

- 暂无高价值评论线程

## 风险与影响

- 风险：该 WebUI 强依赖尚未合并的 #26954（实时视频终端和 LingBot 运行时），若后端接口变更可能导致前端不兼容。CORS 中间件设置为 `allow_origins=["*"]` 在生产环境存在安全风险，但本地开发场景常见。测试仅覆盖静态资产契约，未包含集成测试或 e2e 测试，无法验证实际帧渲染和 WebSocket 交互的正确性。前端使用 OffscreenCanvas 和 Web Worker，需要较新浏览器支持。
- 影响：对用户：提供 diffusion 模型的实时交互界面，方便快速预览和参数调整。对系统：增加约 2.5MB（2480 行新增代码）静态资源打包，运行时无额外开销。对团队：实现了前端与后端的独立开发迭代，未来可扩展更多预设和交互功能。
- 风险标记：依赖未合并后端，缺少 e2e 测试，CORS 配置宽松

## 关联脉络

- PR #26954 [diffusion] realtime video endpoint and LingBot runtime support: 该 PR 依赖的后端实时视频终端和运行时支持，提供 WebSocket 接口和推理能力