

PR #43117 完整报告

vllm-project/vllm

fix(processor): route MiMo-V2-Omni media fetch through MediaConnector

合并时间: 2026-07-11 23:52

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/43117>

执行摘要

- 一句话: 修复 MiMo-V2-Omni 处理器 SSRF 漏洞, 移除字符串输入支持
- 推荐动作: 此 PR 值得关注, 展示了安全补丁如何通过审查演进为架构清理。推荐学习其中的抽象边界讨论以及“最小权限”原则的应用。对于维护类似组件的团队, 建议定期审查处理器是否承担了不应由它负责的输入解析。

功能与动机

根据 PR body 和安全报告 GHSA-4hhp-h66f-j5j7, MiMo-V2-Omni 处理器的 `_fetch_image` 和音频加载器直接使用 `requests.get` 和 `Image.open` 处理用户提供的媒体字符串, 绕过了 `chat_utils.py` 中已实现的 `allowed_media_domains` 和 `allowed_local_media_path` 策略, 存在 SSRF 和任意文件读取风险。需要将字符串解析迁移到 vLLM 媒体管道或通过 `MediaConnector` 统一安全策略。

实现拆解

1. 识别两处 SSRF 入口: `_fetch_image` 函数和音频加载器, 它们直接调用 `requests.get(...)` 和 `Image.open(...)` 处理用户输入的 URL 或路径。
2. 初始方案将字符串源通过 `MediaConnector` 路由, 并传递安全策略参数以复用已有过滤。
3. 审查发现 `MediaConnector` 未接收 `allowed_media_domains` 等参数, 且 bytes / 字符串分支在 vLLM 离线路径中为死代码。
4. 先修复参数传递, 添加 `_normalize_media_str` 支持裸本地路径; 随后采纳 @Isotr0py 建议, 彻底删除字符串处理分支及相关函数 (`_fetch_image`、`_to_rgb`、`_normalize_media_str`)。
5. 最终处理器仅接受已解码的媒体输入 (`PIL.Image` 用于图像, `(waveform, sr)` 元组用于音频), 字符串解析完全交由上游 vLLM 媒体管道处理, 并移除了 `requests`、`torchcodec`、`pybase64` 等依赖。

关键文件:

- `vllm/transformers_utils/processors/mimo_v2_omni.py` (模块 多模态; 类别 source; 类型 core-logic; 符号 `_fetch_image`, `_to_rgb`, `_normalize_media_str`, `preprocess_audio`): 唯一变更文件, 也是漏洞所在和修复的核心。移除了所有字符串输入处理, 清除了相关导入和辅助函数, 实现了更安全专注的处理器接口。

关键符号: `_fetch_image`, `_to_rgb`, `_normalize_media_str`, `preprocess_audio`, `MiMoVLProcessor.init`

关键源码片段

`vllm/transformers_utils/processors/mimo_v2_omni.py`

唯一变更文件，也是漏洞所在和修复的核心。移除了所有字符串输入处理，清除了相关导入和辅助函数，实现了更安全专注的处理器接口。

```
# SPDX-License-Identifier: Apache-2.0
# SPDX-FileCopyrightText: Copyright contributors to the vLLM project
# mypy: ignore-errors
"""MiMo-Omni multimodal processor for vLLM.

Ported from SGLang's MiMoV2OmniProcessor / MiMoVLProcessor implementations.
"""

import contextlib
import logging
import math
from collections import OrderedDict
from concurrent.futures import ThreadPoolExecutor, as_completed
from dataclasses import dataclass, field
from typing import Any, Literal

import numpy as np
import regex as re
import torch
import torch.nn.functional as F
from PIL import Image
from transformers import BatchFeature, TensorType
from transformers.processing_utils import ProcessorMixin

# 同时移除了 io、copy、BytesIO、requests、AudioDecoder、pybase64 等导入

try:
    import torchaudio
    from torchaudio.transforms import MelSpectrogram as _MelSpectrogram
    _HAS_TORCHAUDIO = True
except ImportError:
    torchaudio = None
    _MelSpectrogram = None
    _HAS_TORCHAUDIO = False

logger = logging.getLogger(__name__)

# 数据类简化: ImageInput 仅允许 PIL.Image 或 torch.Tensor
@dataclass
class ImageInput:
```

```

# PIL.Image | torch.Tensor (C,H,W)
image: Any
max_pixels: int | None = None
min_pixels: int | None = None

# 数据类简化: AudioInput 仅允许 tuple、np.ndarray、torch.Tensor
@dataclass
class AudioInput:
    # tuple[waveform_1D, sr] | np.ndarray | torch.Tensor (T,n_vq)
    audio: Any

class MiMoVLProcessor:
    """
    Handles image/video/audio preprocessing and token sequence construction.
    Ported from SGLang's MiMoVLProcessor.

    Media strings (URL / path / data:) are resolved upstream in vLLM's media
    pipeline; this processor accepts only already-decoded inputs (`PIL.Image`
    for images, `(waveform, sr)` tuples for audio).
    """
    # 构造器不再需要 MediaConnector 或 allowed_media_domains 参数
    ...

```

评论区精华

- depthfirst-app[bot] 指出 MediaConnector 未传递 allowed_media_domains 参数，导致 SSRF 防护无效（HIGH 严重性）。维护者随后添加了参数传递。
- @Isotr0py 指出处理器中的 bytes 和字符串分支是死代码，建议删除；最终建议仅保留 PIL.Image 和 (waveform, sr) 路径，因为字符串解析应由上游完成。维护者完全遵循该建议，删除了字符串处理和相关导入。
- 整个过程展示了从快速补丁到设计清理的演进，最终实现了更安全的抽象边界。
- MediaConnector 缺少 allowed_media_domains 参数传递 (security): 维护者添加了从模型配置读取 allowed_media_domains 和 allowed_local_media_path 并传递给 MediaConnector 的逻辑。
- 移除 bytes 分支死代码 (style): 维护者删除了 bytes 分支和相关导入。
- 仅保留已解码输入接口 (design): 处理器最终只接受已解码媒体，字符串处理代码完全移除。

风险与影响

- 风险:
 - 兼容性风险: 离线用户若直接向 LLM.generate 传递字符串路径（如 /data/img.jpg）将不再被处理器接受。vLLM 的 LLM.generate 通常会经过上游媒体管道，影响较小，但依赖处理器直接调用的自定义代码可能受影响。
 - 安全风险: 已消除，但需验证上游媒体管道是否完全覆盖所有字符串到媒体的转换，避免绕过。

- 测试覆盖：未新增测试用例，缺少对处理器输入类型变更的回归测试，可能遗漏接口变更引起的崩溃。
- 影响：
 - 用户：MiMo-V2-Omni 模型的用户如果直接调用处理器（少见）需要调整，但正常使用 vLLM API 无影响。
 - 系统：减少了 76 行代码和多个依赖（requests、torchcodec、pybase64），降低了攻击面和维护成本。
 - 团队：处理器现在职责更清晰，仅处理已解码媒体，与上游管道明确分工，有利于长期维护。
 - 风险标记：安全修复，接口变更，缺少测试覆盖，依赖移除

关联脉络

- 暂无明显关联 PR