

PR #51451 完整报告

vllm-project/vllm

[CI] Guard remote-code Transformers compatibility

合并时间: 2026-08-08 12:05

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/51451>

执行摘要

- 一句话: 限制 Intern-S1-Pro 测试至 Transformers 5.14.1
- 推荐动作: 该 PR 虽小, 但展示了如何为依赖 `trust_remote_code` 的模型设置版本保护, 值得测试维护者参考。建议后续系统性地排查其他远程代码模型在 Transformers v5 下的兼容性, 并考虑将 `max_transformers_version` 与 `transformers_version_reason` 作为远程代码模型的通用防御模式。

功能与动机

Intern-S1-Pro 依赖外部 `trust_remote_code` 代码, 其视频处理器导入了 `BASE_VIDEO_PROCESSOR_DOCSTRING`, 而该符号在 Transformers 5.15 中被移除。PR 正文明确指出该失败已在 Buildkite AMD CI 构建 11817 的初始化和处理任务中暴露, 因此需要限定测试版本以保持 CI 稳定。

实现拆解

1. 定位注册条目: 在 `tests/models/registry.py` 的 `_HfExamplesInfo` 映射中找到 `InternS1ProForConditionalGeneration` 条目, 该条目原本只有模型仓库名和 `trust_remote_code=True`。
2. 添加版本上限: 新增 `max_transformers_version="5.14.1"` 字段, 让测试运行器在 Transformers `>= 5.15` 时跳过该条目的覆盖, 从而避免远程视频处理器导入已删除符号导致的初始化失败。
3. 记录失败原因: 在 `transformers_version_reason` 中补充 vllm 侧的解释, 说明远程代码依赖 `BASE_VIDEO_PROCESSOR_DOCSTRING` 且该符号在 Transformers 5.15 中移除, 方便未来维护者判断何时可以解除限制。
4. 配套与验证: 本 PR 无源码与配置改动, 仅测试注册表变更; 通过 `github-actions[bot]` 触发 Buildkite CI #82932 验证。

关键文件:

- `tests/models/registry.py` (模块 注册表; 类别 `test`; 类型 `test-coverage`): 这是本 PR 唯一修改的文件, 为 `InternS1ProForConditionalGeneration` 注册条目添加 `max_transformers_version="5.14.1"` 和 `transformers_version_reason`, 直接解决了远程代码在 Transformers 5.15 中的兼容性问题, 防止 CI 失败。

关键符号: 未识别

关键源码片段

tests/models/registry.py

这是本 PR 唯一修改的文件，为 `InternS1ProForConditionalGeneration` 注册条目添加 `max_transformers_version="5.14.1"` 和 `transformers_version_reason`，直接解决了远程代码在 `Transformers 5.15` 中的兼容性问题，防止 CI 失败。

```
# tests/models/registry.py 中 Intern-S1-Pro 的注册配置
"InternS1ProForConditionalGeneration": _HfExamplesInfo(
    "internlm/Intern-S1-Pro",
    trust_remote_code=True,
    # 限制 Transformers 版本：远程视频处理器导入的
    # `BASE_VIDEO_PROCESSOR_DOCSTRING` 在 5.15 中被移除
    max_transformers_version="5.14.1",
    transformers_version_reason={
        "vllm": (
            "Remote video processor code imports "
            "`BASE_VIDEO_PROCESSOR_DOCSTRING`, removed in Transformers 5.15."
        )
    },
),
```

评论区精华

本 PR 没有实质性的 review 讨论。[DarkLight1337](#) 直接批准，[claude\[bot\]](#) 仅发布了手册式提醒，未涉及技术争议。[github-actions\[bot\]](#) 成功触发了 Buildkite CI #82932 进行验证。

- 暂无高价值评论线程

风险与影响

- 风险：
 1. 版本上限可能过于严格：如果未来 Transformers 恢复该符号或远程代码更新，测试仍会被跳过，但 `transformers_version_reason` 提供了清晰的追踪线索。
 2. 仅保护单一模型条目，其他 `trust_remote_code` 模型可能仍有类似问题，需要后续系统性排查。
 3. 变更仅涉及测试注册表，不触及运行时路径，回归风险极低。 - 影响：对用户与系统运行无任何影响，因为变更仅在测试注册表。对团队 CI 稳定性有正面作用，避免因 Transformers 升级导致 AMD CI 中断；同时可能减少该模型在新版本上的测试覆盖，但该覆盖依赖版本原因记录可追踪。 - 风险标记：暂无

关联脉络

- PR #51408 [1/N] Harden Transformers modelling backend multi-modal path: 同为 Transformers 兼容性与多模态远程代码问题的加固工作，属于同一演进脉络。

- PR #51427 [Bugfix][models_multimodal] Remote HF python code misses importing class: 同属远程代码兼容性修复, 关注 Transformers 远程代码导入问题。