

PR #24174 完整报告

sgl-project/sglang

ci: temporarily disable multimodal-gen-test-1-b200

合并时间: 2026-05-01 07:16

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/24174>

执行摘要

- 一句话: 临时禁用 CI 中断言失败的多模态任务
- 推荐动作: 该 PR 是一个典型的短期 CI 修复措施, 内容简单直接。建议阅读者确认后续 tokenizer 修复后及时恢复该任务, 避免长期测试覆盖缺失。设计决策亮点在于使用 `if: false` 而非完全移除 job, 保留了任务定义和恢复路径。

功能与动机

该 PR 旨在解决 `multimodal-gen-test-1-b200` 任务在每次运行中确定性失败的问题, 失败原因为 `RobertaProcessing.__new__() got an unexpected keyword argument 'cls'` (tokenizer/transformers 兼容性问题), 该失败阻塞了所有 CI 流程。PR 描述中明确提到 'blocking CI on every run', 并提供了失败任务链接。

实现拆解

1. 定位失败任务: 在 `.github/workflows/pr-test-multimodal-gen.yml` 中找到 `multimodal-gen-test-1-b200` job。
2. 修改 job 条件: 将 job 的 `if` 条件改为 `if: false`, 使其在每次 CI 触发时都被跳过。
3. 保留原始条件作为注释: 在 `if: false` 下方以注释形式保留原始条件, 方便后续通过取消注释 `if: false` 并取消注释原始条件来快速恢复。
4. 添加 TODO 注释: 在代码中明确注释说明临时禁用的原因以及恢复条件 (待 tokenizer 修复后)。
5. 无其他文件修改: 仅此一个文件变更, 无测试、配置或部署配套改动。

关键文件:

- `.github/workflows/pr-test-multimodal-gen.yml` (模块 CI 配置; 类别 `infra`; 类型 `infrastructure`): 唯一变更的文件, 通过修改 job 的 `if` 条件临时禁用了 B200 多模态测试任务。

关键符号: 未识别

关键源码片段

[.github/workflows/pr-test-multimodal-gen.yml](#)

唯一变更的文件, 通过修改 job 的 `if` 条件临时禁用了 B200 多模态测试任务。

```
# ...
# job 定义开始
multimodal-gen-test-1-b200:
  # TODO: Temporarily disabled (`if: false`) — deterministically failing
  # with `RobertaProcessing.__new__()` got an unexpected keyword argument
  # 'cls'` (tokenizer/transformers compatibility) and blocking CI.
  # Restore the original `if:` condition (commented below) once the
  # underlying fix lands.
  if: false
  # 原始条件（待恢复）：
  # if: |
  # (inputs.target_stage == 'multimodal-gen-test-1-b200') ||
  # (
  # !inputs.target_stage &&
  # ((github.event_name == 'schedule' || inputs.test_parallel_dispatch == 'true') || (inputs.caller_
  needs_failure != 'true' && !cancelled())) &&
  # inputs.multimodal_gen == 'true'
  # )
  runs-on: ${{ inputs.b200_runner }}
  timeout-minutes: 240
  steps:
    # ... 步骤保持不变
```

评论区精华

该 PR 的 review 评论数为 0，没有公开的讨论。但提交历史显示从最初的“注释掉整个 job”演变为“使用 `if: false` 并保留原始条件作为注释”，体现了 reviewer 或许建议保持 job 定义完整、仅通过条件跳过的做法。

- 暂无高价值评论线程

风险与影响

- 风险：
 - 回归风险：低。仅禁用一个已知失败的任务，不影响其他任务和代码库。
 - 性能影响：无。
 - 安全风险：无。
 - 兼容性风险：无。
 - 潜在遗漏风险：如果上游 tokenizer 修复未及时集成，该任务可能长期处于禁用状态，导致相关测试覆盖缺失。
- 影响：
 - 用户影响：无，PR 仅涉及 CI 配置。
 - 系统影响：CI 中该任务不再执行，避免阻塞其他任务，提高 CI 流程效率。
 - 团队影响：维护者需关注上游 tokenizer 修复进展，及时恢复该任务。
 - 影响范围：仅影响 CI 中的一个任务，影响程度低。

- 风险标记：低风险，临时禁用

关联脉络

- 暂无明显关联 PR