

# PR #27785 完整报告

sgl-project/sglang

Fix PR-close cancellation skipping workflows beyond the first 30

合并时间: 2026-06-13 06:39

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/27785>

## 执行摘要

- 一句话: 修复 PR 关闭时取消工作流遗漏后续工作流的问题
- 推荐动作: 建议阅读该 PR 的实现, 特别关注如何使用 `gh api` 组合分页和过滤来实现可靠的取消逻辑。该模式可用于其他需要针对分支取消运行或清理资源的场景。代码内联注释清晰, 易于理解。

## 功能与动机

PR body 指出, `style/cancel-workflow-action` 的 `workflow_id: all` fetches only the first page (30) of repo workflows, so runs of newer workflows (e.g. `pr-test-extra.yml`, observed on #23906 holding GPU runners for ~55 min after merge) were never cancelled on PR close. 需要替换为更全面的取消方式。

## 实现拆解

1. 删除旧工作流: 移除 `.github/workflows/cancel-pr-workflow-on-merge.yml`, 该文件使用 `style/cancel-workflow-action` 且 `workflow_id: all`, 只能获取前 30 个工作流。
2. 新建工作流: 创建 `.github/workflows/cancel-pr-workflows-on-close.yml`, 使用 `pull_request_target` 触发器在 PR 关闭时触发, 设置 `actions: write` 权限。
3. 列出并过滤运行: 在 `run` 步骤中通过 `gh api` 列出仓库中指定 branch 的所有运行, 用 `jq` 过滤出 `head_repository.id` 与当前 PR head repo ID 匹配且运行 ID 不等于当前运行 ID 的运行。遍历所有未完成状态 (`queued`, `in_progress`, `waiting`, `pending`, `requested`, `action_required`), 并执行两轮以避免时序问题; 每个状态列表最多重试 3 次。
4. 取消与降级: 对每个运行 ID 先尝试 `gh run cancel`, 若失败则调用 `force-cancel` API (处理被 `approval/deployment protection` 阻塞的情况); 若运行已在间隔中完成则忽略。记录错误到 `FAILURES` 变量。
5. 错误处理: 使用 `set -euo pipefail`, 最后以 `exit $FAILURES` 退出, 确保步骤失败能被工作流捕获。

关键文件:

- `.github/workflows/cancel-pr-workflows-on-close.yml` (模块 CI 工作流; 类别 `infra`; 类型 `infrastructure`): 核心变更文件: 新增工作流, 实现基于 `head branch` 和 `repo id` 的全面取消逻辑, 替换了旧的有限取消方式。

- `.github/workflows/cancel-pr-workflow-on-merge.yml` (模块 CI 工作流; 类别 `infra`; 类型 `deletion`): 旧文件, 使用有问题的 `styfle/cancel-workflow-action`, 已被删除。

关键符号: 未识别

## 评论区精华

无有效 review 讨论。仅有一条来自 `gemini-code-assist[bot]` 的自动回复, 表示无法因文件类型不支持而生成审查。

- 暂无高价值评论线程

## 风险与影响

- 风险: 风险较低。主要风险包括: 1) `gh api` 可能因网络或限频失败, 但代码内置了重试机制; 2) `pull_request_target` 在 `fork` 场景下需确保 `head_repo.id` 可用, 代码中有空值检查并提前退出; 3) `GITHUB_TOKEN` 权限可能不足以 `force-cancel` 某些运行, 但已处理错误记录。该变更仅影响 CI 取消行为, 不涉及生产逻辑。
- 影响: 影响范围限于 PR 关闭后的 CI 运行取消。此前未取消的工作流现在会被及时取消, 释放 GPU Runner 资源, 减少资源浪费。对频繁合并 PR 的仓库节省显著。影响程度中等偏下。
- 风险标记: 简单的 CI 配置变更, API 限频风险, 权限依赖

## 关联脉络

- 暂无明显关联 PR