

PR #22031 完整报告

sgl-project/sglang

[diffusion] CI: temporarily disable accuracy ci

合并时间: 2026-04-03 17:39

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/22031>

执行摘要

本 PR 临时禁用了多模态生成 (diffusion) 相关的准确性 CI 测试, 通过删除三个 CI 配置文件中的作业定义和引用, 以减少因测试不稳定导致的 CI 失败。这是一个基础设施调整, 待测试稳定后需恢复, 风险可控但需团队监控。

功能与动机

由于多模态生成准确性测试当前不稳定, 频繁导致 CI 失败和噪声, 作者 mickqian 在 Issue 评论中指示“@Ratish1 please add it back only when it becomes stable”, 因此此 PR 旨在临时禁用这些测试, 确保 CI 流程顺畅。这反映了团队在开发过程中对测试稳定性的临时处理策略。

实现拆解

改动集中在三个文件, 均为删除操作:

- .github/workflows/pr-test-multimodal-gen.yml: 移除了以下两个作业定义: - multimodal-gen-component-accuracy-1-gpu - multimodal-gen-component-accuracy-2-gpu 这些作业原本负责在 1-GPU 和 2-GPU H100 环境中运行扩散组件准确性测试。
- .github/workflows/pr-test.yml: 删除了对上述作业的引用, 确保主 CI 工作流不触发这些测试。
- scripts/ci/utils/slash_command_handler.py: 更新了 handle_rerun_stage 函数中的作业名称列表, 移除无效项, 防止 CI 命令错误。

评论区精华

没有正式的 review 讨论。在关联 Issue 的评论中, 作者 mickqian 指出:

@Ratish1 please add it back only when it becomes stable

这明确了此 PR 是临时措施, 强调恢复条件依赖于测试稳定性, 为团队提供了明确的操作指导。

风险与影响

风险:

- 禁用准确性测试可能掩盖模型输出回归问题, 尤其是多模态生成功能的代码变更。
- 如果忘记恢复, 长期测试覆盖率下降, 增加潜在 bug 风险。

- 无代码逻辑变更，因此不影响性能、安全或兼容性。

影响：

- CI 执行时间可能缩短，减少团队因测试失败产生的干扰。
- 对最终用户无直接影响，但开发团队需依赖其他测试（如单元测试）确保质量。
- 需要建立跟踪机制，确保在测试稳定后及时恢复，避免技术债累积。

关联脉络

从近期历史 PR 看，此 PR 是团队持续优化 CI 流程的一部分：

- PR 22001 修复 CI 作业名称重复问题，提升安全性。
- PR 22010 调整内核版本更新工作流，减少不必要测试触发。这些 PR 共同反映了团队对 CI 效率、稳定性和安全性的关注，本 PR 作为临时调整，应与这些优化措施结合理解，未来可能通过类似 PR（如恢复测试）完成闭环。