

PR #26985 完整报告

sgl-project/sglang

ci: drop redundant multimodal-server jobs from nightly (Nvidia)

合并时间: 2026-06-02 14:40

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/26985>

执行摘要

PR #26985 从 NVIDIA nightly CI 中移除了两个冗余的多模态 server 测试 job，因为这些 job 已在 per-commit 的 `pr-test-multimodal-gen.yml` 中完全覆盖。变更仅涉及单个 CI 配置文件，删除 123 行，风险极低，有助于节省 nightly 运行资源。

功能与动机

根据 PR 描述，`nightly-test-multimodal-server-1-gpu` 和 `nightly-test-multimodal-server-2-gpu` 这两个 job 执行的 diffusion 测试套件（`multimodal_gen/test/run_suite.py --suite 1-gpu` 和 `--suite 2-gpu`）与 `pr-test-multimodal-gen.yml` 在每次提交时运行的内容完全相同。nightly 中的重复运行不增加任何额外覆盖，却消耗宝贵的 GPU 时间，因此删除它们是合理的 CI 资源精简。

实现拆解

- 删除 job 定义：从 `.github/workflows/nightly-test-nvidia.yml` 中移除了两个完整 job 块，每个包含 checkout、依赖安装、测试执行、指标收集和 artifact 上传等步骤。
- 更新 `workflow_dispatch` 选项：在 `on.workflow_dispatch.inputs.job_filter.options` 列表中删除了对应 job 名称，使手动触发时 UI 不再列出已移除的选项。
- 调整下游依赖：在 `consolidate-metrics` 和 `check-all-jobs` 的 `needs:` 字段中移除了这两个 job，避免依赖等待不存在的任务。

由于本 PR 仅涉及 CI 配置删除，不涉及业务逻辑代码，因此无需展示源码片段。

评论区精华

无实质性讨论，reviewer mickqian 直接批准合并。

风险与影响

- 风险：极低。被删 job 的所有测试已在 per-commit 流程中覆盖，且后续 job 依赖已正确更新，不会导致 CI 中断。
- 影响：减少 nightly 测试时间与 GPU 消耗，提升 CI 整体效率。对普通开发者无影响。

关联脉络

本 PR 是 CI 基础设施优化的延续，与近期 #26990（禁用跨任务快速失败）等 PR 共同体现了项目组对 CI 效率与可维护性的关注。