

# PR #21978 完整报告

sgl-project/sglang

[CI] Remove stale Ascend suite entries from test/srt/run\_suite.py

合并时间: 2026-04-03 07:47

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/21978>

## 执行摘要

- 一句话: 清理 CI 测试套件中过时的 Ascend NPU 条目, 修复 XPU CI 检查失败。
- 推荐动作: 该 PR 变更简单直接, 适合快速浏览以了解 CI 套件清理模式。值得关注的是如何通过 `__not_in_ci__` 处理遗留文件, 以及 sanity check 机制对文件存在性的依赖。

## 功能与动机

根据 PR body 描述, Ascend NPU 测试已迁移到新的注册制 CI 系统 ( `test/registered/ascend/`, 由 `test/run_suite.py` 管理), 但旧条目未清理。这导致 XPU CI 的 sanity check 失败, 因为 `_sanity_check_suites()` 断言所有缺失的 Ascend 测试文件都存在, 即使 XPU 从不运行这些测试。具体失败见 <https://github.com/sgl-project/sglang/actions/runs/23919335318/job/69761428239>。

## 实现拆解

1. 删除 `test/srt/run_suite.py` 中整个 `suite_ascend` 字典定义 (37 行删除), 包括 `per-commit-1-npu-a2`、`per-commit-2-npu-a2`、`per-commit-4-npu-a3`、`per-commit-16-npu-a3` 四个子套件。
2. 将唯一保留的本地文件 `ascend/test_embed_interpolate_unittest.py` 移至 `__not_in_ci__` 列表 (3 行添加)。
3. 移除 `suites.update(suite_ascend)` 调用。

关键文件:

- `test/srt/run_suite.py` (模块 CI/ 测试基础设施): 唯一修改文件, 包含 CI 测试套件定义, 清理过时 Ascend 条目修复 XPU CI 失败。

关键符号: `_sanity_check_suites()` (间接影响)

## 评论区精华

无 review 评论, 但关联 Issue 评论显示: 1. `gemini-code-assist[bot]` 提示配额限制。2. `Fridge003` 执行 `/tag-and-rerun-ci`。3. `Godmook` 确认此 PR 已修复问题, 并提及 #21916 (可能为相关 PR)。

- 修复 CI 失败确认 (other): PR 有效解决了 XPU CI 的 sanity check 失败。

## 风险与影响

- 风险：1. 回归风险：删除 suite\_ascend 可能意外影响依赖旧路径的脚本或文档，但新注册系统已就绪，风险较低。2. 兼容性风险：\_\_not\_in\_ci\_\_列表添加文件可能被误用，但该文件本就不在 CI 中，影响有限。3. 逻辑风险：sanity check 逻辑假设所有套件条目对应真实文件，此次清理正修复此假设，无新增风险。
- 影响：1. 对系统：修复 XPU CI 稳定性，避免因过时条目导致失败；清理代码库，减少技术债务。2. 对团队：CI 运行更可靠，减少调试时间；明确测试套件管理边界。3. 对用户：无直接影响，属内部基础设施改进。
- 风险标记：清理过时条目，CI 依赖假设

## 关联脉络

- PR #21916 未知（从评论提及推断）：Godmook 在评论中提及此 PR，可能为相关 CI 修复或 Ascend 迁移 PR。
- PR #21408 [NPU] Support GLM-4.7-Flash on NPU: 同属 NPU 相关 PR，涉及硬件后端和测试，可能共享类似 CI 上下文。