

# PR #42197 完整报告

vllm-project/vllm

Fix mypy failure on main

合并时间: 2026-05-10 15:55

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/42197>

## 执行摘要

- 一句话: 修复 NVFP4 SP 测试 mypy 失败
- 推荐动作: 合并即可, 无需深入审阅。但值得关注 #41362 的长期修复方案。

## 功能与动机

修复 main 分支上因最近合入的 PR #33322 和 PR #41882 导致的 mypy 失败, 保持 CI 健康。

## 实现拆解

1. 定位问题: PR #33322 在 `_compare_sp` 函数中增加了必需的 `enable_prompt_embeds` 参数, 但 PR #41882 新增的 `test_tp_sp_nvfp4_generation` 调用时未传入该参数, 导致 mypy 静态检查报错。
2. 修复方式: 在 `tests/compile/correctness_e2e/test_sequence_parallel.py` 的 `test_tp_sp_nvfp4_generation` 函数中, 向 `_compare_sp` 调用添加 `enable_prompt_embeds=False` 参数。
3. 无其他文件修改, 仅测试文件中一行变更。

关键文件:

- `tests/compile/correctness_e2e/test_sequence_parallel.py` (模块测试; 类别 test; 类型 test-coverage; 符号 `test_tp_sp_nvfp4_generation`): 修复 mypy 失败的关键文件: 为 `_compare_sp` 调用增加 `enable_prompt_embeds=False` 参数, 对齐函数签名。

关键符号: `test_tp_sp_nvfp4_generation`

## 评论区精华

评论中提到该问题实际上已被 PR #41362 修复, 但由于队列积压, 其他 PR 尚需等待合并。本 PR 作为快速补丁解决了当前 CI 故障。

- mypy 失败修复方案 (question): 本 PR 作为临时补丁, 合并即可。

## 风险与影响

- 风险: 风险极低: 修改仅涉及一行测试参数传递, 不会影响运行逻辑, 且已获得两位维护者批准。

- 影响：影响范围仅限于修复 mypy CI 失败，使后续 PR 能够通过静态检查。对用户无直接影响。
- 风险标记：暂无

## 关联脉络

- PR #33322 [Bugfix] Fix SP pass for multimodal models and PP+SP residual handling: 使 `enable_prompt_embeds` 成为 `_compare_sp` 必需参数，直接导致本 PR 修复的问题。
- PR #41882 Add NVFP4 all-gather GEMM fusion for AsyncTP: 新增了调用 `_compare_sp` 但未传必需参数的测试函数，触发了 mypy 失败。
- PR #41362（未在列表中，但评论区提到）：已修复相同问题，本 PR 是临时补丁。