

# PR #24105 完整报告

sgl-project/sglang

relax the threshold in test\_step3p5\_flash\_chain\_mtp

合并时间: 2026-04-30 07:53

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/24105>

## 执行摘要

- 一句话: 放宽 step3.5 flash chain MTP 测试阈值
- 推荐动作: 可直接合并, 无需精读。这是一个常规的 flaky 测试修复, 值得注意的是一致性: 同仓库近期有多个类似阈值调整 PR (如 PR#24026 修复参数缺失), 但本 PR 仅为测试维护, 无深层技术洞察。

## 功能与动机

PR body 指出此变更用于修复 CI flaky 问题, 具体引用 CI 运行链接:

<https://github.com/sgl-project/sglang/actions/runs/25086309626/job/73673927890?pr=23456>, 说明因阈值过紧导致测试不稳定。

## 实现拆解

1. 修改文件 `test/registered/8-gpu-models/test_step3p5_flash_chain_mtp.py` 第 100 行, 将 `self.assertGreater(metrics["score"], 0.84)` 改为 `self.assertGreater(metrics["score"], 0.83)`。
2. 仅此一处变更, +1/-1, 无新增测试或逻辑改动。

关键文件:

- `test/registered/8-gpu-models/test_step3p5_flash_chain_mtp.py` (模块测试; 类别 test; 类型 test-coverage): 唯一变更文件, 将 GSM8K 准确率断言阈值从 0.84 下调至 0.83。

关键符号: 未识别

## 关键源码片段

`test/registered/8-gpu-models/test_step3p5_flash_chain_mtp.py`

唯一变更文件, 将 GSM8K 准确率断言阈值从 0.84 下调至 0.83。

```
# test_step3p5_flash_chain_mtp.py (partial)
if is_in_ci():
    write_github_step_summary(
        f"### test_gsm8k (step-3.5-flash chain mtp)\n"
        f'{metrics["score"]=.3f}\n'
        f'{avg_spec_accept_length=.2f}\n'
    )
```

```
# 阈值从 0.84 调整为 0.83, 消除 CI flaky 失败
self.assertGreater(metrics["score"], 0.83)
self.assertGreater(avg_spec_accept_length, 2.6)
```

## 评论区精华

无 review 评论, 提交仅一条 message 为 'upd', 表明为快速修复。

- 暂无高价值评论线程

## 风险与影响

- 风险: 极小改动, 仅调整测试断言阈值, 不影响任何生产逻辑或其他测试。风险极低。
- 影响: 仅影响该测试在 CI 中的稳定性。放宽阈值后, 测试通过概率提高, 避免频繁重跑 CI 带来的资源浪费。对用户无影响。
- 风险标记: 暂无

## 关联脉络

- 暂无明显关联 PR