

PR #27082 完整报告

sgl-project/sglang

test: disable test_gemma4_mtp_26b_a4b_extra from CI

合并时间: 2026-06-03 05:00

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/27082>

执行摘要

- 一句话: 禁用 Gemma4 MTP 26B-A4B extra 测试
- 推荐动作: 该 PR 为一次标准的 CI 测试禁用操作, 变更简单明确, 无需精读。

功能与动机

PR body 指出: Temporarily drop the Gemma4 MTP 26B-A4B `extra` test from CI due to an accuracy issue。

实现拆解

1. 在 `test/registered/spec/test_gemma4_mtp_26b_a4b_extra.py` 中找到 `register_cuda_ci` 调用。
2. 将单行调用改为多行, 添加 `disabled="FIXME(kpham-sgl): temporary drop due to accuracies issue"` 参数。
3. 该参数利用 CI 注册的 AST 收集机制: 当 `disabled` 为非 `None` 字符串时, 测试注册被标记为禁用, CI 调度器不再分发该测试。
4. 测试类与 `__main__` 块未修改, 因此本地执行 `python3 test_gemma4_mtp_26b_a4b_extra.py` 仍能正常运行。

关键文件:

- `test/registered/spec/test_gemma4_mtp_26b_a4b_extra.py` (模块 `推测解码`; 类别 `test`; 类型 `test-coverage`): 唯一变更文件, 通过添加 `disabled` 参数禁用 CI 测试。

关键符号: 未识别

关键源码片段

`test/registered/spec/test_gemma4_mtp_26b_a4b_extra.py`

唯一变更文件, 通过添加 `disabled` 参数禁用 CI 测试。

```
# 变更前: register_cuda_ci(est_time=720, stage="extra-a", runner_config="2-gpu-large")
```

```
# 变更后:
```

```
register_cuda_ci(  
    est_time=720,  
    stage="extra-a",
```

```
runner_config="2-gpu-large",
disabled="FIXME(kpham-sgl): temporary drop due to accuracies issue", # 非 None 值使 CI
收集器跳过此注册
)
```

评论区精华

无 review 讨论。

- 暂无高价值评论线程

风险与影响

- 风险：无风险。变更仅影响 CI 注册元数据，不涉及运行时逻辑。若精度问题后续修复，需手动移除 disabled 参数以重新启用 CI 测试。
- 影响：影响范围仅限于 CI 调度：该 extra 测试不再在 PR 或 nightly CI 中执行。本地开发不受影响。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR