

PR #28934 完整报告

sgl-project/sglang

[MUSA][24/N] CI:Fix LLM server smoke test

合并时间: 2026-06-24 13:58

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/28934>

执行摘要

- 一句话: 修复 MUSA CI 烟感测试断言
- 推荐动作: 该 PR 是极小范围的测试健壮性修复, 无需精读。可关注其变更模式: 测试断言中先 strip 再 lower, 是一种通用的字符串匹配防御写法。

功能与动机

MUSA LLM server smoke test 中 chat completion 响应断言过于严格, 模型返回有效输出时可能因前导空格 (如 " Beijing") 导致误判失败, 影响 CI 可靠性。PR body 明确说明: "Avoid false failures when the model returns valid output with leading whitespace."

实现拆解

1. 在 test/registered/musa/test_llm_server_smoke_musa.py 的 test_send_receive_chat_message_contains_beijing 方法中, 将第 85 行的断言语句从 self.assertIn("Beijing", content.lower()) 改为 self.assertIn("beijing", content.strip().lower())。
2. 主要改动: 对 content 字符串先调用 .strip() 去除两端空白, 再转为小写进行子串匹配。预期子串从 "Beijing" 改为全小写 "beijing", 保持大小写一致性。
3. 该文件是 MUSA 硬件后端的 LLM server 烟感测试, 属于端到端测试, 不涉及核心逻辑变更, 仅修正测试断言健壮性。

关键文件:

- test/registered/musa/test_llm_server_smoke_musa.py (模块 CI 测试; 类别 test; 类型 test-coverage) : 唯一变更文件, 修复断言避免前导空格导致误判

关键符号: 未识别

评论区精华

无 review 讨论。Review 由 yeahdongcn 直接批准。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低，仅修改一行测试断言，且由原大写匹配改为全小写匹配并加 strip，误报概率降低。不会影响生产逻辑或其他后端测试。
- 影响：仅影响 MUSA 后端的 LLM server 烟感测试（CI nightly），确保模型返回合法输出时测试通过，提高 CI 稳定性。对其他后端、用户功能无影响。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR