

# PR #21561 完整报告

sgl-project/sglang

test: point DSV3 int8 MLA CI models to lmsys Hugging Face org

合并时间: 2026-03-28 04:04

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/21561>

## 执行摘要

本次 PR 将 DeepSeek V3 int8 MLA CI 测试中使用的 Hugging Face 模型 ID 从 `sgl-project/*` 更新为 `lmsys/*`, 涉及 `channel int8`、`block int8` 权重及 EAGLE draft model 路径。这是一个简单的字符串替换变更, 旨在维护 CI 测试的可用性, 风险低且无功能影响。

## 功能与动机

PR body 说明更新 Hugging Face 模型 IDs, 以适配模型存储位置的变更。具体动机未在材料中详细阐述, 推测是为了将测试依赖的模型迁移到 lmsys 组织下, 确保 CI 测试能正确运行, 避免因模型路径失效导致测试失败。

## 实现拆解

变更仅涉及一个文件: `test/registered/mla/test_mla_int8_deepseek_v3.py`。关键改动如下:

- 在 `TestMLADeepseekV3ChannelInt8` 类的 `setUpClass` 方法中, 将 `cls.model` 从 `"sgl-project/sglang-ci-dsv3-channel-int8-test"` 替换为 `"lmsys/sglang-ci-dsv3-channel-int8-test"`。
- 类似地, 在 `TestDeepseekV3MTPChannelInt8`、`TestMLADeepseekV3BlockInt8` 和 `TestDeepseekV3MTPBlockInt8` 类中更新相应模型路径。
- 在 `TestDeepseekV3MTPChannelInt8` 的 `setUpClass` 中, 更新 EAGLE draft model path 从 `"sgl-project/sglang-ci-dsv3-channel-int8-test-NextN"` 到 `"lmsys/sglang-ci-dsv3-channel-int8-test-NextN"`。

所有修改均为直接的字符串替换, 不涉及任何逻辑调整。

## 评论区精华

review 中仅有一个来自 `gemini-code-assist[bot]` 的自动评论: "I have no feedback to provide.", 无实质性讨论或技术交锋。

## 风险与影响

风险分析:

- 主要风险是模型路径变更后, 如果 lmsys 组织下的模型不可访问或权重不兼容, CI 测试可能失败。但鉴于这是维护性更新, 风险较低。
- 无回归、性能、安全或兼容性问题, 因为变更仅限于测试配置。

影响分析：

- 对用户：无直接影响。
- 对系统：确保 DeepSeek V3 int8 量化相关的 MLA 测试能正常运行，维护 CI 稳定性。
- 对团队：属于常规基础设施维护，影响范围小。

## 关联脉络

从近期历史 PR 看，本 PR 与以下变更相关：

- PR 21485 “Remove redundant DeepSeek V3 FP4 PCG test”：同涉及 DeepSeek V3 测试的清理，优化 CI 效率。
- PR 21547 “[CI] Register missing jit\_kernel test files”：都关注 CI 测试文件的维护和更新，体现团队对测试基础设施的持续改进。

这些关联表明，项目在持续优化测试套件，特别是针对 DeepSeek 模型和量化场景的 CI 配置。