

PR #42176 完整报告

vllm-project/vllm

[CI/Build] Use modelscope's international site for regression test

合并时间: 2026-05-10 10:47

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/42176>

执行摘要

- 一句话: 回归测试改用 ModelScope 国际站
- 推荐动作: 该 PR 是简单的 CI 基础设施调整, 无需深入精读, 但可供关注 CI 稳定性的团队参考。

功能与动机

CI 的 AWS 机器有时难以连接到默认的 modelscope.cn 国内站, 使用国际站 modelscope.ai 对 CI 机器更稳定。

实现拆解

在 tests/test_regression.py 的 test_model_from_modelscope 函数中:

1. 将注释中的模型 URL 从 <https://modelscope.cn/models/qwen/Qwen1.5-0.5B-Chat/summary> 更新为 <https://www.modelscope.ai/models/qwen/Qwen1.5-0.5B-Chat>。
2. 在 monkeypatch 上下文中新增 `m.setenv("MODELSCOPE_DOMAIN", "www.modelscope.ai")`, 显式指定使用国际站域名。

关键文件:

- tests/test_regression.py (模块 回归测试; 类别 test; 类型 test-coverage) : 唯一变更文件, 修改 ModelScope 域名和环境变量设置以提升 CI 稳定性。

关键符号: test_model_from_modelscope

关键源码片段

tests/test_regression.py

唯一变更文件, 修改 ModelScope 域名和环境变量设置以提升 CI 稳定性。

```
# tests/test_regression.py (修改部分)
def test_model_from_modelscope(monkeypatch: pytest.MonkeyPatch):
    # 更新注释中的 URL 指向国际站
    # model: https://www.modelscope.ai/models/qwen/Qwen1.5-0.5B-Chat
    with monkeypatch.context() as m:
        m.setenv("VLLM_USE_MODELSCOPE", "True")
    # 新增: 显式设置 ModelScope 域名为国际站 www.modelscope.ai
```

```
m.setenv("MODELSCOPE_DOMAIN", "www.modelscope.ai")
# 使用 ModelScope 时不设置 HF_TOKEN, 否则会报 400 错误
m.setenv("HF_TOKEN", "")
attn_backend = "TRITON_ATTN" if current_platform.is_rocm() else "auto"
llm = LLM(model="qwen/Qwen1.5-0.5B-Chat", attention_backend=attn_backend)
# ... 后续 generate 和断言逻辑不变 ...
```

评论区精华

无讨论。review 中 bot 评论未提供具体反馈，human reviewer DarkLight1337 直接批准。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低：仅修改单文件中的注释和环境变量设置，不影响业务逻辑。若国际站不可用可能导致 CI 测试失败，但可通过回退环境变量恢复。
- 影响：仅影响 CI 回归测试中从 ModelScope 下载模型的场景，对用户无影响。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR