

PR #45858 完整报告

vllm-project/vllm

[ROCm][CI] fix multimodal run cmds

合并时间: 2026-06-17 04:31

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/45858>

执行摘要

- 一句话: 修复 AMD CI 多模态测试误跑语言测试
- 推荐动作: 建议精读该 PR 以了解 CI 配置修复方式。值得关注的设计决策: 审核者指出这些命令在之前更新 AMD 定义时被忽略, 说明 CI 配置维护需要更仔细的 review 流程。

功能与动机

PR body 指出: Fixes the mismatch between the test naming and the actual test being run: 'Multi-Modal Models (Extended Generation 1)' was mistakenly running the language generation tests。即测试标签与实际运行测试不匹配, 导致多模态测试被跳过或运行错误测试。

实现拆解

1. 修改 .buildkite/test-amd.yaml 中第 1785-1788 行和第 1798-1801 行的 commands 字段。
2. 将 Multi-Modal Models (Extended Generation 1) 的 commands 从 `uv pip install mamba/causal-conv1d + 运行 models/language/generation` 改为 `pip install git+https://github.com/TIGER-AI-Lab/Mantis.git + 运行 models/multimodal/generation -m 'not core_model' --ignore models/multimodal/generation/test_common.py 和 models/multimodal/test_mapping.py`。
3. 将 Multi-Modal Models (Extended Generation 2) 的 commands 从类似的语言测试改为运行 `models/multimodal/generation/test_common.py -m 'split(group=0) and not core_model'`。
4. 同时更新了两个任务的依赖安装, 将之前的 `mamba/causal-conv1d` 替换为 `Mantis`。

关键文件:

- .buildkite/test-amd.yaml (模块 构建脚本; 类别 config; 类型 configuration): 唯一变更文件, 修复了两个多模态测试步骤的测试命令, 使其从运行 language generation 测试改为运行正确的多模态测试。

关键符号: 未识别

评论区精华

审核者 AndreasKaratzas 评论: LGTM I overlooked these ones when updating the AMD definitions last month, thanks for fixing that. 表明这些命令在之前更新 AMD 定义时被忽略, 本次修复是必要的。无其他讨论。

- 暂无高价值评论线程

风险与影响

- 风险: 风险较低。配置修改仅影响 2 个 CI 测试步骤, 且变更方向是使命令与标签匹配, 属于修复性变更。潜在风险: 若 Mantis 安装或测试命令语法有误, 可能导致测试失败; 但已由审核者确认。
- 影响: 影响范围限于 AMD ROCm CI 中的多模态测试步骤。影响程度中等: 修复了 CI 测试覆盖漏洞, 确保多模态模型在 ROCm 上得到正确测试, 提高 CI 可靠性。
- 风险标记: 配置变更, 低风险

关联脉络

- 暂无明显关联 PR