

PR #37723 完整报告

vllm-project/vllm

[ROCm][CI] Stabilize ROCm speech-to-text translation test with lower min acc threshold

合并时间: 2026-03-22 17:32

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/37723>

执行摘要

此 PR 降低了语音转文本翻译测试的准确度阈值，从 0.9 到 0.87，以在 ROCm CI 的 MI355 硬件上稳定测试。这是一个临时修复，旨在减少 CI 失败，但可能掩盖潜在问题，建议关注测试质量。

功能与动机

- 为什么做: 解决在 MI355 上的 CI 测试失败，具体构建链接显示失败 (<https://buildkite.com/vllm/amd-ci/builds/6721/steps/canvas?sid=019d09d4-711d-4fbc-9f40-6ec17a28f286&tab=output>)。
- 跟进 PR: 此 PR 是 #34839 的后续措施，专注于调整测试阈值。

实现拆解

- 改动文件: 仅修改 `tests/entrypoints/openai/speech_to_text/test_translation_validation.py`。
- 关键代码变更: `python assert (sum([x == y for x, y in zip(res_stream, res_no_stream.text.split())]) >= len(res_stream) * 0.87 # 从0.9降低到0.87)`
- 模块影响: 测试模块中的 `entrypoints/openai/speech_to_text` 子模块。

评论区精华

- 核心交锋: reviewer DarkLight1337 指出标题误导，作者 AndreasKaratzas 回应并计划重构测试。

@DarkLight1337 True, sorry, I forgot to do that, initially I thought of adding the rocm args, but then I saw the comment and thought that this flakiness is expected. Btw, I'll probably refactor this test, but I though of first stabilize the CI.

- 结论: 更改被批准，但标题未更新，突出了 CI 稳定性的短期优先。

风险与影响

- 技术风险:
 - 测试阈值降低可能导致假阳性，掩盖 speech-to-text 功能的真实缺陷。

- 缺乏根本原因分析，依赖临时调整。
- 影响评估：
 - 用户：无直接影响。
- 系统：CI 更稳定，减少失败噪声。
- 团队：短期提高效率，但长期需加强测试健壮性。

关联脉络

- 关联 PR：与 #34839 相关，后者可能涉及类似测试稳定措施，但提供的材料中未详述。
- 演进趋势：近期历史 PR 显示频繁的测试和 ROCm 相关调整（如 #36100、#36505），表明团队在优化硬件兼容性和 CI 稳定性。