

PR #38937 完整报告

vllm-project/vllm

[ROCm][CI] Added back missing common deps

合并时间: 2026-04-04 06:58

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/38937>

执行摘要

- 一句话: 恢复 ROCm CI 测试依赖中的通用依赖项, 修复 AMD CI 环境。
- 推荐动作: 对于负责 CI 或 ROCm 维护的工程师, 建议关注此 PR 以理解依赖管理策略。变更较简单, 但讨论揭示了 `-c` 与 `-r` 的微妙区别以及外部依赖兼容性的重要性。

功能与动机

根据 PR body, 作者承认在之前的 PR 中错误移除了通用依赖, 需要添加回来以修复 AMD CI 测试环境。Issue 评论中进一步讨论到 `-c` 和 `-r` 选项的区别以及 `compressed-tensors` 的兼容性风险。

实现拆解

实现方案简单直接: 在 `requirements/rocm-test.in` 文件中添加一行 `'-r common.txt'`, 引用通用依赖文件。这导致 `pip-compile` 工具更新 `requirements/rocm-test.txt`, 添加了 282 行新依赖并删除 10 行, 包括 FastAPI 组件、AI SDKs 和其他工具库。

关键文件:

- `requirements/rocm-test.in` (模块 `ci`): 控制 ROCm 测试依赖的源文件, 添加 `'-r common.txt'` 是关键变更
- `requirements/rocm-test.txt` (模块 `ci`): 编译后的依赖列表, 大量更新反映依赖恢复的影响

关键符号: 未识别

评论区精华

Review 中, hmellor 询问是否因此可以移除 `-c` 选项, 引发了对依赖管理设计的讨论。Issue 评论中深入探讨了 `-c` 与 `-r` 的区别, 以及 `compressed-tensors` pin 到 `transformers<5` 可能导致 AMD CI v5 失败。结论是 hmellor 指出添加所有通用依赖可能破坏 CI, 建议更精细地添加缺失依赖, 但 PR 已合并。

- Dependency management with `-c` vs `-r` (design): hmellor 指出 `-c` 可能多余, 但 AndreasKaratzas 认为仍有作用, 未达成明确结论。
- Compressed-tensors compatibility issue (correctness): 建议需要更精细修复, 而不是添加所有通用依赖; 后续计划修复 `compressed-tensors` 项目。

风险与影响

- 风险：主要风险：引入 compressed-tensors (版本 0.14.0.1) ，其 pin 到 transformers<5，可能与 AMD CI v5 中使用的 transformers v5 冲突，导致测试失败。此外，大量新依赖可能增加环境复杂性和潜在冲突。
- 影响：影响范围：仅限于 ROCm 平台的 CI 测试环境。影响程度：中，修复了依赖缺失问题，但可能引入新兼容性问题，需要监控 CI 运行状态。
- 风险标记：依赖冲突风险，测试环境不稳定

关联脉络

- PR #38184 [ROCm][CI] Run Kernels Core Operation Test On MI325 and mitigate flakiness: 同属 ROCm CI 改进，涉及测试环境配置
- PR #38959 [ROCm][CI] Fix ROCm Dockerfile confctest generation for older Docker parsers: 修复 ROCm CI 相关问题，属于基础设施维护