

PR #52904 完整报告

vllm-project/vllm

[XPU][CI] downgrade sentencepiece

合并时间: 2026-08-19 17:01

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/52904>

执行摘要

本 PR 将 `requirements/test/xpu.txt` 中的 `sentencepiece` 版本从 `0.2.2` 降级至 `0.2.1`，以修复 XPU CI 中 InternVL2-2B 模型 tokenizer 因空字节校验失败的问题。变更虽小，但解决了实际 CI 阻塞，保障了 XPU 测试的稳定性。

功能与动机

XPU CI 构建 (Buildkite #9299) 因 InternVL2-2B 的 `tokenizer.model` 包含 `\x00` token 而失败。`sentencepiece 0.2.2` 引入了对 null-byte pieces 的校验 (见 issue #1308)，导致旧版模型无法加载。为恢复测试，作者选择降级依赖。

实现拆解

仅修改 `requirements/test/xpu.txt` 一行：

- `sentencepiece==0.2.2` → `sentencepiece==0.2.1`
- 无其他文件变更，无测试配套改动。

无源码变更，故省略。

评论区精华

无实质讨论。claude[bot] 提示 fork 仓库需手动 review；维护者 `jikunshang` 直接批准。

风险与影响

风险极低，仅涉及测试依赖版本固定，可能错过后续修复，但不影响生产。影响范围限定在 XPU CI 测试环境，解决了实际的加载错误。

关联脉络

上游 issue [google/sentencepiece#1308](#) 尚在打开状态，v0.2.3 预计会修复。本 PR 是临时规避，未来可关注上游修复后再升级。同仓库 PR #53035 也在调整 XPU CI 测试，均服务于 Intel GPU 测试稳定性。