

# PR #56783 完整报告

vllm-project/vllm

[XPU][CI] update gpt-oss package version

合并时间: 2026-09-15 10:23

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/56783>

## 执行摘要

- 一句话: gpt-oss 回退 0.0.8 并修复 XPU 评估 CI 安装
- 推荐动作: 值得快速浏览, 不深入阅读。关注点: CI 步骤硬编码 gpt-oss[eval]==0.0.5, 与 requirements 锁定 0.0.8 不一致, 建议统一版本来源并重新生成 pip compile 清单; 后续可留意 gpt-oss 0.0.9 准确性问题是否在上游被根因修复。

## 功能与动机

PR body 说明: pip compile 将 gpt-oss 解析到最新 0.0.9, 但该版本存在准确性问题 (accuracy issues), 因此手动固定回 0.0.8, 以与 cuda.txt 保持一致。这意味着 XPU 上的 GPQA 评估若使用 0.0.9, 结果可能失真, 必须锁定到已验证版本。

## 实现拆解

1. 依赖回退: requirements/test/xpu.txt 将 gpt-oss 从 0.0.9 改回 0.0.8, 并同步删除 0.0.9 引入的传递依赖条目 (jupyter-client、jupyter-core、tornado、traitlets), 修正 platformdirs、pyzmq、tqdm 等包的 via 注释。该文件由 pip compile 生成, 回退后应重新编译校验。
2. CI 安装前置步骤: .buildkite/intel\_jobs/lm\_eval\_intel.yaml 的 xpu-gpqa-eval-gpt-oss 步骤, 把命令从直接 cd tests 后运行 pytest, 改为先执行 uv pip install "gpt-oss[eval]==0.0.5", 再进入测试目录运行 GPQA 正确性测试, 确保 eval 附加依赖可用。
3. 提交演进: 4 个 commit 依次为 pin 版本、补充 eval 安装、细节调整、合并 main; 全程无测试文件或源码改动, 体现为一次聚焦的 CI 修复。
4. 测试与配置配套: 无新增测试; requirements 清单的修改依赖 pip compile 保持一致性, CI 行为依赖 run-intel-test.sh 脚本透传命令。

需要注意: CI 安装的是 0.0.5, 而 requirements 锁定 0.0.8, 版本脱节未在 PR 中解释。

关键文件:

- .buildkite/intel\_jobs/lm\_eval\_intel.yaml (模块 构建流水线; 类别 config; 类型 configuration): 在 XPU GPQA 评估命令前增加 uv pip install "gpt-oss[eval]==0.0.5", 保证评估前具备带 eval 附加依赖的 gpt-oss; 这是 CI 行为的关键改动, 也是版本不一致风险的来源。

- requirements/test/xpu.txt (模块 依赖锁定; 类别 config; 类型 dependency-pin) : 核心依赖锁定变更: gpt-oss 从 0.0.9 回退到 0.0.8, 并清理由此不再需要的传递依赖, 是修复准确性问题的主载体。

关键符号: 未识别

## 评论区精华

该 PR 无实质性 review 讨论: claude[bot] 注明 fork 自动审查被禁用, 维护者 jikunshang 直接批准并触发两次 Buildkite CI (#88765、#88976)。分析层面可指出一个未被讨论的疑点: CI 安装 gpt-oss[eval]==0.0.5, 而 requirements/test/xpu.txt 锁定 gpt-oss==0.0.8, 两者不一致。

- CI 安装的 gpt-oss 版本与锁定版本不一致 (question): 未在 PR 评审中澄清, 建议后续统一版本来源。

## 风险与影响

- 风险:
  1. 版本不一致: uv pip install 安装 0.0.5, 与 requirements 锁定的 0.0.8 脱节, CI 实际验证的环境可能不是依赖清单代表的版本, 评估结果可复现性受影响。
  2. 环境覆盖: uv pip install 会在镜像基础上覆盖安装, 若 0.0.5 与 0.0.8 行为不同, GPQA 结果可能漂移。
  3. 传递依赖清理: xpu.txt 删除 jupyter-client、jupyter-core、tornado、traitlets 等条目, 若其他 XPU 测试隐式依赖这些包, 后续 CI 可能暴露缺失。
  4. 影响面: 仅涉及 XPU LM Eval 步骤与测试依赖清单, 不涉及核心推理路径, 风险隔离。
    - 影响: 影响范围集中在 Intel GPU (XPU) 的 Buildkite LM Eval (GPQA) 步骤和 XPU 测试依赖清单。受影响用户为维护 XPU 评估 CI 的工程师; 评估结果的准确性依赖 gpt-oss 版本。对系统整体无核心路径影响, 属于隔离的 CI 修复。
    - 风险标记: CI 安装与锁定版本不一致, uv 覆盖镜像依赖, 传递依赖清理风险

## 关联脉络

- 暂无明显关联 PR