

PR #42150 完整报告

vllm-project/vllm

Bump llguidance to 1.7

合并时间: 2026-05-15 08:35

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/42150>

执行摘要

升级 llguidance 依赖版本从 1.3 至 1.7, 同时同步更新 ROCm 测试 lockfile 中的固定版本。该变更为跨仓库依赖升级 (vllm-metal 项目需求) 的前置条件, 影响所有使用结构化输出功能的用户。变更仅涉及两行依赖声明修改, 无代码逻辑改动。

功能与动机

PR body 明确说明: bump 至 llguidance 1.7 是 vllm-metal 项目安全升级 mlx-vlm (修复 Qwen3-VL deepstack 参考输出) 的前置条件。因为新版 mlx-vlm ($\geq 0.5.0$) 要求 $llguidance \geq 1.7.0$, 而 vLLM 原有的 $llguidance < 1.4.0$ 约束会阻止 vllm-metal 解除版本锁定。

实现拆解

- 更新运行时依赖约束 (`requirements/common.txt`): 将 llguidance 的版本范围从 $\geq 1.3.0, < 1.4.0$ 改为 $\geq 1.7.0, < 1.8.0$ 。该约束适用于 x86_64、arm64、aarch64、ppc64le 平台。
- 同步 ROCm 测试 lockfile (`requirements/test/rocm.txt`): 将 pytest 测试环境的 lockfile 中的 llguidance pin 版本从 1.3.0 更新为 1.7.5, 确保在同时安装 `-r requirements/test/rocm.txt` 的场景下不会因为 lockfile 优先级更高而降级 llguidance。

两个修改均只有 +1/-1 行, 无其他代码或配置文件更改。

无源码逻辑变更。变更内容仅为版本号文本修改, 无需代码片段展示。

评论区精华

chatgpt-codex-connector: "This bump leaves the generated ROCm test requirements stale: `requirements/test/rocm.txt` still pins `llguidance==1.3.0` ... In ROCm test/dev setups that install the test lock after the package requirements, that explicit pin can downgrade llguidance back below the new $\geq 1.7, < 1.8$ runtime constraint."

该建议被采纳, PR 后续补充了 `rocm.txt` 的同步修改。

风险与影响

- 风险: 低。仅依赖版本号变更, 无代码逻辑。llguidance 1.7.x 可能引入与 lm-format-enforcer 或 outlines 的轻微兼容性差异, 但已在审核中评估为可接受。

- 影响：直接解除 vllm-metal 升级 mlx-vlm 的阻塞；所有使用 llguidance 的系统将自动获得 1.7.x 的更新。

关联脉络

- 与 PR#39337 (Model Runner v2 Oracle for Qwen3) 存在间接关联：本 PR body 提及的 vllm-metal 升级正是为了修复 Qwen3-VL 相关问题。
- 无其他历史 PR 直接关联。