

PR #21032 完整报告

sgl-project/sglang

[Deps] Bump xgrammar to 0.1.32

合并时间: 2026-03-26 16:22

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/21032>

执行摘要

本次 PR 将 xgrammar 依赖从 0.1.27 升级到 0.1.32, 以获取最新安全修复。修改了多个 pyproject.toml 文件和 CI 配置文件, 影响所有使用 xgrammar 的模块。变更简单, 风险可控, 建议快速审查版本变更并关注 CI 测试结果。

功能与动机

升级动机明确: PR body 中说明“Bump xgrammar dependency from 0.1.27 to 0.1.32 to pick up latest security fixes.”, 目的是修复安全漏洞, 提升系统安全性。无关联 Issue, 但依赖更新是常规维护的一部分。

实现拆解

实现包括更新 6 个文件中的 xgrammar 版本号:

- 配置文件更新:
 - python/pyproject.toml、python/pyproject_cpu.toml、python/pyproject_npu.toml、python/pyproject_other.toml、3rdparty/amd/wheel/sglang/pyproject.toml: 将所有 xgrammar 版本从 0.1.27 改为 0.1.32。
 - .github/workflows/nightly-test-npu.yml: 修复遗留版本, 从 0.1.25 更新到 0.1.32, 确保 CI 测试使用最新依赖。

所有变更均为字符串替换, 示例代码块:

```
- "xgrammar==0.1.27",  
+ "xgrammar==0.1.32",
```

评论区精华

review 过程中无实质性讨论, 仅有 reviewer Fridge003 的批准, 表明变更直接通过。无争议点或设计权衡, 反映出依赖升级的常规性。

风险与影响

风险分析:

- 兼容性风险: xgrammar 版本跳变可能引入 API 变化, 影响相关功能 (如 MUSA 支持), 需验证无回归。

- 安全修复验证：升级基于安全修复，但未提供测试覆盖，依赖 CI 确保稳定性；若新版本有漏洞，可能反向引入风险。
- CI 配置风险：修改 NPU 测试文件可能影响夜间测试，特别是在依赖安装环节。

影响分析：

- 用户影响：无直接可见变更，潜在提升安全性能。
- 系统影响：所有依赖 xgrammar 的模块（如约束生成）使用新版本，可能行为微调，需监控集成效果。
- 团队影响：增加维护复杂度，需确保 CI 测试通过并关注硬件兼容性（参考 PR #21296）。

关联脉络

与历史 PR #21296 “[MUSA] apply_vocab_mask support musa device” 相关，后者修改了 xgrammar_backend.py 文件，共享相同依赖版本。这表明 xgrammar 在约束生成中扮演关键角色，升级可能影响硬件兼容性趋势。近期 PR 中多涉及 NPU 和 CI 优化，本 PR 是依赖维护链的一部分，支持整体架构演进。