

PR #32354 完整报告

sgl-project/sglang

Fix PyPI release: drop the git-only sgl-eval dep from packaged metadata

合并时间: 2026-07-25 08:00

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/32354>

执行摘要

- 一句话: 修复 PyPI 发布失败, 移除 git-only 依赖
- 推荐动作: 值得快速合并, 已修复关键发布阻塞问题。可关注后续是否有其他 git-only 依赖需要在发布前检查。

功能与动机

v0.5.16 的 PyPI 发布因 wheel 元数据中包含 git direct URL 依赖 (sgl-eval) 而被 PyPI 拒绝, 导致发布失败。

实现拆解

1. 移除 pyproject.toml 中的 git-only 依赖: 在 python/pyproject.toml 的 test extra 中删除了 sgl-eval @ git+... 条目, 并更新注释说明原因和替代安装路径。
2. 将依赖移至 CI 安装脚本: 在 scripts/ci/cuda/ci_install_dependency.sh 的 install_extra_deps() 函数中添加了通过 pip 安装 sgl-eval (指定 SHA) 的步骤, 确保 CI (如夜间 GSM8K 评估) 仍能正常安装。
3. 保持兼容性: 在 pyproject.toml 中保留 antlr4-python3-runtime==4.9.3 固定版本, 确保当 sgl-eval 通过 CI 脚本安装时环境兼容。

关键文件:

- python/pyproject.toml (模块 项目配置; 类别 config; 类型 configuration): 主元数据文件, 移除 sgl-eval 的 git direct URL 依赖, 修复 PyPI 发布失败的根本原因。
- scripts/ci/cuda/ci_install_dependency.sh (模块 CI 脚本; 类别 infra; 类型 infrastructure): CI 安装脚本, 新增 sgl-eval 的 pip 安装步骤, 确保 CUDA CI 测试仍能正常获取该依赖。

关键符号: 未识别

关键源码片段

python/pyproject.toml

主元数据文件, 移除 sgl-eval 的 git direct URL 依赖, 修复 PyPI 发布失败的根本原因。

```
# python/pyproject.toml - test extra 片段 (变更后)
# 4.9.3 is what sgl-eval's math_verify grader needs (latex2sympy2_extended
```

```
# raises ImportError on 4.7.x), so a `sglang[test]` environment is already
# compatible when sgl-eval is installed on top of it.
# Do NOT declare sgl-eval itself here: it is git-only, and PyPI rejects any
# uploaded distribution whose metadata carries a direct URL requirement.
# CUDA CI installs it in scripts/ci/cuda/ci_install_dependency.sh.
"antlr4-python3-runtime==4.9.3",
# "sgl-eval @ git+..." 已删除
```

scripts/ci/cuda/ci_install_dependency.sh

CI 安装脚本，新增 sgl-eval 的 pip 安装步骤，确保 CUDA CI 测试仍能正常获取该依赖。

```
# scripts/ci/cuda/ci_install_dependency.sh 片段（变更后）
install_extra_deps() {
    MOONCAKE_VERSION="0.3.11.post1"
    NIXL_VERSION="1.3.0"
    # sgl-eval is git-only and cannot be declared in python/pyproject.toml (see
    # the note there). The nightly GSM8K eval shells out to the sgl-eval CLI and
    # fails without it. Bumping the SHA can change zero-shot \boxed{} grading, so
    # re-baseline MODEL_SCORE_THRESHOLDS in
    # test/registered/eval/test_text_models_gsm8k_eval.py first.
    SGL_EVAL_REF="b2a2703c42cae379bbcb8b7ff092df6601a61694"
    # ... 其他安装逻辑 ...
    $PIP_CMD install "sgl-eval @ git+https://github.com/sgl-project/sgl-eval.git@${SGL_EVAL_REF}"
    " $PIP_INSTALL_SUFFIX
    # ...
}
```

评论区精华

无 review 讨论。

- 暂无高价值评论线程

风险与影响

- 风险：低风险：仅涉及元数据和 CI 脚本调整，不修改任何核心逻辑。风险在于 CI 脚本中 pip 安装可能因网络问题失败，但已有幂等处理。另外，非 CUDA 的 CI 作业（如 pr-test）将不再安装 sgl-eval，但这不影响现有测试（已 mock）。
- 影响：影响范围：主要影响 PyPI 发布流程（修复阻断问题）和 CUDA CI 中的夜间评估。CPU/AMD/ 其他平台的 CI 不受影响。用户无需变更。
- 风险标记：非核心路径变更，CI 脚本调整可能影响测试环境

关联脉络

- PR #32296 Bump sgl-deep-gemm to 0.1.5: 同为依赖管理和发布相关的 PR，涉及 pyproject.toml 和 CI 脚本的修改。
- PR #32349 Add verbose flag to twine upload command: 同为发布流程改进，该 PR 添加 verbose 标志以帮助调试未来发布问题。