

# PR #21797 完整报告

sgl-project/sglang

Fix killall.py crash when sglang is not yet installed

合并时间: 2026-04-01 08:40

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/21797>

## 执行摘要

本 PR 修复了 CI 中 `killall.py` 脚本因导入 `sglang` 模块而崩溃的问题，通过内联 `nvidia-smi` 查询代码避免依赖，确保 CI 在新 runner 或路径变更时稳定运行。变更范围小，风险低，属常规维护。

## 功能与动机

此修复源于 CI 测试中的确定性失败：`killall.py` 在 `pip install -e` 之前运行，但自 PR #21780 引入后，它导入 `from sglang.srt.utils.common import get_nvidia_driver_version_str`。在未安装 `sglang` 的 runner 上（如新 runner 或 `git clean` 后），该导入会立即失败，导致 `ModuleNotFoundError`。PR body 中详细说明了失败证据，强调这不是 flaky 测试，而是会随 runner 增加而加剧的问题。

## 实现拆解

改动仅涉及 `python/sglang/cli/killall.py` 文件的 `_get_smi_version` 函数。关键变更包括：

- 移除导入依赖：删除 `from sglang.srt.utils.common import get_nvidia_driver_version_str`。
- 内联 `nvidia-smi` 查询：添加 `subprocess.run` 调用，直接执行 `nvidia-smi --query-gpu=driver_version` 命令，并捕获输出。
- 异常处理：使用 `try-except` 块处理 `SubprocessError` 和 `FileNotFoundError`，失败时返回 `None`。

```
代码示例: try: result = subprocess.run( [ "nvidia-smi", "--query-gpu=driver_version", "--format=csv,noheader,nounits", ], capture_output=True, text=True, check=True, timeout=10, ) driver = result.stdout.strip().split("\n")[0].strip() or None except (subprocess.SubprocessError, FileNotFoundError): driver = None
```

## 评论区精华

无 review 讨论或评论，PR 直接由作者合并。这表明变更被认可为简单修复，无需进一步审查。

## 风险与影响

风险分析：

- 低风险：变更逻辑与原函数一致，仅替换导入为命令调用。
- 外部依赖：依赖 `nvidia-smi` 命令的可用性，但已有异常处理。
- 测试覆盖：未新增单元测试，但修复通过 CI 验证。

影响分析：

- CI 稳定性：解决了 runner 未安装 `sglang` 时的崩溃，提高 CI 成功率。
- 系统影响：不涉及核心功能或用户接口，仅影响内部 CI 脚本。
- 长期影响：减少因环境差异导致的 flaky 测试，提升开发效率。

## 关联脉络

此 PR 直接关联到 PR #21780 ("[Fix] Fall back to triton MOE for GPT-OSS on Blackwell with driver  $\geq 595$ "), 后者引入了对 `sglang.srt.utils.common` 的导入，导致本问题。近期历史 PR 中，多个涉及 CI 和 bugfix 的 PR (如 #21778 缓存 Nvidia wheels、#21729 修复 HF API 速率限制) 显示团队正持续优化 CI 流程和稳定性。本修复是这一趋势的延续，聚焦于消除 import 依赖导致的脆弱性。