

PR #2811 完整报告

radixark/miles

openenv: raise the tbench2_env contract floor to OpenEnv#1025

合并时间: 2026-09-02 23:15

原文链接: <http://prhub.com.cn/radixark/miles/pull/2811>

执行摘要

- 一句话: 将 OpenEnv 集成对 tbench2_env 环境包的最低版本要求从 #1012 提升至 #1025, 以修复基础设施失败被误判为测试失败的评分漏洞。
- 推荐动作: 这个 PR 是一个重要的、针对性的正确性修复。它解决了一个能悄无声息地污染训练数据的评分漏洞。变更本身风险较低, 逻辑清晰, 主要是版本门槛提升和文档同步。强烈建议相关维护者和使用 OpenEnv 集成的工程师精读此 PR, 特别是理解从源码探测到 per-episode 守卫的双层防御机制为何必要, 以及 #1025 修复的具体漏洞场景。虽然没有复杂的代码逻辑, 但它展示了在 RL 训练数据流中建立可靠“合约”检查的重要性。

功能与动机

动机源于 PR body 和关联 Issue #1025 中描述的关键评分漏洞。在 #1025 之前, 一个从未写入 `reward.txt` 的 canonical test.sh 运行 (例如因工具链下载失败或 shell timeout 导致) 会返回与真实测试失败字节级相同的 `reward=0.0` 且携带 harness 标记。如 Issue #1025 所述: “RL consumers must be able to drop a no-verdict episode: ingesting it as 0.0 injects a false negative into training that no client-side check can catch (the reply looked completely valid).” mile 侧现有的逐 episode 守卫 (`multi_turn` 中的 marker 检查) 无法捕捉这种“伪造”的 0 分。因此, 必须将环境包的安装门槛提升至已修复此问题的 #1025 版本, 从源头上阻止不合规的环境被用于训练。

实现拆解

本次变更的核心是提高 mile 对 `tbench2_env` 环境包的安装要求, 并同步更新所有相关文档以反映新的合约地板。具体实现拆解如下:

1. 增强启动器源码预检逻辑: 在 `examples/experimental/openenv/openenv_launch_common.py` 的 `apply_optional_env_vars` 函数中, 原有的源码探测仅检查 `TB2_WITHHOLD_TESTS` 关键字是否存在。变更后, 该探测逻辑增加了第二个必要条件: `_require_canonical_verdict` 函数也必须存在于服务器源码中。这是发现不合规安装的唯一可靠方式, 因为 per-episode 的 marker 守卫无法区分一个携带正确标记但从未评分的回复和一个真实的测试失败。此预检覆盖所有三个 sandbox 后端 (Daytona, E2B, Modal)。
2. 更新核心 agent 函数模块文档: 在 `openenv_agent_function.py` 及其三个变体 (`openenv_daytona_agent_function.py`, `openenv_e2b_agent_function.py`, `openenv_modal_agent_function.py`) 的文件头注释和模块 docstring 中, 将 server contract 的引用版本从 #1012 (04d259ea6) 更新为 #1025 (38b2a3135)。同时,

openenv_agent_function.py 的模块 docstring 和其 multi_turn 函数内的守卫注释被精确更新，明确指出了 reward=None / error 分支现在涵盖了“test.sh never wrote reward.txt”这一新场景。

3. 同步更新用户文档：examples/experimental/openenv/README.md 中所有提及 server contract 地板版本的地方（安装说明和远程 Docker 主机部署说明）均从 #1012 更新为 #1025，确保文档与实际代码要求一致。

关键文件：

- examples/experimental/openenv/openenv_launch_common.py（模块 启动器；类别 source；类型 core-logic）：核心变更文件。在启动器预检逻辑中增加了对 _require_canonical_verdict 的源码探测，是阻止使用不合规 tbench2_env 安装的关键防线。
- examples/experimental/openenv/openenv_agent_function.py（模块 Agent 函数；类别 source；类型 documentation）：核心 agent 函数模块。更新了模块文档字符串中的合约描述和 multi_turn 函数内守卫的注释，明确了 #1025 带来的行为变化。
- examples/experimental/openenv/README.md（模块 文档；类别 docs；类型 documentation）：用户文档。确保安装说明和远程部署说明中的合约版本与代码逻辑同步更新。

关键符号：apply_optional_env_vars

关键源码片段

examples/experimental/openenv/openenv_launch_common.py

核心变更文件。在启动器预检逻辑中增加了对 _require_canonical_verdict 的源码探测，是阻止使用不合规 tbench2_env 安装的关键防线。

```
# 在 apply_optional_env_vars 函数中，对 tbench2_env 服务器源码的预检逻辑：
server_src = Path(tbench2_env.__file__).resolve().parent / "server" / "tbench2_env_environment.
py"
src_text = server_src.read_text(encoding="utf-8") if server_src.is_file() else ""
# 增加对 `_require_canonical_verdict` 的检查，这是 #1025 修复缺失 verdict
评分漏洞的关键函数。
# 在 #1025 之前，缺失 verdict 的回复会伪装成 0 分，per-episode 守卫无法识别。
if "TB2_WITHHOLD_TESTS" not in src_text or "_require_canonical_verdict" not in src_text:
    raise RuntimeError(
        "the installed tbench2_env server lacks the native-evaluate "
        "contract (canonical test.sh scoring / TB2_WITHHOLD_TESTS / "
        "missing verdict reported as an error): install from an OpenEnv "
        "checkout at or after the #1025 merge (38b2a3135) — see this "
        "directory's README"
    )
```

评论区精华

本次 PR 的审核讨论非常简短。Claude bot 提供了自动化的初步审查，并标记“未发现问题”。最终由 Shi-Dong 批准并评论“LGTM!”，表示对变更的完全认可。讨论本身未暴露技术争议，

这符合变更性质——它是一个明确的版本升级和文档同步，逻辑清晰且基于上游已确认的修复。

- 自动化代码审查与批准 (question): 变更被批准合并，未发现阻塞性问题。

风险与影响

- 风险：主要风险在于 远程 docker 主机环境的覆盖缺口。如 PR body “Not in this PR” 部分明确指出，通过 `--openenv-env-url` 指向的远程环境服务器无法在此 PR 的预检范围内，其部署仍可能使用低于 #1025 的 `tbench2_env` 版本，从而继续产出带假 0 分的回复。这部分依赖于运维对远程环境进行重新部署。次要风险在于 #1025 将多种错误路径（如 `uvx` 下载失败、`shell timeout`）统一通过 `eval_error` 上报，而 #2802 目前将所有 `eval_error` 映射为 `VerifierError/0` 分，未来是否需要按错误文本进一步区分属于 #2802 的设计决策范畴，与本 PR 无关。
- 影响：本次变更的影响范围主要集中在 OpenEnv 集成路径。它提高了启动时对环境的检查严格度，确保所有由 `mile` 启动的本地 `sandbox` 环境（Daytona, E2B, Modal）都使用了修复了关键评分漏洞的 `tbench2_env` 版本。对用户而言，如果本地安装的 `tbench2_env` 低于 #1025，启动器会立即抛出明确的错误信息，阻止训练运行，避免了错误数据的产生。这是一个向后兼容的变更，但提高了最低版本要求。它不直接影响核心训练逻辑或其他模块，但对于保障基于 OpenEnv 的 `agentic RL` 训练数据的正确性至关重要。
- 风险标记：远程环境未覆盖，合约升级依赖

关联脉络

- PR #2802 agent functions: one failure contract for the OpenEnv, NeMo Gym and Harbor legs: 本 PR 修复了导致 OpenEnv 端产生假 0 分的环境合约漏洞，是 PR #2802 所建立统一失败合约中关键的一环。#2802 处理 `mile` 侧如何统一处理各类失败，而 #2811 确保了 OpenEnv 环境端不会产生无法区分的伪装失败。
- PR #2877 fix(ci): mint the command-App token with pull-requests: write for label commands: 同属仓库近期维护活动，但领域不同（CI 权限 vs. OpenEnv 合约）。