

PR #35024 完整报告

vllm-project/vllm

[Deprecation] Remove old locations of ``get_tokenizer`` and ``resolve_hf_chat_template``

合并时间: 2026-05-15 15:13

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/35024>

执行摘要

- 一句话: 移除旧版 `get_tokenizer` 和 `resolve_hf_chat_template` 导入路径
- 推荐动作: 此 PR 是清理老旧废弃代码的良好实践, 展示了如何协调外部依赖更新以安全移除兼容层。建议维护者确保所有内部代码已不再使用旧导入路径。对于读者, 值得关注其与 `lm-eval` 社区的协作流程。

功能与动机

PR 的动机是 `lm-eval` 库已更新 (PR #33994, 版本 0.4.12), 支持新的导入路径, 因此可以安全移除 `vllm` 中提供的废弃向后兼容代码。

实现拆解

1. 删除旧导入路径文件: 移除 `vllm/transformers_utils/tokenizer.py` 整个文件, 该文件仅包含一个 `__getattr__` 函数, 用于将 `get_tokenizer` 重新路由到 `vllm.tokenizers.get_tokenizer`。
2. 清理 `entrypoints` 模块: 在 `vllm/entrypoints/chat_utils.py` 中删除 `__getattr__` 函数, 该函数为 `resolve_hf_chat_template` 提供后向兼容。同时移除了不再需要的 `warnings` 导入。
3. 更新依赖版本: 在 `requirements/test/cuda.in` 和 `requirements/test/rocm.in` 中将 `lm-eval` 最低版本从 0.4.11 提升到 0.4.12, 确保测试环境使用已更新的 `lm-eval`。
4. 同步 CI 脚本: 在 `.buildkite/lm-eval-harness/` 下的多个 `shell` 脚本以及 TPU CI 脚本中, 将 `lm-eval` 版本要求同步更新为 `>=0.4.12`。
5. 调整锁定文件: `requirements/test/cuda.txt` 和 `requirements/test/rocm.txt` 随之更新, 反映依赖版本变化及部分间接依赖的调整。

关键文件:

- `vllm/transformers_utils/tokenizer.py` (模块 分词器; 类别 `source`; 类型 `deletion`; 符号 `getattr`): 被整体删除的核心文件, 之前提供 `get_tokenizer` 的后向兼容导入。
- `vllm/entrypoints/chat_utils.py` (模块 入口点; 类别 `source`; 类型 `dependency-wiring`; 符号 `getattr`): 删除了 `__getattr__` 函数及相关的 `warnings` 导入, 清除 `resolve_hf_chat_template` 的旧入口。
- `requirements/test/cuda.in` (模块 测试依赖; 类别 `test`; 类型 `test-coverage`): `lm-eval` 最低版本从 0.4.11 提升到 0.4.12, 与外部依赖更新同步。

- requirements/test/rocm.in (模块 测试依赖; 类别 test; 类型 test-coverage) : 与 cuda.in 同步更新 lm-eval 版本要求, 保持测试依赖一致。
- .buildkite/scripts/hardware_ci/run-tpu-v1-test.sh (模块 CI 脚本; 类别 infra; 类型 infrastructure) : TPU CI 脚本中更新 lm-eval 版本, 确保基础设施一致性。

关键符号: `getattr`

关键源码片段

vllm/transformers_utils/tokenizer.py

被整体删除的核心文件, 之前提供 `get_tokenizer` 的后向兼容导入。

```
# SPDX-License-Identifier: Apache-2.0
# SPDX-FileCopyrightText: Copyright contributors to the vLLM project
import warnings

# 此 __getattr__ 函数已被移除, 不再提供 vllm.transformers_utils.tokenizer.get_tokenizer
# 的向后兼容导入。改为直接使用 vllm.tokenizers.get_tokenizer。
def __getattr__(name: str):
    # Keep until lm-eval is updated
    if name == "get_tokenizer":
        from vllm.tokenizers import get_tokenizer

        warnings.warn(
            "`vllm.transformers_utils.tokenizer.get_tokenizer` "
            "has been moved to `vllm.tokenizers.get_tokenizer`. "
            "The old name will be removed in a future version.",
            DeprecationWarning,
            stacklevel=2,
        )

        return get_tokenizer

    raise AttributeError(f"module {__name__!r} has no attribute {name!r}")
```

vllm/entrypoints/chat_utils.py

删除了 `__getattr__` 函数及相关的 `warnings` 导入, 清除 `resolve_hf_chat_template` 的旧入口。

```
# 以下 __getattr__ 函数已从 vllm/entrypoints/chat_utils.py 中移除。
# 用户应改用 vllm.renderers.hf.resolve_chat_template。
# 同时移除了 import warnings。
```

```
def __getattr__(name: str):
    if name == "resolve_hf_chat_template":
        from vllm.renderers.hf import resolve_chat_template

        warnings.warn(
            "`vllm.entrypoints.chat_utils.resolve_hf_chat_template` has been moved to "
            "`vllm.renderers.hf.resolve_chat_template`. "
```

```
"The old name will be removed in v0.16.",
DeprecationWarning,
stacklevel=2,
)

return resolve_chat_template

raise AttributeError(f"module {__name__!r} has no attribute {name!r}")
```

评论区精华

- 作者发现 lm-eval 尚未更新新导入路径，主动在 lm-eval 仓库提交 PR (EleutherAI/lm-evaluation-harness#3595)。
- 随后 lm-eval 发布 v0.4.12，作者确认后请求合并本 PR。
- Reviewers (Isotr0py、tjtanaa) 均批准，无异议。
- lm-eval 更新进度协调 (other): lm-eval v0.4.12 已发布并支持新路径，本 PR 可安全合并。

风险与影响

- 风险：低风险：旧导入路径已废弃一段时间并带有 DeprecationWarning，外部使用者应已迁移。lm-eval 依赖从 0.4.11 升到 0.4.12 为最小 patch 升级，兼容性风险极低。但需注意任何未升级的 lm-eval 版本 (<0.4.12) 将无法通过测试；如外部代码直接引用旧路径，更新后会出现 ImportError。
- 影响：影响面窄：主要影响 vllm 内部维护，减少废弃兼容代码。对用户而言，若仍使用 vllm.transformers_utils.tokenizer.get_tokenizer 或 vllm.entrypoints.chat_utils.resolve_hf_chat_template，将需要改用新路径 vllm.tokenizers.get_tokenizer 和 vllm.renderers.hf.resolve_chat_template。依赖 lm-eval 的测试或用户需确保版本 $\geq 0.4.12$ 。
- 风险标记：废弃兼容层移除可能影响旧用法，lm-eval 版本锁定需同步

关联脉络

- 暂无明显关联 PR