

PR #40907 完整报告

vllm-project/vllm

[Tests] Gate Isaac under Transformers v5

合并时间: 2026-04-27 10:26

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/40907>

执行摘要

- 一句话: 为 Isaac 模型添加 Transformers v5 版本门控
- 推荐动作: 此 PR 是 Transformers v5 兼容性工作的典型增量——对单个模型添加版本门控。建议精读以理解 `_HfExamplesInfo` 的版本门控模式, 便于后续类似问题的快速修复。

功能与动机

关联 Issue #38389 指出, 在 Transformers v5 升级测试中, `IsaacForConditionalGeneration` 会因远程代码不兼容而失败 (如 `SlidingWindowCache` 导入错误)。PR body 明确说明: 'The Isaac checkpoints still rely on custom Hub code that is not fully compatible with Transformers v5', 因此需要在测试注册表中跳过此模型。

实现拆解

1. 定位文件与入口: 在 `tests/models/registry.py` 的 `_HfExamplesInfo` 字典中, 找到 `IsaacForConditionalGeneration` 条目。
2. 添加版本约束: 为 `IsaacForConditionalGeneration` 添加 `max_transformers_version="4.57"` 和 `transformers_version_reason` 字段。
3. 填写原因: 在 `transformers_version_reason` 中为 `vllm` 和 `hf` 两种检查分别提供兼容性说明, 强调自定义代码在 v5 下不可用, 并建议上游化。
4. 无其他影响: 变更仅涉及测试注册元数据, 不修改任何运行时逻辑。

关键文件:

- `tests/models/registry.py` (模块 模型注册表; 类别 `test`; 类型 `test-coverage`): 新增 Isaac 模型的 Transformers v5 版本门控, 阻止其在 v5 下运行测试。

关键符号: 未识别

关键源码片段

`tests/models/registry.py`

新增 Isaac 模型的 Transformers v5 版本门控, 阻止其在 v5 下运行测试。

```
# tests/models/registry.py 中 IsaacForConditionalGeneration 条目变更
# 添加版本门控以在 Transformers v5 下跳过测试
"IsaacForConditionalGeneration": _HfExamplesInfo(
```

```
"PerceptronAI/Isaac-0.1",
trust_remote_code=True,
extras={"0.2-2B-Preview": "PerceptronAI/Isaac-0.2-2B-Preview"},
max_transformers_version="4.57", # 限制最大版本为 4.57, v5+ 跳过
transformers_version_reason={
    "vllm": (
        "Custom Isaac code is not compatible with Transformers v5. "
        "The model should be upstreamed to Transformers for "
        "long-term support."
    ),
    "hf": (
        "Isaac's remote model and processor code import or configure "
        "APIs that changed in Transformers v5."
    ),
},
),
```

评论区精华

讨论较少，但关键点是 [ywang96](#) 标记了 [oscardev256](#) 来关注此模型的 v5 修复 ('Not sure if you have bandwidth to fix this model with v5 but FYI!')，表明上游化是预期方向。

- Isaac v5 修复责任分配 (other): [oscardev256](#) 回应会查看。

风险与影响

- 风险：无风险。变更仅在测试注册表中添加版本门控，不影响任何运行时或核心路径。如果 `IsaacForConditionalGeneration` 在 Transformers v5 下意外兼容，门控仍可能导致误跳过测试，但概率极低。
- 影响：直接影响：CI 中 `IsaacForConditionalGeneration` 相关的测试在 Transformers v5 版本下会被自动跳过，避免因远程代码兼容性问题导致的失败。间接影响：门控提示开发者该模型需要上游化至 Transformers，避免后续重复排查。影响范围：仅限测试基础设施，无用户可见变化。
- 风险标记：暂无

关联脉络

- PR #38389 [Transformers v5] `IsaacForConditionalGeneration`: 直接关联的 Issue，记录了 Transformers v5 下的测试失败，是本 PR 的动机来源。