

# PR #38429 完整报告

vllm-project/vllm

[CI] Fix Ernie4.5-VL initialization test

合并时间: 2026-03-28 22:43

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/38429>

## 执行摘要

本 PR 通过将 Ernie4.5-VL 模型测试的 HF revision 从 [refs/pr/14](#) 更新至 [refs/pr/17](#), 修复了因上游 bug 导致的 CI 初始化测试失败, 变更仅涉及测试配置, 不影响产品核心功能。

## 功能与动机

PR body 引用上游 Hugging Face 仓库的讨论, 指出在 [baidu/ERNIE-4.5-VL-28B-A3B-PT](#) 的 PR #17 中修复了两个关键 bug:

- `Ernie4_5_MoEConfig.__init__` 中 `self.num_hidden_layers` 访问过早导致 `to_diff_dict()` 崩溃
- `processing_ernie4_5_vl.py` 中的 `import decord` 在没有安装 `decord` 时阻塞 tokenizer 加载 这些 bug 导致 vLLM 的 Ernie4.5-VL 模型初始化测试失败, 因此本 PR 旨在更新测试配置以使用修复后的版本。

## 实现拆解

变更集中在 `tests/models/registry.py` 文件的 `check_available_online` 函数中, 仅修改一行配置:

```
"Ernie4_5_VLMoeForConditionalGeneration": _HfExamplesInfo(
    "baidu/ERNIE-4.5-VL-28B-A3B-PT",
    trust_remote_code=True,
    revision="refs/pr/17", # 从 "refs/pr/14" 更新
),
```

无其他文件或逻辑改动, 属于简单的配置调整。

## 评论区精华

review 中, `gemini-code-assist[bot]` 提出重要建议:

```
"For improved test stability and reproducibility, it is best practice to pin the revision to a specific commit hash rather than a pull request reference."
```

这引发了关于 CI 测试中依赖管理设计的讨论: 使用 PR 引用 (如 [refs/pr/17](#)) 可能未来因引用可变导致测试失败, 而 commit hash 提供更稳定的保证。但最终 PR 未采纳该建议, 使用 PR 引用合并, 凸显了短期修复 vs 长期稳定性的权衡。

## 风险与影响

- 风险：使用 PR 引用而非 commit hash 增加未来测试不稳定风险，如果上游 PR 引用被移除或修改，可能再次引发 CI 失败。但鉴于上游 bug 已修复，短期风险可控。
- 影响：仅影响 CI 测试系统，确保 Ernie4.5-VL 多模态模型的初始化测试通过，对用户和产品系统无直接影响，属于低影响维护性变更。

## 关联脉络

从近期历史 PR 分析，PR #38369 (标题 '[CI] Skip failing test') 也修改了 [tests/models/registry.py](#) 文件，处理 Ernie 模型 CI 测试问题，标签相似 (ci, test, multi-modality)。这表明 vLLM 团队正持续优化多模态模型的 CI 测试配置，以应对上游依赖变化，维护测试稳定性。