

PR #43860 完整报告

vllm-project/vllm

[Bugfix] Fix HyperCLOVAX CI failure after upstream removed remote code

合并时间: 2026-05-28 18:37

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/43860>

执行摘要

- 一句话: 修复上游删除远程代码导致的 HyperCLOVAX CI 失败
- 推荐动作: 值得快速合并以解除 CI 阻塞。设计决策简单有效, 无需深度审查。

功能与动机

修复 CI 测试失败: buildkite 上 `test_can_initialize_large_subset[HyperCLOVAXForCausalLM]` 因上游仓库删除配置和模型文件而抛出 `OSError` (`OSError: naver-hyperclovax/HyperCLOVAX-SEED-Think-14B does not appear to have a file named configuration_hyperclovax.py`)。PR body 明确指出该故障由上游变更引起。

实现拆解

1. 注册模型配置: 在 `vllm/transformers_utils/config.py` 中的 `_CONFIG_REGISTRY` 字典中增加条目 `hyperclovax="HyperCLOVAXConfig"`。这样当 vLLM 加载 HyperCLOVAX 模型时, 会使用 vendored 的配置类而非远程配置文件, 绕过上游 `auto_map`。
2. 更新测试注册表: 在 `tests/models/registry.py` 中, 将 `HyperCLOVAXForCausalLM` 的 `_HfExamplesInfo` 中的 `trust_remote_code=True` 替换为 `min_transformers_version="5.9.0"`。这样当 transformers 版本低于 5.9.0 时, 测试会被跳过, 因为老版本不原生支持 HyperCLOVAX。

关键文件:

- `vllm/transformers_utils/config.py` (模块 配置加载; 类别 `source`; 类型 `core-logic`): 在 `_CONFIG_REGISTRY` 中注册 `hyperclovax` 模型类型, 使 vLLM 使用自身 vendored 的 `HyperCLOVAXConfig` 而非远程代码
- `tests/models/registry.py` (模块 测试注册表; 类别 `test`; 类型 `test-coverage`): 移除 `trust_remote_code` 并设置 `min_transformers_version`, 确保测试在 transformers `<5.9.0` 时跳过而非依赖远程代码

关键符号: 未识别

关键源码片段

[vllm/transformers_utils/config.py](#)

在 `_CONFIG_REGISTRY` 中注册 `hyperclovax` 模型类型，使 vLLM 使用自身 vendored 的 `HyperCLOVAXConfig` 而非远程代码

```
# vllm/transformers_utils/config.py (partial)

class LazyConfigDict(dict):
    """A dict that lazily imports config classes from vllm.transformers_utils.configs."""
    def __getitem__(self, key):
        if isinstance(value := super().__getitem__(key), type):
            return value
        import vllm.transformers_utils.configs as configs
        return getattr(configs, value)

# The registry maps model_type string (from HF config.json) to vendored config class names.
# Previously, hyperclovax was missing, so vLLM fell back to HF auto_map, which now points to
# deleted remote code.
_CONFIG_REGISTRY: dict[str, type[PretrainedConfig]] = LazyConfigDict(
    # ... other entries ...
    hyperclovax="HyperCLOVAXConfig", # newly added; bypasses stale HF auto_map
    hyperclovax_vlm="HCXVisionConfig", # vision models are unaffected (separate repos)
    # ... other entries ...
)
```

tests/models/registry.py

移除 `trust_remote_code` 并设置 `min_transformers_version`，确保测试在 `transformers < 5.9.0` 时跳过而非依赖远程代码

```
# tests/models/registry.py (partial)

# ... n. HyperCLOVAX 模型从 HF transformers >= 5.9.0 开始原生支持，
# 因此不再需要 trust_remote_code=True。设置 min_version 以跳过旧版本。
"HyperCLOVAXForCausalLM": _HfExamplesInfo(
    "naver-hyperclovax/HyperCLOVAX-SEED-Think-14B",
    min_transformers_version="5.9.0", # replaces former trust_remote_code=True
),
```

评论区精华

Reviewer `DarkLight1337` 在测试文件行上建议使用 `min_transformers_version` 而非直接移除 `trust_remote_code=True`，以避免在老版本 `transformers` 上出现未定义行为。作者 `khluu` 快速回应并实施该建议，在后续提交中添加了 `min_transformers_version="5.9.0"`。

- 使用 `min_transformers_version` 替代移除 `trust_remote_code` (testing): 作者接受建议，在后续提交中添加 `min_transformers_version="5.9.0"`。

风险与影响

- 风险：风险很低：变更仅增加一行注册并调整一个测试参数。注册 `hyperclovax` 到 `_CONFIG_REGISTRY` 是安全的，因为 `HyperCLOVAXConfig` 已在 vendored configs 中定

义但未注册。测试改用 `min_transformers_version` 可确保旧 transformers 版本跳过测试。不影响其他模型或非 HyperCLOVAX 变体（视觉模型使用独立仓库）。

- 影响：影响范围小：仅修复 HyperCLOVAX 语言模型的初始化测试。视觉模型（如 `hyperclovax_vlm`）不受影响，因为它们使用不同的 HF 仓库且仍保留远程代码。CI 管道将恢复绿色，解除对其他 PR 的阻塞。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR