

PR #47143 完整报告

vllm-project/vllm

[Model] Remove Tarsier, Tarsier2

合并时间: 2026-06-30 21:20

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/47143>

执行摘要

- 一句话: 移除已废弃的 Tarsier 和 Tarsier2 模型
- 推荐动作: 此 PR 属于技术债务清理, 不包含复杂设计决策。但因其涉及多文件联动修改 (模型、配置、注册表、示例、测试、文档), 可作为 vLLM 模型弃用流程的参考样例, 值得相关维护者阅读。

功能与动机

根据 PR 描述, 这两个模型“超过 1.5 年, 在 vLLM 中几乎无人使用, 其 HF 下载量因 CI 下载而虚高”。关联 Issue #38736 进一步指出 Tarsier2 的 config.json 存在嵌套结构, 在 Transformers 最新版下无法初始化, 需要维护 workaround。移除可减少维护负担。

实现拆解

1. 删除核心模型文件: 移除 vllm/model_executor/models/tarsier.py (591 行), 其中包含 TarsierForConditionalGeneration、TarsierImagePixelInputs、TarsierHfConfig、TarsierProcessor 等全部实现。
2. 移除自定义配置: 删除 vllm/transformers_utils/configs/tarsier2.py, 并从 vllm/transformers_utils/configs/__init__.py 和 vllm/transformers_utils/config.py 中清理导入和映射。
3. 清理寄生在 Qwen2VL 中的 Tarsier2 适配代码: 从 vllm/model_executor/models/qwen2_vl.py 移除 Tarsier2MultiModalProcessor、Tarsier2ImageProcessor、Tarsier2Processor、Tarsier2ProcessingInfo、Tarsier2ForConditionalGeneration 等类及不再需要的导入。
4. 更新模型注册表: 在 vllm/model_executor/models/registry.py 中将两个模型从 _ARCH_MAP 中移除, 并添加到 _PREVIOUSLY_SUPPORTED_MODELS (Tarsier: v0.24.0, Tarsier2: v0.23.0); 同时清理模块路径白名单。
5. 删除示例和测试: 移除 examples/generate/multimodal/vision_language_offline.py 和 vision_language_multi_image_offline.py 中的对应运行函数和映射; 删除 tests/models/multimodal/generation/vlm_utils/model_utils.py 中的 tarsier_patch_hf_runner 以及 test_common.py 和 tests/models/registry.py 中的相关测试用例。
6. 更新文档: 从 docs/models/supported_models.md 中移除对这两个模型的引用。

关键文件：

- `vllm/model_executor/models/tarsier.py`（模块 模型定义；类别 `source`；类型 `deletion`；符号 `TarsierImagePixelInputs`, `TarsierImageEmbeddingInputs`, `TarsierHfConfig`, `TarsierProcessorKwargs`）：核心模型文件被删除，包含 `Tarsier` 模型的完整实现。
- `vllm/model_executor/models/qwen2_vl.py`（模块 模型适配；类别 `source`；类型 `data-contract`；符号 `Tarsier2MultiModalProcessor`, `Tarsier2ImageProcessor`, `init`, `Tarsier2Processor`）：移除了寄生在 `Qwen2VL` 模型文件中的 `Tarsier2` 适配类。
- `vllm/transformers_utils/configs/tarsier2.py`（模块 配置系统；类别 `source`；类型 `deletion`；符号 `Tarsier2Config`）：`Tarsier2` 的自定义配置类被删除，该类用于绕过多层嵌套的 `config.json` 问题。
- `vllm/model_executor/models/registry.py`（模块 模型注册；类别 `source`；类型 `data-contract`）：更新模型注册表，移除自动注册并记录过去支持版本。
- `examples/generate/multimodal/vision_language_offline.py`（模块 示例代码；类别 `source`；类型 `core-logic`；符号 `run_tarsier`, `run_tarsier2`）：移除了 `Tarsier` 和 `Tarsier2` 的离线推理示例。
- `examples/generate/multimodal/vision_language_multi_image_offline.py`（模块 示例代码；类别 `source`；类型 `core-logic`；符号 `load_tarsier`, `load_tarsier2`）：移除了 `Tarsier` 和 `Tarsier2` 的多图推理示例。
- `vllm/transformers_utils/configs/__init__.py`（模块 配置系统；类别 `source`；类型 `core-logic`）：从配置导出中移除 `tarsier2` 模块引用。
- `tests/models/multimodal/generation/vlm_utils/model_utils.py`（模块 测试工具；类别 `test`；类型 `test-coverage`；符号 `tarsier_patch_hf_runner`）：移除 `Tarsier` 的 HF 测试补丁。

关键符号：`run_tarsier`, `run_tarsier2`, `load_tarsier`, `load_tarsier2`, `tarsier_patch_hf_runner`, `TarsierProcessor.call`, `Tarsier2ImageProcessor.init`, `Tarsier2Processor.init`, `Tarsier2ProcessingInfo.get_hf_config`, `Tarsier2ProcessingInfo.get_hf_processor`, `Tarsier2ProcessingInfo.get_image_processor`

关键源码片段

`vllm/model_executor/models/qwen2_vl.py`

移除了寄生在 `Qwen2VL` 模型文件中的 `Tarsier2` 适配类。

```
# Tarsier2ImageProcessor 适配了 Tarsier2 非标准的 size 参数格式 (min_pixels/max_pixels) ,
# 将其映射为 Qwen2VLImageProcessor 所需的 shortest_edge/longest_edge。
# Tarsier2 移除后，此类已不再需要。
class Tarsier2ImageProcessor(Qwen2VLImageProcessor):
    def __init__(
        self,
        size: dict[str, int] | None = None,
        **kwargs,
    ) -> None:
        if size is not None and "min_pixels" in size and "max_pixels" in size:
            # 将 Tarsier2 专有格式映射为标准 Qwen2VL 格式
```

```
remapped_size = {
    "shortest_edge": size["min_pixels"],
    "longest_edge": size["max_pixels"],
}
super().__init__(size=remapped_size, **kwargs)
else:
    super().__init__(size=size, **kwargs)
```

vllm/model_executor/models/registry.py

更新模型注册表，移除自动注册并记录过去支持版本。

```
# 记录曾支持但已废弃的模型及其最后兼容版本，协助用户从旧版本升级。
_PREVIOUSLY_SUPPORTED_MODELS = {
    # ... 已有条目 ...
    # Tarsier 最后在 v0.24.0 中受支持
    "TarsierForConditionalGeneration": "0.24.0",
    # Tarsier2 最后在 v0.23.0 中受支持（最后一个使用 Transformers v4 的版本）
    "Tarsier2ForConditionalGeneration": "0.23.0",
}
```

评论区精华

PR 的 review 过程没有实质性讨论；仅有一个自动 bot 评论和一个 Approved 审批。变更直接获得合并。

- 暂无高价值评论线程

风险与影响

- 风险：
 - 向后兼容风险：已使用 Tarsier/Tarsier2 的用户在升级后模型将不可用。但已通过 `_PREVIOUSLY_SUPPORTED_MODELS` 记录最后版本，用户可回退或迁移。
 - 测试覆盖损失：移除了相关测试，但这两个模型 CI 覆盖率原本就低，影响可控。
 - 配置链影响：Tarsier2Config 继承自 Qwen2VLCConfig，移除后不会影响 Qwen2VL 系列模型。
- 影响：
 - 用户影响：少数仍依赖这两个模型的用户需要固定 vLLM 版本或迁移到其他模型。
 - 系统影响：减少代码库大小约 872 行，降低维护成本。
 - 团队影响：减少因 Transformers 升级带来的兼容性修复工作。
 - 风险标记：模型废弃通知不足，用户可能静默失败，测试覆盖移除

关联脉络

- PR #38736 [Transformers v5] Tarsier2ForConditionalGeneration: Issue 报告了 Tarsier2 在 Transformers v5 下的初始化失败，是此 PR 的直接动机之一。