

PR #43784 完整报告

vllm-project/vllm

Deprecate `JAISLMHeadModel`

合并时间: 2026-05-29 02:29

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/43784>

执行摘要

- 一句话: 废弃 JAISLMHeadModel 模型, 移除支持
- 推荐动作: 该 PR 提供了一个良好的废弃模型清理模板, 包括删除源代码、更新注册表、添加废弃记录、更新测试和文档。值得在类似场景下参考。

功能与动机

该架构已近 3 年历史, 已被 Jais2 取代。唯一有显著下载量的检查点是 vLLM CI 使用的检查点。用户可继续使用 vLLM ≤ 0.22 来使用该模型。

实现拆解

1. 删除模型源代码文件 `vllm/model_executor/models/jais.py` 和配置类文件 `vllm/transformers_utils/configs/jais.py`, 移除所有与 JAIS 相关的类定义 (JAISLMHeadModel、JAISAttention、JAISMLP、JAISBlock 等)。
2. 在模型注册表 `vllm/model_executor/models/registry.py` 中, 从 `_TEXT_GENERATION_MODELS` 字典移除 "JAISLMHeadModel" 条目, 并添加到 `_PREVIOUSLY_SUPPORTED_MODELS` 中指定最后版本为 0.22.0, 以便用户在升级后加载旧模型时获得清晰提示。
3. 清理配置加载路径: 在 `vllm/transformers_utils/configs/__init__.py` 和 `vllm/transformers_utils/config.py` 中移除 JAISConfig 的映射, 避免配置类被错误加载。
4. 更新测试文件 `tests/models/registry.py` 和 `tests/distributed/test_pipeline_parallel.py`, 移除 JAIS 相关的测试引用和用例。
5. 更新文档 `docs/models/supported_models.md`, 从支持模型列表中移除 JAIS。

关键文件:

- `vllm/model_executor/models/jais.py` (模块 模型层; 类别 source; 类型 deletion; 符号 SwiGLUActivation, forward, `_get_alibi_slopes`, `get_slopes_power_of_2`): 模型核心实现文件, 删除所有 JAIS 相关的层定义 (SwiGLUActivation、JAISAttention、JAISMLP、JAISBlock 等)。
- `vllm/transformers_utils/configs/jais.py` (模块 配置加载; 类别 source; 类型 deletion; 符号 JAISConfig, init, `_alibi_scaling_validation`): JAIS 模型的配置类定义, 删除 JAISConfig 和相关的验证函数。

- `vllm/model_executor/models/registry.py` (模块 模型注册; 类别 `source`; 类型 `data-contract`) : 模型注册表, 移除了 `JAISLMHeadModel` 的映射, 并添加到已废弃模型列表中。
- `vllm/transformers_utils/configs/__init__.py` (模块 配置加载; 类别 `source`; 类型 `core-logic`) : 配置类映射字典, 移除了 `JAISConfig` 的导入和导出。
- `vllm/transformers_utils/config.py` (模块 配置加载; 类别 `source`; 类型 `core-logic`) : 模型短名到配置类的映射, 移除了 `jais` 条目。
- `tests/models/registry.py` (模块 测试; 类别 `test`; 类型 `test-coverage`) : 测试文件, 移除了 `JAISLMHeadModel` 的测试引用。
- `tests/distributed/test_pipeline_parallel.py` (模块 测试; 类别 `test`; 类型 `test-coverage`) : 分布式测试文件, 移除了 `JAIS` 相关的测试用例。
- `docs/models/supported_models.md` (模块 文档; 类别 `docs`; 类型 `documentation`) : 模型支持列表文档, 移除了 `JAIS` 条目。

关键符号: `SwiGLUActivation.forward`, `_get_alibi_slopes`, `get_slopes_power_of_2`, `JAISAttention.init`, `JAISMLP`, `JAISBlock`, `JAISConfig.init`, `JAISConfig._alibi_scaling_validation`

关键源码片段

`vllm/model_executor/models/registry.py`

模型注册表, 移除了 `JAISLMHeadModel` 的映射, 并添加到已废弃模型列表中。

```
# 活跃模型列表 (部分)
_TEXT_GENERATION_MODELS = {
    # ...
    # "JAISLMHeadModel": ("jais", "JAISLMHeadModel"), # 已移除
    "Jais2ForCausalLM": ("jais2", "Jais2ForCausalLM"),
    # ...
}

# 已废弃模型列表, 记录最后支持的版本
_PREVIOUSLY_SUPPORTED_MODELS = {
    "JAISLMHeadModel": "0.22.0", # 用户在 v0.22.0 之后加载此模型会收到错误提示
    # ...
}
```

`vllm/transformers_utils/configs/__init__.py`

配置类映射字典, 移除了 `JAISConfig` 的导入和导出。

```
# 变更后: 移除 JAISConfig 相关行
_CLASS_TO_MODULE: dict[str, str] = {
    # ...
    # "JAISConfig": "vllm.transformers_utils.configs.jais", # 已删除
    "LagunaConfig": "vllm.transformers_utils.configs.laguna",
    # ...
}
```

```
__all__ = [  
    # ...  
    # "JAISConfig", # 已从 __all__ 删除  
    "LagunaConfig",  
    # ...  
]
```

评论区精华

该 PR 没有实质性的 review 讨论。仅由 Isotr0py 审批通过，无评论。说明变更合规且争议小。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。JAIS 模型已过时且有替代 Jais2。用户若仍需使用旧模型，可锁定 vLLM 版本 ≤ 0.22 。主要风险是用户升级后未注意兼容性说明而误加载失败，但已在 registry 中添加废弃模型提示。
- 影响：影响范围较小。对使用 JAIS 模型的用户是破坏性变更，但提供了明确的降级路径。对 CI 的影响是移除了该模型的测试，减少 CI 负担。对团队来说是代码清理，降低维护成本。
- 风险标记：模型废弃，用户兼容性，降级路径

关联脉络

- 暂无明显关联 PR