

PR #45128 完整报告

vllm-project/vllm

[Model] Remove InternLMForCausalLM registry alias

合并时间: 2026-06-12 04:02

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/45128>

执行摘要

- 一句话: 移除 InternLMForCausalLM 注册别名
- 推荐动作: 值得一读以了解 vLLM 模型别名弃用流程。展示了如何平衡向后兼容与清理遗留代码, 包括使用 `_PREVIOUSLY_SUPPORTED_MODELS` 进行优雅降级。

功能与动机

InternLMForCausalLM 是 InternLM-v1 的遗留别名, 已被 InternLM2/3 完全取代, 过去 6 个月仅 44 个实例使用 (0.007%), 维护成本远大于收益。用户可通过重新标记架构或 `--hf-overrides` 改用 LlamaForCausalLM。

实现拆解

1. 在 `vllm/model_executor/models/registry.py` 的 `_TEXT_GENERATION_MODELS` 字典中删除 `"InternLMForCausalLM": ("llama", "LlamaForCausalLM")` 一行。
2. 在同一文件的 `_PREVIOUSLY_SUPPORTED_MODELS` 字典中新增 `"InternLMForCausalLM": "0.23.0"`, 表示从该版本起不再直接支持。
3. 从 9 个共享的 `LlamaDecoderLayer` 类文件 (`apertus.py`, `exaone.py`, `exaone4.py`, `exaone_moe.py`, `granite.py`, `jais2.py`, `llama.py`, `nemotron.py`, `nemotron_nas.py`, `solar.py`) 中删除 `# Support internlm/internlm-7b with bias` 注释行, 消除误导。
4. 同步更新测试文件 `tests/models/registry.py`, 移除对 InternLMForCausalLM 的引用测试用例; 更新文档 `docs/models/supported_models.md`, 从表中移除该模型行。

关键文件:

- `vllm/model_executor/models/registry.py` (模块 模型注册; 类别 source; 类型 data-contract) : 核心变更: 移除 InternLMForCausalLM 注册别名, 并添加至 `_PREVIOUSLY_SUPPORTED_MODELS`
- `vllm/model_executor/models/apertus.py` (模块 模型实现; 类别 source; 类型 data-contract) : 清理 `internlm/internlm-7b` 遗留注释
- `tests/models/registry.py` (模块 注册测试; 类别 test; 类型 test-coverage) : 测试文件同步更新, 移除 InternLMForCausalLM 相关测试
- `docs/models/supported_models.md` (模块 文档; 类别 docs; 类型 documentation) : 文档表移除 InternLMForCausalLM 行, 同步变更

关键符号：未识别

关键源码片段

vllm/model_executor/models/registry.py

核心变更：移除 InternLMForCausalLM 注册别名，并添加至
_PREVIOUSLY_SUPPORTED_MODELS

```
# vllm/model_executor/models/registry.py

# _TEXT_GENERATION_MODELS 字典（部分）
# 移除了以下键（原来映射到 LlamaForCausalLM）：
# "InternLMForCausalLM": ("llama", "LlamaForCausalLM"),
# 保留 InternLM2/3 等后续模型：
"InternLM2ForCausalLM": ("internlm2", "InternLM2ForCausalLM"),
"InternLM2VEForCausalLM": ("internlm2_ve", "InternLM2VEForCausalLM"),
"InternLM3ForCausalLM": ("llama", "LlamaForCausalLM"),

# _PREVIOUSLY_SUPPORTED_MODELS 字典（部分）
# 添加 InternLMForCausalLM 并从 v0.23.0 起不再直接支持：
_PREVIOUSLY_SUPPORTED_MODELS = {
    "InternLMForCausalLM": "0.23.0",
    # ... 其他已移除模型 ...
}
```

评论区精华

Isotr0py 指出应同时将 InternLMForCausalLM 加入 _PREVIOUSLY_SUPPORTED_MODELS 以保证向后兼容，已采纳。noooop 建议全面清理所有提及 InternLM 的注释，PR 作者做了清理但保留了 bias 字段检查（其他模型如 telechat 仍使用），noooop 同意仅清理注释。

- 添加 InternLMForCausalLM 到 _PREVIOUSLY_SUPPORTED_MODELS (correctness): 作者采纳建议，在 _PREVIOUSLY_SUPPORTED_MODELS 中添加了 "InternLMForCausalLM": "0.23.0"。
- 清理所有 InternLM 相关注释 (style): 仅在 9 个模型文件中删除了注释行，未改变代码逻辑。

风险与影响

- 风险：低风险。别名使用极少，且已通过 _PREVIOUSLY_SUPPORTED_MODELS 提供错误提示。唯一风险是极少数用户仍依赖 InternLMForCausalLM 而不知如何切换，但文档和错误信息已提供指引。注释清理不改变任何行为。
- 影响：对绝大多数用户无影响。InternLMForCausalLM 使用量仅为 InternLM2ForCausalLM 的 0.007%。受影响用户可通过简单配置迁移。团队维护成本降低：移除了一个一行别名和散布的过时注释。
- 风险标记：低使用率别名，向后兼容性，文档更新

关联脉络

- 暂无明显关联 PR