

PR #46396 完整报告

vllm-project/vllm

[Docs] Add Phi-3.5-mini-instruct to batch invariance tested models

合并时间: 2026-07-17 06:05

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/46396>

执行摘要

本 PR 将 `microsoft/Phi-3.5-mini-instruct` 加入批量不变性 (batch invariance) 已测试模型列表, 仅修改文档 `docs/features/batch_invariance.md` 增加一行。作者完成了全部 16 个测试用例并通过, 审核者确认后合并。

功能与动机

Phi-3.5-mini-instruct 是微软的 3.8B 小型模型, 适合边缘部署和需要确定性推理的场景。将其纳入已验证模型列表可扩大覆盖范围, 方便用户参考。本 PR 是 Issue #27433 中“模型覆盖”子任务的一部分。

实现拆解

- 在文档文件 `docs/features/batch_invariance.md` 的第 107 行 (模型列表末尾) 增加一行:
: - **Phi series**:microsoft/Phi-3.5-mini-instruct`
- 提交并推送, 无其他更改。

无可用源码片段 (纯文档变更)。

评论区精华

- 审核者 yewentao256: " 您应该使用 `pytest test_batch_invariance.py` 并收集完整输出。" 作者按要求重新测试, 输出显示 16 个测试全部通过。
- 审核者最终评论: "LGTM, thanks for the work!"

风险与影响

无技术风险。文档更新对用户有轻微正面影响, 增加模型覆盖参考信息。

关联脉络

本 PR 是 Issue #27433 (批量不变性特性跟踪) 中模型覆盖部分的延续, 此前已有多项 PR 添加其他模型到该列表。