

PR #42066 完整报告

vllm-project/vllm

[Docs] Fix OpenAI batch model argument examples

合并时间: 2026-05-08 22:02

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/42066>

执行摘要

该 PR 修复了 `examples/features/openai_batch/README.md` 中两处 OpenAI batch S3 示例命令的重复 `--model` 参数问题。仅修改 2 行, 属于低风险、高清晰度的文档修复。

功能与动机

原文档中的 S3 批量推理命令包含 `--model --model meta-llama/Meta-Llama-3-8B-Instruct`, 这是无效参数, 用户直接复制会出错。PR 旨在解决此问题, 使示例命令可正确运行。

实现拆解

1. 定位文件: `examples/features/openai_batch/README.md`。
2. 定位两处 S3 示例命令 (分别对应 `python -m vllm.entrypoints.openai.run_batch` 和 `vllm run-batch`) 。
3. 将 `--model --model meta-llama/Meta-Llama-3-8B-Instruct` 改为 `--model meta-llama/Meta-Llama-3-8B-Instruct`。
4. 通过 `git diff --check` 和 `markdownlint` 验证。

无源码变更。

评论区精华

reviewer noooop 直接批准并留言 "thanks for your fix", Gemini-code-assist 机器人表示无进一步反馈。

风险与影响

该 PR 无风险, 仅修改文档注释。影响范围小, 仅提升文档正确性和用户体验。

关联脉络

该 PR 是独立的文档修复, 与近期其他 PR 无直接关联。