

PR #38264 完整报告

vllm-project/vllm

[Mypy] Fix adjust_request typing

合并时间: 2026-03-31 12:21

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/38264>

执行摘要

本次 PR 修复了工具解析器中 `adjust_request` 方法的类型注解, 使其支持 `ChatCompletionRequest` 和 `ResponsesRequest` 两种请求类型, 移除了 `servicing.py` 中的类型忽略注释, 并更新了相关文档。变更琐碎, 风险低。

功能与动机

动机是扩大 `adjust_request` 方法的签名, 以接受 `ChatCompletionRequest` | `ResponsesRequest` 类型, 从而增强类型安全性。根据 PR body, 需要移除 `servicing.py` 中的 `# type: ignore[arg-type]` 注释, 因为 `ResponsesRequest` 已被传递。

实现拆解

- 基类调整: 修改 `vllm/tool_parsers/abstract_tool_parser.py` 中的 `adjust_request` 方法, 更新类型注解。
- 具体解析器: 更新多个具体工具解析器文件 (如 `deepseekv32_tool_parser.py` 等) 中的同名方法。
- 服务层修复: 在 `vllm/entrypoints/serve/render/servicing.py` 中移除类型忽略注释。
- 文档更新: 更新 `docs/features/tool_calling.md` 以反映新签名。

评论区精华

review 中没有实质性技术讨论。reviewer `aarnphm` 和 `chaunceyjiang` 直接批准, `chaunceyjiang` 评论: 'Thanks @sfeng33. This has been on my TODO list for a while, but I haven't had the time to work on it yet.'

风险与影响

风险: 变更仅限于类型注解和文档, 无业务逻辑修改, 风险极低。但需确保 `mypy` 静态检查通过。

影响: 对用户无直接影响; 对开发团队, 提升了代码类型安全性和可维护性。

关联脉络

与 PR #38265 相关, 后者统一了工具类型别名, 同属工具调用模块的类型改进系列。