

PR #31917 完整报告

sgl-project/sglang

:busts_in_silhouette: chore(codeowners): Update codeowners for NPU quantization

合并时间: 2026-07-29 17:10

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/31917>

PR#31917 分析报告

执行摘要

本 PR 为 Ascend NPU 量化相关目录添加了 @TallMessiWu 作为 CODEOWNERS，确保 MXFP8/MXFP4 等量化代码的变更自动路由给该领域的主要贡献者。变更简单，风险极低，已获团队批准。

功能与动机

提交者 TallMessiWu 是 Ascend NPU 上 MXFP8/MXFP4 量化支持的主要作者，涉及多个已合并PR（#20922、#22338、#30768等）。通过[gitblame](#)确认其对相关目录有大量代码贡献。添加其为代码负责人可确保未来变更得到及时审核，减少遗漏。

实现拆解

只修改了一个文件: [.github/CODEOWNERS](#)

1. 在 /python/sglang/srt/hardware_backend/npu/quantization 条目后追加 @TallMessiWu
2. 在 /python/sglang/srt/layers/quantization 条目后追加 @TallMessiWu
3. 新增条目 /python/sglang/multimodal_gen/runtime/layers/quantization/
@OrangeRedeng @TallMessiWu（根据 review 意见调整）

评论区精华

审核者 ping1jing2 要求调整扩散量化目录的负责人格式，提交者按要求修改。

风险与影响

- 风险：极低，仅涉及配置文件，不影响代码逻辑
- 影响：新增代码负责人后，相关路径的 PR 将自动请求 @TallMessiWu 审核，有助于维护代码质量和团队分工

关联脉络

本 PR 是对 PR#26784（建立 NPU 量化代码负责人）的扩展，与 #30768（MoE W8A8 MXFP8 量化）等已有量化 PR 相关。