

PR #26270 完整报告

sgl-project/sglang

Lazy-load speculative-naming via skill instead of always-on rule

合并时间: 2026-05-25 15:58

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/26270>

执行摘要

将 speculative-decoding 命名规范从全局加载的规则文件 (.claude/rules/speculative-naming.md) 迁移至按需加载的技能文件 (.claude/skills/speculative-naming/SKILL.md), 非 spec 任务不再浪费约 100 行 context。变更仅涉及文档配置文件, 无运行时影响。

功能与动机

规则文件之前被设置成项目级规则, 每次 Claude Code 在 sglang 仓库中工作时都会加载到 agent 上下文中, 但该规则仅对 speculative-decoding 相关代码有意义。PR body 指出“每个无关任务加载~100行永远用不到的指令”, 因此将其改为按需技能, 使规则只在需要时被加载。

实现拆解

1. 新增技能文件: 在 .claude/skills/speculative-naming/ 下创建 SKILL.md, 包含 YAML frontmatter (name、description), 以及原来规则的 1-6 全部内容、异常说明和范围表。内容与原规则文件保持字节一致。
2. 精简规则文件: 将 .claude/rules/speculative-naming.md 从 106 行详细规则减少为一行注释, 指向技能文件。
3. 无功能变更: 所有规则文本均未改动, 仅重组文件结构。

无需展示, 仅文档 Markdown 重组, 无实质代码片段。

评论区精华

- gemini-code-assist[bot]指出技能文件中“Rule 7”引用不存在 (实际只有 6 条规则), 应为“Rule 6”。该问题在合并前未被修复。

风险与影响

- 风险: 若 Claude 版本更新导致技能自动发现机制失效, 规则可能不会被加载; 另外 Rule 7 引用的错误可能造成读者困惑。但这些风险都很低。
- 影响: 对使用 Claude Code 的开发者有正面影响 (减少不必要 context)。对系统用户无影响。

关联脉络

此 PR 是独立的配置优化, 与近期其他 spec-decoding 或 loRA bugfix 无直接关联。