

PR #21262 完整报告

sgl-project/sclang

Fix customized_info offset truncation

合并时间: 2026-03-26 10:57

原文链接: <http://prhub.com.cn/sgl-project/sclang/pull/21262>

执行摘要 本次 PR 修复了 `customized_info` 在输出因停止令牌截断时的切片偏移错误，通过匹配 `output_ids` 的长度确保数据一致性。变更微小但针对核心调度器输出处理器模块，以预防潜在的数据损坏。

功能与动机 动机源于 PR body 中描述的问题：当生成输出被截断时，`customized_info` 的切片逻辑不正确，导致与 `output_ids` 不匹配。这会影响自定义信息的正确处理，需修复以确保数据对齐。

实现拆解 关键改动位于文件 `python/sclang/srt/managers/scheduler_output_processor_mixin.py` 的 `stream_output_generation` 函数中。将 `v[send_token_offset:]` 修改为 `v[send_token_offset : len(output_ids_)]`，使切片结束索引基于 `output_ids` 的长度，从而在截断情况下正确切片。

评论区精华 Review 过程中无实质性讨论。审核者 Qiaolin-Yu 直接批准了变更，未提出任何评论或疑虑。

风险与影响 风险较低，但需注意切片索引的边界情况，避免数组越界。影响范围限于使用 `customized_info` 的功能，修复后提高数据可靠性，对系统整体影响小。

关联脉络 从近期历史 PR 看，无直接相关变更。此 PR 是独立的 bugfix，可能与其他调度或输出处理改进无直接关联。