

PR #44608 完整报告

vllm-project/vllm

[Bugfix][Responses API] Set id on function_call item in streaming done event

合并时间: 2026-06-11 11:52

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/44608>

执行摘要

- 一句话: 修复函数调用流式事件 id 为空
- 推荐动作: 值得快速合入, 属于明确的字段映射 bugfix, 改动小而精准。

功能与动机

修复 Harmony 流式函数调用完成事件中 id 为 null 的 bug, 确保 output_item.done 事件包含正确的 id 字段。PR body 指出 item_id 被存储为额外字段, 而真正的 id 保持 None。

实现拆解

1. 在 vllm/entrypoints/openai/responses/streaming_events.py 的 emit_function_call_done_events 函数中, 将 ResponseFunctionToolCall 构造参数 item_id 替换为 id。
2. 该修改与已有 emit_simple_tool_call_done 模式一致, 确保 response.output_item.done 事件中 id 字段不再为 null。

关键文件:

- vllm/entrypoints/openai/responses/streaming_events.py (模块 请求路由; 类别 source; 类型 core-logic; 符号 emit_function_call_done_events): 核心修复文件, 修改了 ResponseFunctionToolCall 的构造参数, 将 item_id 改为 id。

关键符号: emit_function_call_done_events

关键源码片段

[vllm/entrypoints/openai/responses/streaming_events.py](#)

核心修复文件, 修改了 ResponseFunctionToolCall 的构造参数, 将 item_id 改为 id。

```
# vllm/entrypoints/openai/responses/streaming_events.py
# 函数调用完成时, 发射 arguments done 事件和 output item done 事件
# 修复前 id 字段始终为 None, 因为传入了错误的 item_id 参数
def emit_function_call_done_events(
    function_name: str,
    arguments: str,
    state: StreamingState,
) -> list[StreamingResponsesResponse]:
```

```

events: list[StreamingResponsesResponse] = []
events.append(
    ResponseFunctionCallArgumentsDoneEvent(
        type="response.function_call_arguments.done",
        arguments=arguments,
        name=function_name,
        item_id=state.current_item_id, # 此 event 使用 item_id, 正确
        output_index=state.current_output_index,
        sequence_number=-1,
    )
)
function_call_item = ResponseFunctionToolCall(
    type="function_call",
    arguments=arguments,
    name=function_name,
    # 修复: 将 item_id 改为 id, 这样 output_item.done 中 id 字段不再为 null
    id=state.current_item_id,
    output_index=state.current_output_index,
    sequence_number=-1,
    call_id=state.current_call_id,
    status="completed",
)
events.append(
    ResponseOutputItemDoneEvent(
        type="response.output_item.done",
        sequence_number=-1,
        output_index=state.current_output_index,
        item=function_call_item,
    )
)
return events

```

评论区精华

仅有 reviewer sfeng33 批准，无其他讨论。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低：仅修改了一个构造参数的键名，从 item_id 改为 id，且同函数内 item_id 在其他位置（ResponseFunctionCallArgumentsDoneEvent）仍保留，不影响其他逻辑。
- 影响：影响范围为 Responses API 流式函数调用事件消费者。修复后，客户端能够正确获取函数调用项的 id，提升兼容性和一致性。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR