

PR #42188 完整报告

vllm-project/vllm

[Bugfix] Gemma 4 chat template crash with missing tool name and tool id

合并时间: 2026-05-11 11:07

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/42188>

执行摘要

修复 Gemma 4 聊天模板中当工具结果名称缺失或 `tool_call_id` 不匹配时, 因 Jinja `default` 过滤器未正确处理 `None` 而触发的硬崩溃。变更仅涉及 1 个模板文件中的 2 处修改, 风险极低。

功能与动机

Gemma 4 的聊天模板在工具结果缺少名称时本应默认使用 "unknown", 但由于 `default` 过滤器未设置 `true` 参数, `None` 值被保留, 导致后续 `format_tool_response_block` 尝试拼接字符串时抛出 `TypeError: can only concatenate str (not "NoneType") to str`。该问题同时影响遗留的 `tool_responses` 路径和 OpenAI Chat Completions 路径 (当消息缺少 `name` 或 `tool_call_id` 不匹配时)。

实现拆解

1. 修复 `tool_responses` 路径 (第 266 行): 在 `tool_response['name'] | default('unknown')` 后加 `true` 参数, 使 `None` 和空字符串均回退为 "unknown"。
2. 修复消息路径 (第 280 行): 同样为 `follow.get('name') | default('unknown')` 添加 `true` 参数, 确保 `name` 缺失或为空时正确回退。

examples/tool_chat_template_gemma4.jinja

唯一变更文件, 修复了 Gemma 4 聊天模板中两处 Jinja `default` 过滤器缺少 `true` 参数的问题, 防止工具名称为空时崩溃。

```
{#- 第 266 行: tool_responses 路径 -#}
{% if message.get('tool_responses') %}
  {% for tool_response in message['tool_responses'] %}
    {{ format_tool_response_block(
      tool_response['name'] | default('unknown', true), # 修复前为 default('unknown'), None
      导致崩溃
      tool_response['response']
    ) }}
  {% endfor %}
{% endif %}

{#- 第 280 行: 消息路径 -#}
{% set ns_tname = namespace(
  name=follow.get('name') | default('unknown', true) # 同样修复, 确保 name 缺失时安全回退
) %}
```

评论区精华

- 作者 (yzong-rh) : 此 bug 是在调试 PR #42189 时发现的边角情况。
- bbrowning 询问是否需要同步上游 HF 模板, 作者表示应协调。

风险与影响

风险: 极低。仅修改两处模板字符串, 不影响其他模型或逻辑。但上游 HF 模板仍存在相同 bug, 需单独跟进。

影响: 仅影响使用 Gemma 4 模型且工具调用结果可能缺少名称或 `tool_call_id` 不匹配的用户。修复后这些场景不再崩溃, 回退为 "unknown"。

关联脉络

该 PR 与 #42189 (未在历史 PR 列表中) 相关, 是调试该 PR 时发现的边角问题。建议与 HuggingFace 上游团队协同修复默认模板。