

# PR #22890 完整报告

sgl-project/sglang

test: fix flaky required function calling assertion

合并时间: 2026-04-17 00:44

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/22890>

## 执行摘要

- 一句话: 修复 OpenAI 函数调用测试中因随机采样导致的 flaky 断言。
- 推荐动作: 建议工程师阅读此 PR 以理解如何设计更鲁棒的测试, 特别是处理非确定性行为时。关注测试断言从具体硬编码到通用验证的转变, 以及消除随机性的策略。review 中提到的 `strict=True` 设置问题值得后续关注。

## 功能与动机

PR body 指出, `tool_choice="required"` 仅保证模型必须调用某个工具, 但不指定具体工具。旧测试硬编码断言为 `get_weather`, 在 `temperature=0.8` 和 constrained decoding 下, 小模型如 `Llama-3.2-1B` 偶尔会选择 `sub` 工具, 导致测试 flaky。此问题由 #21593 引入, 而 #22586 修复了空参数问题但未解决工具选择 flakiness。

## 实现拆解

1. 修复测试逻辑: 在 `test_function_call_required` 函数中, 更新文档字符串以明确 `tool_choice="required"` 只保证调用工具, 不指定哪个。收集所有工具名称到 `valid_tool_names` 集合中。
2. 消除随机性: 将 `temperature` 从 0.8 改为 0, 并移除 `top_p` 参数, 使采样确定性。
3. 通用化断言: 将硬编码的 `assertEqual(function_name, "get_weather")` 改为 `assertIn(function_name, valid_tool_names)`, 只验证工具名称在提供的候选列表中。
4. 验证参数格式: 移除对具体参数内容 (如 `city` 值) 的断言, 只验证参数是可解析的 JSON 对象, 提高测试鲁棒性。
5. 更新用户消息: 将用户消息从 "What is the capital of France?" 改为 "Tell me about Paris", 使上下文更相关。涉及文件: `test/registered/openai_server/function_call/test_openai_function_calling.py`。关键符号: `test_function_call_required` 函数。

关键文件:

- `test/registered/openai_server/function_call/test_openai_function_calling.py` (模块 功能调用测试; 类别 `test`; 类型 `test-coverage`; 符号 `test_function_call_required`): 唯一变更的文件, 修复了 flaky 测试 `test_function_call_required`。

关键符号: `test_function_call_required`

## 评论区精华

在 review 中，gemini-code-assist[bot] 指出新实现移除了设置所有工具 `strict=True` 的循环，仅在部分工具中显式添加 `strict=True`。这可能导致工具间行为不一致，潜在重新引入 flakiness。评论建议恢复循环以确保一致性，但此问题在 PR 合并前未明确解决。

- `strict=True` 设置的一致性 (correctness): 未在 PR 中解决，但 PR 已合并。

## 风险与影响

- 风险：风险较低，主要限于测试逻辑。然而，review 中提到的 `strict=True` 设置不一致可能导致某些工具参数验证不严格，在边缘情况下影响测试稳定性。此外，将 `temperature` 设置为 0 可能掩盖了真实场景中的随机行为，但鉴于这是单元测试，目的是确保协议正确性，此风险可控。
- 影响：对用户无直接影响，因为这只是测试修复。对系统，提高了测试套件的稳定性和可靠性，减少 CI flakiness。对团队，减少了因 flaky 测试导致的 CI 失败和重试开销，提升开发效率。
- 风险标记：测试逻辑变更，潜在不一致设置

## 关联脉络

- PR #21593 未知，根据 PR body 引用：引入了 `tool_choice="required"` 协议语义变更，导致测试 flakiness。
- PR #22586 未知，根据 PR body 引用：修复了空参数失败但未解决工具选择 flakiness。