

PR #2779 完整报告

radixark/miles

test(dsv4): accept ValueError from reasoning-effort validation

合并时间: 2026-08-27 14:53

原文链接: <http://prhub.com.cn/radixark/miles/pull/2779>

执行摘要

- 一句话: 测试适配 sglang 新异常类型
- 推荐动作: 此 PR 值得快速浏览, 但无需精读。它展示了处理依赖库升级导致的测试兼容性问题最佳实践: 通过匹配错误消息而非异常类型, 使测试对依赖版本变化更具鲁棒性。

功能与动机

CI 的 stage-a-cpu 运行在 test_reasoning_effort_invalid_value_raises 上失败, 因为 sglang 修改了 reasoning-effort 校验方式: 从 assert 改为抛出 ValueError ('Invalid reasoning effort ... expected one of [high, max]'), 而测试固定使用 pytest.raises(AssertionError)。该问题在 PR #2777 的 CI 运行中暴露, 但与 #2777 的改动无关。

实现拆解

仅修改单个测试文件 tests/fast/utils/chat_template_utils/test_deepseek_v4.py。核心改动是:

1. 将 pytest.raises(AssertionError) 改为 pytest.raises((AssertionError, ValueError), match="Invalid reasoning effort"), 同时接受两种异常类型。
2. 更新注释, 说明 sglang 不同版本的行为差异 (旧版 assert, 新版 ValueError), 并强调通过消息匹配而非异常类型来追踪校验逻辑。
3. 无配置、部署或 schema 变更, 属于纯测试适配。

关键文件:

- tests/fast/utils/chat_template_utils/test_deepseek_v4.py (模块 测试; 类别 test; 类型 test-coverage; 符号 test_reasoning_effort_invalid_value_raises): 核心测试文件, 修改了 reasoning-effort 校验的异常断言, 使测试兼容不同 sglang 版本。

关键符号: test_reasoning_effort_invalid_value_raises

关键源码片段

[tests/fast/utils/chat_template_utils/test_deepseek_v4.py](#)

核心测试文件, 修改了 reasoning-effort 校验的异常断言, 使测试兼容不同 sglang 版本。

```
# tests/fast/utis/chat_template_utils/test_deepseek_v4.py

def test_reasoning_effort_invalid_value_raises():
    # Value-level validation is delegated to sglang; depending on the sglang
    # version it asserts the allowed set or raises ValueError. Both carry the
    # same "Invalid reasoning effort" message.
    # 这里通过消息匹配而不是固定异常类型，使测试对 sglang 版本变化更健壮。
    with pytest.raises((AssertionError, ValueError), match="Invalid reasoning effort"):
        deepseek.V4.render_messages(_MSGS_BASIC, thinking_mode="thinking", reasoning_effort=
            "ultra")
```

评论区精华

无 review 评论，仅有 claude[bot] 的自动提示和 guapisolo 的批准。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。仅修改测试断言，不影响生产代码。潜在小风险：如果 sglang 未来移除 ValueError 或改变消息文本，测试会失败；但当前匹配消息的方式已足够健壮。
- 影响：影响范围为 CI 稳定性。修复后，即使 sglang 升级，相关测试也不会因异常类型变化而误报失败，确保测试覆盖有效。对用户无直接影响，主要受益者是 CI 流程和后续 PR 的验证效率。
- 风险标记：依赖版本兼容性

关联脉络

- 暂无明显关联 PR