

PR #47619 完整报告

vllm-project/vllm

[Rust Frontend] Add DeepSeek V3.2 roundtrip fixture

合并时间: 2026-07-07 10:47

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/47619>

执行摘要

本 PR 在 Rust chat 前端的 roundtrip 测试中, 新增 DeepSeek V3.2 测试用例, 仅覆盖 tool-call mix 路径, 以验证 DSML tool-call 的渲染、流式解析和重新渲染。仅修改一个测试文件, 无生产代码变更。

功能与动机

为 DeepSeek V3.2 模型提供 Rust 前端的 roundtrip 测试覆盖, 确保 tool-call 路径可靠工作。

实现拆解

1. 新增 `RoundtripCase::deepseek_v32` 方法: 在 `rust/src/chat/tests/roundtrip.rs` 中, 添加配置方法, 模型 ID 为 `deepseek-ai/DeepSeek-V3.2-Exp`, 解析器为 `Auto`, 思考行为为 `Toggleable` (默认关闭), 与现有 DeepSeek V4 配置一致。
2. 注册测试宏: 在 `roundtrip_tests!` 中增加 `deepseek_v32 => [tool_call_mix]`, 仅启用 `tool_call_mix` fixture, 不启用 `reasoning_and_content` fixture。

rust/src/chat/tests/roundtrip.rs

新增 DeepSeek V3.2 测试用例定义和注册, 是本次唯一变更的文件。

`rust/src/chat/tests/roundtrip.rs`

```
// 新增 RoundtripCase 方法, 配置 DeepSeek V3.2 模型和解析器
/// DeepSeek V3.2 DSML tool-call format.
fn deepseek_v32() -> Self {
    Self {
        model_id: "deepseek-ai/DeepSeek-V3.2-Exp",
        assistant_stop_suffix: "< | end__of__sentence | >",
        tool_call_parser: ParserSelection::Auto,
        reasoning_parser: ParserSelection::Auto,
        thinking_behavior: ThinkingBehavior::Toggleable { default: false },
        json_fmt: compact_json_fmt(),
        sort_json_keys: false,
    }
}

// 在 roundtrip_tests! 宏中注册, 仅启用 tool_call_mix fixture
roundtrip_tests! {
```

```
// ... 其他用例 ...  
deepseek_v32 => [tool_call_mix], // 仅 tool_call_mix, 原因见讨论  
// ...  
}
```

评论区精华

BugenZhao: "Have you tried the `reasoning_and_content` case? Is there any blocker preventing us from enabling it?" reidliu41: "I checked this. ... For DeepSeek V3.2, those two paths create a text-only final assistant message. The dedicated renderer rejects that shape ..." BugenZhao: "Thanks for explanation. Let's focus on tool-call roundtrip then, as it also covers reasoning content."

风险与影响

风险：极低，仅测试变更。若模型或渲染器未来变化，此测试可能需要调整。影响：仅影响 Rust chat 前端的 roundtrip 测试集，增加 DeepSeek V3.2 的 tool-call 路径覆盖。

关联脉络

- 与早期 PR #47780 (DeepSeek V4 bugfix) 同属 DeepSeek 系列模型支持。
- 是 Rust 前端扩展模型覆盖的持续工作的一部分。