

# PR #44588 完整报告

vllm-project/vllm

[Reasoning][Structured Outputs] Add Command A plus tags for structural tags

合并时间: 2026-06-05 21:51

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/44588>

## 执行摘要

- 一句话: Cohere2Moe 推理时增加 Command A+ 结构标签支持
- 推荐动作: 该 PR 改动简单直接, 逻辑清晰, 无需精读。但可以作为 Cohere 系列模型结构标签配置的参考示例。

## 功能与动机

Cohere2MoeForCausalLM 架构 (Command A+ 模型) 在推理时需要同时识别两种结构标签: `COMMAND_A_JSON_TAG` 和 `COMMAND_A_PLUS_JSON_TAG`。此前配置仅包含前者, 导致带 `<ISTART_TEXTI>` 标签的响应无法被正确解析, 影响结构化输出功能的正确性。

## 实现拆解

在 `vllm/reasoning/cohere_command_reasoning_parser.py` 中, 修改 `MODEL_TO_TAG_STYLE` 字典中 `Cohere2MoeForCausalLM` 键对应的 `json_tags` 值, 将单元元素元组 (`COMMAND_A_JSON_TAG,`) 扩展为双元素元组 (`COMMAND_A_JSON_TAG,` `COMMAND_A_PLUS_JSON_TAG`), 使模型支持识别两种结构标签。改动仅涉及一行代码。

关键文件:

- `vllm/reasoning/cohere_command_reasoning_parser.py` (模块 推理解析; 类别 `source`; 类型 `core-logic`): 唯一变更文件, 修正 `Cohere2MoeForCausalLM` 的 `json_tags` 配置, 新增 `COMMAND_A_PLUS_JSON_TAG` 支持。

关键符号: 未识别

## 关键源码片段

`vllm/reasoning/cohere_command_reasoning_parser.py`

唯一变更文件, 修正 `Cohere2MoeForCausalLM` 的 `json_tags` 配置, 新增 `COMMAND_A_PLUS_JSON_TAG` 支持。

```
# 文件: vllm/reasoning/cohere_command_reasoning_parser.py
```

```
# 以下为关键部分, 展示模型到标签风格的映射配置
```

```
COMMAND_A_PLUS_JSON_TAG = CohereTagRegistry(  
    trigger="<ISTART_TEXTI>",  
    end="<IEND_TEXTI>",
```

```
)

MODEL_TO_TAG_STYLE: dict[str, CohereTagStyle] = {
    "Cohere2ForCausalLM": CohereTagStyle(
        json_tags=(COMMAND_A_JSON_TAG,), # 仅响应标签
        tools=COMMAND_A_TOOLS_TAG,
    ),
    "Cohere2VisionForConditionalGeneration": CohereTagStyle(
        json_tags=(COMMAND_A_JSON_TAG, COMMAND_A_PLUS_JSON_TAG), # 已有双标签
        tools=COMMAND_A_TOOLS_TAG,
    ),
    # 以下为本次 PR 的变更：为 Cohere2MoeForCausalLM 增加 COMMAND_A_PLUS_JSON_TAG
    "Cohere2MoeForCausalLM": CohereTagStyle(
        json_tags=(COMMAND_A_JSON_TAG, COMMAND_A_PLUS_JSON_TAG), # 新增 <ISTART_
        TEXTI> 标签
        tools=COMMAND_A_TOOLS_TAG,
    ),
}
```

## 评论区精华

无 review 讨论。

- 暂无高价值评论线程

## 风险与影响

- 风险：风险极低。仅修改单个字典条目，属于配置性调整。因为 `COMMAND_A_PLUS_JSON_TAG` 已在同一文件中预先定义并被其他模型（如 `Cohere2VisionForConditionalGeneration`）使用，不会引入未定义符号或依赖问题。该变更不会影响其他模型（如 `Cohere2ForCausalLM` 和 `Cohere2VisionForConditionalGeneration`）的行为。
- 影响：影响范围小，仅针对 `Cohere2MoeForCausalLM` 架构模型（如 `Command A+`）。修复后，该类模型在进行结构化输出解析时能正确识别 `<ISTART_TEXTI>` 和 `<IEND_TEXTI>` 标签，提升推理结果的完整性和正确性。对用户而言，这是无感的向后兼容修复。
- 风险标记：低风险

## 关联脉络

- 暂无明显关联 PR