

PR #36602 完整报告

sgl-project/sglang

[CI] Remove GLM-4.1V-9B-Thinking from nightly VLM MMMU eval

合并时间: 2026-08-27 11:29

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/36602>

执行摘要

- 一句话: 将 GLM-4.1V-9B-Thinking 移出 nightly MMMU 评测
- 推荐动作: 该 PR 为简单的测试配置调整, 不值得精读, 用于了解 CI 评测的阈值管理方式。

功能与动机

PR 描述指出 GLM-4.1V-9B-Thinking 在 nightly VLM MMMU 评测中因延迟 37.31s 超过 35.5s 阈值而失败, 准确率虽达标但仍需移除, 以避免持续告警。

实现拆解

该 PR 只修改了测试文件 `test/registered/eval/test_vlms_mmmu_eval.py`, 删除了 `MODEL_THRESHOLDS` 字典中关于 `zai-org/GLM-4.1V-9B-Thinking` 的条目。具体为删除一行代码, 无其他改动, 没有测试或配置配套变更。

关键文件:

- `test/registered/eval/test_vlms_mmmu_eval.py` (模块 评测; 类别 test; 类型 test-coverage) : 移除 GLM-4.1V-9B-Thinking 的延迟阈值条目, 影响 nightly MMMU 评测的模型列表。

关键符号: 未识别

关键源码片段

`test/registered/eval/test_vlms_mmmu_eval.py`

移除 GLM-4.1V-9B-Thinking 的延迟阈值条目, 影响 nightly MMMU 评测的模型列表。

```
# MODEL_THRESHOLDS 字典的部分内容, 删除了一行模型配置
MODEL_THRESHOLDS = {
    ...
    ModelLaunchSettings("unsloth/Mistral-Small-3.1-24B-Instruct-2503"): (0.30, 43.0),
    ModelLaunchSettings("XiaomiMiMo/MiMo-VL-7B-RL"): (0.28, 40.0),
    # 删除行: ModelLaunchSettings("zai-org/GLM-4.1V-9B-Thinking"): (0.280, 35.5),
    ModelLaunchSettings("zai-org/GLM-4.5V-FP8", extra_args=["--tp=2"]): (0.26, 140.0),
}
```

评论区精华

Review 中 mmangkad 直接批准，无任何评论或讨论。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低，仅影响 nightly MMMU 评测的模型覆盖范围，移除后该模型不再被评测，可能掩盖未来性能退化。但不会影响其他模型的评测或生产代码。
- 影响：对用户无直接影响，对 CI 系统：nightly MMMU 评测不再包含该模型，减少一个失败点。对团队：需要关注该模型后续性能，必要时重新引入。
- 风险标记：评测覆盖缩减

关联脉络

- 暂无明显关联 PR