PR #21241 完整报告

sgl-project/sglang

[bugfix] Fix rope theta config for MiniMax after transformers v5 update

合并时间：2026-04-01 02:37

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/21241

执行摘要

本 PR 修复了 MiniMax 模型在 transformers v5 更新后的 RoPE 配置不兼容问题，通过统一使用 get_rope_config 函数确保参数正确设置，变更影响单一文件，风险低，已获批准合并。

功能与动机

由于 transformers v5 更新，MiniMax 模型的外部配置读取方式出现不兼容，导致 rope_theta 参数错误。PR body 中明确说明 "Fixes for Transformers v5 update"，且代码注释指出 "minimax_m2 config use external config that not compatible with transformers v5"，因此需修复以保持库兼容性。

实现拆解

修改仅涉及 python/sglang/srt/models/minimax_m2.py 文件：

- 在文件头部导入 get_rope_config 函数：from sglang.srt.utils.hf_transformers_utils import get_rope_config
- 在 __init__ 方法中，将 self.rope_theta = config.rope_theta 替换为 self.rope_theta, self.rope_scaling = get_rope_config(config)
- 更新 rope_scaling 引用：从局部变量 rope_scaling 改为 self.rope_scaling，确保一致性。

评论区精华

无实质性讨论，reviewer Fridge003 直接批准，表明变更简洁且被团队接受。

风险与影响

- 风险：依赖 get_rope_config 函数的正确性，若该函数实现有误，可能导致 RoPE 配置错误；缺少本 PR 中的单元测试，需依赖现有测试套件验证。
- 影响：仅影响 MiniMax 模型，修复后避免因 transformers v5 更新引发的运行时错误，提升系统稳定性。

关联脉络

本 PR 与历史 PR #20931 相关，后者可能处理类似配置兼容性问题。结合近期 PR 如 #21752（修复 kimi-linear 配置错误），显示项目在持续调整模型配置以应对库更新，整体趋势是增强与 transformers 等外部库的兼容性。