

PR #27824 完整报告

sgl-project/sglang

docs: Diffusion Gemma cookbook

合并时间: 2026-06-11 00:33

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/27824>

执行摘要

- 一句话: 新增 DiffusionGemma 部署 cookbook 文档
- 推荐动作: 值得阅读, 可作为新模型接入的文档模板参考。建议关注 #27823 合入后实际运行时与文档描述的一致性。

功能与动机

Split out of #27823 (DiffusionGemma model support). This PR carries only the Mintlify cookbook documentation for google/diffusion-gemma-26B-A4B-it, so the docs can be reviewed independently of the runtime code.

实现拆解

1. 新建 cookbook 页面: 创建 docs_new/cookbook/autoregressive/Google/DiffusionGemma.mdx, 包含模型介绍、架构参数表、启动命令、精度验证表等内容 (+306 行)。
2. 注册导航: 在 docs_new/docs.json 的 Google 分组中添加 'cookbook/autoregressive/Google/DiffusionGemma' 页面引用 (+2/-1)。
3. 补充模型支持列表: 在 docs_new/docs/supported-models/diffusion_language_models.mdx 中添加一行记录 DiffusionGemma 的架构特点 (+5 行)。

关键文件:

- docs_new/cookbook/autoregressive/Google/DiffusionGemma.mdx (模块 部署指南; 类别 other; 类型 documentation): 核心文档页, 包含模型部署的全部指导内容
- docs_new/docs.json (模块 导航配置; 类别 config; 类型 configuration): 在文档导航中注册 DiffusionGemma 页面
- docs_new/docs/supported-models/diffusion_language_models.mdx (模块 模型支持表; 类别 other; 类型 data-contract): 在支持模型表中添加 DiffusionGemma 条目

关键符号: 未识别

评论区精华

无公开讨论; PR 获得 1 位 reviewer 批准后直接合入。

- 暂无高价值评论线程

风险与影响

- 风险：纯文档变更，无代码执行风险。文档内容基于已合入的 Gemma4 基础架构和 Gemma4Renoise sampler 的实现，不存在技术信息不准确的风险，但建议后续与运行时代码同步更新以保持一致性。
- 影响：对用户：提供了 DiffusionGemma 模型的部署指导，降低上手成本。对系统：无影响。对团队：新增了需要维护的文档页面。
- 风险标记：暂无

关联脉络

- PR #27823 DiffusionGemma model support: 此 PR 从 #27823 拆分出来，仅包含文档部分；两者共同实现 DiffusionGemma 完整支持。