PR #27171 完整报告

sgl-project/sglang

[Docs] Update unified Text/Vision/Audio model cookbook: install + sgl-eval accuracy

合并时间：2026-06-04 00:32

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/27171

执行摘要

针对 #27167 新增的编码器自由统一 Text/Vision/Audio 模型支持，更新 cookbook 文档：安装部分指向匹配的 transformers commit 并移除已过时的 Docker 参考；新增经 sgl-eval 测量的 MMLU 和 GSM8K 精度数据，解决严格提取导致的分数低估问题。

功能与动机

作为 #27167 的后续，在模型支持落地 main 后更新 cookbook 页面，提供准确的安装指引和可靠的精度数据，因为严格提取会低估推理型模型的真实性能。

实现拆解

1. 安装章节更新：将 transformers commit 从 91b1ab1f... 更新为包含 encoder-free unified 家族的 1423d22f...，移除 pre-merge 的 PR#21952 引用和 Docker 标签，简化安装命令。
2. 精度章节新增：展示 sgl-eval 在 2000 例 MMLU 上获得的 0.878 和 1319 例 GSM8K 上获得的 0.960 结果，并附重现命令；解释严格提取低估原因。
3. 无代码变更。

无可用源码片段。

评论区精华

无 review 讨论。

风险与影响

- 风险：无。仅文档变更，不影响代码逻辑。
- 影响：为用户提供准确安装步骤和可靠精度数据，便于部署评估。

关联脉络

与 #27167 直接关联，该 PR 实现了 Gemma 4 Unified 模型支持，本 PR 补充其文档。