

PR #23905 完整报告

sgl-project/sglang

[docs] fix sglang serve --model-path in cookbooks

合并时间: 2026-04-28 12:54

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/23905>

执行摘要

该 PR 修复了多个模型部署 cookbook 中 `sglang serve` 命令参数 `--model` 未更新为 `--model-path` 的问题，共涉及 5 个文件、9 处修改。变更简单、风险极低，仅影响文档中的命令字符串，确保用户复制的命令可直接使用。

功能与动机

`sglang serve` 的 CLI 参数已从 `--model` 变更为 `--model-path`，但 cookbook 文档中的命令示例和动态生成脚本仍使用旧的参数名，导致用户直接复制命令执行时报错。本 PR 旨在同步文档与最新 CLI 接口。

实现拆解

1. 修改 JSX 动态命令生成:

- `deepseek-v32-deployment.jsx`: 在 `generateCommand` 函数的 `NVFP4`、`MXFP4` 和默认分支中，将 `--model` 拼串改为 `--model-path`。
- `glm-5-deployment.jsx`: 修改主命令中的 `--model` 为 `--model-path`。
- `ministral-3-deployment.jsx`: 修改主命令中的 `--model` 为 `--model-path`。

2. 修改 MDX 文档静态命令示例:

- `DeepSeek-V3_2.mdx`: 更新两个 shell 命令块中的 `--model` 为 `--model-path`。
- `GLM-5.mdx`: 更新两个 shell 命令块中的 `--model` 为 `--model-path`。

所有变更均只修改参数名，不涉及其他参数或逻辑调整。

[docs_new/src/snippets/autoregressive/deepseek-v32-deployment.jsx](#)

核心 JSX 文件，动态生成 DeepSeek-V3.2 部署命令，包含 3 处 `--model` 修改。

关键源码片段

[docs_new/src/snippets/autoregressive/deepseek-v32-deployment.jsx](#)

核心 JSX 文件，动态生成 DeepSeek-V3.2 部署命令，包含 3 处 `--model` 修改。

```
// 文件 : docs_new/src/snippets/autoregressive/deepseek-v32-deployment.jsx
// 在 generateCommand 函数中，将所有 `--model` 参数改为 `--model-path`
```

```
// NVFP4: fixed config
if (isNvfp4) {
```

```
let cmd = 'sglang serve \\n';
cmd += ` --model-path ${modelName}`; // 原为 --model
cmd += ' \
--tp 4';
cmd += ' \
--quantization modelopt_fp4';
cmd += ' \
--moe-runner-backend flashinfer_trtllm';
return cmd;
}

// MXFP4: fixed config for AMD
if (isMx4) {
  let cmd = 'sglang serve \\n';
  cmd += ` --model-path ${modelName}`; // 原为 --model
  cmd += ' \
--tp 8';
  cmd += ' \
--trust-remote-code';
  return cmd;
}

// 通用默认分支
let cmd = 'sglang serve \\n';
cmd += ` --model-path ${modelName}`; // 原为 --model
```

评论区精华

无 review 讨论。

风险与影响

风险极低。直接影响用户从 cookbook 复制命令的使用体验，不涉及任何运行时逻辑、API 或性能。

关联脉络

该 PR 与 #23883 (DeepSeek-V4 cookbook 更新) 属于同一类文档维护任务，均涉及部署命令的准确性。建议在类似文档改动中检查参数一致性。