

PR #28693 完整报告

sgl-project/sglang

docs: add --trust-remote-code to Laguna-M.1 / XS.2 cookbook configs

合并时间: 2026-06-19 10:21

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/28693>

PR 分析报告: Poolside Laguna 模型添加

--trust-remote-code 参数

1. 执行摘要

本次 PR 为 Poolside 家族的 Laguna-M.1 和 Laguna-XS.2 模型的 cookbook 部署命令补全了 `--trust-remote-code` 参数。这两个模型在 Hugging Face Hub 上使用自定义配置代码，缺少该标志会导致服务启动失败。变更覆盖了 JSX 配置文件和 Markdown 文档，全部为纯文档性改动，不涉及任何运行时逻辑。

2. 功能与动机

- 动机: Laguna-M.1 和 Laguna-XS.2 在 Hugging Face Hub 上发布时附带了自定义的 modeling/config 代码 (transformers 原生的 laguna config 不兼容)，而 SGLang 的 transformers 集成默认要求信任远程代码。原先的 cookbook 命令中缺少 `--trust-remote-code`，用户直接复制命令运行时会导致模型加载失败，产生糟糕的用户体验。
- 关联 Issue: 无，但 PR body 明确指出了问题及修复方向。

3. 实现拆解

1. 更新 JSX 配置文件注释: 在 docs_new/src/snippets/configs/poolside/laguna-m1.jsx 中，将原先的注释 'Model is now natively supported → NO --trust-remote-code needed' (该声明不准确，原生支持并不代表自定义配置可被自动信任) 改为 '--trust-remote-code is required'，并增加了解释: Hub 上自定义的 config 与 transformers 原生 laguna config 不兼容。
2. 为所有 deployment cells 添加参数: 在该 JSX 文件中，每个 deployment cell (共 11 个) 的 flags 数组中都插入了 `--trust-remote-code` 参数，覆盖所有硬件 (H200、B200、B300、GB200、GB300) 和量化 (BF16、FP8、NVFP4) 组合。
3. 修改 Laguna-XS.2 的部署组件: 在 docs_new/src/snippets/autoregressive/laguna-xs2-deployment.jsx 中，调整了 generateCommand 函数初始命令行的构建方式: 将 `--tp {tp}` 后添加换行符和跳行，然后新增 '`--trust-remote-code`' 行，使其成为所有生成命令的固定组成部分。
4. 更新 Cookbook Markdown 文档: 分别修改两个 .mdx 文件: 移除原先声称 'no --trust-remote-code needed' 的句子，改称 'the model ships custom config code on the Hub, so --trust-remote-code is required'; 同时在 'Configuration Tips' 小节各自增

加一条关于 'Trust remote code' 的配置提示。

docs_new/src/snippets/configs/poolside/laguna-m1.jsx

核心变更文件，为 Laguna-M.1 的所有 11 个 deployment cells 添加了 `--trust-remote-code` 参数，并更新了相关注释，确保用户复制命令时不会遗漏。

```
// Laguna-M.1 (poolside) — config-driven cookbook page.
// ... (build requirements, hardware details omitted for brevity)
//
// --trust-remote-code is required: M.1 ships custom config code on the Hub (the transformers-
// native
// `laguna` config is incompatible). Carried on every cell.
//
// ...

export const config = {
  // ...
  cells: [
    // Each cell now includes "--trust-remote-code" in its flags array
    {
      match: { hw: "h200", variant: "default", quant: "bf16", strategy: "balanced", nodes: "single" },
      verified: true,
      env: [],
      flags: [
        "--model-path {{MODEL_NAME}}",
        "--trust-remote-code", // ADDED: required for custom config on Hub
        "--reasoning-parser poolside_v1",
        "--tool-call-parser poolside_v1",
        "--tp 8",
        "--host {{HOST_IP}}",
        "--port {{PORT}}",
      ],
    },
    // ... (all other 10 cells follow the same pattern)
  ],
};
```

docs_new/src/snippets/autoregressive/laguna-xs2-deployment.jsx

为 Laguna-XS.2 的启动命令模板添加了 `--trust-remote-code` 参数，确保所有生成的命令都包含该参数。

```
// Inside the component
const generateCommand = () => {
  const { hardware, quantization, reasoning, toolcall, dpAttention } = values;
  // Error checks omitted...

  const tp = 8;
  const lines = [
    'sglang serve \\'
```

```

    ` --model-path ${modelId} \`,
    ` --tp ${tp} \`,
    ' --trust-remote-code' // ADDED: required for custom config on Hub
];

if (dpAttention === 'enabled') {
    lines[lines.length - 1] += ' \';
    lines.push(` --dp ${tp} \`);
    lines.push(' --enable-dp-attention');
}
// remaining flag additions...
return lines.join('\n');
};

```

5. 评论区精华

无 Review 评论。PR 由维护者 JustinTong0323 直接批准合并，未产生讨论。

6. 风险与影响

- 风险：风险极低，仅为文档修复，不涉及运行时代码变更。唯一潜在风险是如果未来模型不再需要此参数（如 SGLang 更改默认行为），需再次清理文档。但概率很低。
- 影响：直接影响使用 cookbook 部署这两个模型的用户，原先复制即失败的命令现在可以正常工作。对现有生产环境无任何影响。

7. 关联脉络

- 关联 PR #28661（新增 Laguna-M.1 cookbook）：本 PR 修复了该 cookbook 中遗漏的 `--trust-remote-code`，属于后续修补。
- 关联 PR #28400（Laguna-M.1 原生支持）：使模型可在 SGLang 上运行，但需要信任自定义配置。
- 暂无进一步扩展计划，属于一次性的文档修正。