

# PR #21756 完整报告

sgl-project/sglang

[diffusion] fix: wire --prompt-path into \_resolve\_prompts

合并时间: 2026-04-01 16:47

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/21756>

## 执行摘要

本次 PR 修复了 diffusion 生成器中 `--prompt-path` 参数被解析但从未消费的 bug，通过将 `prompt_path` 从 `sampling_params_kwargs` 传递到 `_resolve_prompts` 方法并设置优先级，确保用户能正确使用该参数指定提示文件。影响范围小，风险低，测试覆盖充分，是重要的功能修复。

## 功能与动机

问题背景: 在 sglang 的 diffusion 生成器中，`SamplingParams` 的 CLI 参数 `--prompt-path` 被解析但从未在运行时使用，导致用户无法通过此参数指定提示文件。仅 `--prompt-file-path` (`ServerArgs`) 在 `_resolve_prompts()` 中被检查，造成功能缺失。

解决目标: 修复此 bug，使 `prompt_path` 参数正确消费，优先级高于 `server_args.prompt_file_path`，并保持向后兼容性。PR body 中明确指出: "`--prompt-path` (`SamplingParams` CLI arg) was parsed but never consumed by the runtime"。

## 实现拆解

- 核心代码修改: `diffusion_generator.py` 中的 `generate` 方法新增 `prompt_path` 参数传递: `prompts = self._resolve_prompts( sampling_params_kwargs.get("prompt"), sampling_params_kwargs.get("prompt_path"), )` 同时修复 null 安全问题，初始化 `sampling_params_kwargs` 为空字典。
- `_resolve_prompts` 方法更新: 新增 `prompt_path` 参数，优先级逻辑为 `path = prompt_path or self.server_args.prompt_file_path`，并添加文件检查和错误处理:
  - 文件不存在时抛出 `FileNotFoundError`。
  - 从文件读取多行提示（一行一个）。
- 测试增强: 新增 `test_resolve_prompts.py` 单元测试，覆盖场景包括:
  - 内联提示（字符串、列表）。
  - `prompt_path` 单行和多行文件。
  - `prompt_path` 优先于 `server_args.prompt_file_path`。
  - 错误案例（文件缺失、空文件）。在 `run_suite.py` 中集成测试到测试套件。

## 评论区精华

review 讨论中, gemini-code-assist[bot] 提出了两个关键改进:

"The `generate` method signature explicitly allows `sampling_params_kwargs` to be `None`, but the implementation immediately calls `.get()` on it, which will raise an `AttributeError`. Initializing it to an empty dictionary at the start of the method ensures robustness."

"The help text was changed to singular ('the prompt'), but the implementation in `_resolve_prompts` still supports multiple prompts (one per line)... The original help text was more accurate."

这些建议在提交中被采纳, 避免了潜在崩溃并确保了文档准确性。

## 风险与影响

- 技术风险:
  - `sampling_params_kwargs` 为 `None` 的风险已通过初始化修复。
  - 优先级引入可能影响旧版用户, 但测试验证了 `--prompt-file-path` 仍正常工作, 兼容性良好。
  - 文件路径处理中的错误已通过异常处理覆盖。
- 影响范围: 仅限 `diffusion` 生成器的提示解析逻辑, 用户现在能灵活使用 `--prompt-path` 参数。对系统性能、安全无影响, 局部修复提升了功能完整性。

## 关联脉络

与历史 PR 的关联显示 `diffusion` 模块的持续演进:

- PR #21763 (CI 改善) 与本 PR 都聚焦 `diffusion` 测试, 强调了 CI 稳定性和测试覆盖的重要性。
- PR #20379 (添加日志前缀排除) 展示了 `diffusion` 服务器参数的扩展, 本 PR 则修复了参数消费的 bug, 共同体现了参数处理功能的完善。这些 PR 表明团队在强化 `diffusion` 模块的可靠性和用户体验, 本次修复是这一趋势的具体体现。