

Slime 仓库周报：2026 第 33 周（08-10 至 08-16）

THUDM/slime

周期：2026-08-10 至 2026-08-16

来源 PR：29 • 重点 PR：24 • 自动生成

原文链接：<http://prhub.com.cn/THUDM/slime/reports/2026-08-10-to-2026-08-16>

执行摘要

本周（2026-08-10 至 08-16）仓库共有 29 个 PR，其中 24 个为重点分析对象。标签分布显示 bugfix 21 个、test 16 个、backends 14 个，本周主旋律是“围绕 Megatron 后端正确性，并同步补测试”。平均重要性 6.01、平均洞察度 4.52，高信号 PR 占比不低。作者侧 keepkeen 一人贡献 9 个 PR，覆盖 rollout、数据管线、配置校验与 agent sandbox，是本周最主要的修复力量。

从整体变化看，本周没有明显的新功能扩张，更多是“还旧账”：修复了 CP 下 advantage whitening 的进程组误用、DP>1 下 log-correct-samples 的崩溃、eval 配置静默丢键、fully-async rollout 丢弃已完成组等一批影响训练结果或用户体验的缺陷。性能优化也有亮点，REINFORCE++ 回报计算向量化后最高获得 357 倍 CPU 端加速。风险上，“缺少测试覆盖”出现 6 次，加上多处核心训练路径变更，需要警惕修复本身引入新回归。

本周重点变化

第一，训练正确性修复密集落地。#2235 修正了 context parallelism 下 advantage whitening 使用排除 CP 的 DP 组做 all-reduce 的统计错误，修复前后序列两半的 whitened advantage 从 +1.707/+1.274 收敛到一致；#2234 为 --log-correct-samples 增加 DP 本地奖励视图，解决 DP>1 时的 IndexError 与指标错配；#2243 处理了 load_other_checkpoint 未还原 args.ckpt_step 导致的错误迭代加载；#2261 修复 partial rollout 续跑时未扣除已生成 token 数，避免请求超过上下文限制被 SGLang 拒绝。

第二，配置与参数验证成为另一个高频主题。#2242 修复 eval.defaults 只透传 14 个字段导致 rm_type、repetition_penalty 等被静默丢弃，以及 per-dataset 的 stop/min_new_tokens 成为死配置的问题；#2236 修复 --start-rollout-id 被无条件重置为 0；#2247 补上 eps_clip_c 从参数到损失函数的透传，使 dual-clip PPO 真正生效。这类问题都表现为“配置写了但没起作用”，根因是参数在多层函数调用之间丢失，建议后续引入更严格的接线校验。

第三，数据管线修复与测试同步增强。#2237 修复 filter_long_prompt 在混合多模态数据下按组拼接导致的顺序重排；#2241 恢复负索引切片（如 path@[-100:]）的能力；#2213 修复 tau-bench 在 Qwen3 推理模板下 token 增量计算截断用户输入与遗漏生成前缀。这些 PR 都新增了 CPU 单元测试，并把部分测试接入 .github/workflows/pr-test.yml（该文件本周被 7 个 PR 改动）。

模块与主题趋势

按涉及文件看，热点集中在四个区域：

- Megatron 后端: loss.py 涉及 advantage whitening、dual-clip 透传、debug train data 捕获; actor.py 涉及 checkpoint 加载与 debug dump; arguments.py 与 slime/utils/arguments.py 承载参数校验; 模型转换相关改动分布在 megatron_to_hf 与 tools/convert_hf_to_fp8.py, 修复 FP8 量化与最新 Megatron 的兼容性。
- 数据与配置: slime/utils/data.py 被 3 个 PR 修改, 覆盖 process_rollout_data 的本地视图、filter_long_prompt 顺序恢复、read_file 负切片; slime/utils/eval_config.py 是 eval 配置解析的核心修复点。
- Rollout: sglang_rollout.py 与 fully_async_rollout.py 分别修复 partial 续跑预算和异步队列丢组问题。
- 基础设施与兼容性: pr-test.yml 高频改动体现 CI 扩展; docker/npu_patch/slime.patch 和 int4_qat 内核则是面向 NPU 与 ROCm 的适配。

主题趋势上, 仓库明显在为大规模训练做准备: GLM-5 对齐路径 (#2262) 把训练与推理的数值差异压到 1e-7 量级; debug train data 重构为与 rollout dump 对齐的单文件格式并支持 CP 恢复 (#2266) 。这些不是简单修 bug, 而是为可复现训练、问题定位和生产调优打基础。同时, 大量 CP/DP 维度相关修复表明, 并行维度的组合 (尤其是 context parallelism + data parallelism) 是当前正确性薄弱点, 需要更多针对性测试。

风险观察

本周最突出的风险是“核心训练路径变更 + 测试覆盖不足”的组合。#2262、#2266、#2235、#2234 都直接改动损失计算、数据打包或 checkpoint 加载逻辑, 而 top_risks 中“缺少测试覆盖”出现 6 次, 意味着部分修复没有回归测试支撑。例如 #2243 和 #2261 均被明确标注缺少测试, 建议安排跟踪补测。

第二个风险是配置参数静默失效的反复出现。从 eval.defaults 到 eps_clip_c, 再到 --start-rollout-id, 本周至少 3 个 PR 在修同一类问题。这提示仓库缺少“参数声明—消费点”的一致性校验机制, 未来新增配置项时仍可能重蹈覆辙。

第三个风险是数据格式契约的不兼容变更。#2234 在 rollout_data 中新增 local_raw_reward 字段, #2266 重写了 debug train data 的文件布局, 这些都会影响下游日志解析、指标聚合和外部工具。若有外部使用者, 需要确认同步更新。

第四个风险是外部依赖兼容的长期维护。Megatron 新版本 API 变化 (#2267)、ROCm 的 hipcc 差异 (#2274) 以及 NPU 补丁 (#2199) 都属于“跟着上游走”的修法, 缺少统一兼容层, 上游再次变动时相似问题可能反复。

重点 PR 速览

PR	标题	要点
#2262	feat(glm5): align Megatron DeepEP training with SGLang rollout	新增约 1.2 万行确定性对齐路径, 覆盖 DeepGEMM/DeepEP/稀疏注意力 /FP8 KV cache, 训练与推理 logprob MAE 1.89e-7

PR	标题	要点
#2266	Refactor --save-debug-train-data	重写调试数据转储为单文件格式，支持 CP 恢复，并新增训练期 log-prob 捕获钩子
#2213	Fix tau-bench token deltas for reasoning templates	修复多轮对话 token 增量计算在历史重写和生成前缀上的错误
#2242	fix: honor every eval.defaults key...	修复 eval.defaults 丢键与 per-dataset stop/min_new_tokens 失效
#2238	fix: stop the fully-async rollout dropping completed groups	修复异步 rollout 排空队列导致丢组，并把有界队列改为无界 + qsize 门控
#2235	fix: whiten advantages over the DP group that includes context parallel	修复 CP 下 advantage whitening 用错进程组
#2234	fix: pair --log-correct-samples rewards with DP-local samples	修复 DP>1 下 log-correct-samples 崩溃与指标错配
#2205	perf: vectorize REINFORCE++ discounted returns	向量化折扣回报计算，长响应大 batch 下 CPU 加速 357x
#2237	fix: keep dataset order in filter_long_prompt...	修复混合多模态数据顺序重排
#2241	fix: restore negative dataset slice bounds	修复数据集路径负索引切片崩溃
#2267	Fix model convert when use latest megatron	兼容最新 Megatron 的 tokenizer 导入、Gloo 进程组等 API 变化

此外，#2274 让 INT4 QAT 内核可在 ROCm 下构建，#2199 为 NPU 补丁增加 IPv6 地址包裹，#2132 修复多 head MTP 训练日志崩溃，#2247 打通 dual-clip PPO 的 eps_clip_c 透传。文档侧有 5 个 PR，主要澄清 PPO/Critic、OPD 反向 KL 定义等，风险较低。

后续建议

1. 为核心训练路径变更建立回归护栏：针对 CP/DP 组合场景扩充 #2235 中多进程 gloo 测试的覆盖范围；对 #2262 的对齐路径，在 CI 中保留 layerwise 对比任务，防止后续优化破坏数值一致性。
2. 补测无测试覆盖的修复：优先为 #2243 (ckpt_step)、#2261 (partial rollout 预算)、#2246 (gpu_id 排序) 补充单元测试；#2261 明确回归自 1eed2493 重构，应加入针对续跑预算的测试。
3. 构建配置接线的一致性检查：可借鉴 #2242 中反射校验 + setdefault 的做法，将参数校验从运行时发现前移到启动时报错，并为 eval、PPO、rollout 等配置项建立完整字段注册表。
4. 跟踪外部依赖兼容：关注 Megatron 新版本发布节奏，评估是否引入版本适配层；ROCm 内核构建条件编译的正确性建议增加 AMD GPU 冒烟测试。
5. 数据格式变更需通知下游：local_raw_reward、debug dump 新布局等变更应在 changelog 与文档中标注，并检查 log_rollout_data 等消费端是否已正确处理新字段。