

verl 仓库周报 | 2026 年第 37 周 (09-07 至 09-13)

verl-project/verl

周期: 2026-09-07 至 2026-09-13

来源 PR: 18 • 重点 PR: 18 • 自动生成

原文链接: <http://prhub.com.cn/verl-project/verl/reports/2026-09-07-to-2026-09-13>

1. 执行摘要

本周 verl 共合入 18 个 PR，标签分布高度集中在 **npu** (6)、**bugfix** (6)、**test** (5) 和 **trainer** (5)，平均重要性 6.29。变化并非单点修补，而是围绕三条主线推进：一是 Ascend/NPU 适配从功能补丁延伸到镜像、CI 和回归测试；二是 Megatron 前向与 R2 router replay 路径出现修复后快速回滚；三是 Trainer V1 的异步恢复、profiling、DAPO 过滤与 OOM 诊断集中增强。整体看，本周修复密度高、平台覆盖变广，但核心路径测试缺口和临时补丁治理仍是主要不确定性。

2. 本周重点变化

- NPU/Ascend 成为绝对主线：#7827 修正 Ascend 镜像中 Megatron-LM 依赖 tag 并让 NPU 单测默认加载 megatron_adaptor，#7728 修复 Ascend A5 上 IPC 被错误禁用的问题，#7779 为 GLM-5.2 RL 训练引入 vLLM-Ascend 兼容补丁，#7788 新增 32 节点 GRPO 配方，#7808 补充 batch-invariant 回归测试。昇腾侧已从“能跑通”转向“可配置、可回归、可复现”。
- Megatron 前向与路由回放出现波动：#7702 修复 VLM THD 输入未按 TP/CP 对齐导致 SP scatter 崩溃，#7106 先合入 R2 router replay 修复，#7786 次日将其 revert；verl/utils/megatron_utils.py、router_replay_utils.py 各被触及 2 次，说明该路径仍是高风险区。
- Trainer V1 恢复与调度链路集中修复：#7762 修复异步训练 checkpoint 恢复后重复提交 warmup 导致跳样本和窗口溢出，#7722 将 rollout profiler 与训练 profiler 生命周期解耦，#7792 默认使用 rm_scores 做 DAPO 分组过滤。verl/trainer/ppo/v1/trainer_base.py 是本周最热文件。
- 诊断与可观测性增强：#7770 增加 CUDA/NPU OOM 自动快照和多步内存历史，#7722 修复 NPU/GPU 并发 profiler 数据丢失，#7740 让 MLflow run 在训练结束时正确记录 end time。这些改动直接影响大规模训练排障效率。
- 平台抽象与文档补齐 AMD 侧：#7797 将 AMD/ROCm 纳入平台自动检测测试，#7690 在 README 增加 AMD Developer Cloud 的 Fully Async DAPO 教程入口，AMD 生态覆盖继续补位。

3. 模块与主题趋势

主题	代表 PR	趋势判断
NPU/Ascend	#7808、#7827、#7788、#7799、#7779、#7728	从模型适配扩展到镜像、CI、配方和临时补丁，硬件专属代码增多，验证闭环尚未完全建立
Megatron / 模型前向	#7702、#7106、#7786	THD 打包、SP/TP/CP 对齐、router replay 是本周最集中的正确性风险区
Trainer V1	#7762、#7722、#7792、#7770、#7740	重点在恢复、profiling、过滤默认值和诊断能力，核心路径变更较多
Rollout/TOKENIZER	#7643、#7728	Continuous Token 支持 assistant role, A5 IPC 修复，涉及 API 重命名与硬件检测
CI/ 平台 / 文档	#7797、#7827、#7799、#7690	AMD/NPU 平台测试与镜像配置持续补齐，但存在外链失效和教程不同步风险

4. 风险观察

- R2 router replay 快速回滚：#7106 合入后次日被 #7786 revert，输入未给出回滚原因，但同一组文件周内两次变更，建议把该路径列为 Megatron 侧最高优先级跟踪项。
- 核心路径缺少自动化测试：#7702 的 TP/CP 对齐修复只以 Qwen3.6-35B-A3B VLM Geo3K 四节点 E2E 结果佐证，未补单测；#7764 Mooncake 完成槽位修复依赖手动验证；#7762 恢复门控口径变更，后续回归成本较高。
- NPU 验证闭环不足：#7808 作者明确表示 NPU 测试和 pre-commit 未在准备阶段运行，reviewer 空 body 批准；#7827 的 VERL_USE_EXTERNAL_MODULES 在镜像 ENV 中全局无条件生效；#7779 是临时 monkey-patch，缺少上游修复跟踪机制。
- 默认配置与 API 变更影响存量用户：#7792 将 DAPO 过滤默认指标从 null 改为 reward，#7643 将 merge_non_assistant_* 重命名为 merge_context_*，两者都可能改变现有脚本或下游集成行为，需要在发布说明中显式提示。
- 平台检测依赖字符串与内部返回值：#7728 的硬件识别依赖设备名称字符串匹配，并变更 get_npu_versions 返回值口径；#7690 的外部教程链接可能失效或与 trainer.v1 主线不同步。

5. 重点 PR 速览

PR	标题 / 主题	关键观察
#7643	continuous_token 支持 assistant role	本周重要性最高 (9.1)，API 从 merge_non_assistant_* 改为 merge_context_*，子类改为钩子化适配，需关注兼容迁移
#7770	OOM 自动快照与多步内存历史	8.7，新增 OOM observer 与窗口子目录，显著提升难复现 OOM 的诊断能力，但依赖回调容错和设备流状态处理

PR	标题 / 主题	关键观察
#7779	vllm-ascend GLM-5.2 补丁	8.6, 修复 RMSNorm 顺序、kv_b_proj 释放、indexer 缩放和 CaMem 同步, 属临时性 monkey-patch
#7762	V1 异步 warmup 修复	7.8, 改用恢复状态计数而非 reissue 返回值做门控, 修复跳样本与窗口溢出
#7722	profiler 生命周期 解耦	7.8, rollout profiler 在 sample 结束后停止, 覆盖 sync/colocate_async/separate_async 三种模式
#7728	Ascend A5 IPC 默 认开启	7.4, 新增硬件版本枚举并对 A5 短路返回 True, 配套单测较完整
#7808	Ascend batch-inv ariant 测试	7.3, 新增 14 例 NPU 回归, 但未在实机验证即合入, 需确认是否长期 skipped
#7106/ #7786	R2 router replay 修复与回滚	修复与 revert 相邻合入, 涉及 megatron_utils.py 、 router_replay_utils.py , 是本周最需复盘的事件
#7702	VLM THD TP/CP 对齐	6.6, 修复 SP scatter 崩溃, 改动核心前向路径但缺少单测
#7792	DAPO canonical reward 过滤	6.3, 默认指标改为 reward , 简化配置但影响存量默认行为
#7764	Mooncake 完成槽 位 drain	5.8, 修复跨权重版本缓冲区过早释放, 核心同步逻辑变更, 无自动化测试
#7740	MLflow finish 记录	5.9, 训练结束调用 mlflow.end_run() , 修复 UI 状态长期 RUNNING
#7788	GLM-5.2 Ascend GRPO 配方	5.5, 新增 32 节点示例脚本, 含完整硬件检查和并行配置, 但硬编码依赖版本
#7827	Ascend 镜像 Megatron 集成	3.6, 修正 tag 与 MindSpeed 拉取方式, ENV 全局生效缺自动化验证
#7797	AMD 平台自动检 测测试	4.0, 补充 AMD 环境变量覆盖和注册断言
#7799	Ascend nightly CI 清理	3.0, 移除冗余 Megatron Bridge 安装步骤, 降低 CI 冲突
#7690	AMD DAPO Notebook 文档	1.4, README 新增可启动教程, 属于低风险文档变更

6. 后续建议

1. 对 #7106/#7786 做一次快速复盘：明确 revert 原因、受影响模型与并行配置，并决定是重新提交修复还是回退到更保守实现。Megatron THD/router replay 路径建议增加最小 CPU 单测和至少一种真实并行配置的 E2E 门禁。
2. 把核心路径测试补齐作为下周硬性动作：优先覆盖 #7702 的 TP/CP padding 下界、#7762 的 terminal-only 恢复场景、#7764 的跨版本槽位消费。对无法自动化的场景，至少在 PR 模板中记录手动验证环境和命令。
3. 建立临时补丁与 NPU 验证台账：#7779 的 vLLM-Ascend monkey-patch、#7827 的镜像 ENV、#7808 的 NPU 测试都应关联上游 issue 或 CI job，避免合入即遗忘；NPU CI 需要明确哪些用例真实执行、哪些长期 skip。
4. 在发布说明中标注行为变更：#7792 的 DAPO 默认过滤指标、#7643 的 API 重命名、#7827 的 VERL_USE_EXTERNAL_MODULES 全局开启都应以迁移提示形式告知用户。
5. 降低单点依赖：lxb007981 本周贡献 4 个 PR，覆盖 Ascend CI、镜像、配方和 vLLM 补丁。建议安排第二维护者共同 review Ascend 关键路径，减少单点知识风险。