

Verl 项目周报：2026 年第 17 周（04/20 - 04/26）

verl-project/verl

周期：2026-04-20 至 2026-04-26

来源 PR：31 · 重点 PR：24 · 自动生成

原文链接：<http://prhub.com.cn/verl-project/verl/reports/2026-04-20-to-2026-04-26>

执行摘要

本周（04/20-04/26）Verl 项目迎来重大架构调整，两个 Breaking PR 正式废弃了 workers 和 interactions 模块，全面向统一的模型引擎抽象迁移。同时，多教师在线策略蒸馏（OPD）功能落地，VeOmni 和 ROCm/NPU 平台兼容性显著提升。模型支持方面新增 Qwen3.5 MTP 与 VeOmni Qwen3.5 SP。异步训练的多项性能与正确性修复合入，工程质量持续改善，但测试覆盖和部分边缘风险仍需关注。

本周重点变化

1. 架构重构：废弃 workers 与 interactions 模块 - PR #6067 删除了整个 `verl/workers` 目录（FSDP/Megatron Actor/Critic 等），并迁移至 `verl/workers/engine` 下的统一引擎抽象；PR #6074 同时废弃了 `verl/interactions` 模块，将环境主动控制交互的能力交给智能体自主决策。这两项破坏性变更标志着 Verl 从分散的 worker 实现向引擎驱动架构的关键转折，简化了维护路径，但需要所有上层代码适配新接口。
2. 多教师蒸馏（Multi-Teacher OPD）上线 - PR #6051 新增 `MultiTeacherModelManager`，支持基于路由键的样本分发和多教师资源池管理。配置层扩展为字典结构，允许指定多个教师模型及其副本数。该功能在 GSM8K 和 Geo3K 数据集上完成验证，为蒸馏场景提供了更灵活的配置能力。
3. 跨平台兼容性进展 - AMD ROCm (#6062)：修复 ZMQ IPC 句柄不匹配、JSON 序列化错误和套接字残留问题，完全异步训练可在 MI300X 上运行。 - Ascend NPU (#5967, #6102)：通过延迟加载策略处理器修复 MindSpeed 补丁应用顺序问题；升级 Docker 构建流水线支持 CANN 8.5.0 环境变量，并切换 CI 机器至 A3 系列。
4. 模型支持扩展 - 新增 Qwen3.5 MTP（多令牌预测）模型在 Megatron 引擎的 SFT/RL 训练 (#5898)。 - VeOmni 引擎新增 Qwen3.5 系列支持，包括 MoE 变体，并提供了 GRPO 训练示例 (#6061)。
5. 异步训练性能与正确性修复 - 修复 drain loop 因 Condition 阻塞导致样本分发延迟的问题 (#6090)，通过 Event 信号机制允许提前中断。 - 修复参数卸载场景下 Megatron 模型保存 / 恢复的空指针问题 (#6095)。 - 将完全异步训练器中同步 Ray 调用替换为异步版本，消除事件循环阻塞 (#6052)。
6. 工程质量改进 - 工具调用错误信息拆分 (#6055)：从统一 except 改为分阶段细粒度反馈，提升 LLM agent 自校正能力。 - 修复 KL 惩罚函数中带 '+' 后缀估计器的调度 bug (#6058)。 - 修复等长序列嵌套张量布局损坏问题 (#6127)。

模块与主题趋势

- 模型引擎统一化：本周的两大 Breaking PR 是架构长期演进的结果，未来新功能将直接基于 engine 抽象开发，避免 worker 和 engine 双轨维护。训练器（如 ray_trainer.py）已开始适配新引擎，但仍有 _compute_ref_log_prob 等方法中的 worker group 回退逻辑需要完善。
- 异步训练成熟化：完全异步训练（fully_async）是本周最活跃的标签之一（8 个 PR），既有性能优化（drain loop、Ray 调用异步化），也有平台适配（ROCm、NPU）和正确性修复（offload 保存、profiler 初始化）。表明该项目正在加速推进异步训练的生产化。
- 在线蒸馏探索：多教师 OPD 和 VeOmni 蒸馏支持（#6072）的出现，表明蒸馏场景正在从单一教师向多教师、多引擎方向扩展。配置复杂度和节点对齐是主要挑战。
- 平台生态扩展：ROCm 和 NPU 的兼容性修复频繁，表明 Verl 正积极适配非 NVIDIA 硬件，构建更广泛的训练生态。NPU 的 CI 流水线升级（矩阵构建、CANN 版本更新）进一步巩固了持续集成能力。
- 测试覆盖仍需加强：虽然新增了 TRT-LLM 跨节点测试和 KL 惩罚回归测试，但多个高风险修复（如 drain loop、ROCm 兼容）仍缺少独立的单元测试，存在 top_risks 中“缺少测试覆盖”的高频出现。

风险观察

1. 核心路径变更风险：废弃 workers/interactions 模块及 drain loop 重构涉及多处核心逻辑，可能引发旧代码兼容性或新集成路径的回归。建议在合并后加速推进内部依赖迁移，并增加集成测试。
2. 测试覆盖不足的修复：ROCm 兼容性（#6062）、drain loop（#6090）和 offload 保存（#6095）等修复均被标记为“缺少测试覆盖”，一旦触发边缘场景（如多用户环境残留套接字、CPU 存储天然为空），可能导致难以诊断的训练失败。
3. finally 块异常掩盖风险：PR #6055 中 _call_tool 的 finally 块内 tool.release() 可能抛出异常，掩盖原始错误。虽当前 release 方法简单，但未来扩展时需警惕。
4. 硬断言可能生产崩溃：蒸馏 logprob 提取函数（#6120）中的断言假设 SGLang 返回 logprob 数量与 token 数一致，但若不一致则 worker 直接崩溃。建议改为防御性填充。
5. 配置兼容性风险：Qwen3.5 MTP（#5898）中配置转换逻辑可能忽略用户覆盖参数，需关注混合配置场景的正确性。

重点 PR 速览

- #6067 [BREAKING] 废弃 workers 模块：删除 FSDP/Megatron Actor、Critic 等关键类，训练器适配新引擎。是本周最重要的架构变更，影响所有下游。
- #6051 多教师 OPD：引入路由键和资源池管理，配置层重构为字典，支持灵活的蒸馏拓扑。合并前移除了冗余路由进程，提升了资源利用率。
- #6062 ROCm 兼容性修复：通过添加 HSA 环境变量、修复 JSON 序列化（default=str）和 ZMQ IPC 句柄重构，使完全异步训练在 AMD MI300X 上可用。采纳了 IPC 清理的 OSError 捕获建议。
- #5898 Qwen3.5 MTP 支持：新增 MTP 层数查询 / 设置函数和示例脚本，为后续多令牌预测训练铺路。review 建议将辅助函数公开化但未采纳，需注意代码复用。

- #6090 drain loop 性能优化：将 Condition 替换为 Event，允许参数同步中断 drain 循环，避免空闲副本等待。但因锁粒度问题可能带来新的并发风险。
- #6055 工具调用错误信息改进：将异常处理拆分为三个阶段（名称、JSON 解析、执行），为 LLM 提供 actionable 的错误消息。重要 UX 改进。
- #6058 KL 惩罚 bug 修复：修复了长期存在的 `k3+/low_var_kl+` 等带 '+' 后缀估计器无法使用的 bug，新增回归测试。涉及直通梯度技巧的正确性。

后续建议

1. 加速引擎迁移：废弃 workers 后，应尽快更新所有上层训练器、配置和文档，确保 downstream 不再依赖旧 worker API。建议在下一周启动全量适配。
2. 加强测试覆盖：对本周高风险修复（尤其异步训练和平台兼容性）增补单元测试和集成测试，减少生产回归风险。
3. 关注配置安全性：多教师 OPD 和 MTP 的配置逻辑较复杂，应补充配置验证文档，并考虑在 CI 中添加配置兼容性检查。
4. 持续监控异步训练性能：drain loop 重构可能影响并发行为，建议在长稳测试中重点观察吞吐和延迟指标。
5. 推进平台兼容性：ROCm 和 NPU 的修复虽已合入，但建议在相应硬件上补充 nightly CI 测试，防止后续改动引入回归。