verl 2026 第 27 周（06-29 ~ 07-05）周报

verl-project/verl

周期：2026-06-29 至 2026-07-05

来源 PR：26・重点 PR：24・自动生成

原文链接：http://prhub.com.cn/verl-project/verl/reports/2026-06-29-to-2026-07-05

执行摘要

本周（2026 年第 27 周，06-29 至 07-05）verl 项目共合并 26 个 PR，其中 24 个被标记为重点 PR。整体变化集中在三大主轴：一是围绕 Ascend NPU 平台的基础设施与测试体系大幅增强，二是在 Megatron 引擎和蒸馏训练场景中实现了显著的性能优化，三是修复了多个影响稳定性的关键 Bug 并扩展了多模态训练能力。从数据看，NPU 相关 PR 占到了近一半（12 次标签出现），文档和 Bug 修复紧随其后，表明团队正同时推进平台广度和代码健壮性。尽管工作量饱满，部分优化和修复仍缺乏直接单元测试，文档版本一致性也有遗留问题，需在后续迭代中补齐。

本周重点变化

1. NPU 平台全面铺开

- 测试与 CI：PR #6831 新增了 NPU 的单元测试（UT）和系统测试（ST），涵盖 veomni GRPO、engine 兼容性等，并更新了 5 个 CI workflow 文件。PR #6876 添加了基于 RayJob 的多节点 Nightly CI，支持 Qwen3-30B 多节点训练。PR #6877 为 Ascend CI 增加了 Qwen3-0.6B E2E 和 GRPO profiling 用例。
- 文档与示例：PR #6900 重构了 Ascend Quickstart 文档，提供 FSDP2+SGLang、FSDP2+vLLM 等四种训推后端组合脚本。PR #6924 同步更新了三份最佳实践文档的软件版本兼容性。PR #6918 修复了 Quickstart 文档的链接和锚点错误。PR #6890 提供了 Qwen3-32B GSPO 的完整训练脚本。

2. 性能优化

- Megatron BSHD 填充：PR #6901 发现因 micro-batch 填充到各自最大序列长度导致 cuDNN fused-attention 重复编译（约 1s/ 次），通过设置 pad_bshd_to_minibatch_max 标志（默认启用）将同一 mini-batch 内所有 micro-batch 填充到全局最大长度，共享执行计划。该优化在 BSHD 路径下收益显著。
- OPD 蒸馏显存优化：PR #6848 在 forward_kl_topk 蒸馏场景中，通过 trainer 传递 distillation_only 标志使 actor engine 跳过 log_probs 和 PPO 损失计算，并在 Megatron optimizer_step 前清空缓存，降低显存占用。

3. 关键 Bug 修复

- MoE 权重同步：PR #6896 修复了 Transformers 5 将 Qwen MoE 权重存储为打包 3D 张量后，vLLM（≤0.24.0）无法通过传统 per-expert 路径加载的问题。新增 _iter_vllm_compatible_moe_params 迭代器实时拆分权重，并增加版本守卫。

- FP8 忽略层配置：PR #6906 将 SGLang FP8 量化配置构建集中到 build_sglang_fp8_quant_config，支持从 HF 配置、环境变量等多种源合并忽略层列表，并添加正则匹配支持（re: 前缀），解决 Qwen3 GatedDeltaNet 因 block-wise FP8 崩溃的问题。
- 多节点 profile 崩溃：PR #6861 修复了 nnodes>1 时非 master 节点因 engine 为 None 而引发的 AttributeError，在 start_profile/stop_profile 中添加 node_rank 守卫。
- GLM-4V processor：PR #6873 更正了 hf_processor() 中类名 Glm4vImageProcessor→Glm4vProcessor，使多模态模型正确初始化。

4. 功能扩展

- 多模态训练：PR #6849 为 lmms-lab/multimodal-open-r1-8k-verified（图像数学推理）和 TinyLLaVA-Video-R1（视频问答）新增了数据预处理脚本，并提供了 FSDP GRPO 启动脚本。同时修复了 _ragged_idx 损坏问题。
- MoE 可观测性：PR #6853 在 v1 trainer 中集成了基于 routed expert replay 数据的负载均衡指标，包含专家分配计数和 violation 统计，由 rollout.moe_load_balance_metrics_interval 配置控制（默认关闭）。

模块与主题趋势

从标签频次看，npu（12 次）占据压倒性优势，doc（8 次）和 bugfix（7 次）分列二三位，说明本周工作重心是 NPU 平台适配、文档质量和缺陷修复。ci（5 次）和 algo（4 次）也表现活跃，反映出团队在持续集成和算法支持（GSPO、DAPO、多模态）上持续投入。perf（4次）和 fully_async（3 次）虽然数量较少，但都是高重要性 PR，代表性能优化和全异步架构迭代并未停步。

在作者方面，共 10 位核心贡献者产生了 20 个 PR（提交统计前 10），其中多位提交 2 个 PR，贡献分布较为均匀。热点文件集中在 docs/ascend_tutorial/ 路径下的文档和 tests/special_npu/ 下的测试脚本，表明当前核心活动区域是 Ascend 相关的文档和测试。另外，verl/workers/engine/megatron/transformer_impl.py 出现 2 次，与 PR #6901 和 #6848 相关，显示 Megatron 引擎正在吸收关键性能优化。

综合来看，本周的变化呈现“平台扩充 + 深度优化”的双轨模式：一方面大幅提升 Ascend NPU 的测试覆盖和用户文档，让 NPU 训练链条更完整；另一方面在已有模块中挖掘性能潜力，修复边界问题。全异步和蒸馏等新兴功能也在稳步演进。

风险观察

1. 测试覆盖缺口：尽管新增了 NPU 测试，但多个核心优化 / 修复模块（如 #6896 MoE 权重同步中的 _iter_vllm_compatible_moe_params、#6901 BSHD 填充的逻辑、#6848 OPD 优化）缺少单元测试，仅依赖 E2E 或手动验证，回归风险较高。
2. 文档版本一致性：PR #6924 在合并时留下了 DAPO 和 GSPO 文档中 git checkout main 命令与文本建议“指定 commit ID”的矛盾，用户若直接复制命令可能使用错误版本。该问题在 Review 中指出但未被修正。
3. 配置兼容性敏感：PR #6906 中配置合并优先级（环境变量、HF 配置、正则模式）缺乏显式说明，用户可能误解；PR #6882 中通过 hasattr 判断配置存在，若配置结构变化（如

OmegaConf 启用 struct）可能失效。

4. 空 batch 异常：PR #6901 在分布式训练空 batch 时 .max() 会触发 RuntimeError，Review 已指出但未被修复，对生产环境构成潜在威胁。
5. 环境硬编码：PR #6900 的启动脚本中直接硬编码 ASCEND_RT_VISIBLE_DEVICES 等变量，在多租户或动态资源环境下可能引发冲突。

重点 PR 速览

PR #	标题	作者	重要性	影响简述
6901	[megatron, perf] pad BSHD micro-batches to mini-batch max seq_len	YaoweiFan	7.6	性能优化：统一填充避免 cuDNN 重复编译，默认启用，但空 batch 风险未修复
6848	[fsdp, megatron, trainer] enhance mem footprint for forward_kl_topk OPD	dimjava	8.6	显存优化：蒸馏场景跳过非必要计算，设计精巧，建议补充 GPU 端到端测试
6896	[fsdp] fix Qwen3 MoE FSDP weight sync for vLLM rollout in Transformers 5	lxb0070981	7.4	关键兼容性修复，处理 Transformers 5 打包张量格式，避免权重加载失败
6906	[rollout] fix: support SGLang FP8 ignored layers	gem-mint	8.3	FP8 配置集中化 + 正则匹配，解决 Qwen3 GatedDeltaNet 崩溃，可推广至其他引擎

PR #	标题	作者	重要性	影响简述
6849	[data, rollout, worker] add Open-R1 multimodal and TinyLLaVA-Video preprocessing and training	lihanwen7	8.6	多模态训练扩展，提供完整数据预处理和训练脚本，奖励函数需用户自行实现
6853	[trainer, rollout] log rollout moe load-balance metrics	Luosuu	9.0	可观测性：MoE 负载均衡指标累计器设计优秀，配置驱动，默认关闭
6831	[ci] chore: add some NPU's UT/ST	daikang6	6.0	基础设施：填补 NPU 测试空白，新增 engine 兼容和设备抽象统一函数

此外，PR #6861（多节点 profile 崩溃修复）、PR #6882（partial_rollout 禁用修复）、PR #6867（OPD 温度配置修复）等也值得关注。这些修复虽小，但对全异步和蒸馏场景的稳定性至关重要。

后续建议

1. 补强测试覆盖：为近期合入的优化与修复（特别是 #6896、#6901、#6848、#6906）补充单元测试，重点验证边界条件（空 batch、配置缺失、版本阈值等）。建议在 tests/ 目录下建立与功能模块对应的测试文件。
2. 文档自动化验证：引入步骤检查文档中的 git 命令、版本号是否与代码实际一致。可考虑在 CI 中增加文档 lint 或版本引用检查。
3. 推广配置驱动优化模式：PR #6848 和 #6906 展示了通过配置标志（distillation_only、ignored_layers）灵活控制行为，这种模式应在其他模块推广，逐步替代硬编码条件。
4. 全异步鲁棒性提升：随着 fully_async 模块增多，应统一异常处理和条件守卫模式，避免类似 #6882 中 hasattr 和 #6883（隐含）的条件依赖问题。
5. 多模态规范化：基于 #6849 的模板，制定多模态数据集接入的标准流程（数据预处理→奖励函数→训练脚本），并推动奖励函数接口化。

6. 性能优化自动化：将 BSHD 填充优化集成到默认配置，并评估在 THD 路径中的适用性。同时关注空 batch 安全修复。