

PR #7537 完整报告

verl-project/verl

[doc] refactor: reframe Megatron Lite doc as Megatron Agent Compose preview

合并时间: 2026-08-24 15:02

原文链接: <http://prhub.com.cn/verl-project/verl/pull/7537>

[doc] Megatron Lite 文档重定位为 Agent Compose 预览

执行摘要

该 PR 是纯文档变更，只修改 docs/advance/megatron_lite_backend.rst 一个文件 (+37/-41)。核心动作是把文档从开发期旧名 Megatron Lite 重新框定为 Experimental Megatron Agent Compose preview，新增命名映射说明，将 mlite、megatron.lite、verl_mlite、MLITE_ROOT、megatron_lite 等遗留标识符对应到上游 megatron.experimental.agent_compose 命名空间；同时修正示例中 MLITE_ROOT 的指向（应为 Megatron-LM checkout）并移除快速变动的 DSA 依赖精确版本号。实质技术内容（OPTIMIZER 取值、dist_opt 位对齐验证、DeepSeek-V4 launcher 默认配置）全部保留，文件名刻意不变以维护入站链接。

功能与动机

PR body 明确说明了动机：Megatron Agent Compose 是 Megatron-LM 上游预期的实验路径，Megatron Lite 只是热开发期间的临时名字，原文档只用旧名描述组件，已与上游代码方向不符。此外还有两个内容修正：

- MLITE_ROOT 示例指向 /path/to/mlite 有误导性，实际应指向 Megatron-LM 仓库根目录，launcher 自己会拼接 experimental/lite 路径；
- DSA 相关依赖（nvidia-cutlass-dsl、nvidia-cudnn-frontend）版本变化快于文档更新节奏，继续固定精确版本只会让文档加速腐化，应把版本验证交给实际校验它们的 launcher / 环境文件。

实现拆解

1. 重写文档定位与章节结构：标题从 Megatron Lite backend 改为 Experimental Megatron Agent Compose preview，更新最后更新日期为 07/28/2026，章节从 Install the backend / Run an example 调整为 Evaluate the development preview / Run an evaluation example。
2. 新增 Current naming note 命名映射：显式声明 Agent Compose 是上游预期路径，列出开发预览仍在使用的遗留标识符，提示读者不要把这些名字当作稳定 API。
3. 修正 MLITE_ROOT 示例：两处示例占位符从 /path/to/mlite 改为 /path/to/Megatron-LM，并说明 launcher 会在运行时把预览路径追加到 PYTHONPATH。

4. 收紧 DSA 版本说明：DeepSeek-V4 DSA note 只保留依赖名（nvidia-cutlass-dsl、nvidia-cudnn-frontend），删除精确版本号和 IndexerForwardSm90 等细节，把版本验证职责外移。
5. 保留实质技术内容与文件名：OPTIMIZER 说明、128-GPU 网格默认值（PP4、EP8、CP4、全激活重计算）、dist_opt 位对齐验证结论、Further reading 全部原样保留；文件名刻意不随标题重命名，避免破坏 docs/index.rst toctree 与外部链接。
6. 本地验证配套：检查 tests/special_sanity/check_docs_time_info.py（时间戳存在）、命名约定钩子（全文使用 verl 小写形式）、sphinx-build --keep-going（零警告零错误），确认满足 doc_test 工作流门槛。

以下为重新整理后的 RST 文档核心片段，展示命名映射与修正后的示例：

```
Experimental Megatron Agent Compose preview
=====
```

Last updated: 07/28/2026.

```
.. 命名说明：Agent Compose 是上游 Megatron-LM 的目标实验路径，
.. Megatron Lite 只是开发期间的旧名。文档先给读者一张新旧名称对照表，
.. 避免把 mlite / MLITE_ROOT 等预览标识符当作稳定 API 使用。
```

```
Current naming note
-----
```

```
`Megatron Agent Compose <https://github.com/NVIDIA/Megatron-LM/tree/main/experimental/
agent_compose>`_
is the intended upstream experimental path. Its previous name, under hot
development, was
`Megatron Lite <https://github.com/NVIDIA/Megatron-LM/tree/dev/experimental/lite>`_.
```

The current development preview still uses legacy names such as ``mlite``,
``megatron.lite``, ``verl\_mlite``, ``MLITE\_ROOT``, and ``megatron\_lite`` script
names. Treat those as preview implementation details that will be changed. The
reviewed upstream namespace is ``megatron.experimental.agent\_compose``.

```
.. 修正点：MLITE_ROOT 应指向 Megatron-LM 仓库根目录，launcher 会自行拼接
.. experimental/lite 路径，因此示例占位符从 /path/to/mlite 改为 /path/to/Megatron-LM。
```

```
Run an evaluation example
-----
```

```
.. code-block:: bash
```

```
MODEL_PATH=/path/to/deepseek-v4
MLITE_ROOT=/path/to/Megatron-LM
OPTIMIZER=fsdp2
bash examples/grpo_trainer/run_deepseek_v4_megatron_lite.sh
```

评论区精华

该 PR 没有 review 评论（`comments_count` 与 `review_comments_count` 均为 0），合并者 wuxibin89 直接 APPROVED 且无评论正文，纯文档变更未产生实质技术交锋。PR body 中作者主动论证的「文件名保留」取舍是唯一值得关注的决策：上游重命名尚在进行中，立即重命名会破坏入站链接，待 `agent_compose` 路径落地 main 后再跟随更名。

风险与影响

- 命名时效性风险：上游 `agent_compose` 仍在演进，本文的命名映射可能在重命名落地后过期，需要后续 PR 跟进。
- 文件名与标题不一致的困惑：URL 仍是 `megatron_lite_backend.rst`，侧边栏标签却是新标题，按 URL 直达的读者可能短暂困惑，但这是维护入站链接的主动取舍。
- DSA 版本指引外移：文档不再给出精确版本，用户绕过 launcher 手动装依赖时可能无法复现 DSA 环境，信息有效性依赖 launcher 或环境文件及时更新。
- 影响面：零运行时影响，用户侧获得更准确的 Megatron 后端预览指引，团队侧降低对快速变动版本号的文档维护负担。

关联脉络

该 PR 是 verl 对上游 Megatron-LM `experimental/lite` → `experimental/agent_compose` 重命名的文档侧跟进。在 verl 仓库内，它与近期 Megatron 后端演进（#7407 路由回放与 `transformer_impl` 重构、#7335 Megatron 示例脚本扩展）处于同一脉络；#7536 清理 `megatron/veomni` 遗留配置与本次收敛 `mlite` 等遗留标识符，都是「命名收紧」工作在不同层的表现。作者已在 PR body 中预告：一旦上游 `agent_compose` 路径落地 main，文档文件名可随之从 `megatron_lite_backend.rst` 更名，届时需同步更新 `docs/index.rst` 的 toctree 与外部入站链接。