

PR #2383 完整报告

radixark/miles

[readme]: rewrite the README

合并时间: 2026-08-12 15:53

原文链接: <http://prhub.com.cn/radixark/miles/pull/2383>

执行摘要

- 一句话: 按 SGLang 风格重写 README 并同步文档首页
- 推荐动作: 值得快速浏览, 无需精读: 本 PR 展示了大型 RL 项目如何把 README 从 " 三行链接 " 升级为可导航的项目门户。值得借鉴的决策: 一是 " 引用而非内联 "——安装细节只指向 Quick Start, 避免多副本漂移; 二是 News 与功能 bullet 全部绑定真实存在的文档页, 降低死链概率; 三是 README 与文档首页保持同一份口径。若团队后续继续重组文档, 建议补充链接有效性检查 (如 markdown-link-check) 到 CI。

功能与动机

PR body 明确说明重写目标是 "Reworks the README against the finished v0.1 docs, following the shape of SGLang's README": 旧 README 只有一行标语和三个落地页链接, 与已完工的 v0.1 文档体系严重脱节。另一个关键考量是防漂移——"The snippet deliberately does not inline the download and conversion steps; it points at Quick Start for those rather than shipping a copy that can drift", 并且要求 "Every docs link points at a page that exists on main today", 用具体页面替换掉旧的 /docs/advanced 与 /docs/developer 落地页链接。

实现拆解

1. 重构 README 顶层结构。以 SGLang README 为模板, 将 logo 从 550px 缩小到 340px, 新增 GitHub/Docs/License/Slack 徽章行和 Documentation/Quick Start/Models/Blog/Slack 导航链接行; 新增 News 区块 (9 条 LMSYS 博客, 统一「日期标题 (blog 链接)」格式、最新在前); 新增 Acknowledgment 区块 (致谢 slime、SGLang、Megatron-LM 与 torch_memory_saver, 预留 TODO 图位)。
2. About 功能清单重组并绑定真实文档页。功能被归为三组: Performance (Fully async RL、Fast agentic rollout、Fast weight updates、Low-precision training、LoRA and multi-LoRA)、Correctness and resilience (TITO、R3、Fault tolerance、Miles dashboard)、What Miles runs (Day-0 model、Hardware、Recipe、Coding-agent environments), 每个 bullet 都链接到对应的 docs 页面; 同时明确 FSDP2 后端的定位——recipes、parallelism 与最大模型都跑在 Megatron-LM 上, FSDP2 面向希望按 HuggingFace 原版训练的场所。
3. Getting Started 保持 " 链接为主 "。只放 6 个文档链接与安装页的 Docker pull/run 命令块, 刻意不内联数据下载与模型转换步骤, 而是指向 Quick Start, 避免多副本内容漂移。

4. 同步 docs/index.md 并反复收敛。commit 744a81df 将首页的简介、功能分组与 Supported models/hardware 列表对齐 README；commit 5ba01db 与 6e46e145 把原来含糊的 Latest updates 替换为与 README 完全一致的 News 列表（含 MXFP8/NVFP4 条目）。18 个 commit 体现多轮打磨：先删掉 Contact Us 与 citation 块，两次重组功能分组命名（efficiency/design → performance/correctness/coverage → What Miles runs），恢复 About 开头段、缩小 logo、移除 mbridge 致谢，最终以 "Miles integrates its upstreams" 定稿表述。

本 PR 无测试、配置或部署配套改动。

关键文件：

- README.md（模块 仓库首页；类别 docs；类型 documentation）：仓库最核心的门面文件，从三段式简介重写为 SGLang 风格的完整 README：新增徽章行、导航链接行、News 时间线、About 三大功能分区与 Getting Started、Acknowledgment，是本 PR 的主体。
- docs/index.md（模块 文档首页；类别 docs；类型 documentation）：文档站首页与 README 信息口径同步：简介段落、功能分组、Supported models/hardware 列表与 README 对齐，Latest updates 替换为 News 列表，避免两个入口出现两套说法。

关键符号：未识别

评论区精华

唯一的 review 评论是 yueming-yuan 在 README.md 第 19 行（News 区块）给出的 suggestion，建议把 2026-07-29 的 "End-to-End MXFP8 and NVFP4 RL in Miles" 博客加入 News 列表。作者随后以 commit 6e46e145 ("docs: add the MXFP8 and NVFP4 post to the news list") 将其采纳为 News 首条，整体 review 状态为 APPROVED。该交互体现了文档 PR 的小闭环：reviewer 补遗漏条目，作者即改即合。

- News 列表补充 MXFP8/NVFP4 博客条目 (documentation): 作者以 commit 6e46e145 (docs: add the MXFP8 and NVFP4 post to the news list) 采纳该建议，最终 README 第一条即为该条目；review 状态 APPROVED。

风险与影响

- 风险：纯文档变更无运行时风险，但有三点文档侧隐患：
 1. 双份维护成本：README 与 docs/index.md 现在共享同一套 News 与功能清单，后续任一侧单独更新都会产生口径漂移；仓库正处于文档密集重组期（2375、2391、2396、2400 等），两侧同步压力不小。
 2. 硬编码链接脆弱：README 大量使用完整 URL (miles.radixark.com/docs/...)，目前虽逐个核对过 "exists on main today"，但一旦 docs.json 或页面路径再调整，这些链接会静默 404，且无自动化链接检查兜底。
 3. 时效性内容：硬件支持（GB300、MI355X 等）、模型列表（Kimi-K3、Nemotron 3 Ultra 等）与 News 日期都随发布节奏快速过期，需要团队形成定期刷新机制。- 影响：影响对象是 GitHub 访客（潜在用户、贡献者、评估者）的第一印象与上手路径，以及文

档站门户页的信息一致性。影响范围限定在 docs 层，无代码、构建、CI 或运行时变化；对团队而言，新增了一块需要与发布节奏同步维护的门面内容（News 与硬件 / 模型列表）。整体影响程度中等偏低，但对开源项目的对外形象是正收益。 - 风险标记：文档双份维护，链接漂移风险，时效性内容易过时

关联脉络

- PR #2375 docs: reorganize the Training Backends section, drop the experimental FSDP framing: README 与首页新增的 Training Backends 说明（FSDP2 作为备选后端）正来自该 PR 确立的文档口径，training-backend 页面是本 PR About 段落的直接引用目标。
- PR #2396 docs: refresh MXFP8 and NVFP4 RL guide: README News 首条与 Low-precision training 功能 bullet 均指向该 PR 维护的 low-precision 文档，新闻素材同源。
- PR #2376 docs(developer): rewrite the developer guide against the code: README 的 Getting Started 链接到 contributor-guide 页面，该页面由本 PR 按代码重写，保证了 README 链接的目标存在且内容最新。
- PR #2356 Replace all the .sh launch scripts with .py launch script: README 的 Launch Script Walkthrough 链接对应当前 launch-script.md，该 PR 完成了启动脚本从 .sh 到 .py 的迁移，使文档与实际用法一致。