

# PR #2271 完整报告

radixark/miles

docs: refresh the homepage supported-models table

合并时间: 2026-08-11 04:30

原文链接: <http://prhub.com.cn/radixark/miles/pull/2271>

## 执行摘要

- 一句话: 首页模型表刷新: 增新删旧并声明非穷尽
- 推荐动作: 纯文档 PR, 无需精读实现; 体小 (1 个文件 +11/-9), 建议快速浏览。值得注意的两个设计决策: 一是把主页模型表从“全量承诺”改为“近期高亮 + not exhaustive 声明”, 降低首页与代码库持续同步的维护成本; 二是对暂无文档页的模型采用“脚本 /PR 链接”过渡, 属于务实的渐进式文档策略。若团队后续维护首页模型列表, 可复用这套 9 个月时效线与三层链接规则。

## 功能与动机

PR body 明确指出此前表格承诺 Each model name links to its recipe page, 却混入大量发布超过 9 个月的旧模型, 会让用户误判 Miles 当前支持的模型范围; 因此本次要刷新首页 supported-models 表格: 显式声明 not exhaustive、突出近期发布 (highlights recent releases), 并说明老代模型仍开箱即用。同时交叉核对仓库内 recipe 与开放 PR (#1825、#2216 等) 后新增 6 个以上新模型、按 2025-11-07 时效线移除旧行, 保证首页与主干代码 / 文档同步。PR 还 stacked on #2264, 与首页 Core features 重构同步演进。

## 实现拆解

1. 调整表格定位与引导语: 在 docs/index.md 的 Supported models 段落, 把 Each model name links to its recipe page. 改为 recipe page or launch script, 并新增两句说明——表格 not exhaustive、只突出 recent releases、老代模型仍开箱即用。这是本 PR 的核心语义变更, 后续所有增删行都建立在这个“高亮而非清单”的定位上。
2. 新增近期发布模型行: 交叉核对仓库内 recipe 与开放 PR 后加入 GLM-5.2 ( /models/glm/glm5-2, 页面由 #2216 提供)、Inkling-Small、Nemotron-3-Ultra-550B-A55B、Gemma-4 26B-A4B / 31B、JoyAI-LLM-Flash 与 Kimi-K3。链接分三层: 有文档页的行走 /models 相对链接; 暂无文档页的行指向仓库 main 上的启动脚本 (如 scripts/run\_nemotron\_3\_ultra\_550b\_a55b.py、scripts/run\_gemma\_4\_26b\_a4b.py); Kimi-K3 临时指向 PR#1825, 待其合并后切换。
3. 按 9 个月时效线删除旧模型: 以 2025-11-07 为截止, 移除 DeepSeek-R1/V3、Qwen3-Next-80B-A3B-Thinking、Qwen3-30B-A3B / 235B-A22B、Qwen3 dense 全系、GLM-4.5、GLM-Z1-9B-0414、Kimi-K2 Instruct / Thinking、Moonlight-16B-A3B、MiMo-7B-RL 与 gpt-oss-20b, MiMo 与 GPT-OSS 两个族的整行消失。被删模型的 recipe 页与启动脚本仍保留在仓库和 /models 目录, 主页只是不再高亮它们。

4. review 确认与合入 main: Zhichenzzz 质疑 Qwen3 删除后与 quick-start 示例脱节, Shi-Dong 以“引导语已声明老代模型仍受支持”为由确认不保留; 随后解决与 main 的 docs/index.md 冲突 (最后一个 merge commit), 并重跑一次超时的 CodeQL 检查后合入。

无测试、配置或 schema 配套改动; commit 历史显示前 3 个提交属于上游 #2264 的内容 (Core features 重构), 本次实际改动集中在后 3 个提交。

关键文件:

- docs/index.md (模块 文档; 类别 docs; 类型 documentation): 全 PR 唯一变更文件: 首页 Supported models 表格的定位、增删与链接策略全部集中于此, 定义后续文档维护约定。

关键符号: 未识别

## 评论区精华

本次 review 仅有一个实质分歧线程, 位于 docs/index.md 第 80 行 (Qwen 表格行):

- Zhichenzzz 提问: still keep qwen-3? because the quick example is qwen3 —— 担心删除 Qwen3 后与 quick-start 示例脱节。
- Shi-Dong 回应: I'd suggest dropping it here since we mentioned above that "including older generations of the families below." —— 即引导语已声明老代模型仍受支持, 无需在表中重复保留。

结论: 按维护者意见删除 Qwen3 全系旧行, 问题闭环; 除该线程外无其他 review 评论。另有两条 reviewed 记录 (APPROVED / COMMENTED) 均为空 body, 无实质内容。该讨论的价值在于明确了“主页高亮近期发布、老代模型由 /models 承接”的文档分工。

- 是否在表中保留 Qwen3 旧系列 (question): 按维护者意见删除 Qwen3 全系旧行, 由引导语承担“老代仍支持”的说明职责, 问题闭环。

## 风险与影响

• 风险:

1. 链接时序依赖: GLM-5.2 行指向 /models/glm/glm5-2 页面, 该页面依赖 #2216 合并后才存在, 若 #2216 长期不合并, 主页会出现悬空链接; Kimi-K3 行临时指向 PR#1825, 后者合并后链接内容会从“PR 上下文”变为“已合并 PR”, 需要二次更新。
2. 脚本链接失效: Gemma-4、Nemotron-3-Ultra、JoyAI-LLM-Flash 均直接链接仓库 main 上的 scripts 文件, 脚本路径一旦重命名或移动即 404, 且这种失效不会被常规 review 发现。
3. 时效性主观: 以约 9 个月 (2025-11-07) 作为裁剪线是人为约定, Kimi-K2-Thinking (2025-11-06) 仅差一天即被移除, 读者对“新 / 旧”边界可能产生疑问。
4. 影响面: 纯文档变更, 无运行时回归风险; 风险集中在链接可用性与信息时效性。- 影响: 对用户: docs 主页是新用户了解 Miles 支持模型的第一入口, 调整后信息更贴近当前主干版本, 避免按旧清单误判; 老代模型用户仍可通过 /models 目录找到 recipe 页与启动脚本。对团队: 确立了“主页高亮近期发布 + 老代模型由 /models 承接”的文档分工, 以及“无文档页先链脚本 /PR、落地后回改”的过渡约定; 合入后需记住两处 follow-up (#2216 合并后 rebase GLM-5.2 行、#1825 合并后切换 Kimi-K3 链接)。影响程度较

低，局限在文档导航层。 - 风险标记：链接依赖未合并 PR，时效性裁剪线主观，脚本链接存在失效风险

## 关联脉络

- PR #2264 docs: update homepage Core features section per the v0.1 feature list: 本 PR 的直系前序 (stacked) : 同一分支先重构首页 Core features, 本 PR 继续刷新同一文件 docs/index.md 的模型表, PR body 明确标注 stacked on #2264。
- PR #2220 Add the GLM-5.2 744B x terminal-bench-2 Daytona example: GLM-5.2 文档页 #2216 叠加在此 PR 之上, 而本 PR 的 GLM-5.2 表行依赖 #2216 的页面链接, 构成三层依赖链。
- PR #1825 Kimi-K3 (标题未在上下文中提供) : PR body 与表格中 Kimi-K3 行临时指向 #1825 (提供 scripts/run\_kimi\_k3.py / run\_kimi\_k3\_lora.py), 其合并后需把链接切换为正式文档或脚本链接。
- PR #2216 docs: add GLM-5.2 model page and update supported-models tables: 关联 Issue/ 上游文档 PR: 它新增 docs/models/glm/glm5-2.md 并同步更新 supported-models 表, 与本 PR 的 GLM-5.2 行有显式 rebase 依赖 (PR body 说明 whichever PR lands second takes a trivial table-row rebase) 。