

PR #32835 完整报告

sgl-project/sglang

[docs] Rotate popular models on the landing pages, lead the Cookbook nav with Kimi

合并时间: 2026-07-30 08:46

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/32835>

执行摘要

本 PR 将文档首页和 Cookbook 首页的硬编码模型展示替换为可配置的轮播组件，同时重新排序导航使 Kimi、Thinking Machines、GLM 等热门模型前置。所有模型数据集中在单一文件管理，后续新增模型只需编辑数据列表。

功能与动机

PR body 指出当前文档的三个痛点：

1) 首页 `index.mdx` 硬编码了一个约 180 行的 Kimi-K3 hero，更换模型必须重写页面； 2) Cookbook 首页完全没有模型入口，读者看不到值得打开的模型； 3) 导航以 Qwen 为首，团队希望把最想展示的模型放在前面。为解决这些问题，PR 创建了可复用的轮播组件和统一数据源，并调整了导航顺序。

实现拆解

- 数据合约：新建 `docs_new/src/snippets/configs/popular-models.jsx`，导出 `popularModels` 数组，每项包含模型名称、厂商、链接、logo、badge、标签和 hero 块（用于首页 banner）。当前种子数据包括 Kimi-K3、Inkling 和 GLM-5.2。
- 轮播组件：新建 `docs_new/src/snippets/_popular_models.jsx`，导出 `PopularModels` 组件，支持 `variant="hero"`（全宽 banner）和默认 `variant="strip"`（紧凑行）。组件使用 React hooks 管理索引、轮播定时器、暂停状态和 `prefers-reduced-motion` 监听。轮播在悬停 / 聚焦时暂停，动画可访问性良好。
- 首页替换：修改 `docs_new/index.mdx`，删除原 180 行内联 Kimi-K3 hero，导入并渲染 `<PopularModels models={popularModels} variant="hero" />`。
- Cookbook 首页嵌入：在 `docs_new/cookbook/intro.mdx` 导入组件并渲染 `<PopularModels models={popularModels} />`（默认 strip 样式），为 Cookbook 提供模型入口。
- 导航重新排序：调整 `docs_new/docs.json`，将 Kimi (Moonshot AI)、Thinking Machines、GLM 三组提到 Autoregressive Models 导航的顶部（Qwen 之前）；同步更新 `docs_new/cookbook/autoregressive/intro.mdx` 中的卡片顺序以匹配导航。

`docs_new/src/snippets/_popular_models.jsx`

核心轮播组件，支持 hero 和 strip 两种变体，处理轮播逻辑、可访问性和动画降级。

```
// Auto-rotating carousel shared by docs home (variant="hero") and Cookbook home (variant=
```

```
"strip").  
// Rotation pauses on hover/focus and respects prefers-reduced-motion.  
// One export for both shapes: Mintlify evaluates each component independently at hydration,  
// so shared state (timer, index, controls) must sit inside one component.
```

```
export const PopularModels = ({  
  models = [],  
  variant = "strip",  
  interval = variant === "hero" ? 9000 : 6000, // hero blurb needs more time  
  label = "Popular models",  
}) => {  
  const [index, setIndex] = useState(0);  
  const [paused, setPaused] = useState(false);  
  const [reduceMotion, setReduceMotion] = useState(false);
```

```
  // Listen for prefers-reduced-motion changes  
  useEffect(() => {  
    if (typeof window === "undefined" || !window.matchMedia) return;  
    const mq = window.matchMedia("(prefers-reduced-motion: reduce)");  
    const sync = () => setReduceMotion(mq.matches);  
    sync();  
    mq.addEventListener("change", sync);  
    return () => mq.removeEventListener("change", sync);  
  }, []);
```

```
  const count = models.length;  
  const isHero = variant === "hero";
```

```
  // Functional update to avoid stale closure over index  
  useEffect(() => {  
    if (count < 2 || paused || reduceMotion) return;  
    const id = window.setInterval(  
      () => setIndex((i) => (i + 1) % count),  
      Math.max(2000, interval)  
    );  
    return () => window.clearInterval(id);  
  }, [count, paused, reduceMotion, interval]);
```

```
  // Clamp active index in case list shortened mid-rotation  
  const active = count ? Math.min(index, count - 1) : 0;
```

```
  if (!count) return null;
```

```
  // Only active slide is opaque to prevent seeing two half-drawn cards mid-transition  
  const slideStyle = (i) => ({  
    flex: "0 0 100%",  
    minWidth: 0,  
    opacity: i === active ? 1 : 0,  
    pointerEvents: i === active ? "auto" : "none",
```

```

    transition: reduceMotion ? "none" : "opacity 0.3s ease",
  });

// Tag chip component used inside slides
const tagChip = (t, big) => (
  <span key={t} style={{ padding: big ? "0.35rem 0.65rem" : "0.15rem 0.45rem", borderRadius:
    "999px", background: big ? "rgba(255, 255, 255, 0.1)" : "rgba(255, 255, 255, 0.12)", color:
    "rgba(255, 255, 255, 0.9)", fontSize: big ? "0.78rem" : "0.72rem", fontWeight: 650 }}>{t}</
    span>
  );

// ... (heroSlide, stripSlide, controls, and render logic omitted for brevity)
return ( /* layout */ );
};

```

docs_new/src/snippets/configs/popular-models.jsx

数据源文件，定义三个热门模型（Kimi-K3、Inkling、GLM-5.2）的展示信息，统一维护。

```

// Single literal export: Mintlify re-evaluates at hydration, so no spreads/calls/IIFE.
// Both docs home (variant="hero") and Cookbook home (strip) walk this list in order.

export const popularModels = [
  {
    name: "Kimi-K3",
    vendor: "Moonshot AI",
    href: "/cookbook/autoregressive/Moonshotai/Kimi-K3",
    logo: "/cards/logos/moonshotai.png",
    badge: "New",
    tags: ["8 platforms", "PD disagg", "DSPARK"],
    hero: {
      eyebrow: "Featured model • New",
      headline: "Meet Kimi-K3 on SGLang",
      blurb: "SGLang natively implements and deeply optimizes K3's new architecture with fused
        KDA decode kernels, DP attention, MTP, PD disaggregation, and KDA-aware prefix caching.
        Kimi-K3 is supported on both NVIDIA and AMD GPUs.",
      tags: ["2.8T parameters", "Fused KDA decode", "NVIDIA + AMD"],
      cta: "Open the Kimi-K3 cookbook",
      caption: "Kimi-K3 deployment guide",
    },
  },
  // Inkling and GLM-5.2 entries follow a similar structure
];

```

评论区精华

PR 无 review 评论，但 PR body 作者自述了关键设计决策：

"Mintlify evaluates each exported component on its own at hydration, so a module-level helper shared by two components is out of scope at runtime (this fails at render, not at `mint validate`)."

"Only the active slide is opaque, so mid-slide you

never see two half-drawn cards." "Dots and prev/next live in the card chrome, never in the slide body — a → inside a slide reads as 'next' when clicking it would in fact navigate to the model page."

这些决策体现了对 Mintlify 平台限制的适应和良好的可访问性实践。

风险与影响

风险：纯文档变更无功能风险，但无测试覆盖，依赖 Mintlify 的 JSX 兼容性。导航重新排序虽不改变 URL，但未来新增组时需注意保持 Nav 和卡片顺序一致。

影响：首页和 Cookbook 首页的展示从静态硬编码变为动态轮播，用户体验提升；团队维护成本降低。

关联脉络

本 PR 与近期 #32834 (Kimi-K3 文档更新) 和 #32799 (Inkling cookbook 修复) 同属文档改进线，共同提升热门模型的文档展示效果。未来计划可能在数据列表中继续添加新模型。