

PR #28668 完整报告

sgl-project/sglang

docs(minimax-m3): add MMMU-Pro accuracy to B200 benchmark card

合并时间: 2026-06-19 02:40

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/28668>

执行摘要

为 MiniMax-M3 cookbook 的 B200 基准卡添加 MMMU-Pro (10 选多模态) 精度数据, 包括复现命令和实测 72.7%。纯文档变更, 无运行时影响。

功能与动机

MiniMax-M3 cookbook benchmark 卡在 B200 行仅列出了 GPQA Diamond 和 GSM8K, 缺少 MMMU-Pro (10 选多模态评测)。PR 旨在补全多模态评测维度, 使用户能参考和复现 M3 在 MMMU-Pro 上的表现。

实现拆解

1. 复现命令注册: 在 minimax-m3.jsx 的 benchmarkCommands.accuracy 中新增 mmmu_pro_pct 键, 值为 sgl-eval run mmmu_pro 命令 (含 --thinking 及 M3 推荐采样参数)。
2. UI 标签注册: 在同文件的 accuracyLabels 数组追加 ["mmmu_pro_pct", "MMMU-Pro", "%"] 行。
3. 数据填入: 在 minimax-m3-benchmarks.jsx 的 B200 行 accuracy 对象加入 mmmu_pro_pct: 72.7, 并更新注释记录测量条件。

docs_new/src/snippets/configs/MiniMaxAI/minimax-m3.jsx

注册 MMMU-Pro 精度复现命令和 UI 标签, 是文档数据展示的核心配置入口。

```
// 在 benchmarkCommands.accuracy 中新增 mmmu_pro_pct 命令
accuracy: {
  gsm8k_pct: `pip install git+https://github.com/sgl-project/sgl-eval
sgl-eval run gsm8k ...`,
  gpqa_pct: `pip install git+https://github.com/sgl-project/sgl-eval
sgl-eval run gpqa ...`,
  // 新增: MMMU-Pro 10 选多模态评测复现命令
  mmmu_pro_pct: `pip install git+https://github.com/sgl-project/sgl-eval
sgl-eval run mmmu_pro \
  --base-url http://{{CURL_HOST}}:{{CURL_PORT}}/v1 \
  --model {{MODEL_NAME}} \
  --temperature 1.0 --top-p 0.95 \
  --thinking`,
},
```

```
// 在 accuracyLabels 数组末尾追加 MMMU-Pro 标签 (UI 显示为 "MMMU-Pro")
accuracyLabels: [
  ["gpqa_pct", "GPQA Diamond", "%"],
  ["gsm8k_pct", "GSM8K", "%"],
  ["mmmu_pro_pct", "MMMU-Pro", "%"],
],
```

[docs_new/src/snippets/configs/MiniMaxAI/minimax-m3-benchmarks.jsx](#)

B200 行精度数据更新，实际填入 MMMU-Pro 数值。

```
// B200 行的 accuracy 对象新增 MMMU-Pro 精度值
accuracy: {
  gpqa_pct: 89.1,
  gsm8k_pct: 96.5,
  mmmu_pro_pct: 72.7, // 新增: MMMU-Pro 标准 10 选测试分片, 1730 题单次 72.66%
},
```

评论区精华

无 review 讨论。

风险与影响

纯文档更新，无技术风险。影响仅限于 MiniMax-M3 cookbook 页面 B200 行新增一个精度指标。对系统运行无影响。

关联脉络

本 PR 补充的 MMMU-Pro 数据基于之前 #27944 (MiniMax-M3 初始基准卡) 的 B200 测量结果，是同一 cookbook 的增量完善。sgl-eval 的 MMMU-Pro 支持由 sgl-eval 仓库 PR #19 提供。