

LA-Bench 提出アーキテクチャ @ O_AI team

事前準備: 参考文献要約

入力の references

Web UI ChatGPT(gpt-5.1)で要約
提示文献の日本語要約と重要ポイントを作成

*_references_rag.jsonl
sample_idごとの固定コンテキスト

実行時: gpt-5.1 パイプライン

入力 JSONL

- instruction
- mandatory_objects
- source_protocol_steps
- expected_final_states
- references

+

参考文献
要約

+

General
criteria

2. 手順生成

3. 品質
チェック
制約違反と
内部整合性を確認

needs_fix?

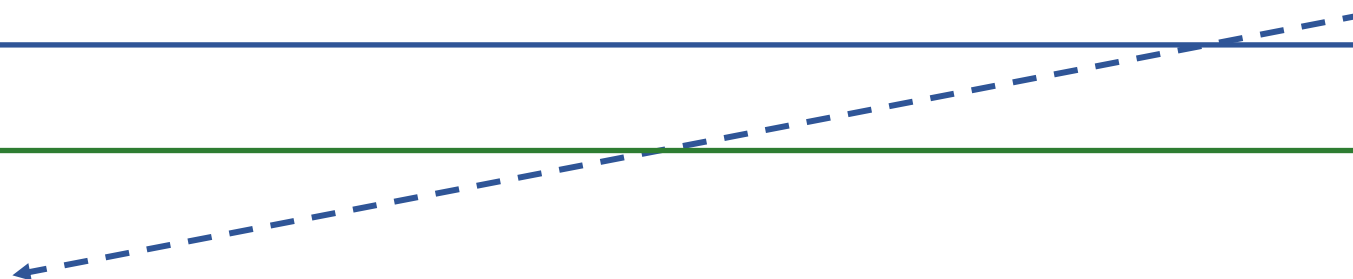
No

提出JSONL

Yes

1. 制約抽出

数値 / 順序 / 安全 /
物品使用 / 最終状態



恐らく効いたもの

- 制約抽出
 - gpt-5.1はまだ数値や単位の扱いが危うかったり、必要な物品を飛ばしたりしていた
 - 事前に構造化することで出力が安定した？
- 品質チェック
 - 出力が安定した

試したが採用しなかったもの

- 動的RAG：protocols.io APIで公開プロトコルを検索し、embedding DBを作成 → 類似検索してcontextに追加
- few-shot：public/exampleのoutput.procedure_stepsをcontextに追加
 - 良くも悪くもcontextに追加したプロトコルの書き方に従う
- Gemini-3.0-pro
 - 計算間違いやハルシネーションがgpt-5.1より頻出