TraceSIR

TraceFormat · ReportEval

Input File Format

- Single Trace Instance in JSON Format
- Multiple Trace Instances in ZIP Format

Optional Additional Requirements for Report Generation

- Language, Template, Structure, Format, Scope .....

Single Trace Instance Format

- [Required] oid messages (OpenAI)
- [Optional] task gold_score gold_judge other

StructureAgent

TraceFormat — messages (OpenAI)

	Thought	Action	Observation
Round1	We have a puzzle:	search{"query":"....."}	【0† Tribute to late actor .....
Round...	.....	.....	.....
Round*	Explanation: - The	finish{"answer":"....."}	/

- id ▶ TraceBench-{1, 2, ... , *}
- oid
- thought
- action
- observation
- task

▶ len(thought)=len(action)=len(observation)=*

▶ Abstract *length threshold per field per step*

- length
- gold_score
- gold_judge
- other

InsightAgent

- > Overall Assessment ➤ 0~100
- > Error Detection ◆ gold_score ◆ gold_judge
  - ➤ Main Error: Core fatal error in the trace
  - ➤ Other Errors: Other minor errors in the trace
- > Weakness Identification
  - ➤ Disadvantages: Shortcomings revealed in the trace
  - ➤ Advantages: Commendable aspects in the trace
- > Root Cause Analysis ➤ Insight
- > Optimization Suggestions
  - ➤ Suggestions: Concrete and actionable suggestions
  - ➤ Fine-tuned Samples: Targeted SFT samples

ReportAgent

- > Error Label
- > Ready Time
- > Error Frequency Estimation
- > Score Distribution Modeling
- > Report Generation ➤ Markdown ➤ Requirement
- > Report Refinement ➤ Appendix ➤ TraceBench

ReportEval

Overall Structure		Overall Impact
Error Analysis	Root Cause Analysis	Optimization Analysis