
Objective: LLM should adhere to format and/or content priorities stated in the conversation.

Rubric pattern:

- Use `instruction_being_tested` (the explicit instruction) and decompose `expected_compliance` into atomic criteria

Example JSON:

```
{
  "question": "How should I allocate my funds for the upcoming quarter?",
  "instruction_being_tested": "Always provide detailed financial breakdowns when I ask about budgeting decisions.",
  "expected_compliance": "Response should include itemized costs, specific amounts for different categories, and detailed breakdown rather than just a total estimate",
  "source_chat_ids": [],
  "rubric": [
    "LLM response should contain: itemized list of costs",
    "LLM response should contain: category-by-category breakdown",
    "LLM response should contain: detailed cost analysis"
  ]
}
```

6. Knowledge Update

Objective: LLM must reflect updated values when prior values have changed over time.

Rubric pattern:

- Derive criteria from the ideal answer, using the stem “LLM response should state/mention:” for the updated value(s).

Example JSON:

```
{
  "question": "How much initial capital have I allocated for real estate investing?",
  "answer": "$60,000",
  "source_chat_ids": {
    "original_info": [
      ],
    "updated_info": [
      ]
    },
  "rubric": [
    "LLM response should state: $60,000"
  ]
}
```

7. Multi-hop Reasoning

Objective: LLM must aggregate or compare information spanning multiple sessions.

Rubric pattern:

- Instantiate criteria from the ideal answer for each required intermediate or aggregated fact.

Example JSON:

```
{
  "question": "How many different banks did I consider for my mortgage options across my conversations?",
  "answer": "Two banks: Halkbank and Ziraat Bank.",
  "source_chat_ids": [],
  "rubric": [
    "LLM response should state: Two banks",
    "LLM response should state: Halkbank",
    "LLM response should state: Ziraat Bank"
  ]
}
```

8. Preference Following