

Analysis and Evaluation of DeepSeek r1:7b

Overview of DeepSeek r1:7b

DeepSeek r1:7b is a distilled reasoning model fine-tuned for handling real-time reasoning tasks and solving complex mathematical problems.

Its base model is Qwen2.5-Math-7B, which serves as the foundation for its reasoning and mathematical capabilities.

Experimental Setup

- **Local Execution:** The model was run locally via the terminal.
 - Query and response data were saved in a file named "**terminal_deepseek.txt**".
- **Custom Model Creation:** A new model or manifest named "**testrig_testing**", was created from DeepSeek r1:7b.
 - Query and response data for this model are stored in "**testrig_testing.txt**".
 - Model components were prepared, and the manifest file was successfully created before running the new model "testrigtesting.txt".

Steps Taken:

1. Pulled the DeepSeek r1:7b model

2. Created a new model using the command:

Command: `ollama create testrigtesting <path_to_modelfile>`

3. Executed the new model using the command:

`ollama run testrigtesting`

Saved and analyzed the responses from model.

- **Google Cloud Platform:** Although DeepSeek r1:7b is available on Google Cloud's Model Garden, it could not be used due to quota limitations. Quota renewal was required.

Evaluation Metrics

1. Latency

- **Observed Latency:** High, ranging between 60 to 90 seconds per query.
- **Probable Cause:** The model was run locally; system configurations may have contributed to slower processing times compared to cloud-based execution.

2. Accuracy

- **Rating:** 4.5/5
- The model demonstrated a high degree of precision in its reasoning and mathematical problem-solving capabilities.

3. Relevance

- **Rating:** 4.5/5
- Responses were contextually appropriate and aligned with the queries posed.

4. Coherence and Fluency

- **Fluency:** Good.
- Responses were well-structured, easy to understand, and grammatically accurate.

5. Factual Correctness

- **Rating:** Good.
- The model exhibited strong factual accuracy in its responses.

6. Response Length

- **Nature of Responses:** Lengthy and detailed.
- While the responses were comprehensive, they may require summarization for certain use cases.

7. Bias

- **Observed Bias:** Minimal.

- No significant biases were detected during the evaluation.

8. Perplexity

- **Rating:** Good.
- The model's ability to maintain low perplexity indicates a strong understanding of the input context.

9. Toxicity

- **Observation:** None.
- All responses were free from toxic or harmful content.

Summary

DeepSeek r1:7b is a robust model for real-time reasoning and mathematical problem-solving tasks. Despite high latency during local execution, it delivers accurate, relevant, and fluent responses. The minimal bias and high factual correctness further establish its reliability for diverse applications. However, cloud-based deployment can be used for improved performance and reduced latency.

Files

- **terminal_deepseek.txt:** Contains query and response data from the DeepSeek r1:7b model.
It contains queries and response for following queries:
 1. I have an equation $y=mx+c$, where $c=.3$ and slope is $.2$, what will be value of y ?
 2. write manual test cases for login in scenario of an ecommerce application
 3. how to solve $\sin(x)+\cos(x)=0.45*x$
- **testrig_testing.txt:** Contains query and response data from the custom "testrigtesting" model.
The model "testrigtesting" is prompted to generate test cases for different scenarios only.
It contains queries and response for following queries:
 1. write manual test cases for login in scenario of an ecommerce application
 2. Can u write test cases for a chatbot.

- **modelfile:** Includes code and components used to create the custom model.

Concerns:

1. [Researchers Uncover Prompt Injection Vulnerabilities in DeepSeek and Claude AI](#)
2. [DeepSeek Limits Sign-Ups Citing 'Malicious Attacks' - Business Insider](#)