Max Eduardo Lazarini Wienandts

Project Topic: Mistral vs Llama2

Problem statement: Is Mistral better than Llama?

Abstract:

This study delves into the performance evaluation of Mistral and Llama2 LLMs, two prominent language models hailed for their advancements in natural language understanding. Despite Mistral's initial claims of superiority over Llama, this research aims to ascertain whether it genuinely outperforms its predecessor.

This study involves comparing both models using 11 distinct categories of questions or tasks. Initially, simple instructions were provided to assess the models' ability to complete the tasks. In the event of model failure or hallucination, prompt engineering techniques were applied to address the issue. If the model continued to face challenges in completing the task despite prompt engineering interventions, additional experiments were conducted exploring alternative inference parameters.

Results reveal a notable discrepancy between the models' capabilities. Llama exhibited competence across most categories without requiring prompt engineering. In contrast, Mistral consistently demanded prompt engineering and specific inference parameters across all tasks, indicating a higher degree of finetuning for optimal performance.

This comparative analysis sheds light on the nuanced performance dynamics between Mistral and Llama LLMs, offering valuable insights into their respective strengths, limitations, and applicability across diverse use cases.

The categories analyzed were:

1. Knowledge-based questions.
2. Reasoning questions.
3. Language understanding.
4. Inference questions.
5. Contextual understanding.
6. Creative tasks.
7. Ethical dilemmas.
8. Commonsense reasoning.
9. Translation tasks.
10. Summarization tasks.
11. Evaluation tasks.

Mistral was only able to correctly solve without any further instruction the categories 5, and 6.

Llama2 was able to correctly solve without any further instruction the categories 1, 2, 3, 4, 5, 6, 8, 10, and 11.

Youtube URL, short video: <https://youtu.be/c-0x3z-bO2M>

Youtube URL, full video: <https://youtu.be/QU1ZPKXKtM4>