Technical Report: Retrieval Augmented Generation

Kim Nguyen

October 2, 2024

1 Introduction

This is the technical report of the project: Develop a chatbot that can answer queries related to a document that has not been included in the training data.

2 Results

Below are the results of multiple queries to the chatbot including questions and answers along with the time taken to generate in seconds.

1. **Query 1:**
   - **Question**: *What is novel in the paper Llama 2?*
   - **Answer**: Based on the provided context, what appears to be novel in the paper Llama 2 is:

     1. The open release of Llama 2, a large language model, to encourage responsible AI innovation and draw upon the collective wisdom, diversity, and ingenuity of the community. 2. The Responsible Use Guide provided, which offers guidelines for safe development and deployment of Llama 2, acknowledging potential risks and emphasizing the need for safety testing and tuning in specific applications. 3. The comparison of Llama 2-Chat models with ChatGPT generations, including a human evaluation without system prompts, which shows an increase in win rate for Llama 2-Chat. 4. The competitiveness of the largest Llama 2-Chat model (70B) with ChatGPT, with a win rate of 36%. 5. The exceptional performance of Llama 2-Chat 34B, which achieves an overall win rate of over 75% against equivalently sized Vicuna-33B and Falcon 40B models.

     These aspects may be considered novel contributions in the paper Llama 2.
   - **Time taken**: 67.8 seconds
2. **Query 2:**
   - **Question**: *Can you explain the scaling trends that is mentioned in the paper Llama 2?*
   - **Answer**: Based on the provided context, the paper Llama 2 discusses scaling trends in terms of data and model size for the reward model. The authors fine-tuned different model sizes on an increasing amount of the reward model data collected each week (see Table 26 for details on volume per batch). Figure 6 reports these trends, showing that:

     * Larger models generally obtain higher performance for a similar volume of data.

     * The scaling performance has not yet plateaued given the existing volume of data.

     This indicates that increasing the model size and data volume can lead to improved performance, but the study suggests that there is still room for further scaling.
   - **Time taken**: 51.2 seconds
3. **Query 3:**
   - **Question**: *Can you explain the rejection sampling strategy that is mentioned in the paper Llama 2?*

- **Answer**: Based on the provided context, the paper Llama 2 describes a rejection sampling strategy used for fine-tuning and exploring the model's output space. Here's a summary:

  * For each prompt, the model generates K outputs (samples) and scores each sample using a reward model.

  * The best sample (output) is selected based on the reward score.

  * This selected output is then used as the starting point for a gradient update, rather than simply re-ranking the original samples.

  * The authors use this rejection sampling strategy to collect a new dataset before applying fine-tuning, similar to SFT (Self-Finetuning). However, iterative model updates were applied, which reduced the differences between Rejection Sampling and PPO (Proximal Policy Optimization).

  Additionally, Figure 8 illustrates the impact of temperature on the rejection sampling strategy for Llama 2-Chat-SFT (left) and Llama 2-Chat-RLHF (right). The optimal temperature is found to vary depending on the number of sampled outputs (N) and the specific algorithm used. For Llama 2-Chat-RLHF, the optimal temperature when sampling between 10 and 100 outputs is $T \in [1.2, 1.3]$.

- **Time taken**: 82.3 seconds

4. **Query 4:**

   - **Question**: *How did the authors evaluate the model?*

   - **Answer**: Based on the provided context, the authors evaluated the model using a human-centric benchmark called Agieval, which involves presenting human annotators with prompts and generations from two models (Model A and Model B) side-by-side. The annotators were asked to rate which model is better (helpful while also being safe and honest) on a 7-point scale, with categories such as "A is much better", "About the same", "B is slightly better", and so on.

     Additionally, the authors also used a test set of prompts for both helpfulness and safety, and asked three annotators to judge the quality of the answers based on a 7-point Likert scale (the higher the better). They observed that the reward models were well-calibrated with human preference, enabling them to use the reward as a point-wise metric.

     Furthermore, the authors validated major model versions with human evaluations, which involved comparing model responses and presenting them to GPT-4 in a randomized order to alleviate bias.

   - **Time taken**: 78.2 seconds

5. **Query 5:**

   - **Question**: *Do you know about the Demographic representations in the training data?*

   - **Answer**: Based on the provided context, I can answer the question about Demographic representations in the training data as follows:

     1. The authors analyzed the representation of different demographic groups in the pretraining data by measuring the rates of usage of demographic identity terms from the HolisticBias dataset (Smith et al., 2022).

     2. They computed frequencies for each descriptor term in the pretraining corpus, grouping descriptors into 5 axes: Religion, Gender and Sex, Nationality, Race and Ethnicity, and Sexual Orientation.

     3. Table 9b shows the top 5 terms in each demographic axis, along with the percentage of all documents that mention any of the descriptor terms in that axis. The percentages listed for each demographic descriptor represent the proportion of documents mentioning that specific term among those that mention any term in the given axis.

     From this analysis, I can summarize the demographic representations in the training data as follows:

     * There is a potential skew in the representation of Western demographics.

* The top 5 terms in the Gender and Sex axis show a higher representation of "He" pronouns compared to "She" pronouns, which could lead to potential biases in model generations.

I do not have information about the representation of other demographic groups, such as Nationality, Race and Ethnicity, and Sexual Orientation, beyond the top 5 terms listed in each axis.

In summary, while the analysis provides some insight into the demographic representation in the pretraining data, further examination of specific groups and potential biases may be needed to fully understand the impact on model performance and safety.

- **Time taken**: 86.8 seconds

6. **Query 6:**

- **Question**: *"What does it mean by Safety Data Scaling?*
- **Answer**: Based on the provided context, Safety Data Scaling refers to the process of adjusting the amount of safety training data used during the RLHF (Reinforcement Learning from Human Feedback) stage of training Large Language Models (LLMs).

  Specifically, it involves gradually increasing the amount of safety data while keeping the helpfulness training data constant (0.9M samples). This allows the researchers to investigate how the addition of safety training data affects the model's performance, particularly in terms of helpfulness.

  In this context, Safety Data Scaling means modifying the proportion of safety data in the training dataset to evaluate its impact on the model's behavior and output. The goal is to understand how more safety data can influence the model's ability to respond safely to prompts, without compromising its helpfulness.

  By scaling the safety data, the researchers aim to observe trends and patterns in the model's behavior, such as:

  1. Changes in the mean safety RM score as a function of increased safety data.

  2. Shifts in the distribution of safety RM scores, including a reduction in the number of unsafe responses (left tail) as more safety data is added.

  By analyzing these trends, the researchers hope to identify the optimal amount of safety data required to achieve a balance between helpfulness and safety in the model's output.
- **Time taken**: 75.5 seconds

7. **Query 7:**

- **Question**: *""What is safety RLHF? and what is the difference compared to regular RLHF?*
- **Answer**: Based on the provided context information, I'll break down the concept of safety RLHF and its differences with regular RLHF:

  1. **Regular RLHF**: RLHF (Reinforcement Learning with Human Feedback) is a training method that uses human preference data to guide the model's learning process. In regular RLHF, the model is fine-tuned to align with human preferences for both helpfulness and quality of responses. The training process includes gathering human preference data, constructing a reward model, and then using this reward model to train the model in an adversarial setting (Section 3.2.2).

  2. **Safety RLHF**: Safety RLHF is an extension of regular RLHF, specifically designed to focus on safety. It incorporates safety considerations into the RLHF pipeline, aiming to improve the model's ability to generate safe and appropriate responses. The key differences from regular RLHF are:

  * **Safety-specific reward model**: Safety RLHF introduces a separate safety-specific reward model, in addition to the regular helpfulness reward model. This new model is trained to evaluate the safety of generated responses and provides an additional signal for the model to follow during training (Section 3.2.2).

  * **Adversarial prompts for safety**: Safety RLHF involves gathering more challenging adversarial prompts designed to test the model's safety capabilities (Section 3.2.2). These

prompts are intended to simulate real-world scenarios where the model might be faced with unsafe or offensive input.

3. **Impact on model behavior**: By incorporating safety considerations, Safety RLHF aims to improve the model's recall of unsafe responses, increase its accuracy on safety-related subcategories, and make it more robust to "jailbreak" attempts (Bai et al., 2022a). While regular RLHF focuses on overall helpfulness and quality, Safety RLHF adds an additional layer of emphasis on safety.

In summary, Safety RLHF is a variant of RLHF that prioritizes safety by introducing a safety-specific reward model and adversarial prompts to test the model's ability to generate safe responses. This extension does not replace regular RLHF but rather supplements it to ensure the model's overall safety and appropriateness, while preserving its helpfulness capabilities.

- **Time taken**: 115.5 seconds