Reinforcement Learning for Generative AI Large Language Models

Data Science Institute
COLUMBIA UNIVERSITY

Yanni Chen, Junyuan Huang, Yi Lu, Xiaolin Sima, Michelle Sun
Satish Banka, Subha Seshagiri

Data Science Capstone Project with Accenture

Introduction

We want to unlock significant insights into the Oil and Gas Industry through the use of generative AI. Company documentations such as annual reports and sustainability reports can range from 50 to 500 pages, and often contain large amount of numeric and textual information. With the help of Large Language Models (LLMs), the process of understanding how the company performs can be really efficient.

Prompt Optimizer (Black-Box Prompt Optimization)

The BPO algorithm operates by enabling the model to optimize the prompt itself by learning what is a good prompt that can produce ideal result. In our approach, we utilized our dataset to fine-tune the THUDM/BPO model, which is based on Llama2. A comparison of outputs pre and post fine-tuning revealed a consistent improvement, as evidenced by higher average Rouge and Bleu scores in the generated results.

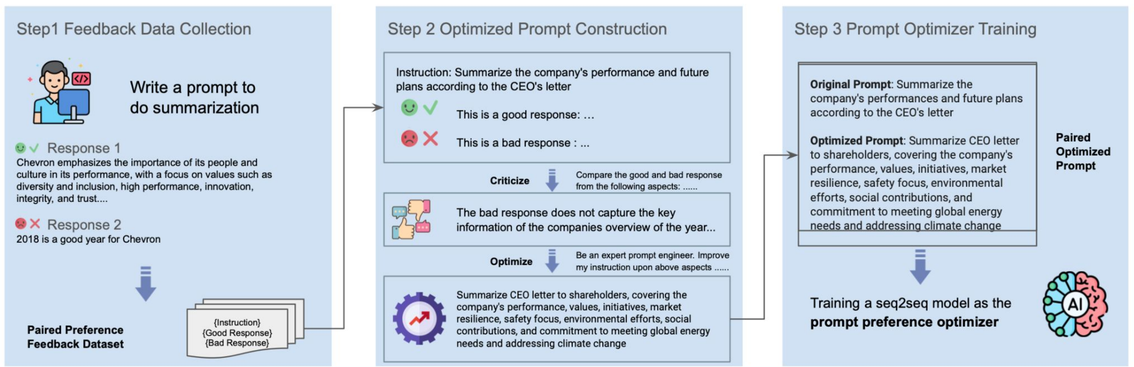

Figure 1. Alignment Stage of Black-Box Prompt Optimization

Report Analyzer (Q&A and Summarization)

Our aim is to comprehensively analyze a targeted company's report, focusing on two key aspects: summarizing annual highlights and extracting major KPIs. To accomplish this, we fine-tuned the Llama2 model using a custom dataset that includes informative summarizations and accurate KPI extractions. Specifically, the model adeptly identifies the CEO report—a key source of annual highlights—and produces concise and insightful summarizations. In terms of Q&A, the Rouge score increased by 20% after fine-tuning.

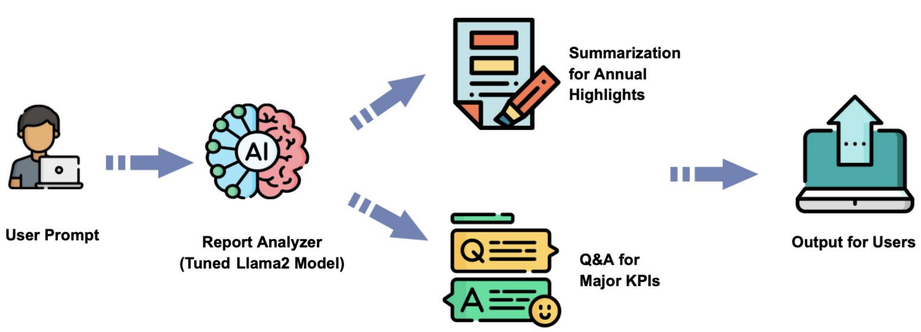

Figure 2. Report Analyzer Workflow

Streamlit Web App

As a final deliverable, we deploy our Prompt Optimizer and Report Analyzer into a web app through Streamlit. On the left-hand side bar, users are able to drop their files and specify which model they would like to use. On the right-hand side, users can enter their message and start a conversation with the chatbot.

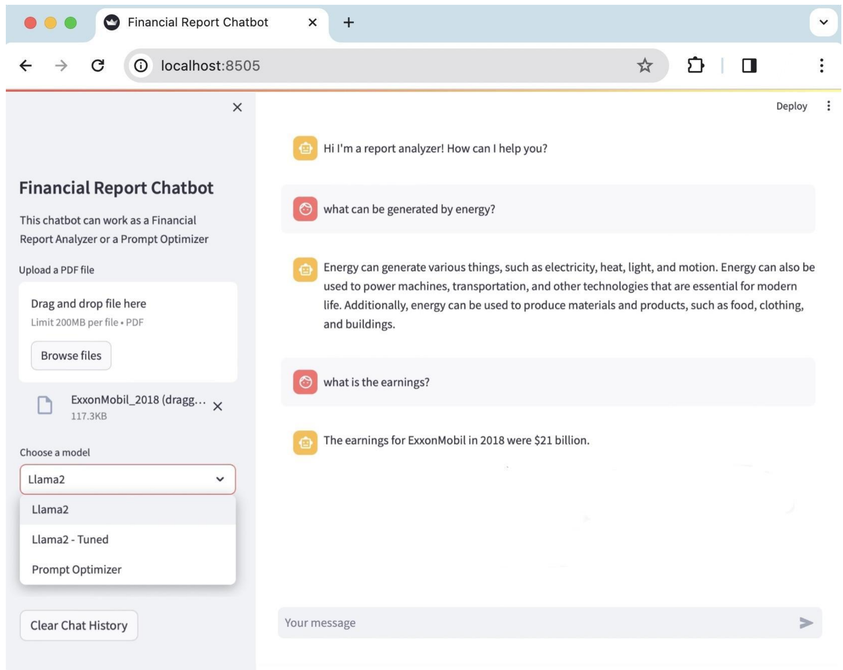

Figure 3. Financial Report Chatbot Web App interface through Streamlit

Achievement & Recommendation

Our report analyzer system, comprising a prompt optimizer and a report analyzer, serves as a compelling proof of concept for the effective application of large language models in the consulting industry, particularly within the field of report analysis. For the next step, utilizing a larger dataset for fine-tuning the model and designing a more comprehensive rewarding system for reinforcement learning during the training process can be considered.

Acknowledgments

We would like to acknowledge our industry mentors Satish Bank and Subha Seshagiri, and our course instructors Vivian Zhang and Ju-Chin Chao, who contributed to and ensured the progress of this project.

References

Cheng, Jiale, et al. Black-Box Prompt Optimization: Aligning Large Language Models without Model Training. arXiv:2311.04155, arXiv, 7 Nov. 2023. arXiv.org, http://arxiv.org/abs/2311.04155.

Nantasenamat, Chanin. “How to Build a Llama 2 Chatbot.” Streamlit, 31 Aug. 2023, blog.streamlit.io/how-to-build-a-llama-2-chatbot.