ID: TP55

Primary Contact:
Muhammad Anwar
Pickering

All Authors:
Muhammad Anwar **(Primary Presenter)**
Mishca De Costa, Ontario Power Generation
Dan Lau, Ontario Power Generation
Issam Hammad, Dalhousie University

Title:
Evaluating ChatGPT on Nuclear Domain-specific data

Abstract:

Evaluating ChatGPT on Nuclear Domain-specific data

M. Anwar, M. de Costa and D. Lau

Enterprise Digital Technology, Digital Technology and Services, Ontario Power Generation, Pickering, Ontario, Canada

Department of Engineering Mathematics and Internetworking, Faculty of Engineering, Dalhousie University, Halifax, Nova Scotia

muhammad.anwar@opg.com; mishca.decosta@opg.com; daniel.lau@opg.com

I. Hammad

Department of Engineering Mathematics and Internetworking, Faculty of Engineering, Dalhousie University, Halifax, Nova Scotia

issam.hammad@dal.ca

Abstract

This paper examines the application of ChatGPT, a large language model (LLM), for question-and-answer (Q&A) tasks in the highly specialized field of nuclear data. The primary focus is on evaluating ChatGPT's performance on a curated test dataset, comparing the outcomes of a standalone LLM with those generated through a Retrieval Augmented Generation (RAG) approach. LLMs, despite their recent advancements, are prone to generating incorrect or 'hallucinated' information, which is a significant limitation in applications requiring high accuracy and reliability. This study explores the potential of utilizing RAG in LLMs, a method that integrates external knowledge bases and sophisticated retrieval techniques to enhance the accuracy and relevance of generated outputs.

In this context, the paper evaluates ChatGPT's ability to answer domain-specific questions, employing two methodologies: A) direct response from the LLM, and B) response from the LLM within a RAG framework. The effectiveness of these methods is assessed through a dual mechanism of human and LLM evaluation, scoring the responses for accuracy and relevance. The findings underscore the improvement in performance when incorporating a RAG pipeline in an LLM, particularly in generating more accurate and contextually appropriate responses for nuclear domain-specific queries. Additionally, the paper highlights alternative approaches to further refine and improve the quality of answers in such specialized domains.

Keywords: ChatGPT, nuclear domain data, large language model (LLM), question-and-answer (Q&A), performance evaluation, Retrieval Augmented Generation (RAG), external knowledge bases, retrieval techniques, LLM evaluation, contextually appropriate responses

Keywords:
ChatGPT, nuclear domain data, large language model (LLM), question-and-answer (Q&A), performance evaluation, Retrieval Augmented Generation (RAG), external knowledge bases, retrieval techniques, LLM evaluation, contextually appropriate responses

Track Assigned:
Enhancing Nuclear Safety and Security

Final Paper:
View File - FinalPaper_TP55_0520021432.pdf

Draft Paper:
View File - DraftPaper_TP55_0303104829.docx

Upload Presentation:
View File - UploadPresentation_TP55_0605102450.ppt