OKLLM: Online Knowledge Search for LLM Innovations

Huan Chen
Insight Centre for Data Analytics,
University of Galway,
Galway, Ireland
huan.chen@universityofgalway.ie

Andy Donald
Insight Centre for Data Analytics,
University of Galway,
Galway, Ireland
andy.donald@universityofgalway.ie

***Abstract*— Nowadays, a breakthrough era in online content discovery and search is being ushered in by the increased availability of Large Language Models (LLMs). However, LLMs are resource and computational-intensive. Additionally, Artificial Intelligence (AI) generated material can occasionally contain bias or false information. With our proposed framework, Online Knowledge Search for Large Language Models (OKLLM), we seek to fill these gaps through the use of knowledge distillation, knowledge graph generation and verification, bias detection, and transfer learning via the development of three distinct components. The first component will concentrate on employing knowledge distillation to handle bias detection tasks in order to enhance search results and lessen computationally taxing tasks. The usage of the knowledge graph to solve the hallucination phenomenon to improve search results will be the focus of the second component, and the third component will make it possible to handle the explainability challenge by utilizing information such as the path gathered from the knowledge graph and visualizing it, thus enhancing the search results output. The intention is to present these components using open-source principles.**

Keywords - Ethical AI, LLMs, Knowledge Graph Generation, eXAI, Bias Detection, Transfer Learning.

I. INTRODUCTION

Large Language Models are resource and computational-intensive, which restricts their use and applicability in some circumstances [1]. Additionally, Artificial Intelligence-generated material can occasionally contain bias or false information [2]. In this research, we seek to fill these gaps. The proposed work focuses on addressing research challenges around computational demands, hallucination, and explainability of large language models in online content discovery through knowledge distillation, knowledge graph generation and verification, bias detection, and transfer learning techniques. We will look at embedding political misinformation or bias detection as a domain focus for the project.

The rest of the paper is structured as follows. In Section 2, we will introduce the proposed methodology for construction of the OKLLM framework. Section 3 describes the various datasets that we are proposing to be used as part of the OKLLM development. Section 4 details the associated projects that we will leverage as part of the research. Section 5 will lay out future work that the research will bring and, finally, we conclude in Section 6.

II. PROPOSED RESEARCH METHODOLOGY

The proposed work is comprised of three primary, individually deployable components. They are designed so that each can be interacted with in isolation or within a constructed pipeline. The first component concentrates on employing knowledge distillation to handle bias detection tasks [3] in order to enhance search results and lessen computationally taxing procedures.

The usage of the knowledge graph to solve the hallucination phenomenon [4] to improve search results is the focus of the second component. This component utilizes the knowledge extraction framework Saffron [5] to automatically extract entities and generate the knowledge graph. The third component will make it possible to handle the explainability challenge by utilizing the path information gathered from the knowledge graph and visualizing it. Figure 1 describes the components and how they will interact with each other whilst also describing in more detail the sub-components within the full high-level architecture.

For the first component, we are going to employ the pre-trained language model Bidirectional Encoder Representations from Transformers (BERT) and Large Language Model Meta Artificial Intelligence (LLaMa) as the teacher model to perform the task of bias detection, and for the student model, Distilled Bidirectional Encoder Representations from Transformers (DistillBERT) and logistic regression will be used to predict bias.

The second component consists of two phases, the first of which uses Saffron to generate a knowledge graph. The second stage in addressing the hallucinatory phenomena will be Knowledge Graph (KG) verification. Verification of entities and relationships is required for this phase as well as the construction of an evidence-based knowledge graph. The output of this step will be a Resource Description Framework (RDF), which will be applied as the first complement to find the bias once more. The final RDF will be used for the third component in terms of visualization and explainability.

The third element will focus on visualising the KG, thus improving the explainability of the search results via the KG path traversal. In this step, Neo4j will be utilised.

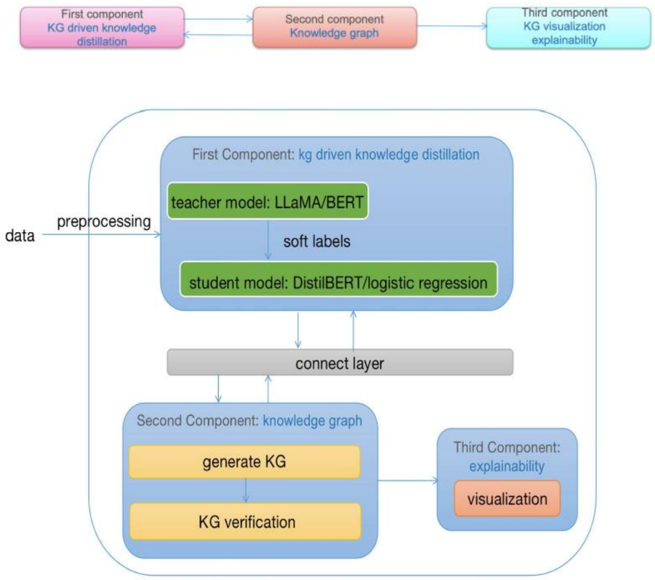
First component
KG driven knowledge distillation
Second component
Knowledge graph
Third component
KG visualization explainability
data
preprocessing
First Component: kg driven knowledge distillation
teacher model: LLaMA/BERT
soft labels
student model: DistilBERT/logistic regression
connect layer
Second Component: knowledge graph
generate KG
KG verification
Third Component: explainability
visualization

Figure 1. OKLLM Component interactions & architecture.

III. Datasets

Dataset selection and curation is a critical part of the development of the OKLLM framework. We have identified and utilised two initial datasets including the Search Engine Result Pages (SERP) [7] and Microsoft MAchine Reading COmprehension (MS MARCO) [6] passage ranking datasets.

A. *Search Engine Result Pages (SERP)*

Search Engine Result Pages (SERP) Datasets: The term "SERP Data" refers to information and data gathered from search engine results pages (SERPs), which may contain details about a website's position in search results, the number of searches for keywords, and other Search Engine Optimization (SEO) related metrics.

B. *MS MARCO Passage Ranking Dataset*

MS MARCO Passage Ranking Dataset: A popular dataset in the fields of Natural Language Processing (NLP) and information retrieval, the Microsoft Machine Reading Comprehension (MS MARCO) dataset's purpose is to rank passages (short text excerpts) in response to a query.

A particularly important task will be to enable a preprocessing of the data to allow identification of text relevant to the politics domain. This will allow us to tailor the project to focus specifically on this domain.

IV. Related Projects

This project is connected to some previous studies that we conducted. First, Saffron is an extremely configurable open-source program that extracts knowledge from structured and unstructured text using natural language processing. Saffron will produce the first iteration of the knowledge graph in this suggested endeavor. Second, we have identified the shortcomings in the available tools and methods for bias detection in the Customer Interaction Data project [3]. Thirdly, the website Practice Ecosystem for Standards (PEERS); PEERS is an EU Horizon project which aims to produce a knowledge-based repository detailing standards and connecting experts within the Chemical, Biological, Radiological, Nuclear, and high yield Explosives (CBRN-E) domain. In the future, this proposed work will be integrated into this knowledge base website. Figure 2 describes the related projects which will contribute to the OKLLM project. Saffron is released under Apache 2.0 license. All contributions made to Saffron code as part of this project will be distributed under the same license at the time of a new software release. Other components implementation will be released as open-source software under an Apache 2.0 license contingent.

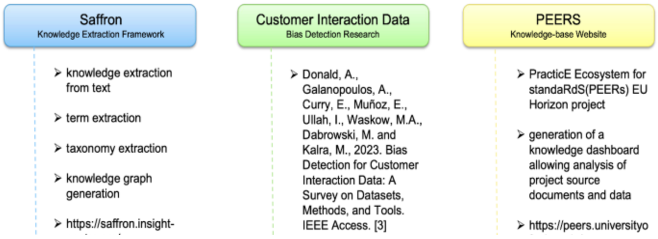
Saffron
Knowledge Extraction Framework
➢ knowledge extraction from text
➢ term extraction
➢ taxonomy extraction
➢ knowledge graph generation
➢ https://saffron.insight-
Customer Interaction Data
Bias Detection Research
➢ Donald, A., Galanopoulos, A., Curry, E., Muñoz, E., Ullah, I., Waskow, M.A., Dabrowski, M. and Kalra, M., 2023. Bias Detection for Customer Interaction Data: A Survey on Datasets, Methods, and Tools. IEEE Access. [3]
PEERS
Knowledge-base Website
➢ PracticE Ecosystem for standaRdS(PEERs) EU Horizon project
➢ generation of a knowledge dashboard allowing analysis of project source documents and data
➢ https://peers.universityo

Figure 2. Related projects.

V. Conclusions

This paper introduces the core concepts behind the OKLLM project by detailing the approaches that will be taken to address the identified gaps in enterprise level large language models. In addition, detail has been provided as to the methods that will be utilised in the development of the OKLLM framework, including datasets and related projects.

VI. Future Work

The next steps for the OKLLM framework is to make an initial open source version available to support more implementations of domain specific deployments to support the various different hypotheses around the large language model gaps that we have identified. In particular, the focus on identified types of bias detection will feature heavily in future work.

Acknowledgment

This project is funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them. It is funded within the framework of the NGI Search project under grant agreement No 101069364 and the project is run at the Insight SFI Research Centre at the University of Galway.

REFERENCES

[1] C. Kachris, “A Survey on Hardware Accelerators for Large Language Models” in arXiv preprint, arXiv:2401.09890, 2024.

[2] J. Zybaczynska, M. Norris, S. Modi, J. Brennan, P. Jhaveri, T.J. Craig, and T. Al-Shaikhly, “Artificial Intelligence–Generated Scientific Literature: A Critical Appraisal.” in The Journal of Allergy and Clinical Immunology: In Practice, 12(1), pp.106-110, 2024.

[3] A. Donald *et al*., "Bias Detection for Customer Interaction Data: A Survey on Datasets, Methods, and Tools," in *IEEE Access*, vol. 11, pp. 53703-53715, 2023, doi: 10.1109/ACCESS.2023.3276757.

[4] S. Athaluri, V. Manthena, M. Kesapragada, V. Yarlagadda, T. Dave, and S. Duddumpudi, “Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References” in Cureus, vol. 15, 10.7759/cureus.37432, 2023.

[5] J. P. McCrae, P. Mohanty, S. Narayanan, B. Pereira, P. Buitelaar, S. Karmakar, and R. Sarkar, “Conversation Concepts: Understanding Topics and Building Taxonomies for Financial Services” in *Information*, vol. 12, pp. 160, 2021.

[6] D. F. Campos *et al*., "MS MARCO: A Human Generated Machine Reading Comprehension Dataset" *arXiv preprint arXiv:*611.09268, 2016.

[7] N. Höchstötter and D. Lewandowski, What Users See - Structures in Search Engine Results Pages. Information Sciences. 179. 1796-1812. 10.1016/j.ins.2009.01.028, 2009.