Large Language Models and Knowledge Graphs: Opportunities and Challenges

Khalil Ibrahim

April 22, 2024

Abstract

This bachelor's thesis explores the intersection of Large Language Models (LLMs) and Knowledge Graphs (KGs), two of the most promising technologies in the field of artificial intelligence. Guided by the seminal insights from the paper *'Large Language Models and Knowledge Graphs: Opportunities and Challenges,'* this work aims to uncover how the sophisticated linguistic capabilities of LLMs and the structured knowledge representation of KGs can be combined to push the boundaries of AI.

Part I

1 Introduction

The way we historically shared knowledge has evolved from oral traditions to the written word, and now, to the digital systems where it is based on complex algorithms and data structures. Systems like ELIZA in the 1960s marked the beginning of machine interaction with human language, a field that has expanded rapidly with the creation of advanced Large Language Models (LLMs) like GPT and BERT.

At the same time as the progress in natural language processing, Knowledge Graphs (KGs) have also become very important for organizing large amounts of data in a very user-friendly and intuitive manner. Later, concepts like the semantic web in the early 2000s utilized technologies such as RDF (Resource Description Framework) and OWL (Web Ontology Language). KGs structured data in a manner aligned with human thought processes. This alignment enabled machines to perform complex reasoning and inferences, based on the relationships and attributes defined within the graphs. In other words, KGs has provided a framework for organizing information in a way that mimics human understanding, enabling machines to reason, infer, and even learn from the interconnected web of data.

2 Explicit and Parametric Knowledge

As we begin to explore the complex relationship between KGs and LLMs, it is important to establish a foundational understanding of the specific types of knowledge these technologies use. Understanding these specific types is important because they form an important foundation for further research and deepen the insights presented in this paper Pan et al. (2023a).

Explicit knowledge, like KGs, is informations and knowledge that is documented, structured and accessible in a format way that both humans and machines can understand. It is often represented through entities, relationships, and attributes that are organized in a graph structure. It includes facts, rules, and relationships that are directly stated, making it possible for AI-systems to perform logical reasoning and decision-making based on this well-defined and saved information.

Parametric knowledge, on the other hand, is knowledge that is explicitly stored within the parameters (like weights) of a model during training fase. In LLMs, this knowledge is embedded in the vast network of connections adjusted during the training process on a wide array of text data. This form of knowledge is implicit and is a part of model's architecture, enabling the model to make predictions or generate content based on the pattern and relationship that reflects the learned information.

3 A Contemporary Debate in Knowledge Computing

The intersection of parametric and explicit knowledge forms a central debate within the Knowledge Computing community, engaging a range of perspectives on the efficacy and feasibility of this integration. Here, we summarize the critical points of agreement and dispute shaping this dialogue:

- *Knowledge Representation and Reasoning*: KGs provide structured and precise representation of knowledge that enables logical reasoning. LLMs, however, excel in language understanding and data generalization, but may not be as efficient in logical reasoning due to their reliance on statistical patterns.
- *High Precision Methods*: The success of KGs is largely due to their high accuracy in providing detailed information about entities, such as YAGO's claimed accuracy rate of over 95 %. However, LLM-based methods for KG enhancement often struggle to reach such high levels of precision, particularly in tasks that require factual accuracy.
- *Numerical Values*: LLMs face problems with numerical values, including simple arithmetic and KG completion tasks entailing numerical facts. This limitation questions the current LLMs capability to precisely compute numbers during pretraining to use them in KG completion effectively.
- *Long-tail Knowledge*: LLMs typically struggle with recalling information about less common "long-tail" entities or knowledge, as their performance

decreases with unknown and unclear facts not frequently present in training datasets. In contrast, KGs inherently provide detailed knowledge about these long-tail entities, which could help enhance LLMs performance in knowledge tasks.

- *Bias, Fairness, and Beyond*: One primary concern remains that LLMs may learn biases present in their training data, which could lead to biased outputs. KGs also face bias challenges, especially in manual and automated decisions about which facts to include and exclude and how to represent them. In addition, LLMs also face challenges with copyright violation and misinformation.

- *Explainability and Interpretability*: KGs are preferred for tasks requiring clear explanations, due to their transparent structure that clearly depicts relationships between entities. Conversely, LLMs are criticized for their lack of transparency in how decisions are made. Efforts to improve LLM interpretability include research on attention mechanisms and model introspection. Techniques such as the Chain-of-Thought (CoT) are also being developed to clarify LLM operations, although challenges remain in dissecting complex queries and providing accurate responses.

4 Parametric knowledge and the new opportunities

One of the central questions this paper (Pan et al. (2023a)) sought to answer is, what new opportunities does the emergence of parametric knowledge bring, especially with today's explicit knowledge? In the following, we will explore how the combination of parametric knowledge with explicit knowledge could change the way we share and use knowledge in a new way.

Big language models (LLMs) give us quick access to huge amounts of text on demand. This makes it easier for AI developers to work with data without worrying about collecting, preparing, storing, or searching through it on a large scale. It cuts down on the need for traditional information retrieval methods, especially when integrated with technologies like Retrieval-Augmented Generation (RAG). LLMs also simplify the handling of tasks such as sentence parsing, name or thing recognition, and relationships between entities. With structured knowledge and LLMs, we can utilize knowledge optimally and effectively in doing specific tasks and supporting the idea that "knowledge is power".

LLMs have already made big improvements in how well computers and models can understand language. This includes figuring out if one sentence follows from another, summarizing texts, spotting paraphrases, and more. By combining parametric knowledge with explicit knowledge, it could help to deal with the challenges of varying language, typos, and redundant or messy text.

5 Key Research Themes and Their Challenges in Enhancing Knowledge Graphs with LLMs

The following sections is aimed to elaborate on how the synergy of parametric and explicit knowledge is outlining new pathways in terms of knowledge development for sharing and application. I will focus on four critical research areas, with an overview of the recent innovations realized and the challenges posed.

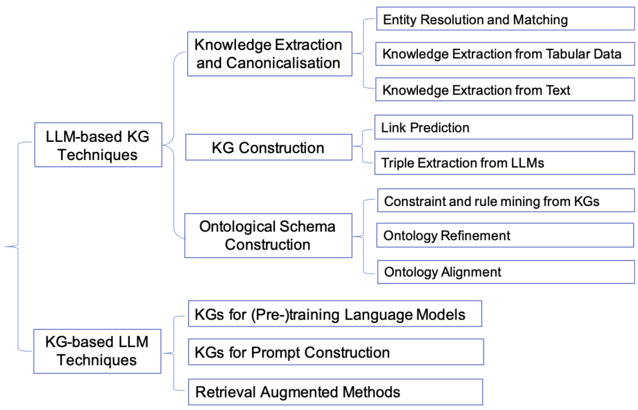

Figure 1: Pan et al. (2023b)

5.1 Extracting and Standardizing Knowledge

KGs construction is a task that presents significant challenges as it involves the massive fusion of data from different sources, while LLMs when trained over a vast information spectrum from all sources, perform knowledge extraction tasks very well.

5.1.1 Entity Resolution and Matching

Entity resolution aims at identifying and linking instance data that refer to the same real-world entity from different datasets. It generally focuses on flat structured data, although research is relatively new for entity resolution in semi-structured data in KGs. Entity resolution approaches can be broadly classified into two main categories: general and embedding-based. The general ones usually adopt graph neural networks to consider a wide range of entity types. While the embedding-based ones transform entity similarities into vector spaces to mitigate the impact of heterogeneity and enable reasoning. LLMs will also help deal with the problem of providing an efficient means of automation for the creation of labeled samples for KGs, and the development of entity-matching rules given a logical development framework. However, practical prompt engineering is critical to deriving valuable outputs in the context of large-scale KGs, such as DBpedia or Wikidata.

5.1.2 Knowledge Extraction from Tabular Data

Building KGs now involves gathering knowledge from databases, web tables, and CSV files. When we understand the semantics, the structures and how data lies underneath these tables. We can set rules to convert the same into valuable facts for KGs. But this is hard with the tables in the wild. This may sound somewhat imprecise in semantics, but quite essential details, for instance, the name of the table and the title heads of columns, are not well defined. There is much groundwork before pulling out the knowledge needed: the data needs to be found, collected, made sense of, the pieces integrated, and the information refined.

5.1.3 Knowledge Extraction from Text

Text knowledge extraction is a crucial step in the KG construction process that brings to use the ability of LLMs to allow converting raw text into structured, actionable knowledge. In this way, LLM identifies the entities, their relationships, and the categories of named entities by extracting relations and events from text without needing for domain-specific training. It also works very well for generating high-quality synthetic data to fine-tune models for specific tasks, helping overcome the challenge of limited training data. However, there are still many other challenges in processing long documents and achieving comprehensive information extraction. Future efforts will focus on enhancing extraction from long texts and optimizing the balance between precision and recall, crucial for more effective knowledge extraction systems.

5.2 Bulding Knowledge Graphs

5.2.1 Link Prediction

Link prediction is the function of predicting triples given the other two elements. In this case, there is a predicting head entity (?, r, t), relation prediction (h, ?, t), and tail prediction (h, r, ?). Traditional approaches to link prediction view the KG as a static snapshot, assuming a training phase wherein the current state of the KG is utilized to learn embeddings and other model parameters. Most of these, however, struggle with predicting unseen entities, such as newly added people or products, simply because they can only operate on entities for which an embedding was learned during the training phase. The inductive link prediction (ILP), on the other hand, goes a step further to predict links for brand-new entities not previously included in the KG. They are the key to keeping KG updated, as new information will be inserted, particularly in the case of an entity that did not exist in the original dataset. The problem always has been, *"How do you integrate and make sense out of a huge amount of textual data which describes these entities and their relationships?"*. And that's where LLMs provide assistance. By extracting initial embeddings from the textual descriptions of the entities and their relationships, language models such as BERT and SBERT can generate rich and more descriptive representations of the entities. Then, these embeddings have been used to predict new links more accurately, even for the entities newly added to the graph.

5.2.2 Challenges and Opportunities in Link Prediction

Meanwhile, models like BERT, GPT for KG enhancement, LLM, and Inductive Link Prediction (ILP) pose a list of challenges and opportunities such that they:

- LLMs sometimes struggle with the outcomes that they generate in trying to fit them into the correct KG entry when the entity has more than one name.
- First, it may be challenging to determine in the first place if prior knowledge learned by the model informs the predictions by LLM or if they are inferred directly from KG.
- While the LLMs enrich ILP with vast knowledge, the curriculum depends so much on extras that may not always be available.
- Effective use of the LLMs for KG tasks would further call for the process of well-crafted prompts in a key and resource-intensive way.

5.2.3 Triple Extraction from LLMs

Instead of manually building KGs, we're now using LLMs like BERT and GPT to automatically find and understand relationships in data and build it in KGs. But at the same time, LLMs struggle with accurately retrieving rare or complex information. Tools like LAMA have tried to measure how well LLMs do this, often finding they can't yet match the depth of traditional knowledge bases. Furthermore, these models tends sometimes to just memorize data rather than truly understanding it, which leads to biased knowledge representations based on the predispositions in their training data.

Thus, the use of Prompt Engineering and effective crafting of prompts improves the accuracy of the same, giving useful answers from LLMs. Now, the task is to be clear and unbiased yet guide the LLMs through more in-depth insights without being perceived as leading them astray. Also, to get better answers from LLMs, like GPT, we need to ask the right questions or give right prompts. This task is problematic, for the reason that by the way we ask, the quality of the answers that we can get can be significantly influenced.

5.2.4 Challenges and Opportunities in LLMs Triple Extraction

While LLMs can find and use this relational knowledge, it struggle at the same time with the distinction of the types of knowledge (like commonsense vs. factual) and relation patterns. This remains a question of whether LLMs can even infer new knowledge beyond what they've seen during training. Therefore, the challenges presented in section 3 remain relevant [1].

5.3 Ontological Schema Construction

Due to the sheer volume of data within KGs, manually inspecting and rectifying errors and update knowledge and data in it is costly and complex process. Even though domain experts and well-versed with data, are likely to have skills

[1] A Contemporary Debate in Knowledge Computing

or resources shortcomings for efficiently formalizing these rules. Therefore, we rely on rules and constraints that can be automatically applied to ensure data accuracy and consistency. Such rules would establish the condition and dependency that KG has to go on, making sure that adding or removing facts is not over these guidelines. In KGs, such rules and constraints could be in the form of Graph Functional Dependencies, declarative first-order logic rules, or validating shapes. Once these rules are established, the next steps for LLMs will be to identify any entities or facts that violate the rules, eliminate erroneous information, and infer missing details. Subsequently, LLMs will:

1. Use inductive reasoning to generate new rules by inferring patterns and relationships from the data they learn, thereby proposing new logical rules not explicitly given.
2. Understand words in context to handle synonyms and polysemy (words with multiple meanings), which is vital for reconciling rules that may use different terms to express the same concept.
3. Provide rules or generate validating examples that confirm the rules and counter-examples that demonstrate deviations from these rules.

For example, a company's internal documents might outline specific protocols that define relationships between entities such as departments, employees, and projects. An LLM can process the natural language in these documents and translate it into formal constraints that can be applied to a KG to ensure that the graph accurately reflects the company's operational structure and rules. And now, a new question brings whether it is possible to train LLMs to treat the task of rule generation as a summarization task. This would require then the ability to perform both inductive and abductive reasoning and treat rules as summaries from the set of facts in the KG.

▶ **Example 2.** We ask ChatGPT "Please list three rules that could help to infer one person's father. And output it in a first-order rule form." The following is the output we get:

"Certainly! Here are three rules expressed in first-order logic form that could help infer a person's father:

Rule 1: $\forall x, y\ (\texttt{child}(x, y) \wedge \texttt{male}(y) \rightarrow \texttt{father}(y, x))$ This rule states that for all individuals x and y, if there exists a 'child' relationship between x and y and y is male, then y is the father of x.

Rule 2: $\forall x, y\ (\texttt{biologicalChild}(x, y) \rightarrow \texttt{father}(y, x))$ This rule states that for all individuals x and y, if there exists a 'biologicalChild' relationship between x and y, then y is the father of x.

Rule 3: $\forall x, y\ (\texttt{fatherOf}(y, x) \rightarrow \texttt{father}(y, x))$ This rule states that for all individuals x and y, if there exists a 'fatherOf' relationship between y and x, then y is the father of x.

These rules, expressed in first-order logic, demonstrate how the relationships between individuals can be used to infer the father of a person within a KG."

Figure 2: Pan et al. (2023a)

This example shows that the GPT can be able to produce a valid rule to extract father-child relations from the KG. Rule 1 is appropriate to correctly define the father by the relations with the child and the male. Rule 2 is biased with the biological relationship while inferring a high likelihood of paternity, while Rule 3 correctly parsing "father" for the fatherhood relationship. LLM formulated rules can be accurately and formally put forward, helpful in automating the creation of these rules within KGs. However, for optimal results, integration with the existing data and context within the KG is essential for such rules developed by LLM.

6 Enhancing LLMs with KGs

6.1 Utilizing Knowledge Graphs to Enhance Large Language Models

In the earlier sections, we saw how LLMs can be helpful in KGs. Now, let us detail how KGs can be very useful for LLMs:

1. KGs can be a valuable source of structured information to provide factual and relational data to LLMs during their training phase. This helps improve the understanding of real-world entities and the relationships they hold.

2. In prompt Construction, the triples (entity-relationship-entity) found in KGs can be used to construct prompts for LLMs. These prompts can guide the models to generate more accurate and contextually relevant responses by leveraging the structured knowledge contained within KGs.

3. KGs can further act as an external knowledge base to be referred to by retrieval-augmented language models. These models can query KGs to retrieve, with augmentation, specific and accurate information not included in their pre-trained data.

6.1.1 KGs for "Pre"-Training LLMs

KGs typically use post-processed information from reliable sources, verified through human evaluations. By integrating KGs into the pre-training corpus of LLMs, they help overcome the limitations of relying solely on natural language text. This integration of structured, fact-based knowledge from KGs is crucial bea It ensures the effective use of prompts to inject global knowledge, thereby enhancing the models' capabilities. Additionally, knowledge from high-resource language Knowledge Bases is transferred to models designed for low-resource languages, improving their functionality.

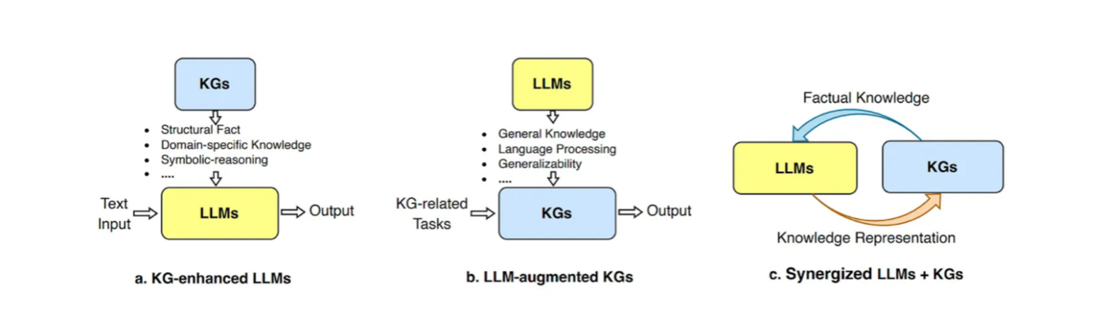

Figure 3: Pan et al. (2023b)

6.1.2 KG for Prompt Construction

Integrating KGs with LLMs in prompt construction further improves LLMs by refining knowledge extraction and quality improvement in the prompt. This will help enhance not only the contextuality and accuracy of the LLM predictions but also introduce a way for developing dynamic, reliable prompts to help improve the interaction of users with the model. With techniques as KnowPrompt that harness KGs for semantic prompt-tuning and bring revolutionary improvements to tasks like relation extraction. This integration, however, faces challenges such as the requirement to adapt dynamically, the relevancy of the prompt, trust based on user feedback, and conducting complex reasoning through prompts.

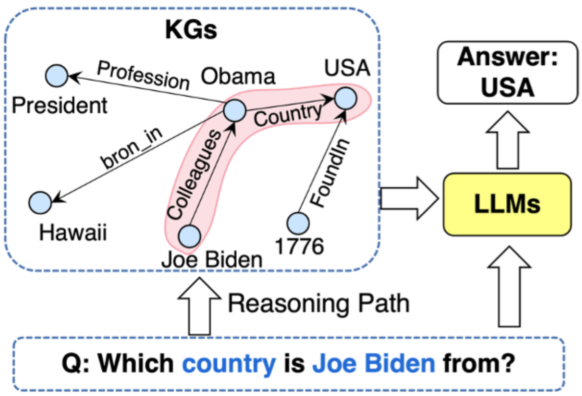

Figure 4: Pan et al. (2023b)

At the same time, such integration does not come without challenges. Developing context-aware prompts that respond to user feedback adds complexity. It makes maintaining content relevance and trustworthiness difficult. Additionally, to achieve KGs' ambition for fostering complex reasoning in LLMs, special attention is needed. The methods used in creating and refining the prompts must be carefully considered. This ensures they effectively reveal the structured knowledge within KGs.

With this, meeting these challenges will open several research avenues. Especially those designed to enhance LLMs functionality through KGs, generating dynamic prompts that could adjust to user interactions, and improving AI-generated information. This will also facilitate advanced reasoning. According to [Pan et al. (2023a)] and [Luo et al. (2023)], this integration offers solutions to ongoing AI challenges such as correcting LLMs' inaccuracies (hallucinations), updating LLMs to have the newest knowledge without retraining, and at the same time embracing the limitation by nature through KG-based prompts.

6.1.3 Retrieval Augmented Methods

Retrieval augmented methods are essential for updating LLMs with the latest information, addressing the knowledge cutoff issue where LLMs lack awareness of recent events. According to [Lewis et al. (2020)] and giving the high costs associated with retraining LLMs with costs of training—such as 4.6 million dollar for GPT-3 with 175 billion parameters, and 17 million dollar for PaLM with 540 billion parameters—restrict frequent updates. Additionally, these models often capture knowledge mainly about popular entities, leaving domain-specific, private, or critical business information inadequately represented. In other word, it means that LLMs can't naturally update their knowledge about recent events or specific domain information after they are trained, due to the high costs of retraining [2].

6.2 Applications

As we discussed before, the integration of KGs and LLMs holds significant potential by merging the precise, explicit knowledge from KGs with the broad understanding capabilities of LLMs. KGs contribute exact knowledge, invaluable in sectors like healthcare, where accuracy is critical. In contrast, LLMs, despite their challenges with factual accuracy and explainability, can enhance KGs by processing unstructured text to overcome their limitations. And the real potential in it, is the synergy has already been applied in healthcare assistants, converting text to triples, support natural language queries, representing graphs as text and question-answering systems, showcasing their complementary advantages.

The digital healthcare industry is one of the most important sector for appying LLMs, focusing on automation of clinical documentation, synthesis of patient histories, and, in general, the selection of candidates for clinical trials. The fact that they are prone to errors and sometimes even mislead information makes it necessary for this kind of data to be handled with care, especially where patient safety is concerned. Integration of KGs would reduce these, to some extent, associated risks by providing accurate, domain-specific knowledge. On the privacy front, LLMs do need to transmit potentially sensitive data, raising questions under regulations such as GDPR. KGs could help with a privacy management plan that includes special rules about data sharing and techniques for deidentification. Another option for enhancing privacy could be using open-source LLMs within the organization. They offer their users control over data, but it would be hard to strengthen sophistication, which may require new resources.

[2] For more details, please read further at Lewis et al. (2020)

Part II

7 Comparsion

In the first part of the thesis, we delved into the synergistic potential of LLMs and KGs and there challenges and oppurtunities. As we now pivot to the next segment, the *"Large Language Models and Knowledge Graphs: Opportunities and Challenges"* (Pan et al. (2023a)) will be referred to as the "first paper,". Similarly, /emph "Unifying Large Language Models and Knowledge Graphs: A Roadmap" (Pan et al. (2023b)) will be termed the "second paper." This distinction aims to provide a structured and focused comparison, highlighting the similarities and differences in their approaches, findings, and contributions to the field.

7.1 Similarities

Similarities in the first and second papers resonate, highlighting the need to integrate LLMs into KGs, which may leap toward developing more sophisticated, human-like artificial intelligence systems. In effect, both papers agree that combining deep linguistic capabilities with structured, explicit knowledge can address long-standing AI challenges such as improving reasoning, factual accuracy, and interpretability of models. Further, both papers utilized extensive literature reviews, case studies, and theoretical analysis to formulate their arguments and support the evidence presented.

7.2 Divergent Perspectives and Contributions

The strength of paper 1 is that it describes the field very well, with much value in terms of community perspectives and development in theory. It addresses broader implications following an integration of LLMs with KGs, making it a must-read for the researcher interested in understanding the field's current state and future directions. The speculative nature of this could limit the applicability for anyone looking for concrete methodological guidance.

The second paper is a more technical and detailed exploration of the algorithmic and computational models that facilitate this integration. It provides concrete tips on how to improve the scalability, efficiency, and accuracy of a system. Its technical focus makes it reasonably necessary for researchers and practitioners interested in putting the integrations into practice. Most importantly, it also includes visual elements and conceptual diagrams rather widely, giving the reader more distinct visual understanding of the concepts and potential implementations of LLM and KG integrations.

When it comes to integration strategies and future directions, and in summary, two scholarly contributions signal a substantial enrichment of the discussion about knowledge graphs. The first paper is concerned with the theoretical challenges and opportunities that come with this integration, while the second paper outlines a practical roadmap toward the achievement of a synergistic blending of KGs and LLMs. While the first paper gives a conceptual base, emphasizing the importance of explicit and parametric knowledge, the second paper provide actionable and coherent plan on how to merge these technologies to foster systems that are not only intelligent but also intuitive and aligned with human reasoning [3].

Part III

8 Conclusions

In conclusion, the paper has elaborated on the possible benefits of integrating knowledge graphs into LLMs for quality improvement, diversity, and ethical integrity of the content developed. The three important research questions, which are very central in the discussion leading to this integrated model. First: How do KGs increase the efficiency of information extraction from LLMs? Second, how can KGs be used in exploring better overreaching the ethical boundaries of AI-generated content? And third, how do KG-based prompt strategies increase reasoning capabilities within LLMs?

The paper also outlines broader research trajectories that could benefit from the synergy between KGs and LLMs. These include using KGs to detect and mitigate hallucinations in LLM outputs, refining LLM knowledge bases to reflect current realities, and overcoming the challenges of injecting knowledge into the 'black-box' nature of LLMs through KG-driven prompts. The combination is likely to open up new avenues for research and application, and not only highlights the potential for significant technological advancements but also underscores the need for careful consideration of ethical and practical implications.

This integration had its challenges. It involved ensuring the relevance and correctness of the data extraction courtesy of the LLMs, the integrity and trustworthiness of the KGs, and managing complexities introduced by the combination of these mighty technologies. These challenges will have to be addressed with innovative solutions enabling the faculty to improve, such as prompting in engineering, enhanced ways of knowledge retrieval, and better ways of maintaining data accuracy and consistency.

8.1 Personal Opinion

I would say that it successfully achieved its goals which was to identify the intersection of LLMs and KGs, highlighting significant synergies and the challenges of integrating these technologies.

But despite its foundational insights, the paper could extend its utility with several refinements. It could benefit from a more granular, step-by-step approach to the methodologies proposed. This would be particularly valuable for practitioners who need precise guidelines to apply the theoretical insights in varied operational contexts. There's a noticeable gap between the theoretical

[3]The entire content of this section (7.2) is based on Pan et al. (2023a) and Pan et al. (2023b)

framework presented and practical deployable strategies.

All in all, *"Large Language Models and Knowledge Graphs: Opportunities and Challenges"* Pan et al. (2023a) is a seminal contribution to the discursive environment concerning the future that AI holds and outlines a course for explorations that can be carried out with the highest rigor in the integration of LLMs and KGs. However, the journey is far from complete. It has left many questions unanswered and opened research lines for future exploration.

References

P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

L. Luo, T.-T. Vu, D. Phung, and G. Haffari. Systematic assessment of factual knowledge in large language models. *arXiv preprint arXiv:2310.11638*, 2023.

J. Z. Pan, S. Razniewski, J.-C. Kalo, S. Singhania, J. Chen, S. Dietze, H. Jabeen, J. Omeliyanenko, W. Zhang, M. Lissandrini, et al. Large language models and knowledge graphs: Opportunities and challenges. *arXiv preprint arXiv:2308.06374*, 2023a.

S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu. Unifying large language models and knowledge graphs: A roadmap. *ArXiv*, abs/2306.08302, 2023b. URL `https://api.semanticscholar.org/CorpusID:259165563`.