Data Management Opportunities in Unifying Large Language Models+Knowledge Graphs

Arijit Khan
Aalborg University
Denmark
arijitk@cs.aau.dk

Tianxing Wu
Southeast University
China
tianxingwu@seu.edu.cn

Xi Chen
Platform and Content Group, Tencent
China
jasonxchen@tencent.com

ABSTRACT

Large Language Models (LLMs), e.g., ChatGPT, PaLM, and LLaMA are transforming natural language processing (NLP) and artificial intelligence (AI). Recent LLMs browse Web knowledge and learn from external knowledge bases, unifying LLMs and knowledge graphs (KGs). The possibility of bridging KGs with LLMs has garnered attention in knowledge engineering. On the one hand, LLMs can be enhanced with KGs to provide answers with more contextualized facts. On the other hand, downstream tasks, e.g., KG curation, embedding, and search can also benefit by adopting LLMs. It remains an interesting direction to explore effective interactions between LLMs and KGs, where many recent advances arise from NLP, deep learning, information retrieval, and computer vision domains. The workshop, titled "LLM+KG: Data Management Opportunities in Unifying Large Language Models+Knowledge Graphs", is targeted at data management researchers, aiming to discuss interesting opportunities, e.g., data cleaning, modeling, designing of algorithms and systems, scalability, fairness, privacy, usability, and explanation.

VLDB Workshop Reference Format:
Arijit Khan, Tianxing Wu, and Xi Chen. Data Management Opportunities in Unifying Large Language Models+Knowledge Graphs. VLDB 2024 Workshop: LLM+KG.

1 WORKSHOP TOPICS AND GOALS

The advent of large language models (LLMs), such as ChatGPT [42], PaLM [13], and LLaMA [61], provides promising capabilities in artificial general intelligence (AGI), demonstrating excellent performance in natural language processing (NLP), e.g., comprehension and generation of human-like texts, sentiment analysis, language translation, question-answering (QA), document classification, summarization, content generation, and virtual assistants, in domains including customer support, healthcare, finance, law, education, engineering, etc. [4, 20, 33, 55, 71, 74, 77, 80]. LLMs are pre-trained on massive text corpora and then fine-tuned through task-specific objectives. Additionally, prompting enables a novel interaction mode with LLMs that does not involve training of model parameters. Prompt engineering designs the inputs given to a model to guide the desired outputs – either via zero-shot prompting, where the model is not provided with any direct examples; or via few-shot prompting, when a few examples consisting of sample inputs and expected outputs are provided to the model, along with the user's query, in order to adopt the model to a certain response format, also known as the in-context learning. Furthermore, the Retrieval-Augmented Generation (RAG) is a paradigm in which a large language model references authoritative knowledge sources outside of its training data before generating a response, thereby optimizing its output. Major technology companies, such as Google, IBM, Microsoft, Meta, Amazon, and Baidu have engaged in competitive rivalries for creating larger and better LLMs, as well as deployed them across commercial products and services to numerous business functions [81, 89].

LLMs, pre-trained on large-scale web and enterprise corpus, encode significant knowledge implicitly in their parameters without human supervision, which can be probed for various QA and querying tasks, thus LLMs act as knowledge bases (KBs) [3, 22, 46, 67]. They generalize from the training corpus. However, LLMs are skilled at learning stochastic language patterns and may not explicitly store consistent representations of knowledge, hence they can output unreliable and incoherent responses, and often experience hallucinations by generating factually incorrect statements, or even harmful content [17, 38, 58]. Like other deep neural networks, LLMs are complex "black-box" systems; knowledge in LLMs is difficult to interpret, update, and is prone to bias, rendering it hard to deploy them in decision-critical applications [43, 88].

Knowledge graphs (KGs), in contrast, enable a structured, highly-curated, and reliable representation of knowledge via explicit relationships, supporting symbolic reasoning and inference, with explainability [12, 24, 25, 72]. KGs such as DBpedia [5], Freebase [10], YAGO [57], Wikidata [65], and NELL [11] store real-world facts as ⟨subject, predicate, object⟩ triples. They may also be represented as large-scale graphs with entities as nodes and relationships between these entities as edges. Almost all big data companies, e.g., Google, Microsoft, IBM, Meta, Amazon, and eBay have proprietary KGs [41]. Commonsense knowledge graphs [28, 29, 87], KGs for synonyms and translations in different languages [40, 56], domain-specific KGs [1], and multi-modal KGs [18, 37, 68] are created. They offer accurate explicit knowledge in many downstream applications including web search, QA [50], semantic search [70], personal assistants [9], fact-checking [60], and recommendation [78]. KGs can also be updated dynamically with new knowledge via the addition or deletion of triples [75].

However, knowledge graphs are difficult to construct and are often incomplete. Non-professional users find it challenging to write an accurate query, e.g., via SPARQL, Cypher [19], Gremlin [49], GSQL [15], etc., since users must have full knowledge of the query language, schema, and the vocabulary used in a KG, besides the schema can be large and complex due to heterogeneity. Current KG

Proceedings of the VLDB Endowment. ISSN 2150-8097.

querying approaches generally lack language understanding, are inadequate to deal with unseen entities and new facts, and often ignore multi-modal information in KGs. Moreover, existing methods are tailored for specific KGs or downstream tasks, referred to as the interoperability issues [26, 34].

In summary, LLMs and KGs offer parametric vs. explicit knowledge, respectively, and can complement each other in knowledge engineering. Recently, efforts have been made to unify LLMs and KGs by leveraging their advantages [2, 39, 43, 44, 53, 82]. KGs assist in the pre-training and inference phases of LLMs, e.g., through retrieval-augmented methods, to provide external knowledge for reducing hallucinations, thus improving accuracy and offering interpretability. LLMs, on the other hand, facilitate knowledge extraction, KG creation, completion, embedding, and various downstream tasks over KGs. In the following, we briefly discuss the synergy between LLMs and KGs, and how they benefit each other.

KGs for LLMs. LLMs may fail to understand a question due to lack of context, might suffer from a knowledge gap, or simply cannot recall facts. Therefore, offering external knowledge through knowledge graphs is becoming prevalent for enhancing the accuracy, consistency, transparency, and the overall capabilities of LLMs.

• *KG-enhanced Pre-training:* Adding knowledge graphs to the training corpus improves pre-training data quality and context, thereby improving LLMs' accuracy. Notable works include KnowBERT [45] which embeds multiple KGs into LLMs by updating contextual word representations with relevant entity embeddings via word-to-entity attention. K-BERT [36] injects KG triples into texts to construct sentence trees for training, thus incorporating domain knowledge into LLMs. KEPLER [69] encodes textual entity descriptions with language models as embeddings, and jointly optimizes KG embeddings and language modeling objectives. DRAGON [84] pre-trains a joint language-knowledge foundation model from KG and text.

• *KG-enhanced Fine-tuning:* Knowledge graphs can assist in fine-tuning LLMs to update their internal knowledge for domain-specific tasks over KGs [8, 32]. RuleBERT [51] fine-tunes an LLM utilizing the Horn rules to incorporate commonsense knowledge. However, it is also costly to fine-tune LLMs to update their knowledge.

• *KG-enhanced Inference:* Sequeda et al. show that using KGs attains higher accuracy for LLM-powered QA systems with zero-shot prompting [53]. The Knowledge Prompts approach trains soft prompts via self-supervised learning based on KGs; the resulting soft knowledge prompts inject world knowledge and new evolving information into LLMs [16]. Baek et al. propose KAPING [7], which first retrieves KG facts relevant to the input question, then prepends the retrieved facts to the question as a prompt to LLMs for the desired output. Wu et al. rewrite the extracted KG triples into well-textualized statements to enhance the accuracy of LLMs [76]. Advanced prompting techniques such as chain-of-thought and graph-of-thought can facilitate retrieving relevant external knowledge for LLMs to improve their reasoning capacity [23, 62, 73].

• *KG-enhanced Validation and Explainability:* KGs provide explanations and fact-checking to justify LLMs' decisions. LAMA [46] probes LLMs by using KGs – it converts KG triples into cloze statements following a prompt template and exploits LLMs to predict the missing entity. Autoprompt [54] generates prompts automatically for various tasks via a gradient-guided search. QA-GNN [85] develops an end-to-end QA model leveraging language models and KGs, and performs interpretable reasoning.

LLMs for KGs. LLMs augment KGs via knowledge extraction, auto-completion, and by considering multi-modal information, as well as enhance the usability and performance of downstream tasks with natural language understanding and generalization capabilities.

• *LLM-enhanced KG Creation:* KGs are difficult to construct due to information extraction and integration from diverse sources. Multi-modal LLMs are well-equipped to extract knowledge from heterogeneous data including text, images, tables, etc. [14, 64, 66]. LLMs are also employed in entity and relation discovery, typing, resolution, linking, and end-to-end construction of KGs [31, 66, 79].

• *LLM-enhanced KG Completion:* LLMs are extensively adopted for KG completion via link prediction. LLMs encode textual information along with KG facts for better link prediction [83]. Recently, LLMs have been used as generators that predict the missing entity in a KG triple directly [52].

• *LLM-enhanced KG Embedding:* LLMs are used for KG+text embedding, such as KEPLER [69] and K-BERT [36]. LLMs with graph and image encoders are combined to train multi-modal KG embedding in [27].

• *LLM-enhanced KG Querying:* The language understanding capacity of LLMs makes them suitable for processing natural language questions (NLQs) over structured KGs. LLMs assist in extracting entities and relations from NLQs, as well as in the answer reasoning process (e.g., QA-GNN [85]). Avila et al. evaluate the ability of ChatGPT to translate the user's NLQs to SPARQL queries on the KG [6]. Relevant facts from KGs can be employed as external knowledge in retrieval-augmented LLMs to answer queries [7, 23, 76].

• *LLM-enhanced KG Analytics:* LLMs are also employed in more complex analytic tasks over graph-structured data (including KGs), commonly known as "graph reasoning", such as computing graph sizes, node degrees, node connectivity, centrality and position of nodes, etc. Various prompting-based approaches have been developed for solving natural language graph problems [86].

• *LLM-enhanced Domain-specific KG Applications:* The synergy between KGs and LLMs is also exploited in multi-disciplinary domains including healthcare, biomedical [59], education [35], e-commerce [47], and spatio-temporal data [30].

Opportunities for Data Management Research. The unification of LLMs and KGs provides exciting data management research opportunities across multiple dimensions.

• *Data and Input Modeling:* The graph structures need to be serialized as part of LLMs' input, either by verbalizing the graph structure in natural languages, or by encoding the sparse structure in dense vector forms. How to integrate graph structure with other multi-modal data, e.g., text, tables, and images as input to LLMs, how to extract relevant subgraphs from KGs for specific downstream tasks, and how to design and learn prompts with graph data for better generalization, are interesting open problems.

• *Data Cleaning, Integration, and Augmentation:* Data cleaning (e.g., error detection and repairing) and integration (e.g., entity and relation extraction, entity resolution, linking) are fundamental to data management. The unification of LLMs and KGs provides new opportunities in this domain. Additionally, LLMs as generators assist in KG auto-completion and domain-specific synthetic data generation.

• *Multi-modal Data Management:* Data are multi-modal, consisting of texts, images, tables, key-values, graphs, and other multimedia data. KGs can serve as a unified data model for cross-domain and diverse data. For example, nodes and edges in a KG may contain features of different modalities. Multi-modal LLMs are better suited to extract knowledge from heterogeneous data.
• *Vector Data Management:* With the emergence of multi-modal KG embedding and multi-modal LLMs, vector data management is critical. Querying vectors is challenging, since they are dense and high-dimensional, rendering many indexing approaches ineffective due to the curse of dimensionality. The data management community can contribute to this field with high-dimensional data indexing, join, and geometric querying. Vector representations are also used in retrieval-augmented LLMs for efficient top-k retrievals and prompt learning with gradient-based search.
• *Accuracy and Consistency:* Enhancing LLMs' accuracy, consistency, reducing hallucinations and harmful content generation, fake news detection, fact-checking, etc. with knowledge-grounded techniques are emerging research directions.
• *Efficiency and Scalability:* LLM scaling laws are based on empirical observations that a larger number of parameters and tokens in a model improves performance across various downstream tasks. Consequently, the computer systems (e.g., servers, GPUs, TPUs) used to train, run, and serve predictions from these models have high-performance requirements and are expensive to procure and operate – in terms of monetary costs and environmental impacts, e.g., they consume megawatt-hours of electricity and emit tons of greenhouse gasses – limiting access for smaller organizations and researchers. It is, therefore, critical to optimize the performance of LLM systems, characterize resource management and techniques for training and inference, their trade-offs on accuracy requirements, compress model size, and effectively deploy these systems in production environments (e.g., at the edge).
• *Bias and Fairness:* LLMs retain and amplify biases present in training data. LLMs' performance can be biased against long-tail entities, in comparison to popular entities. KGs can mitigate biases by providing explicit knowledge about long-tail entities. Bias in KG embeddings could be mitigated via data augmentation using LLMs.
• *Explainability and Provenance:* KGs offer explainability to LLMs' responses by probing them and grounding their reasoning with external knowledge. It is critical to develop techniques that can associate LLM-generated content with its provenance information.
• *Usability:* LLMs improve the interoperability of KG downstream tasks through their natural language interfaces, transferability, and generalization capacity. It would be interesting to analyze the expressiveness of KG-enhanced LLM models.
• *Security and Privacy:* LLMs, trained on proprietary datasets, can inadvertently reveal confidential information in their responses, increasing the risk of unauthorized data access and security breaches. As the usage of graph data in LLMs expands, so does the concern for privacy and security. Ensuring the confidentiality of sensitive graph information, while still extracting knowledge for LLMs, poses an exciting challenge.
• *Optimizing KG Databases and Systems:* Recent advances in LLM-enhanced database systems have showcased the potential to optimize querying tasks. However, complex graph structures require specialized attention. By exploiting historical usage data and graph topology with in-context learning, LLMs can autonomously adapt storage strategies and predict access patterns.
• *Data and AI Model Market Challenges:* Data and AI model markets enable multiple organizations to sell, discover, share, and purchase high-quality data and AI models for better training and inference. For instance, regardless of fine-tuning or the RAG paradigm, an LLM model's success depends on the nobility and fitness of the data post-training. Analogously, certain fine-tuned LLMs would be more suitable for specific downstream tasks. With the integration of LLMs and KGs, data and AI markets provide potential opportunities for effective sharing at scale, coupled with novel challenges associated with graph data and model pricing.
• *Benchmarking and Ground Truth:* In many emerging domains such as healthcare, biomedical, education, finance, cyber security, coding, personal assistants, e-commerce, etc., the integration of LLMs and KGs have depicted incredible promises. It is important to have ground truth datasets and experimental benchmarks to facilitate future research and developments in these domains.

Goals: Why is the Workshop Important? Why Now? Large language models (LLMs) recently emerged to mainstream, and already became a powerful tool for interacting with data. LLM adoption is rapidly accelerating in the industry – prominent players include Open AI (ChatGPT), Google (PaLM), Amazon (Titan, Olympus), Meta (LLaMA), Huawei (Pangu), Tencent (Hunyuan), Anthropic (Claude), Microsoft (Turing-NLG, Orca), etc. Most teams using LLMs are investing in prompt engineering, vector databases, and LLMs' monitoring (e.g., Responsible AI). The global LLM market size in terms of revenue is projected to reach 259,886.45 Million USD by 2029 from 1,302.93 Million USD in 2023, with a compound annual growth rate (CAGR) 141.72% during 2023-2029 [48]. Given that the space is so new, it is an exciting time to be working at the cutting edge of LLMs. This technology also created several opportunities for applications in the general data management [21, 63].

However, LLMs on graph data and in particular, the synergy between LLMs and knowledge graphs (KGs) has received less attention from graph data management, and by the DB community in general. This workshop's objective is to draw attention to this emerging topic, which has the potential to not only deepen LLMs' impact in real-world KG and graph data applications, but also to enhance the performance of LLMs using external knowledge from KGs. Therefore, our workshop is timely and relevant.

2 WORKSHOP PROGRAM

This workshop consists of nine accepted papers, three keynote talks, one industry talk, and one panel about cutting-edge research and novel directions for open problems in the LLM+KG area.

Session 1

Keynote 1. *Integrating Knowledge Graph with Large Language Model: From the Perspective of Knowledge Engineering* – Guilin Qi (Southeast University, China).

Keynote 2. *Industry-level Knowledge Graph Platform for Large-scale, Diverse and Dynamic Scenarios* – Haofen Wang (Tongji University, China).

Session 2

Keynote 3. *Knowledge Graph-Based Large Language Model Fine-tuning and Its Applications* – Wei Hu (Nanjing University, China).

Paper Presentation: *OneEdit: A Neural-Symbolic Collaboratively Knowledge Editing System* - Ningyu Zhang, Zekun Xi, Yujie Luo, Peng Wang, Bozhong Tian, Yunzhi Yao, Jintian Zhang, Shumin Deng, Mengshu sun, Lei Liang, Zhiqiang Zhang, Xiaowei Zhu, Jun Zhou, and Huajun Chen.
Paper Presentation: *Leveraging LLMs Few-shot Learning to Improve Instruction-driven Knowledge Graph Construction* - Yongli Mou, Li Liu, Sulayman Sowe, Diego Collarana, and Stefan Decker.
Paper Presentation: *SPIREX: Improving LLM-based Relation Extraction from RNA-focused Scientific Literature using Graph Machine Learning* - Emanuele Cavalleri, Mauricio Soto-Gomez, Ali Pashaeibarough, Dario Malchiodi, Harry Caufield, Justin Reese, Chris J Mungall, Peter Robinson, Elena Casiraghi, Giorgio Valentini, and Marco Mesiti.
Session 3
Industry Talk: *Integrating GenAI with Graph: Innovations and Insights from NebulaGraph* – Siwei Gu and Yihang Yu (NebulaGraph).
Paper Presentation: *Enhancing Large Language Models with Multimodality and Knowledge Graphs for Hallucination-free Open-set Object Recognition* - Xinfu Liu, Yirui Wu, Yuting Zhou, Junyang Chen, Huan Wang, Ye Liu, and Shaohua Wan.
Paper Presentation: *From Instructions to ODRL Usage Policies: An Ontology Guided Approach* - Daham M. Mustafa, Abhishek Nadgeri, Diego Collarana, Benedikt T. Arnold, Christoph Quix, Christoph Lange, and Stefan Decker.
Paper Presentation: *Knowledge Graph Efficient Construction: Embedding Chain-of-Thought into LLMs* - Jixuan Nie, Xia Hou, Wenfeng Song, Xuan Wang, Xingliang Jin, Xinyu Zhang, ShuoZhe Zhang, and Jiaqi Shi.
Session 4
Paper Presentation: *Benchmarking and Analyzing In-context Learning, Fine-tuning and Supervised Learning for Biomedical Knowledge Curation: A Focused Study on Chemical Entities of Biological Interest* - Yusuf Abdulle, Emily Groves, Minhong Wang, Holger Kunz, Jason Hoelscher-Obermaier, Ronin Wu, and Honghan Wu.
Paper Presentation: *Research Trends for the Interplay between Large Language Models and Knowledge Graphs* - Hanieh Khorashadizadeh.
Paper Presentation: *InfuserKI: Enhancing Large Language Models with Knowledge Graphs via Infuser-Guided Knowledge Integration* - Fali Wang , Runxue Bao, Suhang Wang, Wenchao Yu, Yanchi Liu, Wei Cheng, and Haifeng Chen.
Panel: *Large Language Models, Knowledge Graphs, and Vector Databases: Synergy and Opportunities for Data Management* - Panelists: Wei Hu (Nanjing University), Shreya Shankar (UC Berkeley), Haofen Wang (Tongji University), and Jianguo Wang (Purdue University).

3 PROGRAM COMMITTEE

Sheng Bi - Southeast University, China
Angela Bonifati - University of Lyon, France
Yongrui Chen - Southeast University, China
Yubo Chen - Institute of Automation, Chinese Academy of Sciences, China
Jiaoyan Chen - The University of Manchester, UK
Peng Fang - Huazhong University of Science and Technology, China
Jonathan Fürst - Zurich University of Applied Sciences, Switzerland
Rainer Gemulla - Universität Mannheim, Germany
Lei Hou - Tsinghua University, China
Ernesto Jimenez-Ruiz - City, University of London, UK
Xiangyu Ke - Zhejiang University, China
Wolfgang Lehner - TU Dresden, Germany
Bohan Li - Nanjing University of Aeronautics and Astronautics, China
Chuangtao Ma - Aalborg University, Denmark
Essam Mansour - Concordia University, Canada
Sharad Mehrotra - UC Irvine, USA
Arash Termehchy - Oregon State University, USA
Xin Wang - Tianjin University, China
Haofen Wang - Tongji University, China
Meng Wang - Tongji University, China
Yuxiang Wang - Hangzhou Dianzi University, China
Shiyu Yang - Guangzhou University, China
Wen Zhang - Zhejiang University, China
Xiang Zhao - National University of Defense Technology, China

4 WORKSHOP CO-CHAIRS

Arijit Khan is an IEEE senior member, an ACM distinguished speaker, and an associate professor in the Department of Computer Science, Aalborg University, Denmark. He earned his Ph.D. from UC Santa Barbara, USA and did a post-doc at ETH Zurich, Switzerland. He has been an assistant professor at NTU Singapore. Arijit is the recipient of the IBM Ph.D. Fellowship (2012-13), a VLDB Distinguished Reviewer award (2022), and a SIGMOD Distinguished PC award (2024). He published over 80 papers in premier data management and mining venues, e.g., SIGMOD, VLDB, TKDE, ICDE, WWW, SDM, EDBT, CIKM, WSDM, and TKDD. Arijit co-presented tutorials on graph queries, systems, applications, and machine learning at VLDB, ICDE, CIKM, and DSAA; and is serving in the program committee/ senior program committee of KDD, SIGMOD, VLDB, ICDE, ICDM, EDBT, SDM, CIKM, AAAI, WWW, and an associate editor of TKDE and TKDD. Arijit served as the co-chair of Big-O(Q) workshop co-located with VLDB 2015, and wrote a book on uncertain graphs in the Morgan & Claypool's Synthesis Lectures on Data Management. He contributed invited chapters and articles on big graphs querying and mining in the ACM SIGMOD blog and in the Springer Encyclopedia of Big Data Technologies. More information at https://homes.cs.aau.dk/~Arijit/index.html.

Tianxing Wu is an associate professor working at School of Computer Science and Engineering of Southeast University, China. He is one of the main contributors to build Chinese large-scale encyclopedic knowledge graph: Zhishi.me and schema knowledge graph: Linked Open Schema. He was awarded 2019 Excellent Ph.D. Degree Dissertation of Jiangsu Computer Society, 2020 Excellent Ph.D. Degree Dissertation of Southeast University, and CCKS 2022 Best Paper Award. His research interests include knowledge graph, knowledge representation and reasoning, and data mining. He has published over 50 papers in top-tier conferences and journals, such as ICDE, AAAI, IJCAI, ECAI, ISWC, TKDE, TKDD, JWS, WWWJ, and etc. He is the editorial board member of International Journal on Semantic Web and Information Systems, Data Intelligence, and etc. He also has served as the (senior) program committee member of AAAI, IJCAI, ACL, TheWebConf, EMNLP, ISWC, ECAI, and etc. More information at https://tianxing-wu.github.io/.

Xi Chen is the director of the algorithm Team of Platform and Content Group, Tencent. He received the Ph.D. Degree Dissertation of Zhejiang University and won good results in many KG and LLM competitions, such as CCKS2020 NER Task, CHIP2020 Relation Extraction Task, SuperGLUE Challenge, Semeval and so on. He has published over 40 papers in top-tier conferences and journals, such as ACL, EMNLP, NeurIPS, WWW, AAAI, IJCAI, TKDE, JWS, and etc. He was awarded the PAKDD 2021 Best Paper Award. More information at https://scholar.google.com/citations?user=qy0QX0MAAAAJ&hl=zh-CN

REFERENCES

[1] Bilal Abu-Salih. 2021. Domain-specific Knowledge Graphs: A Survey. *J. Netw. Comput. Appl.* 185 (2021), 103076.

[2] Garima Agrawal, Tharindu Kumarage, Zeyad Alghami, and Huan Liu. 2024. Can Knowledge Graphs Reduce Hallucinations in LLMs? : A Survey. In *NAACL*.

[3] Badr AlKhamissi, Millicent Li, Asli Celikyilmaz, Mona T. Diab, and Marjan Ghazvininejad. 2022. A Review on Language Models as Knowledge Bases. *CoRR* abs/2204.06031 (2022).

[4] Sihem Amer-Yahia, Angela Bonifati, Lei Chen, Guoliang Li, Kyuseok Shim, Jianliang Xu, and Xiaochun Yang. 2023. From Large Language Models to Databases and Back: A Discussion on Research and Education. *SIGMOD Rec.* 52, 3 (2023), 49–56.

[5] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In *ISWC+ ASWC*.

[6] Caio Viktor S. Avila, Vânia M. P. Vidal, Wellington Franco, and Marco A. Casanova. 2024. Experiments with Text-to-SPARQL based on ChatGPT. In *ICSC*.

[7] Jinheon Baek, Alham Fikri Aji, and Amir Saffari. 2023. Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering. In *MATCHING*.

[8] Teodoro Baldazzi, Luigi Bellomarini, Stefano Ceri, Andrea Colombo, Andrea Gentili, and Emanuel Sallinger. 2023. Fine-Tuning Large Enterprise Language Models via Ontological Reasoning. In *RuleML+RR*.

[9] Krisztian Balog and Tom Kenter. 2019. Personal Knowledge Graphs: A Research Agenda. In *SIGIR*.

[10] Kurt D. Bollacker, Colin Evans, Praveen K. Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge. In *SIGMOD*.

[11] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., and Tom M. Mitchell. 2010. Toward an Architecture for Never-Ending Language Learning. In *AAAI*.

[12] Vinay K. Chaudhri, Chaitanya K. Baru, Naren Chittar, Xin Luna Dong, Michael R. Genesereth, James A. Hendler, Aditya Kalyanpur, Douglas B. Lenat, Juan Sequeda, Denny Vrandecic, and Kuansan Wang. 2022. Knowledge Graphs: Introduction, History and, Perspectives. *AI Mag.* 43, 1 (2022), 17–29.

[13] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. 2023. PaLM: Scaling Language Modeling with Pathways. *J. Mach. Learn. Res.* 24 (2023), 240:1–240:113.

[14] Xiang Deng, Huan Sun, Alyssa Lees, You Wu, and Cong Yu. 2022. TURL: Table Understanding through Representation Learning. *SIGMOD Rec.* 51, 1 (2022), 33–40.

[15] Alin Deutsch. 2018. Querying Graph Databases with the GSQL Query Language. In *SBBD*.

[16] Cícero Nogueira dos Santos, Zhe Dong, Daniel Cer, John Nham, Siamak Shakeri, Jianmo Ni, and Yun-Hsuan Sung. 2022. Knowledge Prompts: Injecting World Knowledge into Language Models through Soft Prompts. *CoRR* abs/2210.04726 (2022).

[17] Yanai Elazar, Nora Kassner, Shauli Ravfogel, Abhilasha Ravichander, Eduard H. Hovy, Hinrich Schütze, and Yoav Goldberg. 2021. Measuring and Improving Consistency in Pretrained Language Models. *Trans. Assoc. Comput. Linguistics* 9 (2021), 1012–1031.

[18] Sebastián Ferrada, Benjamin Bustos, and Aidan Hogan. 2017. IMGpedia: A Linked Dataset with Content-Based Analysis of Wikimedia Images. In *ISWC*.

[19] Nadime Francis, Alastair Green, Paolo Guagliardo, Leonid Libkin, Tobias Lindaaker, Victor Marsault, Stefan Plantikow, Mats Rydberg, Petra Selmer, and Andrés Taylor. 2018. Cypher: An Evolving Query Language for Property Graphs. In *SIGMOD*.

[20] Yu Gai, Liyi Zhou, Kaihua Qin, Dawn Song, and Arthur Gervais. 2023. Blockchain Large Language Models. *IACR Cryptol. ePrint Arch.* (2023), 592.

[21] Alon Y. Halevy, Yejin Choi, Avrilia Floratou, Michael J. Franklin, Natasha F. Noy, and Haixun Wang. 2023. Will LLMs reshape, supercharge, or kill data science? *Proc. VLDB Endow.* 16, 12 (2023), 4114–4115.

[22] Shibo Hao, Bowen Tan, Kaiwen Tang, Bin Ni, Xiyan Shao, Hengzhe Zhang, Eric P. Xing, and Zhiting Hu. 2023. BertNet: Harvesting Knowledge Graphs with Arbitrary Relations from Pretrained Language Models. In *Findings of the Association for Computational Linguistics: ACL*.

[23] Hangfeng He, Hongming Zhang, and Dan Roth. 2023. Rethinking with Retrieval: Faithful Large Language Model Inference. *CoRR* abs/2301.00303 (2023).

[24] Nicolas Heist, Sven Hertling, Daniel Ringler, and Heiko Paulheim. 2020. Knowledge Graphs on the Web - An Overview. In *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*. Vol. 47. 3–22.

[25] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. 2021. Knowledge Graphs. *ACM Comput. Surv.* 54, 4 (2021), 37.

[26] Katja Hose. 2023. Knowledge Engineering in the Era of Artificial Intelligence. In *ADBIS*.

[27] Ningyuan Huang, Yash R. Deshpande, Yibo Liu, Houda Alberts, Kyunghyun Cho, Clara Vania, and Iacer Calixto. 2022. Endowing Language Models with Multimodal Knowledge Graph Representations. *CoRR* abs/2206.13163 (2022).

[28] Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. 2021. (Comet-) Atomic 2020: On Symbolic and Neural Commonsense Knowledge Graphs. In *AAAI*.

[29] Filip Ilievski, Pedro A. Szekely, and Bin Zhang. 2021. CSKG: The CommonSense Knowledge Graph. In *ESWC*.

[30] Ming Jin, Qingsong Wen, Yuxuan Liang, Chaoli Zhang, Siqiao Xue, Xue Wang, James Zhang, Yi Wang, Haifeng Chen, Xiaoli Li, Shirui Pan, Vincent S. Tseng, Yu Zheng, Lei Chen, and Hui Xiong. 2023. Large Models for Time Series and Spatio-Temporal Data: A Survey and Outlook. *CoRR* abs/2310.10196 (2023).

[31] Mandar Joshi, Omer Levy, Luke Zettlemoyer, and Daniel S. Weld. 2019. BERT for Coreference Resolution: Baselines and Analysis. In *EMNLP-IJCNLP*.

[32] Minki Kang, Jinheon Baek, and Sung Ju Hwang. 2022. KALA: Knowledge-Augmented Language Model Adaptation. In *NAACL*.

[33] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences* 103 (2023), 102274.

[34] Arijit Khan. 2023. Knowledge Graphs Querying. *SIGMOD Rec.* 52, 2 (2023), 18–29.

[35] Xiu Li, Aron Henriksson, Martin Duneld, Jalal Nouri, and Yongchao Wu. 2024. Evaluating Embeddings from Pre-Trained Language Models and Knowledge Graphs for Educational Content Recommendation. *Future Internet* 16, 1 (2024).

[36] Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. 2020. K-BERT: Enabling Language Representation with Knowledge Graph. In *AAAI*.

[37] Ye Liu, Hui Li, Alberto García-Durán, Mathias Niepert, Daniel Oñoro-Rubio, and David S. Rosenblum. 2019. MMKG: Multi-modal Knowledge Graphs. In *ESWC*.

[38] Yang Liu, Yuanshun Yao, Jean-Francois Ton, Xiaoying Zhang, Ruocheng Guo, Hao Cheng, Yegor Klochkov, Muhammad Faaiz Taufiq, and Hang Li. 2023. Trustworthy LLMs: A Survey and Guideline for Evaluating Large Language Models' Alignment. *CoRR* abs/2308.05374 (2023).

[39] Justin Lovelace and Carolyn P. Rosé. 2022. A Framework for Adapting Pre-Trained Language Models to Knowledge Graph Completion. In *EMNLP*.

[40] Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a Very Large Multilingual Semantic Network. In *ACL*.

[41] Natalya Fridman Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, and Jamie Taylor. 2019. Industry-scale Knowledge Graphs: Lessons and Challenges. *Commun. ACM* 62, 8 (2019), 36–43.

[42] OpenAI. 2022. Introducing ChatGPT. https://openai.com/blog/chatgpt.

[43] Jeff Z. Pan, Simon Razniewski, Jan-Christoph Kalo, Sneha Singhania, Jiaoyan Chen, Stefan Dietze, Hajira Jabeen, Janna Omeliyanenko, Wen Zhang, Matteo Lissandrini, Russa Biswas, Gerard de Melo, Angela Bonifati, Edlira Vakaj, Mauro Dragoni, and Damien Graux. 2023. Large Language Models and Knowledge Graphs: Opportunities and Challenges. *TGDK* 1, 1 (2023), 2:1–2:38.

[44] Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. 2024. Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Transactions on Knowledge and Data Engineering* 36, 7 (2024), 3580–3599.

[45] Matthew E. Peters, Mark Neumann, Robert L. Logan IV, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A. Smith. 2019. Knowledge Enhanced Contextual Word Representations. In *EMNLP-IJCNLP*.

[46] Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick S. H. Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander H. Miller. 2019. Language Models as Knowledge Bases?. In *EMNLP-IJCNLP*.

[47] André Gomes Regino, Rodrigo Oliveira Caus, Victor Hochgreb, and Julio Cesar dos Reis. 2023. From Natural Language Texts to RDF Triples: A Novel Approach to Generating e-Commerce Knowledge Graphs. In *Knowledge Discovery, Knowledge Engineering and Knowledge Management*.

[48] GII Research. 2024. Global Large Language Model (LLM) Market Research Report 2024. https://www.giiresearch.com/report/qyr1384359-global-large-language-model-llm-market-research.html.

[49] Marko A. Rodriguez. 2015. The Gremlin Graph Traversal Machine and Language (Invited Talk). In *DBPL*.

[50] Rishiraj Saha Roy and Avishek Anand. 2020. Question Answering over Curated and Open Web Sources. In *SIGIR*.

[51] Mohammed Saeed, Naser Ahmadi, Preslav Nakov, and Paolo Papotti. 2021. Rule-BERT: Teaching Soft Rules to Pre-Trained Language Models. In *EMNLP*.

[52] Apoorv Saxena, Adrian Kochsiek, and Rainer Gemulla. 2022. Sequence-to-Sequence Knowledge Graph Completion and Question Answering. In *ACL*.

[53] Juan Sequeda, Dean Allemang, and Bryon Jacob. 2023. A Benchmark to Understand the Role of Knowledge Graphs on Large Language Model's Accuracy for Question Answering on Enterprise SQL Databases. *CoRR* abs/2311.07509 (2023).

[54] Taylor Shin, Yasaman Razeghi, Robert L. Logan IV, Eric Wallace, and Sameer Singh. 2020. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. In *EMNLP*.

[55] Karan Singhal, Shekoofeh Azizi, Tao Tu, S. Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Kumar Tanwani, Heather Cole-Lewis, Stephen Pfohl, Perry Payne, Martin Seneviratne, Paul Gamble, Chris Kelly, Nathaneal Schärli, Aakanksha Chowdhery, Philip Andrew Mansfield, Blaise Agüera y Arcas, Dale R. Webster, Gregory S. Corrado, Yossi Matias, Katherine Chou, Juraj Gottweis, Nenad Tomasev, Yun Liu, Alvin Rajkomar, Joelle K. Barral, Christopher Semturs, Alan Karthikesalingam, and Vivek Natarajan. 2023. Large Language Models Encode Clinical Knowledge. *Nature* 620 (2023), 172–180.

[56] Robyn Speer and Catherine Havasi. 2012. Representing General Relational Knowledge in ConceptNet 5. In *LREC*.

[57] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: A Core of Semantic Knowledge. In *World Wide Web*.

[58] Kai Sun, Yifan Ethan Xu, Hanwen Zha, Yue Liu, and Xin Luna Dong. 2023. Head-to-Tail: How Knowledgeable are Large Language Models (LLM)? A.K.A. Will LLMs Replace Knowledge Graphs? *CoRR* abs/2308.10168 (2023).

[59] Mujeen Sung, Jinhyuk Lee, Sean S. Yi, Minji Jeon, Sungdong Kim, and Jaewoo Kang. 2021. Can Language Models be Biomedical Knowledge Bases?. In *EMNLP*.

[60] Andon Tchechmedjiev, Pavlos Fafalios, Katarina Boland, Malo Gasquet, Matthäus Zloch, Benjamin Zapilko, Stefan Dietze, and Konstantin Todorov. 2019. ClaimsKG: A Knowledge Graph of Fact-Checked Claims. In *ISWC*.

[61] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. *CoRR* abs/2302.13971 (2023).

[62] Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. 2023. Interleaving Retrieval with Chain-of-Thought Reasoning for Knowledge-Intensive Multi-Step Questions. In *ACL*.

[63] Immanuel Trummer. 2022. From BERT to GPT-3 Codex: Harnessing the Potential of Very Large Language Models for Data Management. *Proc. VLDB Endow.* 15, 12 (2022), 3770–3773.

[64] Liane Vogel, Benjamin Hilprecht, and Carsten Binnig. 2023. Towards Foundation Models for Relational Databases [Vision Paper]. *CoRR* abs/2305.15321 (2023).

[65] Denny Vrandecic and Markus Krötzsch. 2014. Wikidata: A Free Collaborative Knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.

[66] David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, Relation, and Event Extraction with Contextualized Span Representations. In *EMNLP-IJCNLP*.

[67] Chenguang Wang, Xiao Liu, and Dawn Song. 2020. Language Models are Open Knowledge Graphs. *CoRR* abs/2010.11967 (2020).

[68] Meng Wang, Haofen Wang, Guilin Qi, and Qiushuo Zheng. 2020. Richpedia: A Large-Scale, Comprehensive Multi-Modal Knowledge Graph. *Big Data Res.* 22 (2020), 100159.

[69] Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, and Jian Tang. 2021. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation. *Trans. Assoc. Comput. Linguistics* 9 (2021), 176–194.

[70] Yuxiang Wang, Arijit Khan, Tianxing Wu, Jiahui Jin, and Haijiang Yan. 2020. Semantic Guided and Response Times Bounded Top-k Similarity Search over Knowledge Graphs. In *ICDE*.

[71] Yuqing Wang, Yun Zhao, and Linda R. Petzold. 2023. Are Large Language Models Ready for Healthcare? A Comparative Study on Clinical Language Understanding. In *MLHC*.

[72] Gerhard Weikum, Xin Luna Dong, Simon Razniewski, and Fabian M. Suchanek. 2021. Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases. *Found. Trends Databases* 10, 2-4 (2021), 108–490.

[73] Yilin Wen, Zifeng Wang, and Jimeng Sun. 2024. MindMap: Knowledge Graph Prompting Sparks Graph of Thoughts in Large Language Models. In *ACL*.

[74] Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David S. Rosenberg, and Gideon Mann. 2023. BloombergGPT: A Large Language Model for Finance. *CoRR* abs/2303.17564 (2023).

[75] Tianxing Wu, Arijit Khan, Melvin Yong, Guilin Qi, and Meng Wang. 2022. Efficiently Embedding Dynamic Knowledge Graphs. *Knowl. Based Syst.* 250 (2022), 109124.

[76] Yike Wu, Nan Hu, Sheng Bi, Guilin Qi, Jie Ren, Anhuan Xie, and Wei Song. 2023. Retrieve-Rewrite-Answer: A KG-to-Text Enhanced LLMs Framework for Knowledge Graph Question Answering. In *IJCKG*.

[77] Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, and Maosong Sun. 2021. Lawformer: A Pre-trained Language Model for Chinese Legal Long Documents. *AI Open* 2 (2021), 79–84.

[78] Da Xu, Chuanwei Ruan, Evren Körpeoglu, Sushant Kumar, and Kannan Achan. 2020. Product Knowledge Graph Embedding for E-commerce. In *WSDM*.

[79] Hang Yan, Tao Gui, Junqi Dai, Qipeng Guo, Zheng Zhang, and Xipeng Qiu. 2021. A Unified Generative Framework for Various NER Subtasks. In *ACL/IJCNLP*.

[80] Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. 2023. FinGPT: Open-Source Financial Large Language Models. In *FinLLM Symposium@IJCAI*.

[81] Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Bing Yin, and Xia Hu. 2024. Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. *ACM Trans. Knowl. Discov. Data* 18, 6, Article 160 (2024), 32 pages.

[82] Linyao Yang, Hongyang Chen, Zhao Li, Xiao Ding, and Xindong Wu. 2023. ChatGPT is not Enough: Enhancing Large Language Models with Knowledge Graphs for Fact-aware Language Modeling. *CoRR* abs/2306.11489 (2023).

[83] Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. KG-BERT: BERT for Knowledge Graph Completion. *CoRR* abs/1909.03193 (2019).

[84] Michihiro Yasunaga, Antoine Bosselut, Hongyu Ren, Xikun Zhang, Christopher D. Manning, Percy Liang, and Jure Leskovec. 2022. Deep Bidirectional Language-Knowledge Graph Pretraining. In *NeurIPS*.

[85] Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. 2021. QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering. In *NAACL-HLT*.

[86] Ruosong Ye, Caiqi Zhang, Runhui Wang, Shuyuan Xu, and Yongfeng Zhang. 2024. Language is All a Graph Needs. In *Findings of the Association for Computational Linguistics: EACL*.

[87] Hongming Zhang, Daniel Khashabi, Yangqiu Song, and Dan Roth. 2020. TransOMCS: From Linguistic Graphs to Commonsense Knowledge. In *IJCAI*.

[88] Haiyan Zhao, Hanjie Chen, Fan Yang, Ninghao Liu, Huiqi Deng, Hengyi Cai, Shuaiqiang Wang, Dawei Yin, and Mengnan Du. 2024. Explainability for Large Language Models: A Survey. *ACM Trans. Intell. Syst. Technol.* 15, 2, Article 20 (2024), 38 pages.

[89] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A Survey of Large Language Models. *CoRR* abs/2303.18223 (2023).