Evaluation of Language Models in Knowledge Engineering (ELMKE) Workshop

Bohui Zhang[1], Reham Alharbi[2] and Yuan He[3]

[1]King's College London, London, United Kingdom
[2]University of Liverpool, Liverpool, United Kingdom
[3]University of Oxford, Oxford, United Kingdom

Abstract

This preface outlines the scope of the Second Workshop on Evaluation of Language Models in Knowledge Engineering (ELMKE 2025) and provides essential details for inclusion in the ESWC 2025 joint proceedings. Held in Portorož, Slovenia, and co-located with ESWC 2025, ELMKE focused on topics such as evaluation methodologies, knowledge and ontology engineering, large language models, benchmark creation, human-centered evaluation, and aspects of trustworthiness, interpretability, and explainability.

1. Introduction

Language models (LMs) have been considered promising in numerous knowledge engineering (KE) tasks, such as ontology engineering [1, 2], knowledge base construction [3, 4], and curation [5]. However, their adoption introduces new challenges for evaluation [6]. The assessment of LM-generated results remains limited, lacking a comprehensive and formally defined framework, and relies heavily on human effort, making it difficult to compare methods and reproduce experiments [7, 8, 9]. The ELMKE 2025 workshop series[1] aims to address this critical gap by fostering a community-driven effort to automate and standardize evaluation. It brings together diverse expertise to advance novel evaluation paradigms. This year's workshop placed special emphasis on the paper topics including:

- Novel evaluation approaches for LMs in KE tasks
- Human-centered evaluation: principles, strategies, paradigms, and interfaces
- Creation of datasets and benchmarks for LM evaluation in KE
- Metrics and methods for assessing trustworthiness, interpretability, and explainability
- Techniques for detecting and mitigating bias in LM outputs
- Detection and evaluation of hallucinations and inconsistencies; enhancing replicability
- Impact of data leakage on evaluation validity

2. Accepted Papers

- From Experts to LLMs: Evaluating the Quality of Automatically Generated Ontologies
  Majlinda Llugiqi, Fajar J Ekaputra and Marta Sabou
- Assessing the Capability of Large Language Models for Domain-Specific Ontology Generation
  Anna Sofia Lippolis, Eva Blomqvist, Mohammad Javad Saeedizade, Andrea Giovanni Nuzzolese and Robin Keskisärkkä
- How do Scaling Laws Apply to Knowledge Graph Engineering Tasks? The Impact of Model Size on Large Language Model Performance
  Desiree Heim, Lars-Peter Meyer, Markus Schröder, Johannes Frey and Andreas Dengel

ESWC 2025: Extended Semantic Web Conference, June 01–05, 2025, Portorož, Slovenia

bohui.zhang@kcl.ac.uk (B. Zhang); R.Alharbi@liverpool.ac.uk (R. Alharbi); yuan.he@cs.ox.ac.uk (Y. He)
https://bohuizhang.github.io/ (B. Zhang); https://www.yuanhe.wiki/ (Y. He)
0000-0001-5430-1624 (B. Zhang); 0000-0002-8332-3803 (R. Alharbi); 0000-0002-4486-1262 (Y. He)

[1] https://sites.google.com/view/elmke

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

- Large Language Models as Knowledge Evaluation Agents
  George Hannah, Jacopo de Bernardinis, Terry Payne, Valentina Tamma, Andrew Mitchell, Ellen Piercy, Ewan Johnson, Andrew Ng, Harry Rostron and Boris Konev

3. Program Committee

The workshop organizers would like to express their sincere appreciation to the Program Committee members for their valuable contributions and support: Terry Payne (University of Liverpool), Nitisha Jain (King's College London), Daniel Dobriy (Vienna University of Economics and Business), Hamed Babaei Giglou (TIB), Chiara Di Bonaventura (King's College London), Ariana Graciotti (University of Bologna), George Hannah (University of Liverpool), Arunav Das (King's College London), Anna Sofia Lippolis (University of Bologna), Mohammad Javad Saeedizade (Linköping University), Yihang Zhao (King's College London)

References

[1] Y. He, J. Chen, D. Antonyrajah, I. Horrocks, BERTMap: A BERT-based Ontology Alignment System, CoRR abs/2112.02682 (2021). URL: https://arxiv.org/abs/2112.02682. arXiv:2112.02682.

[2] B. Zhang, V. A. Carriero, K. Schreiberhuber, S. Tsaneva, L. S. González, J. Kim, J. de Berardinis, OntoChat: A Framework for Conversational Ontology Engineering Using Language Models, in: The Semantic Web: ESWC 2024 Satellite Events: Hersonissos, Crete, Greece, May 26–30, 2024, Proceedings, Part I, Springer-Verlag, Berlin, Heidelberg, 2025, p. 102–121. URL: https://doi.org/10.1007/978-3-031-78952-6_10. doi:10.1007/978-3-031-78952-6_10.

[3] D. Alivanistos, S. B. Santamaría, M. Cochez, J.-C. Kalo, E. van Krieken, T. Thanapalasingam, Prompting as Probing: Using Language Models for Knowledge Base Construction, 2023. URL: https://arxiv.org/abs/2208.11057. arXiv:2208.11057.

[4] B. Zhang, I. Reklos, N. Jain, A. Meroño-Peñuela, E. Simperl, Using Large Language Models for Knowledge Engineering (LLMKE): A Case Study on Wikidata, CoRR abs/2309.08491 (2023). URL: https://doi.org/10.48550/arXiv.2309.08491. doi:10.48550/arXiv.2309.08491. arXiv:2309.08491.

[5] S. Tsaneva, D. Dessì, F. Osborne, M. Sabou, Knowledge graph validation by integrating LLMs and human-in-the-loop, Information Processing & Management 62 (2025) 104145. URL: https://www.sciencedirect.com/science/article/pii/S030645732500086X. doi:https://doi.org/10.1016/j.ipm.2025.104145.

[6] B. Zhang, E. Koutsiana, Y. Zhao, A. Meroño-Peñuela, E. Simperl, Trustworthy Knowledge Graphs: Practices and Approaches, in: Handbook on Neurosymbolic AI and Knowledge Graphs, IOS Press, 2025, pp. 363–384. doi:10.3233/FAIA250215.

[7] B. P. Allen, L. Stork, P. Groth, Knowledge Engineering Using Large Language Models, Transactions on Graph Data and Knowledge 1 (2023) 3:1–3:19. URL: https://drops.dagstuhl.de/entities/document/10.4230/TGDK.1.1.3. doi:10.4230/TGDK.1.1.3.

[8] E. Koutsiana, J. Walker, M. Nwachukwu, A. Meroño-Peñuela, E. Simperl, Knowledge Prompting: How Knowledge Engineers Use Large Language Models, 2024. URL: https://arxiv.org/abs/2408.08878. arXiv:2408.08878.

[9] J. Frey, L.-P. Meyer, F. Brei, S. Gründer-Fahrer, M. Martin, Assessing the Evolution of LLM Capabilities for Knowledge Graph Engineering in 2023, in: A. Meroño Peñuela, O. Corcho, P. Groth, E. Simperl, V. Tamma, A. G. Nuzzolese, M. Poveda-Villalón, M. Sabou, V. Presutti, I. Celino, A. Revenko, J. Raad, B. Sartini, P. Lisena (Eds.), The Semantic Web: ESWC 2024 Satellite Events, Springer Nature Switzerland, Cham, 2025, pp. 51–60.