A Transparent and Adaptive AI Assistant for Teaching Knowledge Engineering

Stefani Tsaneva*, Laura Waltersdorfer, Majlinda Llugiqi and Marta Sabou

Vienna University of Economics and Business, Vienna, Austria

Abstract

As generative AI tools become more widespread, students are increasingly using them for assistance with complex tasks such as modeling ontology constraints. While the success of large language models have been widely explored, in-use applications remain underdeveloped, and experimental findings are often inaccessible to students or novice engineers. As a result, learners do not fully benefit from AI-assisted support or fail to critically engage with AI generated outputs. To bridge this gap, we propose a transparent, research-informed AI Assistant framework that follows hybrid intelligence principles and aims to support Knowledge Engineering education, with a focus on modeling logical ontology constraints. Preliminary results suggest that such a system can improve the accuracy of student-generated ontology models by over 10%.

Keywords

Knowledge Engineering, Ontology Evaluation, Hybrid Intelligence, Transparent AI Systems, Large Language Models

1. Introduction

Knowledge Engineering (KE) comprises a range of activities such as knowledge acquisition and its representation in semantic models, such as ontologies [1]. Traditionally, KE demands high manual efforts to define, implement, and validate domain-specific requirements. Yet, there is a lack of tool support for many KE tasks [2, 3], increasing the risk of modeling errors, particularly when curators lack advanced KE training or deal with the modeling of complex logical constraints [4]. For example, the statement *"Every Professor supervises at least one Student"* can be modeled by defining the class Professor as equivalent to individuals who supervise at least one Student. While such modeling is logically consistent, it implies that anyone who supervises a student is, by definition, a professor—an unintended consequence. Such semantic inaccuracies cannot be detected by logical reasoners and traditionally require validation by domain experts or skilled knowledge engineers [4, 5]. Given the complexity of the ontology modeling task and limited specialized tools to support ontology curation, students and novice knowledge engineers often turn to generative AI tools [6].

While large language models (LLMs) can provide support, they also come with inherent challenges and limitations such as lack of reasoning skills [7] and hallucinations leading to inaccurate or misleading claims [8]. Yet students frequently over-rely on AI-generated outputs, accepting them without sufficient critical evaluation, especially when they lack knowledge in the subject [9, 10].

In the context of KE, the usage of an LLM which lacks the necessary capabilities to perform a concrete KE task, can fail to improve the quality of the developed resource and may even degrade it [11]. Although some research has explored which LLMs perform best [11] or how to prompt for ontology-evaluation tasks [12], these insights often remain inaccessible to the stakeholders who need them most. Students and novice practitioners are rarely exposed to such experimental results, in part because tools based on these findings are seldom developed. As a result, students frequently rely on familiar LLM-based applications such as ChatGPT even when better models might be available [9, 6]. This gap highlights

1st Workshop on Bridging Hybrid (Artificial) Intelligence and the Semantic Web (HAIBRIDGE 2025)

*Corresponding author.

stefani.tsaneva@wu.ac.at (S. Tsaneva); laura.waltersdorfer@wu.ac.at (L. Waltersdorfer); majlinda.llugiqi.rexha@wu.ac.at (M. Llugiqi); marta.sabou@wu.ac.at (M. Sabou)

0000-0002-0895-6379 (S. Tsaneva); 0000-0002-6932-5036 (L. Waltersdorfer); 0000-0002-5008-6856 (M. Llugiqi); 0000-0001-9301-8418 (M. Sabou)

the need for developing tools and educational resources that bridge the divide between research findings and practical applications in KE.

To address these limitations and extend the state of the art, we propose an AI Assistant framework relying on hybrid (Human-AI) intelligence principles, designed to support students in ontology creation and evaluation with focus on correct constraint (i.e., cardinality, universal and existential quantifiers) modeling. By combining multiple LLMs and assigning them to sub-tasks according to their experimentally observed strengths, the application offers context-sensitive support throughout different stages of the modeling process. Moreover, we aim to build trace-based transparency into the framework by advanced system tracing. Such a setup would allow for contextual explanations of generated outputs, a computation of custom confidence scores based on past performance and collection of traces which can support future adaptations of the system. The envisioned framework would allow students to benefit from AI-generated assistance while minimizing the risk of incorrect or misleading outputs. Ultimately, the proposed approach seeks to improve both the quality of student-generated ontologies and the learning experience itself. Preliminary result suggest that the AI-assisted approach can increase the accuracy of created ontology models by over 10%.

2. Envisioned Application

We propose an AI Assistant designed to (1) support students in detecting and classifying constraint-related modeling errors, (2) explain detected mistakes in an accessible and educational manner and (3) generate possible corrections. The system, illustrated in Fig. 1, comprises several key components: First, multiple LLMs are integrated to create a multi-functional AI Assistant that leverages the complementary strengths of different models. Second, an audit layer is incorporated to ensure system transparency and trustworthiness. Third, a feedback loop involving both student and instructor input fosters a hybrid human–AI co-learning process and supports continuous system adaptation.

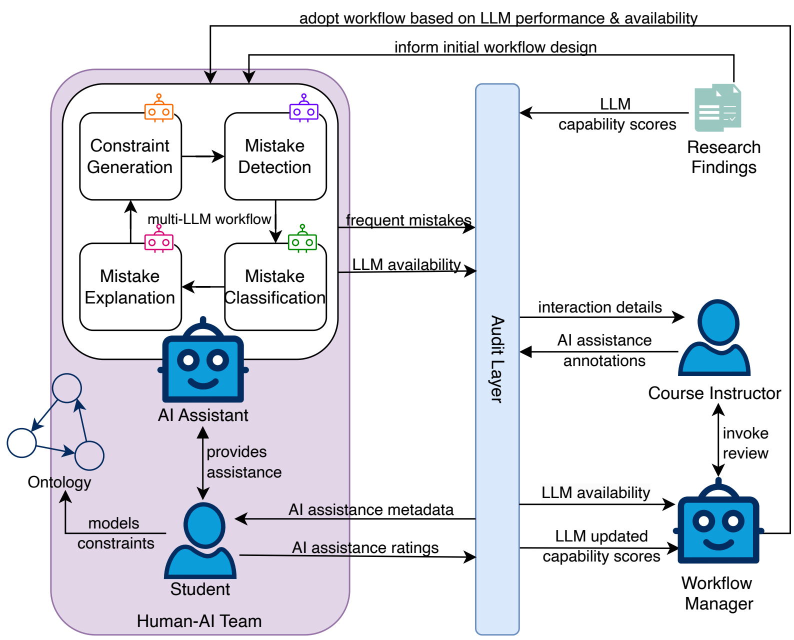

Figure 1: Collaborative Knowledge Engineering framework, incorporating an audit layer for transparency and traceability of generated outputs. Co-learning and adaptation are enabled through student feedback, course instructor reviews and a workflow manager.

Multi-LLM Workflow. The system distributes KE task responsibilities according to research-reported strengths of individual models in [11]. For instance, GPT4o has demonstrated strong adherence to instructions, making it well suited for modeling mistake classification. In contrast, Claude Sonnet excels in generating corrected models due to its generative flexibility. For the modeling mistake detection task, optimal performance varies across models depending on the constraint type. Integrating multiple LLMs therefore enables a task-specific, performance-driven workflow that leverages the unique advantages of each model.

Trace-based Transparency. To mitigate the challenges of transparency and traceability associated with multi-agent workflows, we integrate a Semantic-Web-based audit layer into the system following the AuditMAI methodology [13]. Based on identified audit questions, this layer captures detailed traces, including results of experimental investigations (research finding) and outcomes of prior executions (e.g., detected mistake types, LLM availability, or student feedback).

These audit traces can be leveraged to provide contextual information for each AI suggestion, such as which LLM generated the output and why it was chosen (e.g., superior past performance, fallback in case of unavailability). In addition, the audit layer can be utilized to compute custom confidence scores, based on model consensus and performance history, offering users insight into the reliability of AI outputs. This design encourages critical engagement with AI-generated content, rather than passive acceptance.

Co-learning and adaptation. Building on ideas from human–AI delegation frameworks [14], the system includes a Workflow Manager that operates outside the human–AI team, consisting of the student and AI Assistant. When recurring issues or misclassifications are flagged by the auditing layer, workflows can be adapted, such as switching LLMs for specific tasks, or escalated to instructors for review. This enables targeted intervention and the resulting feedback loop fosters co-learning: students gain from AI guidance, while the system improves through ongoing human oversight and real-world educational interaction.

3. Preliminary Results and Outlook

We leverage findings and collected annotations from a recent experimental assessment of LLM capabilities [11] to simulate initial AI-assisted workflows within the collaborative KE framework. We outline two concrete AI-supported workflows: in the first, students describe their intended constraint, and Claude Sonnet (claude-3-7-sonnet-20250219) generates the corresponding modeling. In the second, students create their own modelings, and various LLMs are used to detect potential errors. In particular, we select LLMs with the highest mistake detection recall scores reported in [11] - Claude Sonnet for cardinality constraints, Llama 3.3 (Llama-3.3-70B-Instruct-Turbo) for existential restrictions and DeepSeek V3 for universal restrictions. To enable AI-assisted correction, constraints flagged as incorrectly modeled, are replaced by alternative modelings generated by Claude Sonnet. Simulation results suggest that students' standalone performance in modeling logical ontology constraints (68.29% accuracy) can be improved by both AI-assistance workflows with over 10%. The constraint generation achieves 79.27% accuracy while the constraint validation and correction reaches 81.71%.

In comparison, AI-assisted workflows using GPT-4o, a model most familiar to students [15, 6], result in lower performance, failing to leverage the full potential of capability-informed LLM workflows. Constraint generation with GPT-4o results in 67.07% accuracy, under-performing standalone student modeling, while the GPT-assisted constraint validation and correction reaches 71.95%, offering only a slight improvement.

It should be noted that the simulated workflows do not yet incorporate AI-generated mistake explanations and do not consider cases in which students would revise their own models instead of fully relying on AI-generated suggestions. The LLM-generated explanations and contextual information provided by the auditing layer have the potential to further improve ontology modeling accuracy by fostering deeper understanding and enabling informed student revisions. In future work, we will implement the proposed framework and utilize it in Knowledge Engineering university courses. To assess the effectiveness and usability of the system, we plan a number of comprehensive user studies, including feedback surveys and interviews with students and instructors. This evaluation will inform future refinements of the framework and provide empirical insights into how LLM-based AI Assistants can be responsibly and meaningfully integrated into Knowledge Engineering education.

Acknowledgments

This research was funded in whole or in part by the Austrian Science Fund (FWF) BILAI (10.55776/COE12) and HOnEst (V 745) projects. For open access purposes, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission. Additionally, the work was supported by the PERKS (101120323) project, co-funded by the European Union. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Declaration on Generative AI

During the preparation of this work the authors used ChatGPT4 in order to suggest improvements to the readability and language of the manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

References

[1] R. Studer, V. Benjamins, D. Fensel, Knowledge engineering: Principles and methods, Data & Knowledge Engineering 25 (1998) 161–197.

[2] B. Zhang, V. Carriero, K. Schreiberhuber, S. Tsaneva, L. Gonzalez, J. Kim, J. de Berardinis, Ontochat: A framework for conversational ontology engineering using language models, in: The Semantic Web: ESWC 2024 Satellite Events, Springer Nature Switzerland, Cham, 2025, pp. 102–121.

[3] S. Tsaneva, K. Käsznar, M. Sabou, Human-centric ontology evaluation: Process and tool support, in: Int. Conf. on Knowledge Engineering and Knowledge Management, Springer, 2022, pp. 182–197.

[4] A. Rector, N. Drummond, M. Horridge, J. Rogers, H. Knublauch, R. Stevens, H. Wang, C. Wroe, Owl pizzas: Practical experience of teaching owl-dl: Common errors & common patterns, Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science) 3257 (2004) 63–81.

[5] M. Poveda-Villalón, A. Gómez-Pérez, M. C. Suárez-Figueroa, Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation, International Journal on Semantic Web and Information Systems (IJSWIS) 10 (2014) 7–34.

[6] T. Strubreiter, Larger Language Model usage when learning Ontology Engineering, Bachelor's thesis, Vienna University of Economics and Business, 2025.

[7] M. Amirizaniani, E. Martin, M. Sivachenko, A. Mashhadi, C. Shah, Can llms reason like humans? assessing theory of mind reasoning in llms for open-ended questions, in: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, 2024, pp. 34–44.

[8] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, P. Fung, Survey of hallucination in natural language generation, ACM Comput. Surv. 55 (2023).

[9] J. Freeman, Student generative ai survey 2025, Higher Education Policy Institute: London, UK (2025).

[10] G. Pitts, N. Rani, W. Mildort, E.-M. Cook, Students' reliance on ai in higher education: Identifying contributing factors, arXiv preprint (2025).

[11] S. Tsaneva, G. B. Herwanto, M. Llugiqi, M. Sabou, Knowledge Engineering with Large Language Models: A Capability Assessment in Ontology Evaluation, Submitted to Semantic Web Journal (Under Review).

[12] S. Kim, D. Kim, S. Hwang, K.-H. Lee, K. Lee, Llm-assisted ontology restriction verification with clustering-based description generation, IEEE Access 13 (2025).

[13] L. Waltersdorfer, F. J. Ekaputra, T. Miksa, M. Sabou, AuditMAI: Towards an infrastructure for continuous AI auditing, in: Austrian Symposium on AI, Robotics and Vision (AIRoV), Innsbruck University Press, 2025, pp. 4–14.

[14] A. Fuchs, A. Passarella, M. Conti, Optimizing delegation in collaborative human-ai hybrid teams, ACM Trans. Auton. Adapt. Syst. 19 (2024).

[15] H. Johnston, R. F. Wells, E. M. Shanks, T. Boey, B. N. Parsons, Student perspectives on the use of generative artificial intelligence technologies in higher education, International Journal for Educational Integrity 20 (2024) 2.