

# Combining Large Language Models & Knowledge Graphs for Energetic Renovation Suggestions

Emma Nuyts<sup>1,\*</sup>, Mathias De Brouwer<sup>1</sup>, Arne Vervae<sup>1</sup>, Bertrand Fontaine<sup>2</sup>, Sofie Van Hoecke<sup>1</sup>, Marc Delghust<sup>2,3</sup> and Femke Ongena<sup>1</sup>

<sup>1</sup>IDLab, Ghent University - imec, Technologiepark-Zwijnaarde 126, 9052 Ghent, Belgium

<sup>2</sup>Scone, Grotehondstraat 44, 2018 Antwerp, Belgium

<sup>3</sup>Building Physics Research Group, Ghent University, Sint-Pietersnieuwstraat 41 - B4, 9000 Ghent, Belgium

## Abstract

Meeting climate targets requires accelerating building energy renovations through tailored renovation suggestions. Large Language Models (LLMs) could support this process, reducing the time and effort required from practitioners, provided the LLM's outputs are grounded in domain expertise. Knowledge Graphs (KGs) provide a structured means of representing such expertise and offer a promising approach for supplying it to LLMs. In this context, KGs are not used to support multi-hop reasoning, but rather to ground a generation task by providing expert knowledge. However, it remains unclear to what extent LLMs benefit from the incorporation of structured expert knowledge represented as KGs in the energetic renovation domain. This short paper investigates the impact of input representation on the quality of LLM-generated renovation paths. Domain knowledge is captured using the Renovation and ENergy Optimization (RENO) ontology and provided to the model as five structured KG serializations (JSON-LD, Turtle, N-Triples, RDF/XML, and linearized) and one unstructured textual representation. Performance is evaluated using an LLM-as-a-judge approach, assessing completeness and ordering of the paths. The results indicate that structured KG representations can improve the completeness of generated renovation paths compared to the textual baseline, suggesting that structured knowledge can support the grounding of LLM outputs. However, the effectiveness of serialization formats varies across models, and no consistent best-performing format is identified. This short paper provides initial insights into how structured knowledge representations influence LLM performance on domain-specific generative tasks and highlights the need for further research into the interactions between ontology design, input representation, and prompt engineering.

## Keywords

Large Language Models, Knowledge Graphs, Energetic Renovations

## 1. Introduction

Meeting climate targets requires a substantial increase in the rate and quality of residential renovations [1]. At the same time, the availability of energy coaches and renovation advisors is limited [2]. Their expertise, however, is essential for ensuring that renovation paths are both correct and logically ordered. Automating parts of the renovation recommendation process using LLMs could therefore help reduce this workload, making it a solution for companies involved in renovation guidance that are seeking to scale their services. Yet, the energy renovation domain is characterised by complex, only partially documented expertise and strong context dependency, with practices varying across countries and even regions [3].

Large Language Models (LLMs) are increasingly being adopted to automate and support complex business processes. Their effectiveness is particularly evident in domains where extensive, well-structured documentation is available, allowing relevant domain knowledge to be directly provided as

*SKGi 2026: Scaling Knowledge Graphs for Industry Workshop, co-located with SEMANTICS'26: 22nd International Conference on Semantic Systems, September 15–17, 2026, Ghent, Belgium*

\*Corresponding author.

✉ Emma.Nuyts@UGent.be (E. Nuyts); Mathias.DeBrouwer@UGent.be (M. De Brouwer); Bertrand.Fontaine@Scone.ai (B. Fontaine); Sofie.VanHoecke@UGent.be (S. Van Hoecke); Marc.Delghust@Scone.ai (M. Delghust); Femke.Ongena<sup>1</sup>@UGent.be (F. Ongena<sup>1</sup>)

ORCID: 0000-0002-9206-9392 (E. Nuyts); 0000-0001-8769-6861 (M. De Brouwer); 0009-0009-4443-6977 (A. Vervae<sup>1</sup>); 0000-0003-4449-8683 (B. Fontaine); 0000-0002-7865-6793 (S. Van Hoecke); 0000-0003-4364-2700 (M. Delghust); 0000-0003-2529-5477 (F. Ongena<sup>1</sup>)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

input to the LLM generation task. In such cases, LLMs can leverage rich contextual information, such as protocols or guidelines, to produce reliable and domain-specific outputs [4].

However, many real-world domains lack such comprehensive and structured documentation. Instead, domain expertise is often fragmented across heterogeneous sources, or is even implicit, as is the case within the energetic renovation domain. This poses a challenge for the application of LLMs, as the absence of structured knowledge limits their ability to generate accurate and consistent results [5].

Knowledge Graphs (KGs) offer a way to model domain knowledge in a structured manner, especially when their underlying ontologies are co-developed with domain experts, such as the Renovation and ENergy Optimization (RENO) ontology. This structured representation may enhance LLM-generated outputs by grounding them in explicitly captured knowledge, in contrast to relying solely on unstructured natural language input.

Therefore, this work addresses the following research question: *To what extent does providing domain knowledge as a KG improve the ability of LLMs to generate correct and well-ordered energetic renovation paths, compared to unstructured textual input?* To evaluate this, this work investigates the impact of the KG input representation using five different serialization methods, as well as in a purely textual format. This enables a comparison of how structured versus unstructured inputs influence LLM performance within the renovation domain. The formulated research involves a generative task, where the model must produce a solution grounded in the provided information and does not involve retrieving answers through multi-hop reasoning.

## 2. Related Work

Prior work has investigated how the representation of structured knowledge influences the performance of LLMs, particularly when comparing KG inputs to unstructured natural language.

Chi and Hsieh [6] compare linearized triples with natural language descriptions in the context of temporal knowledge graphs. Their results indicate that unstructured textual descriptions can yield slightly higher performance on the investigated tasks, although no statistically significant differences were observed. This suggests that LLMs are capable of effectively leveraging fluent natural language, even when it encodes the same information as structured data.

In contrast, Dai et al. [7] find that linearized triples outperform natural language representations when LLMs are tasked with answering fact-intensive questions through multi-hop reasoning. Their findings indicate that explicitly structured inputs can improve the model’s ability to interpret and reason over knowledge graph data. However, the answers to the evaluated tasks can be directly retrieved from the KG through multi-hop reasoning, meaning the task largely evaluates the model’s ability to extract information rather than to generate novel, context-dependent outputs.

Wu and Tsioutsoulou [8] explore multi-hop reasoning through alternative representations of KGs, including natural language descriptions, JSON, and Python-based encodings with both static and dynamic relationships. Their results show that Python-based representations outperform the more conventional formats. However, such representations are not standardised and may be difficult to adopt.

Despite these efforts, existing work predominantly focuses on tasks where the correct answer can be directly derived from the provided knowledge, such as question answering or fact retrieval. As a result, these approaches primarily evaluate the ability of LLMs to access and extract information from structured inputs. Less attention has been given to scenarios where KGs serve as a form of contextual grounding for generative tasks, rather than as a source of explicit answers.

This work addresses this gap by evaluating how different KG serialization formats influence LLM performance in a generative task, namely the creation of energy renovation paths. In this setting, the ontology encodes the space of valid suggestions, but not the reasoning process or final solution. Consequently, the task cannot be solved through direct retrieval or multi-hop reasoning alone. Instead, it requires the LLM to generate outputs that adhere to the constraints and concepts defined in the ontology, allowing for a more realistic assessment of how structured knowledge supports generation in complex, domain-specific applications.

### 3. Methodology

To allow for a comparison between structured and unstructured representations, the building (or entity) information is provided to the model using six different input formats. More specifically, five structured representations are considered, i.e., four widely used RDF serializations, namely JSON-LD<sup>1</sup>, Turtle<sup>2</sup>, N-Triples<sup>3</sup>, and RDF/XML<sup>4</sup>; and a linearized triple format, where each fact is represented as a textual triple (s, p, o). In addition, a purely textual representation of the same information is included, enabling the evaluation of the added value of structured knowledge compared to unstructured natural language. Other representations are left for future work.

The methodology leverages the domain expertise captured in the RENO ontology (<https://w3id.org/re-no#>). This ontology was co-created with renovation domain experts within the RE-ENNOVATE project<sup>5</sup>, following the methodology of Ongenae et al. [9], and captures concepts that are relevant when designing energetic renovation suggestions, such as motivation, heating patterns, and previous renovation interventions. While the ontology specifies which overall renovation actions are possible, it does not encode the reasoning process to decide which sequence of renovations actions is ideal for a given building context. As such, the LLM must interpret the provided information and generate a solution that is consistent with the ontology, rather than extracting a predefined answer. The focus of this work is thus grounding the LLM's generations in domain-specific contextual knowledge, and not multi-hop reasoning or lookup.

A fixed zero-shot prompt template is used across all experiments, shown in Listing 1, to serve as a baseline. In this prompt, the ontology and the contextual building-specific information are provided in the same format. The model is instructed to analyse the provided information and generate a concise, actionable renovation path, which follows a logical sequence of actions and that maximises energy efficiency while remaining cost-effective. For the textual representation, all explicit references to the ontology are omitted from the prompt (i.e., line 5-9 deleted, line 15 adapted), and only a natural language description of the building is provided. This ensures that any observed differences in performance can be attributed to the presence and structure of the KG, rather than differences in task formulation.

```
1 You are an expert in energy-efficient and cost-effective home renovations.
2 Your task is to analyze the provided building-related information and propose a renovation
  path that maximizes **energy efficiency** while remaining **cost-effective**.
3 Focus on practical upgrades that reduce energy consumption and improve long-term operational
  performance. Take into account the most optimal sequence to maximize the benefits of each
  renovation step.
4
5 ### Description of the Input:
6 The RENO ontology, provided in {serialization} format, is designed to comprehensively capture
  all relevant data for generating informed renovation suggestions. This is provided
  together with a building mapped to this ontology, also provided in {serialization} format.
7
8 ### RENO Ontology:
9 {ontology in serialization: JSON-LD, Turtle, N-Triples, RDF/XML or linearized}
10
11 ### Entity Information:
12 {entity information in serialization: JSON-LD, Turtle, N-Triples, RDF/XML, linearized or text}
13
14 ### Renovation Plan:
15 Based on the information above, provide a concise and actionable renovation plan that is
  grounded in the ontology and tailored to the entity.
```

**Listing 1:** Prompt template to generate renovation paths across different input representations for a building.

<sup>1</sup><https://www.w3.org/TR/json-ld11/>, last visited 23/06/2026.

<sup>2</sup><https://www.w3.org/TR/turtle/>, last visited 23/06/2026.

<sup>3</sup><https://www.w3.org/TR/n-triples/>, last visited 23/06/2026.

<sup>4</sup><https://www.w3.org/TR/rdf-syntax-grammar/>, last visited 23/06/2026.

<sup>5</sup><https://www.buildwise.be/nl/onderzoek-innovatie/onderzoeksprojecten/herdenk-het-proces-van-energetische-renovaties-re-ennovate/>, last visited 28/05/2026.

## 4. Experiments

The experiments are designed to systematically evaluate the impact of input representation across multiple buildings. Each of the six input formats is evaluated on five buildings, using two different LLMs, and repeated three times to account for variability, resulting in 180 experiments. All code and the results of the experiments are available on GitHub: <https://github.com/predict-idlab/LLM-KG-for-EER>.

### 4.1. Data

The used entities stem from an artificially generated set of single-family homes in Flanders, developed and used by Scone<sup>6</sup> in its own platform, and was provided for this study. This parquet dataset is mapped to the RENO ontology through YARRRML<sup>7</sup>. This mapping enables the application of Jena reasoning rules, developed within the RE-ENNOVATE project, to derive the ground truth renovation paths. Importantly, these rules are not provided to the LLM, meaning the task cannot be solved through retrieval. As shown in Table 1, five entities with varying building characteristics were chosen to ensure a representative sample for the evaluation.

| Entity | Typology      | Construction Year | EPC | Renovation Actions |
|--------|---------------|-------------------|-----|--------------------|
| A      | Semi-Detached | 1951              | 388 | 9                  |
| B      | Attached      | 1832              | 262 | 9                  |
| C      | Attached      | 2023              | 50  | 3                  |
| D      | Semi-Detached | 2017              | 87  | 1                  |
| E      | Detached      | 1862              | 165 | 6                  |

**Table 1**

Overview of the selected building entities.

### 4.2. LLM-as-a-judge

As manual judgement of 180 outputs is too time-intensive and prone to human-evaluator bias, an LLM-as-a-judge component is added. The LLM is prompted to compare the ground truth renovation path (Listing 2a) with the generated renovation path (Listing 2b), and produce its verdict. The evaluation is based on two metrics: (i) completeness, reflecting missing and additional actions, and (ii) sequence, reflecting ordering errors. As shown in Listing 2c, the model produces a list of errors instead of numerical scores, to allow explainability and mitigate counting errors. In the post-processing of the results, the penalty count for each generated renovation path is calculated as the length of the corresponding list(s).

### 4.3. Technical Setup

All experiments were run locally through the Ollama API<sup>8</sup> and all prompts were posted using `/generate` instead of `/chat`, to prohibit data leakage. To generate renovation suggestions, both Gemma 4 (31b)<sup>9</sup>, and Mistral Small 3.2 (24b)<sup>10</sup> were used. Gemma 4 (31b) was selected for its strong performance, while Mistral Small 3.2 was included as a European-developed model to address data sovereignty considerations and compliance with a European provider, increasingly relevant in the renovation sector [10]. To generate the renovation paths, the temperature was set to 0.7 and the top-p to 0.95. For the LLM-as-a-judge, Gemma 4 (31b) was used with a temperature of 0.0 and top-p of 1.0 to ensure a deterministic output.

<sup>6</sup><https://www.scone.ai/>, last visited 28/05/2026.

<sup>7</sup><https://rml.io/yarrml/spec/>, last visited 28/05/2026.

<sup>8</sup><https://docs.ollama.com/api/introduction>, last visited 28/05/2026.

<sup>9</sup><https://ai.google.dev/gemma/docs/core>, last visited 28/05/2026.

<sup>10</sup><https://mistral.ai/news/mistral-small-3>, last visited 28/05/2026.

(a) Ground truth.

1. insulate flat roof
2. insulate wall on interior
3. replace glazing
4. insulate floor above basement
5. replace ventilation system
6. place PV-panels on roof

(b) Generated renovation path.

1. insulate flat roof
2. insulate wall on interior
3. insulate floor above basement
4. replace glazing
5. place air-water heat pump
6. replace ventilation system

(c) Output of the scoring by the LLM-as-a-judge.

```
{
  "completeness":{
    "missing_actions":["place PV-panels on roof"],
    "extra_actions":["place air-water heatpump"]
  },
  "sequence":{
    "order_errors":["insulate floor above basement"]
  }
}
```

**Listing 2:** Example of the LLM-as-a-judge evaluation, showing the ground truth renovation path, the generated renovation path, and the extracted errors for completeness and sequence, resulting in a penalty count of 2 for completeness and 1 for sequence.

## 5. Results and Discussion

The median penalty results across the three runs are presented in Table 2, where entities A–E correspond to the buildings described in Table 1. For Mistral Small 3.2, Turtle achieves the best performance for completeness across most entities, while the textual representation consistently yields the highest penalties. This suggests that, for this model, explicitly structured input supports the generation of more complete renovation paths. For the sequence metric, the differences between formats are more limited, with JSON-LD and text achieving perfect scores. When considering both metrics jointly, Turtle emerges as the best-performing format, whereas the textual representation performs the worst.

For Gemma 4, a different pattern can be observed. JSON-LD and Turtle achieve the best results in terms of completeness, while the textual representation again performs worse overall, although with more variation across entities. For the sequence metric, almost all formats achieve perfect or near-perfect scores, indicating that this model is generally capable of producing logically ordered renovation steps, regardless of the input representation. When combining both metrics, JSON-LD performs best overall, while the linearized and textual formats result in higher penalties.

Across both models, an observation is that structured representations tend to outperform the purely textual baseline in terms of completeness. This indicates that providing explicitly structured knowledge can support LLMs in generating outputs grounded in the domain knowledge defined in the ontology. However, the impact of the specific serialization format appears to be model-dependent. Another observation is that the sequence metric shows relatively limited variation across formats and models. In most cases, the generated renovation steps are ordered correctly, even when completeness is suboptimal. This may indicate that LLMs are inherently capable of producing plausible action sequences, whereas correctly identifying all required actions benefits more directly from structured domain knowledge.

It is, however, important to interpret these results with caution. A notable variability is observed between different runs of the same experiment, with differences of up to nine penalty points for a single model. This variability persists despite identical experimental conditions. As a result, the observed differences between input formats should not be generalised. Finally, it should be noted that the experiments are conducted on a limited set of synthetic entities, and that the ontology modelling choices may influence the results. Consequently, while the findings provide initial presumptions that structured KG representations can improve the grounding of LLM outputs when generating renovation paths, further experimentation is required to draw more general conclusions.

| (a) Median scores for Mistral Small 3.2 (24b). |            |   |   |    |     |   |      | (b) Median scores for Gemma 4 (31b). |            |   |   |   |   |   |      |
|------------------------------------------------|------------|---|---|----|-----|---|------|--------------------------------------|------------|---|---|---|---|---|------|
|                                                |            | A | B | C  | D   | E | Mean |                                      |            | A | B | C | D | E | Mean |
| Completeness                                   | JSON-LD    | 6 | 6 | 7  | 3   | 4 | 5.2  | Completeness                         | JSON-LD    | 4 | 4 | 1 | 2 | 1 | 2.4  |
|                                                | RDF/XML    | 5 | 4 | 7  | 5   | 3 | 4.8  |                                      | RDF/XML    | 5 | 4 | 1 | 2 | 2 | 2.8  |
|                                                | N-Triples  | 4 | 5 | 7  | 5   | 4 | 5.0  |                                      | N-Triples  | 5 | 4 | 1 | 2 | 2 | 2.8  |
|                                                | Turtle     | 4 | 4 | 4  | 4   | 2 | 3.6  |                                      | Turtle     | 4 | 5 | 1 | 1 | 1 | 2.4  |
|                                                | Linearized | 6 | 4 | 8  | 6   | 3 | 5.4  |                                      | Linearized | 6 | 4 | 1 | 2 | 2 | 3.0  |
|                                                | Text       | 9 | 5 | 10 | 7.5 | 8 | 7.9  |                                      | Text       | 2 | 2 | 5 | 5 | 2 | 3.2  |
| Sequence                                       | JSON-LD    | 0 | 0 | 0  | 0   | 0 | 0.0  | Sequence                             | JSON-LD    | 0 | 0 | 0 | 0 | 0 | 0.0  |
|                                                | RDF/XML    | 3 | 3 | 0  | 0   | 2 | 1.6  |                                      | RDF/XML    | 0 | 0 | 0 | 0 | 0 | 0.0  |
|                                                | N-Triples  | 1 | 0 | 1  | 0   | 0 | 0.4  |                                      | N-Triples  | 0 | 0 | 0 | 0 | 0 | 0.0  |
|                                                | Turtle     | 0 | 1 | 0  | 0   | 0 | 0.2  |                                      | Turtle     | 0 | 1 | 0 | 0 | 0 | 0.2  |
|                                                | Linearized | 0 | 1 | 0  | 0   | 3 | 0.8  |                                      | Linearized | 0 | 0 | 0 | 0 | 1 | 0.2  |
|                                                | Text       | 0 | 0 | 0  | 0   | 0 | 0.0  |                                      | Text       | 0 | 0 | 0 | 0 | 0 | 0.0  |

**Table 2**

Median penalty scores for completeness and sequence across input formats, evaluated on two LLMs and five entities (A–E), aggregated over three runs (lower scores indicate better performance; **bold** is best).

## 6. Conclusions and Future Work

This short paper explores whether providing domain knowledge through a KG benefits LLMs in the energetic renovation domain, where knowledge is not explicitly documented, by comparing different KG serialization formats with an unstructured textual representation. The results suggest that structured representations can improve the completeness of generated renovation paths compared to the textual baseline, although the effectiveness of individual serialization formats varies across models and no single format consistently outperforms the others. While the ordering of renovation steps is generally correct across all input formats, indicating that LLMs can derive plausible action sequences from the ontology, the main contribution of structured knowledge appears to be improved coverage of relevant actions rather than better sequencing. As this is an exploratory study, the findings should be interpreted as indicative trends rather than definitive conclusions, particularly given the limited set of synthetic entities and potential dependence on ontology modelling choices. Future work can extend this study in several directions. First, different prompt engineering techniques could be explored in combination with the input formats to assess their impact on performance. Second, the role of the ontology structure itself could be investigated, for example by filtering the TBox to capture only relevant concepts, or reducing the metadata in the ontology. Finally, validating the approach on real-world data and a larger set of entities would be an important step towards assessing its applicability in practical renovation scenarios.

## Acknowledgments

This work has been financially supported by the Flux 50 Icon project RE-ENNOVATE (grant number HBC.2022.0116) and the Flanders AI Research programme.

## Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

## References

- [1] P. O. of the European Union, Delivering the EU green deal, 2025. doi:10.2760/3105205.
- [2] Groen, Wachttijden voor energie- en renovatieadvies pieken, 2022. URL: [https://www.groen.be/wachttijden\\_voor\\_energie\\_en\\_renovatieadvies\\_pieken](https://www.groen.be/wachttijden_voor_energie_en_renovatieadvies_pieken), last visited 29/06/2026.
- [3] M. Hummel, R. Büchele, A. Müller, E. Aichinger, J. Steinbach, L. Kranzl, A. Toleikyte, S. Forthuber, The costs and potentials for heat savings in buildings: Refurbishment costs and heat saving cost curves for 6 countries in Europe, *Energy and Buildings* 231 (2021) 110454. doi:10.1016/j.enbuild.2020.110454.
- [4] J. Lee, W. Jung, S. Baek, In-House Knowledge Management Using a Large Language Model: Focusing on Technical Specification Documents Review, *Applied Sciences* 14 (2024). doi:10.3390/app14052096.
- [5] K. Liu, Z. Fu, C. Chen, W. Zhang, R. Jiang, F. Zhou, Y. Chen, Y. Wu, J. Ye, Enhancing LLM's Cognition via Structurization, in: A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, C. Zhang (Eds.), *Advances in Neural Information Processing Systems*, volume 37, Curran Associates, Inc., 2024, pp. 133710–133742. doi:10.52202/079017-4249.
- [6] J.-A. Chi, S.-K. Hsieh, Structured vs. unstructured inputs in LLMs: Evaluating the semantic and pragmatic predictive power in abnormal event forecasting, in: K.-W. Chang, K.-H. Lu, C.-K. Yang, Z.-R. Tam, W.-Y. Chang, C.-C. Wang (Eds.), *Proceedings of the 37th Conference on Computational Linguistics and Speech Processing (ROCLING 2025)*, Association for Computational Linguistics, National Taiwan University, Taipei City, Taiwan, 2025, pp. 237–248. URL: <https://aclanthology.org/2025.rocling-main.25/>.
- [7] X. Dai, Y. Hua, T. Wu, Y. Sheng, Q. Ji, G. Qi, Large language models can better understand knowledge graphs than we thought, *Knowl. Based Syst.* 312 (2025) 113060.
- [8] X. Wu, K. Tsioutsoulis, Thinking with Knowledge Graphs: Enhancing LLM Reasoning Through Structured Data, 2024. URL: <https://arxiv.org/abs/2412.10654>. arXiv:2412.10654.
- [9] F. Ongenae, P. Duysburgh, N. Sulmon, M. Verstraete, L. Bleumers, S. De Zutter, S. Verstichel, A. Ackaert, A. Jacobs, F. De Turck, An ontology co-design method for the co-creation of a continuous care ontology, *Appl. Ontol.* 9 (2014) 27–64. doi:10.3233/AO-140131.
- [10] B. Weber, M. Achenbach, Data sovereignty within the construction process, in: *ECPPM 2022 - eWork and eBusiness in Architecture, Engineering and Construction 2022*, 2023.