Neuro-Symbolic AI for Evidence-Based Renewable Energy Planning in Sub-Saharan Africa

Janice Anta Zebaze[1] , **Azanzi Jiomekong**[2,3] , **Germaine Djuidje Kenmoe**[1] and **Maria-Esther Vidal**[3]

[1]Department of Physics, University of Yaounde 1, Yaounde, Cameroon
[2]Department of Computer Science, University of Yaounde I, Yaounde, Cameroon
[3]TIB Leibniz Information Centre For Science and Technology, Hannover, Germany
antazebaze@gmail.com, fidel.jiomekong@facsciences-uy1.cm, kdjuidje@yahoo.fr, maria.vidal@tib.eu

Abstract

Renewable energy is a key alternative to fossil fuels and their environmental harms. However, sub-Saharan African countries remain behind in their implementation, due to poor investment in the appropriate resource, among other causes. This information is dispersed as unstructured knowledge across scientific papers, making the task manually demanding for decision and policy makers. In this work, we combine the strength of Large Language Models (LLMs) and structured knowledge (Knowledge Graphs) in a neuro-symbolic framework to tackle this issue. We implement this approach in the hydropower field, building a question-answering system that uses Retrieval Augmented Generation (RAG) on a knowledge graph. Using LLMs such as LLaMA 3.1 8B, LLaMA 3.3 70B, LLaMA 4 Scout 17B, and Kimi K2, we observe an improvement in exact match of $48 - 60\%$ across all models. These results support the ability of this framework to participate in evidence-based energy planning in developing countries.

1 Introduction

Renewable energy is a key alternative to fossil fuels and their environmental harms. Several countries are actively switching to these alternatives, however, sub-Saharan African countries are still behind. This is due to lack of knowledge on the potential of renewable sources in given locations [Mbasso and others, 2025] and also on the poor investment on the appropriate resource among several causes. This information is however available in some cases and dispersed as unstructured knowledge across scientific papers. This spread makes the task manually demanding for decision and policy makers to manually browse through this corpus of information across different fields. Also in this current state, the information can not be processed nor accessed by a random person in a remote village. Such a person may not have the skills to interpret this information, nor the knowledge on how to implement them. Solving this issue, could foster economic, structural and industrial development, through access to electricity and creation of new jobs in these fields.

In parallel, LLMs offer a promising way to generate and explain information. A capacity interesting and sought in this work, since they could analyze and breakdown the complex scientific notions into a simpler, comprehensible and accessible version. Such that a commoner can process it and understand why a particular resource is appropriate in his given location. However, despite their performances, LLMs are prone to hallucinations, and may generate non-factual information. An issue which can be overcome through a RAG approach which sources the information generated by the LLM. We push this further by associating this framework a symbolic component such as a Knowledge graph, which contains structured manually curated information from research papers. A combination which is neuro-symbolic, and which has shown promising results in several fields [Bhuyan *et al.*, 2024]. To our knowledge this approach combining RAG and a scholarly knowledge graph is yet to be implemented in the renewable energy domain for decision making.

2 Research Objectives

This work addresses the design of neuro-symbolic question-answering systems for accurate and explainable renewable energy assessment. This objective can be broken down into the following research questions:

- Combine a knowledge graph with a LLM and measure the impact of this neuro-symbolic grounding.
- Define traversal strategies to extract from the Knowledge graph the data and knowledge that will be used to enhance the LLMs' behavior.
- Integrate domain-specific symbolic constraints such as wind shear equations for wind energy, and head equations for hydropower into the symbolic layer.
- For locations with sparse or absent knowledge base coverage, explore embedding-based and cross-domain transfer approaches.

3 Methodology

The methodological framework is broken down into the following steps:

Knowledge curation. To go along with this step, research questions and problems were first identified. Next properties to answer them were defined, following a manual extraction and documentation in the Open Research Knowledge Graph [1] [Jiomekong and Tiwari, 2024]. Papers were then curated on the topic of wind, solar and hydropower [2].

System architecture. The architecture consists of three main components: A neural layer provided by the LLM, a symbolic layer provided by the knowledge graph and a RAG mechanism that links them all. For the experimentation, four LLMs were chosen: LLaMA 3.1 8B, LLaMA 3.3 70B, LLaMA 4 Scout 17B, and Kimi K2.

Evaluation. To evaluate this framework, we compared the RAG + knowledge graph vs LLM only performance. This was done by defining a set of benchmark questions with ground truth answers taken directly from the knowledge graph. Several metrics such as exact match, faithfulness, relevance, numerical recall, and numerical hallucination rate were used.

4 Preliminary Results

A first implementation of the RAG + KG approach was on the topic of hydropower. 50 papers covering countries such as Nigeria, Ethiopia, Ghana, and India made the symbolic layer in ORKG. Each study subgraph is verbalized by template into four chunks: overview, key findings, location and capacity, and methodology. These verbalizations are embedded using `multilingual-e5-large` (1024 dimensions) into a Pinecone vector store (199 vectors), each vector keeping its ORKG resource URI for provenance. For each query, the user's question is matched to the stored chunks through cosine similarity and the best 5 results given to the LLM, with the instruction: *"Answer using ONLY the provided studies. If the answer is not found, say: I don't have enough information."*, This instruction explicitly handles knowledge gaps and reduces hallucinations. Questions the system cannot answer point to gaps in the knowledge base, which we can then fill by curating more papers in ORKG.

Proceeding with the evaluation, we defined a set of 50 questions with different complexity levels, and including out of knowledge content. Next, we chose four LLM: LLaMA 3.1 8B, LLaMA 3.3 70B, LLaMA 4 Scout 17B, and Kimi K2 and evaluated LLM only and RAG + KG. Results show consistent improvements with RAG grounding: exact match improves by $48-60\%$, numerical recall by $33-51\%$, and faithfulness by $22-30\%$. This increase is observed across all four models, suggesting an increase due to the knowledge graph rather than the model's capacity. When asked questions outside the knowledge base, the grounded models said they did not know based on the instruction given in all cases, while the ungrounded models made up answers 75% of the time. Looking at which question types the system struggled with, country and methodology questions were mostly answered well, with only around 20% of them failing. However, questions on economic data and policy had failure rates between 75 and 100%, showing that these topics need more papers to be added to the knowledge base. The codes and question set used in this work are available on GitHub [3].

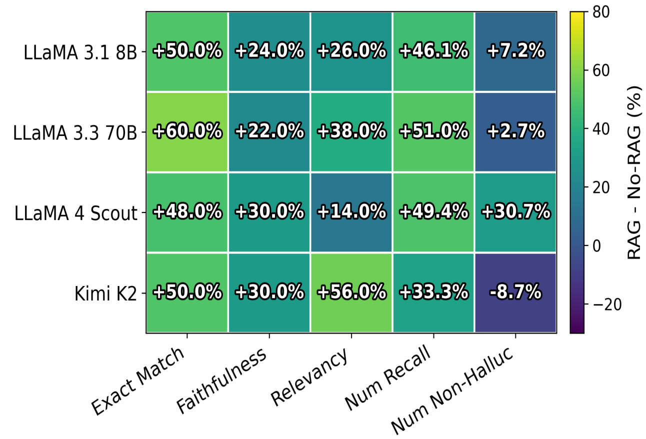

Figure 1: Heatmap of the gain in performance of the RAG + KG approach

5 Conclusion

Neuro-symbolic AI is a promising direction for making scientific knowledge actionable in high-stakes domains. Applied to renewable energy assessment, it could be a breakthrough for energy planning in developing countries, providing sourced, location-specific guidance without expert training. This framework could motivate evidence-based investment in regions still unaware of their renewable energy potential, contributing to affordable and clean energy (SDG 7), decent work and economic growth (SDG 8), and reduced inequalities (SDG 10) across Sub-Saharan Africa.

6 Acknowledgments

This work is partially supported by the THIRDWAVE project, funded by the European Union under the Horizon Europe MSCA Staff Exchanges programme (Grant Agreement No. 101236394).

References

[Bhuyan *et al.*, 2024] Bikram Bhuyan, Ravi Tomar, and Tourangbam Singh. Neuro-symbolic artificial intelligence: a survey. *Neural Computing and Applications*, 36:1–36, 06 2024.

[Jiomekong and Tiwari, 2024] Azanzi Jiomekong and Sanju Tiwari. An approach based on open research knowledge graph for knowledge acquisition from scientific papers. *The Electronic Library*, 42(3):413–442, 2024.

[Mbasso and others, 2025] Wulfran Fendzi Mbasso et al. Policy-driven expansion of renewable energy in cameroon: A technical and sustainability-centered analysis of growth trends and cross-sectoral impacts (2015–2024). *Energy Strategy Reviews*, 62:101912, 2025.

[1] https://orkg.org/

[2] https://orkg.org/fields/R267/Energy_Systems

[3] https://github.com/AntaZebaze/RAG-KG-Hydropower