

Semantic enrichment of causal graphs for strategic foresight

Jože M. Rožanec¹[0000-0002-3665-639X] and Gaël Gendron^{2,3}[0000-0002-2457-934X]

¹ Jožef Stefan Institute, Jamova ulica 39, Ljubljana, Slovenia
joze.rozanec@ijs.si

² NAOInstitute, University of Auckland, New Zealand
gael.gendron@auckland.ac.nz

Abstract. Strategic foresight has been identified as a key tool to enhance policymaking and guide decision-making in private and public organizations. While the potential of artificial intelligence has been recognized in this domain, it has only been applied to specific tasks, and no AI-first approach has been developed yet. Among the multiple strategic foresight methodologies, one of the most frequently used ones is scenario planning. Nevertheless, creating such scenarios requires specific domain knowledge, particularly about causal relationships, to understand how forces of change may affect potential future outcomes. In this research, we describe some early results we obtained from semantic enrichment performed on causal graphs extracted from media news. The experiments were performed using ChatGPT 4o on 50 media events that correspond to oil prices in the first quarter of 2023. The results show that when performing semantic linking, different results are obtained if the extracted causal variables or the causal relationships are considered. While there is complete agreement regarding the assigned wiki concepts in 9% of the cases, such agreement falls below 33% in most of the cases. Furthermore, nearly 23% of the proposed wiki concepts do not correspond to real ones.

Keywords: Causal discovery · NLP · Graphs · Semantic enrichment · Strategic foresight.

1 Introduction and related work

Strategic Foresight aims to provide a structured approach to gathering information regarding plausible futures to prepare for change adequately. It frequently leverages experts' knowledge regarding trends and emerging issues to understand better how decisions and policies may influence the future and guide strategic planning and policy-making [4]. The ability to lead to better future outcomes has promoted an increasing adoption in the private and public sectors [12].

While artificial intelligence is increasingly being used across different domains, it has only been adopted to aid in specific strategic foresight tasks [11], and no end-to-end tool exists to (semi-)automate the complex process performed

by domain experts. Among the most frequently used strategic foresight methods, we find scenario planning [1], which aims to foresee relevant scenarios based on trends and factors of influence to understand better how actions can influence the future [14]. Building such scenarios requires identifying the driving forces of change and possible outcomes while acknowledging the accompanying uncertainties. Knowledge about such causal relationships and context is currently owned by experts, who then estimate plausible outcomes. Artificial intelligence could automate this process using LLMs and causal inference frameworks to extract and test causal relationships and infer future scenarios. In particular, we envision using LLMs to extract candidate causal relationships from media news, assess whether such relationships effectively mean a causal relationship, and merge them into a causal graph [2]. The resulting causal graph could be analyzed, and subgraphs of interest could be identified to build scenarios about plausible futures for human examination.

LLMs have shown promising results but face challenges in causal inference from text due to the inherent complexity of natural language, which is often unstructured, high-dimensional, and semantically ambiguous [9]. LLMs have been used for causal discovery, effect estimation, and tasks like counterfactual reasoning. Research on causal discovery focuses on pairwise causal direction [8], though issues arise with LLMs repeating embedded knowledge [15] or inferring causal relations from entity order [7]. For full causal graph discovery, LLMs have been mostly applied on datasets, where they have even outperformed baselines [10]. Among the few studies focusing on extracting causal links from the text, we can mention [3] and [5]. Jin et al. [6] proposed an alternative approach using a chain-of-thought prompting strategy to extract a causal graph and context (e.g., conditional or interventional probabilities) to perform correct causal inferences and answer causality questions.

Gendron et al. [2] have recently described an approach leveraging LLMs for causal relationship extraction from media news. Their approach shows promise in extracting observed and hidden causal variables and their causal relationships. Nevertheless, the experiments were executed only on a handful of cases and the causal variables and relationships they extract are encoded as strings. While encoding such knowledge as strings can sometimes provide rich expressiveness, we consider that they could be further enriched associating them with a semantic concept. This would allow for a normalized understanding of the causal variables and enable further interoperation with ontologies and knowledge bases. This work, therefore, explores on how such enrichment can be performed, and provides a quality assessment of the causal graphs described above.

2 Experiments and results

This research is part of a wider research effort on how artificial intelligence can be applied to enhance strategic foresight [13]. The manuscript aims to describe experiments and results we obtained when pursuing two research goals (a) semantically enrich the causal variables identified in media news by following the

prompts described by Gendron et al. [2] and (b) assess how reliable is the semantic linking obtained through LLM prompting.

Methodology The experiments were performed considering 50 media news events related to oil prices which were reported in the first quarter of 2023 and extracted from EventRegistry. A total of 169 causal graphs from the events were obtained by applying the prompt described in Gendron et al. [2]. Their observed edges and nodes were further processed to semantically enrich them. In particular, two enrichments were performed: (i) considering the strings describing the already extracted causal variables and (ii) considering the edge description that explains how two causal variables are related to each other. The outcomes were then compared to understand to what extent did (i) and (ii) lead to the same understanding. To that end, we measured the Jaccard similarity of the extracted wiki concept URLs for causal variable pairs associated through a causal relationship. Finally, we measured how many of the wiki concept URLs issued by the LLM corresponded to real ones. The work was performed using the OpenAI ChatGPT 4o model.

Results The results we obtained from the semantic enrichment show that in most cases the wiki concepts retrieved from the causal variables do not match the ones described when obtained from the causal relationships. In fact, in 54% of the cases, the Jaccard similarity was zero, in 36% of the cases the Jaccard similarity was 0.33, and in just 9% of the cases we could observe a perfect match. We randomly picked some cases to understand the quality of the extracted wiki concepts and understand whether it would be better to consider the extracted causal variables or rely on the causal relationships. One such example was »*Influence of Russian supply cuts on oil prices*«. The causal graph prompt correctly identified that »*Russian supply cuts*« influence »*Oil prices*«. When executing the additional prompts, the causal variables were mapped to *Economy of Russia: Natural resources and energy exports* and *Oil price* and the causal relationship to *Oil supply* and *Oil prices*, which is not tied to a particular country and more accurately describes the key resource being considered in the market dynamics. Finally, when assessing how many unique wiki links were valid, we found that 204 unique links corresponded to real wiki concepts, while 59 entries were misleading.

Conclusion and Future work Our research regarding causal extraction and counterfactuals has not been tested on an extensive dataset. Further research is required to understand (i) how causal graph extraction prompts can be enhanced to yield more accurate results, (ii) how to enhance the semantic enrichment to avoid LLM hallucination scenarios. Future work will address this gap by (a) refining the causal graph extraction, (b) enrich the semantic linking considering domain-specific ontologies, and (c) testing the proposed approach on a dataset spanning multiple years of media news.

Acknowledgments. This research was developed as part of the Graph-Massivizer project funded under the Horizon Europe research and innovation program of the

European Union under grant agreement 101093202 and supported by the Slovenian Research Agency.

Disclosure of Interests. The authors declare that they have no competing interests.

References

1. Ebadi, A., Auger, A., Gauthier, Y.: Detecting emerging technologies and their evolution using deep learning and weak signal analysis. *Journal of Informetrics* **16**(4), 101344 (2022)
2. Gendron, G., Rožanec, J.M., Witbrock, M., Dobbie, G.: Counterfactual causal inference in natural language with large language models. arXiv preprint arXiv:2410.06392 (2024)
3. Gopalakrishnan, S., Garbayo, L., Zadrozny, W.: Causality extraction from medical text using large language models (llms). arXiv preprint arXiv:2407.10020 (2024)
4. Greenblott, J.M., O’Farrell, T., Olson, R., Burchard, B.: Strategic foresight in the federal government: a survey of methods, resources, and institutional arrangements. *World futures review* **11**(3), 245–266 (2019)
5. Hobbhahn, M., Lieberum, T., Seiler, D.: Investigating causal understanding in llms. In: *NeurIPS ML Safety Workshop* (2022)
6. Jin, Z., Chen, Y., Leeb, F., Gresele, L., Kamal, O., Zhiheng, L., Blin, K., Adaoto, F.G., Kleiman-Weiner, M., Sachan, M., et al.: Cladder: Assessing causal reasoning in language models. In: *Thirty-seventh conference on neural information processing systems* (2023)
7. Joshi, N., Saparov, A., Wang, Y., He, H.: Llms are prone to fallacies in causal inference. arXiv preprint arXiv:2406.12158 (2024)
8. Kıcıman, E., Ness, R., Sharma, A., Tan, C.: Causal reasoning and large language models: Opening a new frontier for causality. arXiv preprint arXiv:2305.00050 (2023)
9. Ma, J.: Causal inference with large language model: A survey. arXiv preprint arXiv:2409.09822 (2024)
10. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al.: Training language models to follow instructions with human feedback. *Advances in neural information processing systems* **35**, 27730–27744 (2022)
11. Pratt, L., Bisson, C., Warin, T.: Bringing advanced technology to strategic decision-making: The decision intelligence/data science (di/ds) integration framework. *Futures* **152**, 103217 (2023)
12. Rosa, A.B., Gudowsky, N., Repo, P.: Sensemaking and lens-shaping: Identifying citizen contributions to foresight through comparative topic modelling. *Futures* **129**, 102733 (2021)
13. Rožanec, J., Šircelj, B., Cochez, M., Leban, G.: Back to the future: predicting causal relationships influencing oil prices. In: *The Second Tiny Papers Track at ICLR 2024*
14. Wilkinson, A.: *Strategic foresight primer*. European Political Strategy Centre (2017)
15. Zečević, M., Willig, M., Dhami, D.S., Kersting, K.: Causal parrots: Large language models may talk causality but are not causal. arXiv preprint arXiv:2308.13067 (2023)