

EGU24-21883, updated on 14 Oct 2024

<https://doi.org/10.5194/egusphere-egu24-21883>

EGU General Assembly 2024

© Author(s) 2024. This work is distributed under the Creative Commons Attribution 4.0 License.



## Large Language Models for Causal Discovery in the Earth Sciences

**Gustau Camps-Valls**, Kai-Hendrik Cohrs, Emiliano Diaz, Vasileios Sitokonstantinou, and Gherardo Varando

Image Processing Laboratory (IPL), Universitat de València, València, Spain

Causality is essential for understanding complex systems like the Earth and climate, where a plethora of intertwined variables and processes happen in the wild. Constructing causal graphs often relies on either data-driven or expert-driven approaches, both fraught with challenges. The former methods, like the celebrated Peter-Clark (PC) algorithm, face issues with data requirements and assumptions of causal sufficiency, while the latter demand substantial time and expertise.

This work explores the capabilities of Large Language Models (LLMs) as an alternative to domain experts for causal graph generation. We frame conditional independence queries as prompts to LLMs and employ the PC algorithm with the answers. The performances of the LLM-based conditional independence oracle on systems with known causal graphs show a high degree of variability. We improve the performance through a proposed statistical-inspired voting schema that allows control over false-positives and false-negatives rates. We apply our chatPC algorithm to understand the causal relations between complex sets of variables (social, economic, conflicts, environmental, and climatic factors) in two pressing problems: population displacement and food insecurity in Africa. We find plausible graphs as corroborated by experts in the humanitarian sector, finding traces of causal reasoning in the model's answers. We posit that LLM-based causality is a new, promising, alternative avenue for automated causality, especially indicated for rapid response and data-scarce regimes.