

LLM-Based Extraction of Scenario-Relevant Dependencies for Emerging Energy Technologies

Samuel Kieling¹, Jan-Maris Göpfert¹, Noor Putri Titan Hartono¹, Jochen Linßen¹, Jann Michael Weinand¹

¹ Institute of Climate and Energy Systems (ICE) – Jülich Systems Analysis (ICE-2), Forschungszentrum Jülich GmbH, Jülich, Germany

ABSTRACT

Emerging energy technologies, such as fusion energy, carry uncertainty regarding their feasibility, timelines, and techno-economic characteristics. Energy system models typically rely on assumptions about these aspects, which are inherently hard to evaluate with regard to their likelihood or consistency. Structured foresight methods, such as cross-impact analysis, morphological analysis, and expert elicitation, target internally consistent sets of assumptions. However, they rely on small expert panels, making them costly, slow to adapt to fast-evolving research fields, and their results difficult to trace to primary evidence. Additionally, these methods force a single consensus value per relationship, obscuring genuine disagreement and exceptional cases.

This work builds on the premise that the expert knowledge these methods draw on is already dispersed across the scientific literature, in the form of causal and conditional claims linking a technology's parameters to internal and external factors. We therefore present a method based on large language models for the provenance-preserving, machine-readable extraction of scenario-relevant causal claims from scientific literature. Each extracted claim captures the cause and effect entities, the affected parameters and the polarity of their coupling, the conditions and scope under which it holds, and its epistemic strength and evidence basis. Extracted entities are mapped to a shared vocabulary and enriched with taxonomic relationships. Notably, every element remains traceable to its originating text.

The resulting knowledge graph is designed to make agreement and conflict among literature-derived claims explicit. Mapping scenario assumptions onto the graph could then reveal assumptions that corroborate or contradict one another, and thereby help constrain the space of jointly plausible parameter sets.

We demonstrate the method on the full texts of 500 open-access articles relevant to fusion energy commercialization. Extraction recall and precision is assessed on a hand-annotated subset. The method is transferable to other emerging energy technologies

(e.g., direct air capture, next-generation photovoltaics) and represents the first step toward creating knowledge graphs that explicitly model causal relationships between technological developments, characteristics, and external factors. In the future, these graphs could inform the development of more comprehensive and consistent scenario sets, enabling more reliable analyses of energy transformation pathways.

Keywords: fusion energy, causal knowledge graph, large language models, scenario analysis, scientific literature mining, technology assessment