SUPSI

Causal Graph Identification by Large Language Models

Studente/i	Relatore	Correlatore	Committente
Piqué Gregorio	Antonucci Alessandro	Zaffalon Marco	-

Corso di laurea	N° Progetto	Anno	Data
Ingegneria informatica	C10681	2023	01.09.2023

"Smoking increases the risk of respiratory disease, and both can cause lung cancer."

STUDENTSUPSI
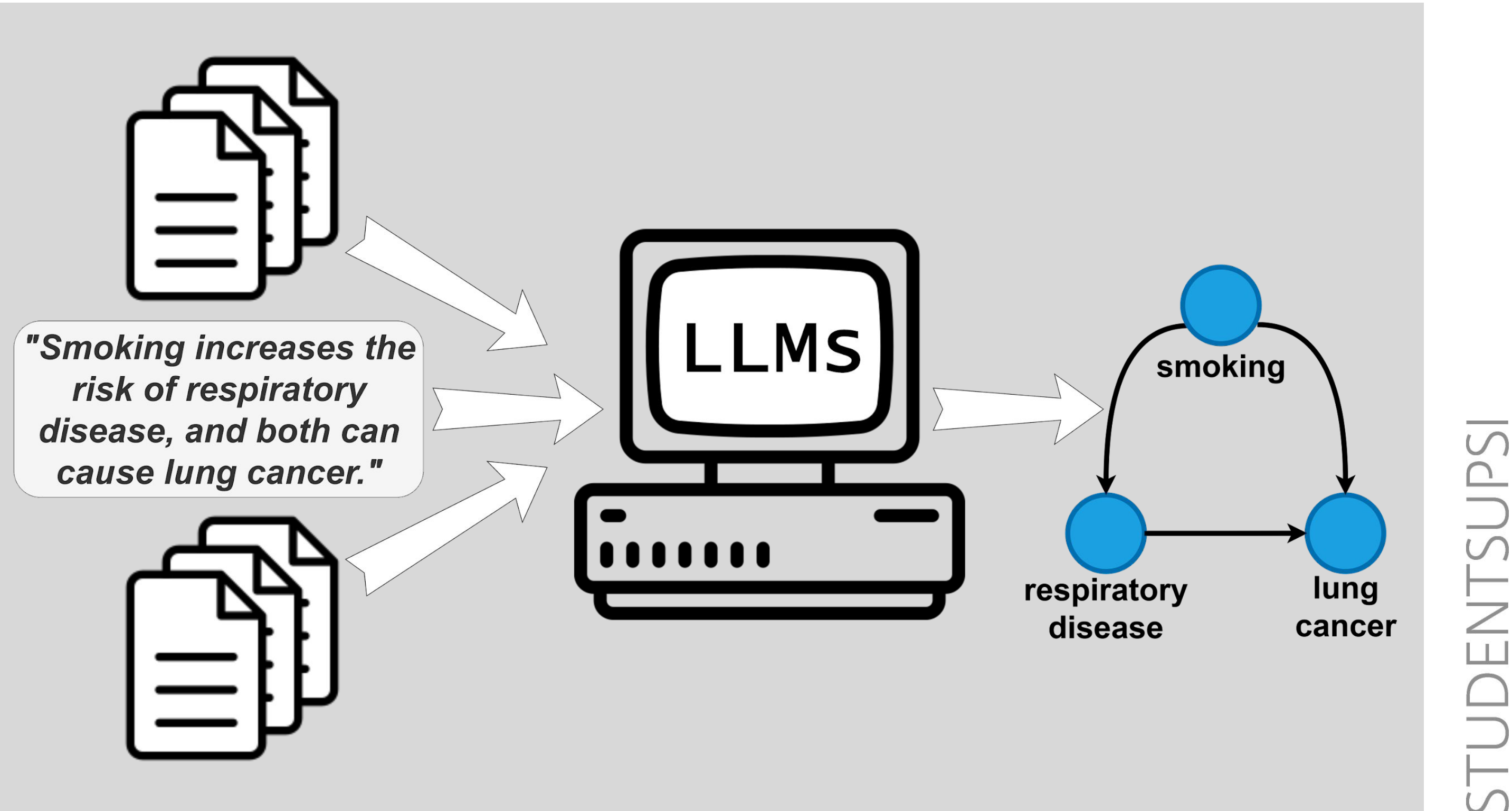
Abstract

Advances in causal inference is vital across multiple fields and contexts. A correct and complete understanding of the causal relationships behind the system of interest is a fundamental requirement for making accurate decisions.

Several methods and techniques can be used to identify causal relationships with the task called causal discovery, but many of these approaches present different flaws and weaknesses. *Large Language Models* (LLMs) can be used as a new assistant to aid human efforts and contributing to the task of causal analysis.

This project aims to evaluate the ability of LLMs in identifying causal relationships and causal graphs from natural language texts, using also a set of techniques that can be used to improve the accuracy of LLM results. The project required the implementation of a software infrastructure to collect textual data and interact with the GPT API to process it to conduct causal discovery.

Objectives

- Conduct an empirical study to assess the possibility of performing causal analysis using LLMs.

- Extract textual data from the web with a data-collection process.

- Extract named entities from text.

- Focus on the operation of causal discovery, which is the task of learning the structure of causal relationships between variables and entities.

- Analyse results with benchmark tests and evaluate effectiveness with appropriate metrics.

- Experiment with LLM usage techniques by using different models and applying prompt-engineering.

Conclusion

Large language models (LLMs) are very powerful tools and have the potential to assist human efforts in causal discovery and the identification of causal relationships.

This project showed that LLMs can achieve an average F1 score of 0.93 on benchmark tests, suggesting that they can be a valuable tool for these tasks.
However, LLMs are not without flaws, and their output should be verified by experts.

LLMs can be useful in extracting common knowledge from medical text, and when combined with expert insights, they offer the potential to efficiently generate more comprehensive causal graphs.

The application of LLMs to tasks of this nature is still in its early stages and has some limitations, but it has achieved some promising results and revealed new opportunities.