School of CIT –
Social Computing
Research Group

Technical
University
of Munich

Advanced Natural Language Processing
CIT4230002

Prof. Dr. Georg Groh

M.Sc. M.Sc. Fabienne Marco

Lecture 7.1
Causality 101

Lecture Outline – Lecture 1

- Motivation
- Defining Causality
- Brief Idea of Structural Causal Models
- Benchmarking Causality
- Current State of LLMs and Causality

Lecture Outline – Lecture 1

- Motivation
- Defining Causality
- Brief Idea of Structural Causal Models
- Benchmarking Causality
- Current State of LLMs and Causality

Motivation

Counterfactual reasoning is one of humans' high-level cognitive capabilities, used across a wide range of affairs, including determining how objects interact, assigning responsibility, credit and blame, and articulating explanations. (Bareinboim 2018)

"Fortunate is he, who is able to know the causes of things." (Virgil 29 BC)

„The development of Western science is based on two great achievements: the invention of the formal logical system (in Euclidean geometry) by the Greek philosophers, and the discovery of the possibility to find out causal relationships by systematic experiment (during the Renaissance)" Albert Einstein (1953)

Lecture Outline – Lecture 1

- Motivation
- Defining Causality
- Brief Idea of Structural Causal Models
- Benchmarking Causality
- Current State of LLMs and Causality

Defining Causality – A Multidisciplinary Concept

- Aristoteles understood causality as explanatory: The search for causes was a search for ‚First Principles‘
- Later, Causality played a minor role, specifically caused by the influential work on correlation and Bayesian networks by Pearson, Mach and Russell
- At the beginning of the twenty-first century, structural causal models and the theories of Judea Pearl became quite influential
- In the current state of science, we have different approaches:
  - Deterministic view:

    Smoking causes Cancer

  - Non-deterministic: Concepts like Structural Causal Models (SCMs), which are highly dependent on the non-observables. However, they still assume fully observable models (deterministic if we have all the information needed)
  - Quantum-Mechanics: The probabilistic nature of these models leads to traditional causality models not working anymore

Defining Causality – Different Disciplines

Causality in Health Sciences

- Measurement of Cause and Effect
- Highly used specifically to measure the effect of different drugs on curing different Diseases

Causality in Social Sciences

- Establishing more rigorous methods to deepen understanding and well-informed interventions through social policy.
- Mechanisms vs. Causal structures

Causality in Natural Sciences

- Causality in terms of mechanical laws (deterministic)
- Causal mechanisms within quantum mechanical processes
- Pendergraft's Causal Factors Requirement

Causality in Computer Science

- Structural Equation Models to handle Causal Relationships (Pearl 2000)
- Minimal Bayesian Nets (Finding the least number of arrows from all those that fit the data and to interpret those arrows as causality

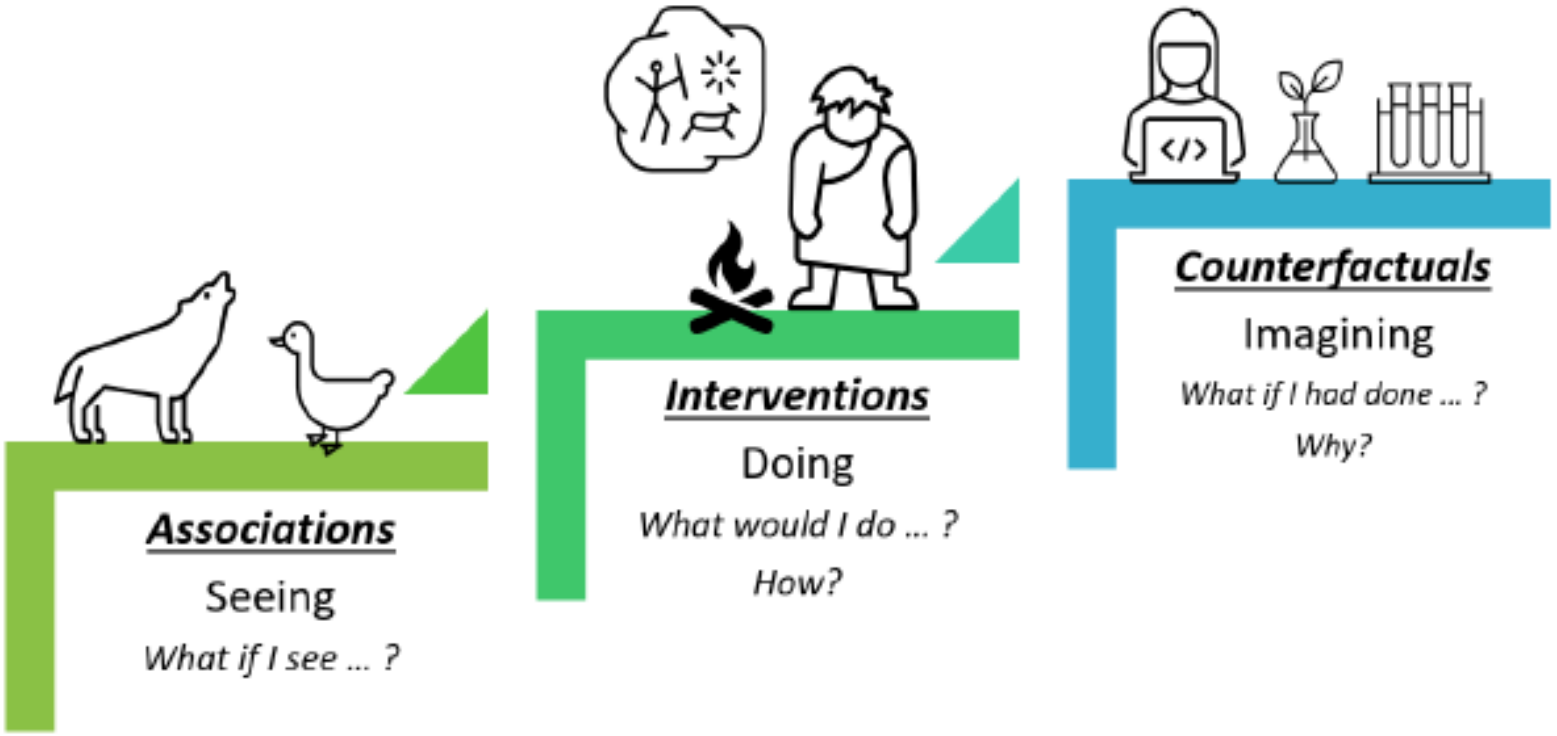
Defining Causality – Domain Depended Definition 1

Table 3: Definitions of legal causation terminology.

Term	Definition
Causation	The causing or producing of an effect.
Factual ("but for") Causation	An act or circumstance that causes an event, where the event would not have happened had the act or circumstance not occurred.
Proximate Causation	A cause that is legally sufficient to result in liability.
Intervening Factor	An event that comes between the initial event (in a sequence of events) and the end result, thereby altering the natural course of events that might have connected a wrongful act to an injury.
Superceding Factor	An intervening act that the law considers sufficient to override the cause for which the original actor is responsible, thereby relieving the original actor of liability for the result.

- Causality as used within legal frameworks (Carey, Wu 2022)

Defining Causality – Pearl's Ladder of Causality



The ladder of causation by Judea Pearl. Illustrated by Carey and Wu (2022).

Lecture Outline – Lecture 1

- Motivation
- Defining Causality
- Brief Idea of Structural Causal Models
- Benchmarking Causality
- Current State of LLMs and Causality

Structural Causal Models in a Nutshell

Set of Variables:

- $\mathbf{V} = \{V_1, V_2, \ldots, V_n\}$: A set of endogenous variables whose values are determined within the model.
- $\mathbf{U} = \{U_1, U_2, \ldots, U_m\}$: A set of exogenous variables representing external factors that are not influenced by other variables in the model.

Directed Acyclic Graph (DAG):

- $\mathcal{G} = (\mathbf{V}, \mathbf{E})$: A DAG where each node corresponds to a variable in $\mathbf{V}$, and each directed edge $(V_i \to V_j) \in \mathbf{E}$ represents a direct causal effect from V_i to V_j.

Structural Equations:

- For each endogenous variable $V_i \in \mathbf{V}$, there is an associated structural equation:

$$V_i = f_i(\mathrm{PA}_i, U_i)$$

  where $\mathrm{PA}_i \subseteq \mathbf{V}$ is the set of parent variables of V_i in the DAG $\mathcal{G}$, and f_i is a deterministic function that combines the effects of the parent variables PA_i and the exogenous variable U_i.

Probability Distribution:

- A joint probability distribution $P(\mathbf{U})$ over the exogenous variables, capturing the uncertainty and stochastic nature of external factors.

Structural Causal Models in a Nutshell

Formally, an SCM can be represented as:

$$\mathcal{M} = (\mathbf{V}, \mathbf{U}, \mathcal{G}, \mathbf{F}, P(\mathbf{U}))$$

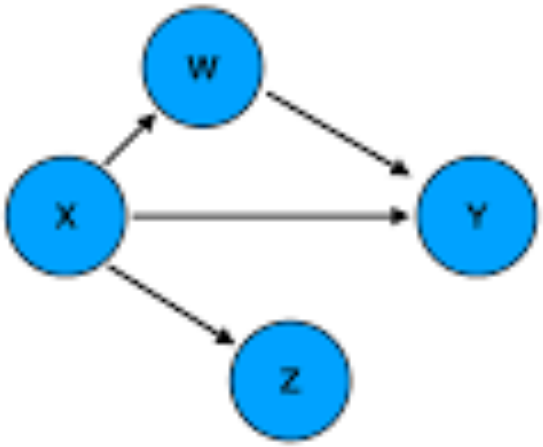
where:

- $\mathbf{V}$ is the set of endogenous variables.
- $\mathbf{U}$ is the set of exogenous variables.
- $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ is a directed acyclic graph (DAG) representing the causal structure.
- $\mathbf{F} = \{f_1, f_2, \ldots, f_n\}$ is the set of structural equations.
- $P(\mathbf{U})$ is the joint probability distribution over the exogenous variables.

Structural Causal Models in a Nutshell - Example

Structural Causal Models can be seen as a combination of Directed Acyclic Graphs and SEMs:

Directed Acyclic Graphs (DAGs)



Structural Equation Models (SEMs)

$$W := f_1(X)$$

$$Z := f_2(X)$$

$$Y := f_3(X, W)$$

Structural Causal Models in a Nutshell - Intervention

Intervention with the do-operator

Given an SCM $\mathcal{M} = (\mathbf{V}, \mathbf{U}, \mathcal{G}, \mathbf{F}, P(\mathbf{U}))$, where:

- $\mathbf{V}$ is the set of endogenous variables.
- $\mathbf{U}$ is the set of exogenous variables.
- $\mathcal{G}$ is the causal graph.
- $\mathbf{F}$ is the set of structural equations.
- $P(\mathbf{U})$ is the joint distribution of exogenous variables.

An intervention $do(X = x)$ involves setting the variable X to a specific value x and modifying the model accordingly.

You drink regular coffee, your productivity is somewhat boosted, but your sleep quality suffers.

You switch to decaf, your productivity drops to zero, but your sleep quality is perfect.

Structural Causal Models in a Nutshell - Counterfactuals

Counterfactual Definition

Given an SCM $\mathcal{M} = (\mathbf{V}, \mathbf{U}, \mathcal{G}, \mathbf{F}, P(\mathbf{U}))$, a counterfactual question typically takes the form: "What would the value of Y have been, had X been x' instead of x, given that we observed $X = x$ and $Y = y$?"

$$P(Y_{x'}|X = x, Y = y) = \int P(Y_{x'}|\mathbf{U})P(\mathbf{U}|X = x, Y = y)d\mathbf{U}$$

You're a dragon trying to understand your popularity levels. Normally, dragons in your community believe that eating spicy food helps them breathe fire, which in turn makes them popular at dragon parties. You, however, have been avoiding spicy food because you can't handle the heat. You want to know: "What would my popularity be if I could breathe fire like the other dragons?"

Structural Causal Models in a Nutshell – Evaluation

- Advantages
  - Clear and understandable, mathematically formalized model for causality
  - Clear Representation
  - Counterfactual Analysis
  - Interventional Predictions
  - High Interpretability

- Problems
  - Highly Dependent on Assumptions
  - Bad Generalization
  - Model Specification can be a complex task
  - High Data Quality needed
  - Causal Discovery
  - Limited to Observable Variables

Lecture Outline – Lecture 1

- Motivation
- Defining Causality
- Brief Idea of Structural Causal Models
- Benchmarking Causality
- Current State of LLMs and Causality

Benchmarking Causality

- The benchmarking of causal models is usually done task-based (Ashwani et al. 2024)
  - Causal Relationship Identification
  - Causal Discovery
  - Causal Explanation
  - Counterfactual Reasoning

- One option to test the different categories within LLMs:
  - Answering Causal Questions
    - How good is the model in answering causal questions of different types? (Zhang et al. 2024)

Benchmarking Causality

- Causal Relationship Identification (also actual causality (Halpern 2016)
  - determines whether and how one variable influences another, distinguishing it from mere correlation.
- Causal Discovery
  - uncovers the underlying causal structure or model from data, often using algorithms and statistical methods.
- Causal Explanation
  - provides a clear and detailed account of how and why a particular causal relationship exists, often using causal models to illustrate the mechanisms involved.
- Counterfactual Reasoning
  - considers hypothetical scenarios to determine what the outcome would have been if a different action or condition had occurred, given the observed data.

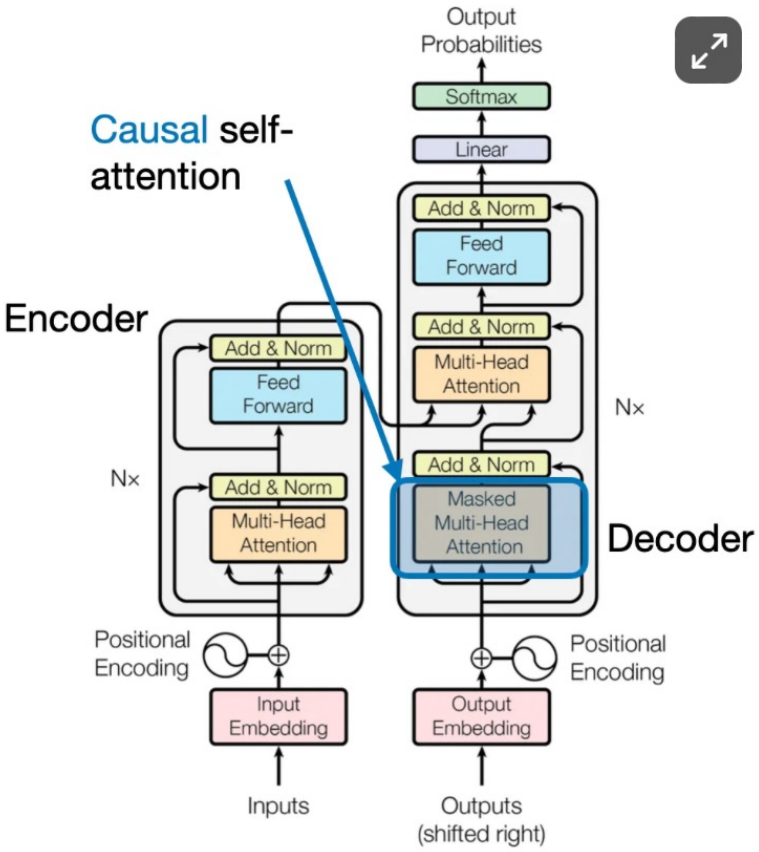
Lecture Outline – Lecture 1

- Motivation
- Defining Causality
- Brief Idea of Structural Causal Models
- Benchmarking Causality
- Current State of LLMs and Causality

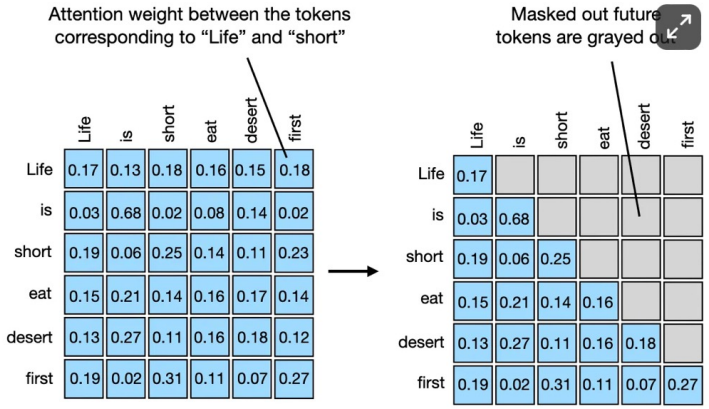
Context Lengths – One Explanation

To first address a major misleading name…

Current State of LLMs and Causality – Causal Attention Heads



The causal self-attention module in the original transformer architecture (via "Attention Is All You Need", https://arxiv.org/abs/1706.03762)



Attention weights above the diagonal should be masked out

Current State of LLMs and Causality I Overview

- LLMs can achieve competitive performance in determining pairwise causal relationships with accuracies up to 97% but their performance varies depending on prompt engineering and may occasionally be inconsistent.

- Knowledge Discovery is more challenging for LLMs

- One reason for the improvement in argumentation and causal reasoning is the increased capacity of Neural Networks in the last years (x1000 between Chat GPT 4 and Chat GPT 4o)

Current State of LLMs and Causality I Context Length

One reason why LLMs improved their ability to discover causality is the ability to deal with longer contexts:

Accurate Causal Inference

Long Context: Provides a more comprehensive view of the preceding information, enabling the model to make more accurate causal inferences. For example, understanding that a specific event in a text is caused by an earlier event can be crucial for accurate interpretation and prediction.

Capturing Long-Range Dependencies

Long Context: Allows the model to recognize and utilize long-range dependencies, leading to a better understanding of the causal relationships in the text. This is particularly important in narratives or complex documents where key causal events might be separated by long passages.

Disambiguating Causes and Effects

•**Long Context**: Helps disambiguate causes and effects by providing additional context that clarifies the relationships between events. This reduces the risk of misattributing causes or effects to the wrong entities or actions.

Current State of LLMs and Causality I Context Length

Enhancing Interpretability and Coherence

Long Context: Enhances the interpretability and coherence of the model's outputs by ensuring that the cause-and-effect relationships are well-grounded in a broader context.

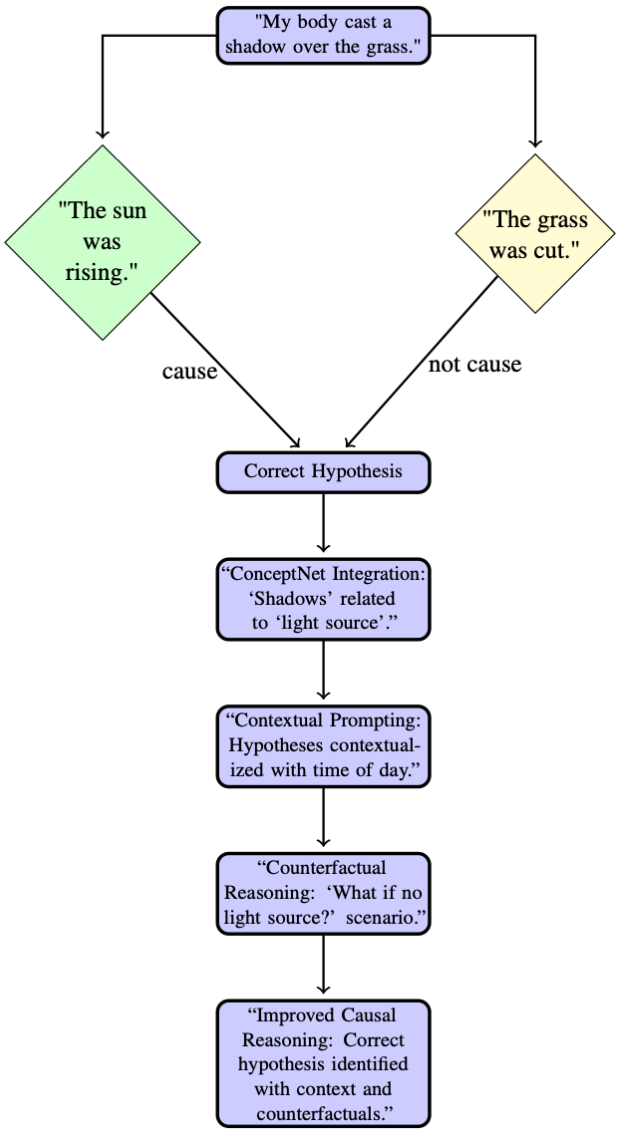
Improving Model Robustness and Generalization

Long Context: Provides a richer dataset for the model to learn from, improving its ability to generalize and apply learned causal relationships to new contexts.

Current State of LLMs and Causality I Overview

Criteria	LLMs
Causal Relationship Identification	LLMs enable knowledge-based causal dis- covery, and achieve competitive performance in de- termining pairwise causal relationships between vari- ables, across datasets from multiple domains, includ- ing medicine and climate science.
Causal Discovery	LLMs can achive a comparable performance to graph based neural networks, but is highly dependend on the prompt engineering
Causal Explanation	Large Language Models mimic causal explanation/argumentation, but are not able to dicover completely new insights.
Counterfactual Reasoning	LLMs struggle can still struggle with ambiguity and exhibit unpre- dictable failure modes, requiring additional context for accurate answers

Current State of LLMs and Causality I Hybrid Models

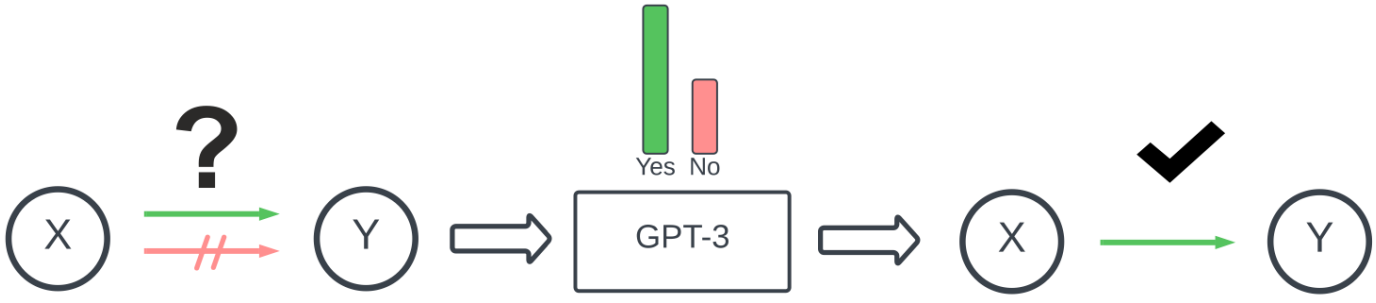


Causal Reasoning enhanced with CARE-CA

- Includes external knowledge from ConceptNet for further understanding
- Uniting explicit and implicit causal mechanisms

- Contextual Knowledge Integrator (CKI): Deep Contextual Backdrop
- Counterfactual Reasoning Enhancer (CRE): What-if Scenarios
- Context-Aware Prompting Mechanism (CAPM): Crafting tailored promts encapsulating enriched context and counterfactual reasoning

Current State of LLMs and Causality I Graph-Based Approach



Using LLMs to build knowledge graphs (Long et al. 2024)

- Current Models make the conception of DAGs more efficient
- LLMs are not prone to oversee cofounders
- Ensuring the acyclic property of GPT-generated graphs is still challenging

Research and Applications

- Causal fairness measurements and the trustworthiness of models
- Knowledge Graphs and Knowledge Discovery
- Explainability
- Linguistic Complexity: In linguistics, causality is embedded in the way language is used to convey relationships between events. The subtleties of natural language, including idiomatic expressions, context-dependent meanings, and pragmatic inferences, are not easily formalized within Pearl's mathematical framework. For example, the phrase "let him go" can imply causation but also carries social and emotional connotations that go beyond a simple causal model.
- Political Multifacetedness: In political science, causality often involves multiple layers of influence, including historical events, power dynamics, and human behavior, which are difficult to model. Political outcomes are rarely the result of a single cause but rather an interplay of various factors.

Final Takeaways

- Causality is a challenge, which can be seen as highly domain-dependent
- Even Attention Networks work with „causal attention heads“, they are not working causally and are therefore more frequently called masked multi - head-attention
- Structural Causal Models are theoretically a good base for implementation, but specifically hard to implement for longer text and abstract matters, which are not event-based.
- Benchmarking Causal Models lacks a standardized definition of causality
- LLMs perform well for some causal tasks, but the most efficient solutions at the moment are hybrid models.
- LLMs seem to be good solutions for substituting necessary expert knowledge

Next Week I Outview

- Graph-Based Models for Causality
- CausalBERT
- Quantum Natural Language Processing and Compositionality (of Language) for Causality – do we need inherently causal models for XAI?

References

[1] Ashwani, S., Hegde, K., Mannuru, N. R., Jindal, M., Sengar, D. S., Kathala, K. C. R., ... & Chadha, A. (2024). Cause and Effect: Can Large Language Models Truly Understand Causality?. *arXiv preprint arXiv:2402.18139*.

[2] Carey, Alycia & Wu, Xintao. (2022). The Fairness Field Guide: Perspectives from Social and Formal Sciences.

[3] Khetan, V., Ramnani, R., Anand, M., Sengupta, S., & Fano, A. E. (2022). Causal bert: Language models for causality detection between events expressed in text. In *Intelligent Computing: Proceedings of the 2021 Computing Conference, Volume 1* (pp. 965-980). Springer International Publishing.

[4] Kıcıman, E., Ness, R., Sharma, A., & Tan, C. (2023). Causal reasoning and large language models: Opening a new frontier for causality. *arXiv preprint arXiv:2305.00050*.

[5] Long, S., Schuster, T., & Piché, A. (2023). Can large language models build causal graphs?. *arXiv preprint arXiv:2303.05279*.

References

[6] Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic books.

[7] Tull, S., Lorenz, R., Clark, S., Khan, I., & Coecke, B. (2024). Towards Compositional Interpretability for XAI. *arXiv preprint arXiv:2406.17583*.

[8] Zhang, C., Bauer, S., Bennett, P., Gao, J., Gong, W., Hilmkil, A., ... & Vaughan, J. (2023). Understanding causality with large language models: Feasibility and opportunities. *arXiv preprint arXiv:2304.05524*.

Study Approach

Minimal

- Work with the Slides

Standard

- Work with the Slides + Read into Ashwani et al. (2024) and

In-Depth

- Standard Approach + Read into Kiciman et al. (2024)