

Causal Inference for Robust, Reliable, and Responsible NLP



Zhijing Jin
zhijing-jin.com

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS



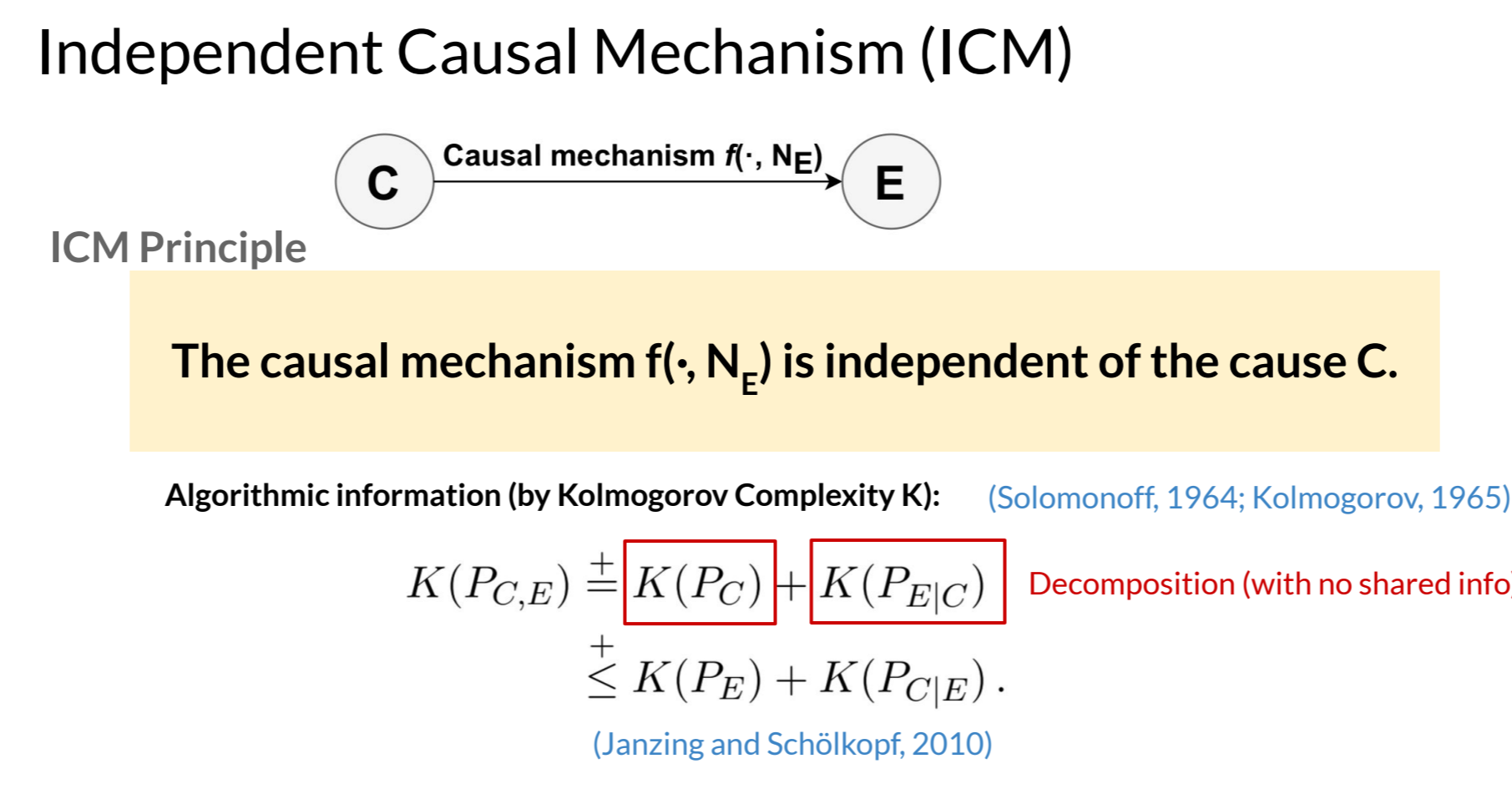
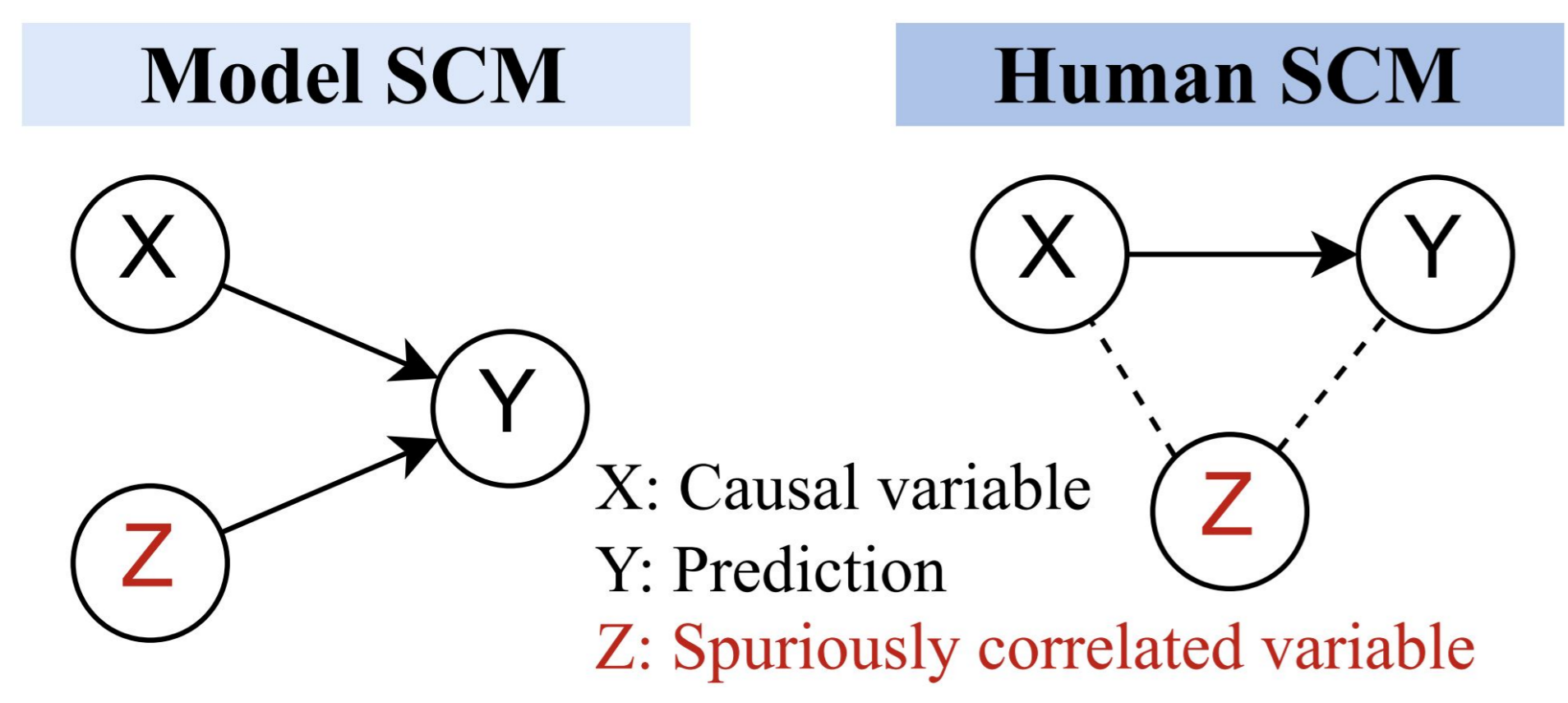
ETH zürich

1. Robust NLP Models by Causality

Core Insight: Robustness = Alignment in Structural Causal Models (SCMs)

Causal vs. Anti-Causal Learning in NLP

(EMNLP 2021 Oral)
arxiv.org/abs/2110.03618



Causal Learning Has Shorter Minimal Description Length (MDL)

The inequality of Kolmogorov complexity can be approximated by

$$MDL(e_{1:n}, e_{1:n}) = MDL(e_{1:n}) + MDL(e_{1:n}|e_{1:n})$$

$$\leq MDL(e_{1:n}) + MDL(e_{1:n}|e_{1:n}), \quad (2)$$

Experiment on MT data confirms the inequality

Dataset	Size	Note
En→Es	81K	Original English, Translated Spanish
Es→En	81K	Original Spanish, Translated English
En→Fr	16K	Original English, Translated French
Fr→En	16K	Original French, Translated English
Es→Fr	15K	Original Spanish, Translated French
Fr→Es	15K	Original French, Translated Spanish

Table 2: Details of the CausalMT corpus.

Takeaway: Achieving robustness by enforcing causal learning (in the same direction as the underlying data generating process)

A Causal Framework for Robustness

(ACL 2023)
arxiv.org/abs/2210.12023

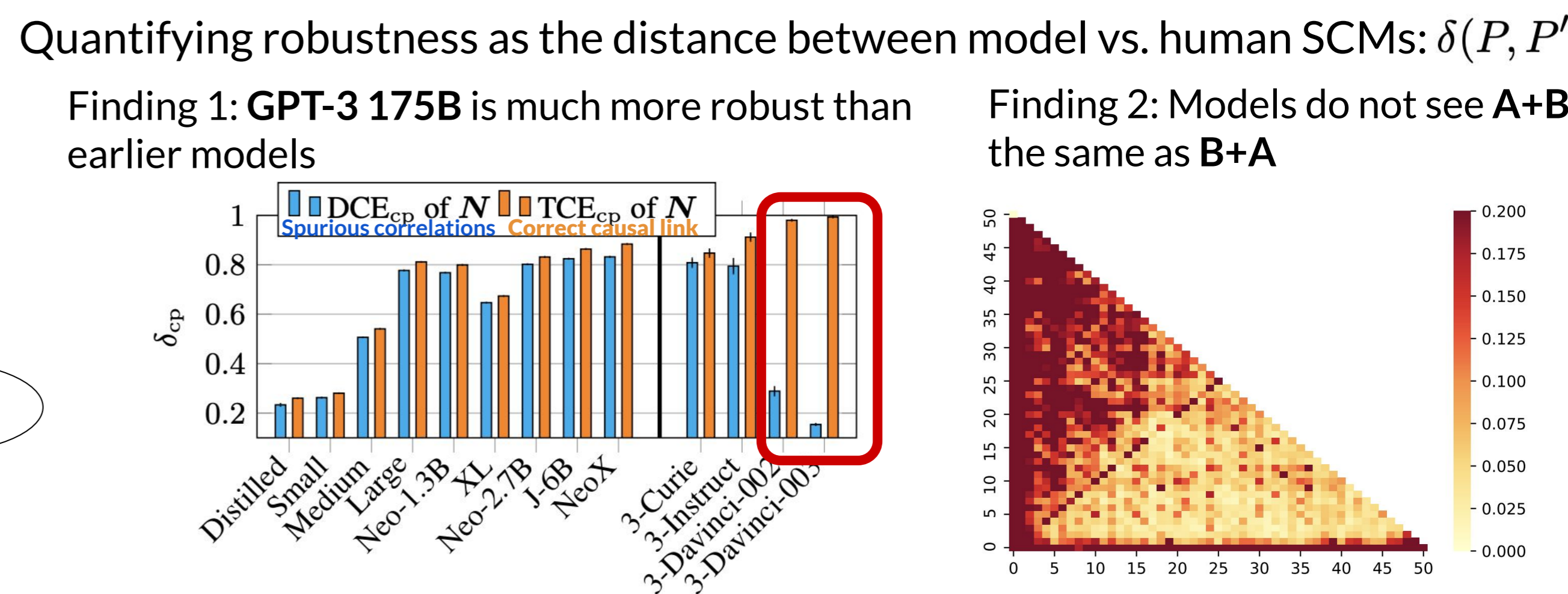
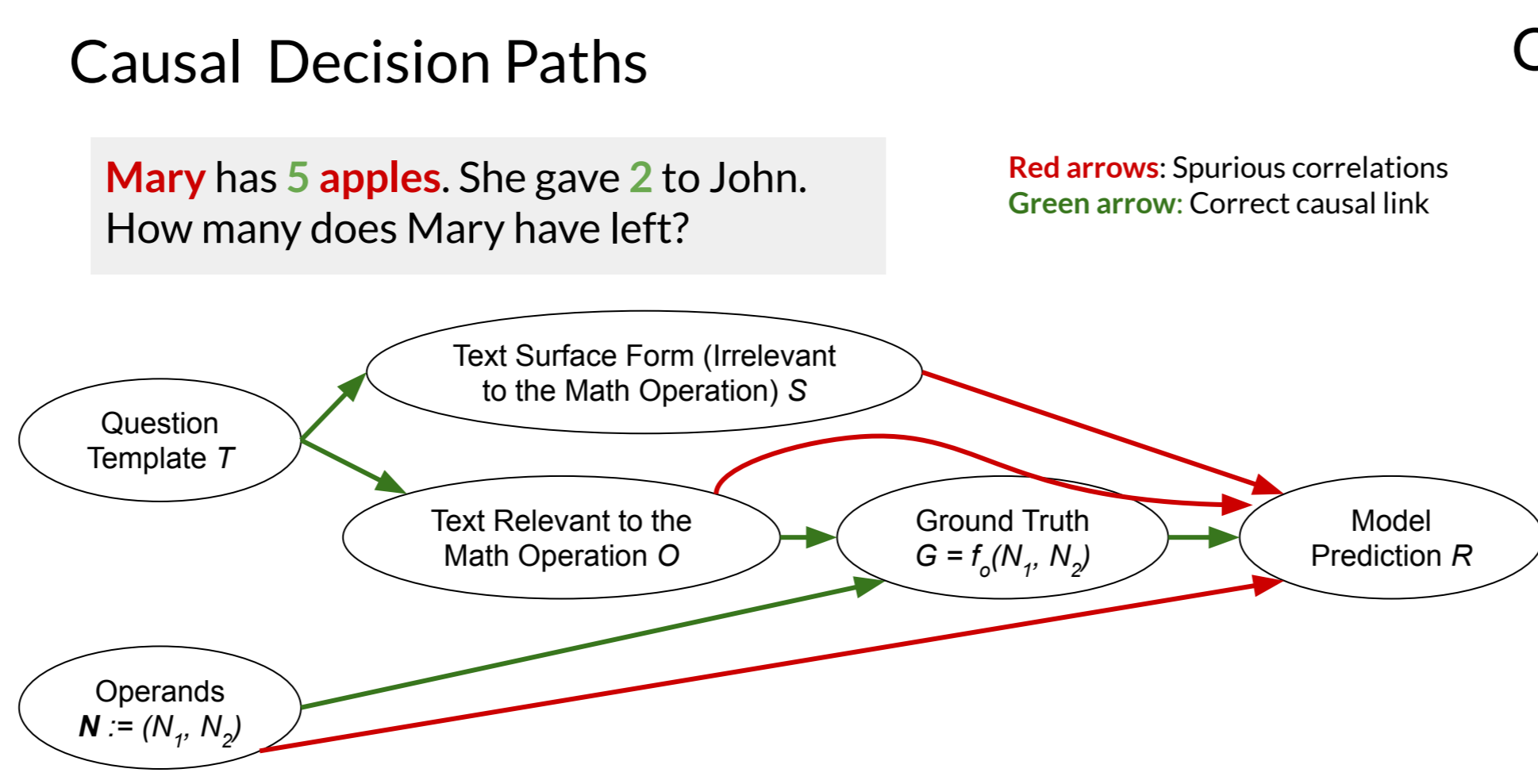
A Causal Reformulation for Robustness Testing

Is a model (e.g., GPT-3) robust on math word problems?

Question: "What is the sum of 10 and 12?"
Answer by the model InstructGPT: "22"

Question: "What is the sum of 91234567 and 47?"
Answer by the model InstructGPT: "91234514" vs. Correct answer: "91234614"

How are the causal mechanisms of the model's decisions different from the causal mechanisms of human reasoning?

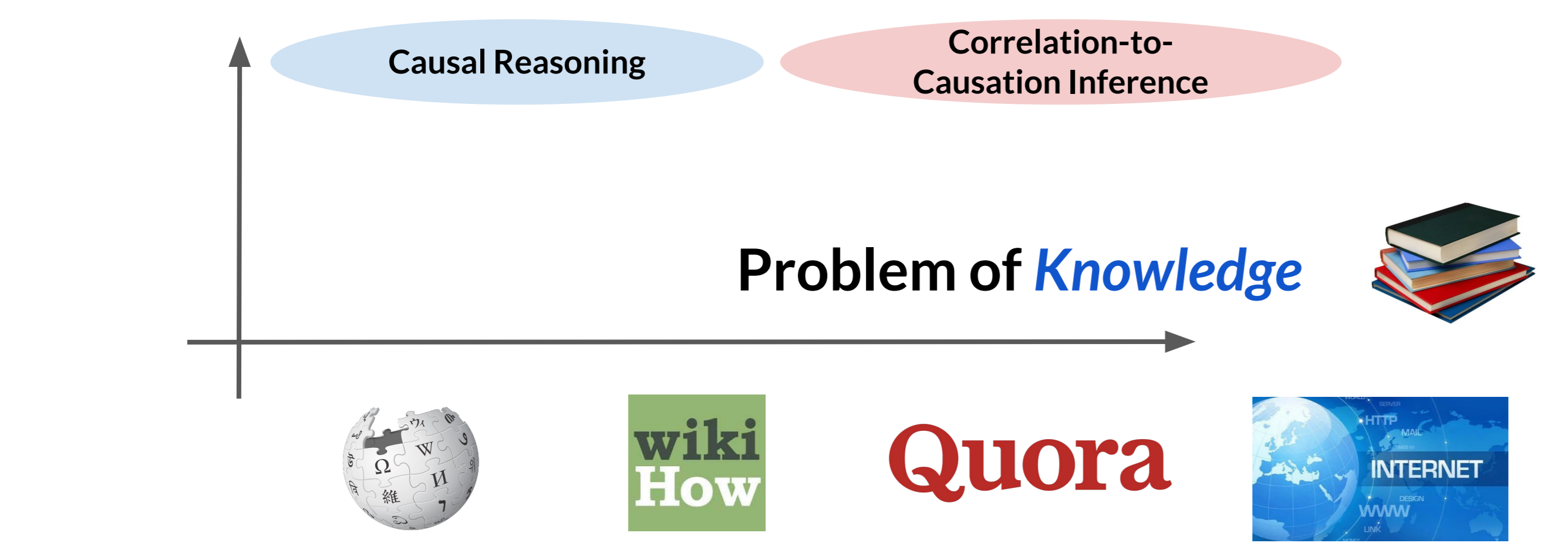


2. Reliable Causal Reasoning in Text

Problem Space
Problem of Reasoning

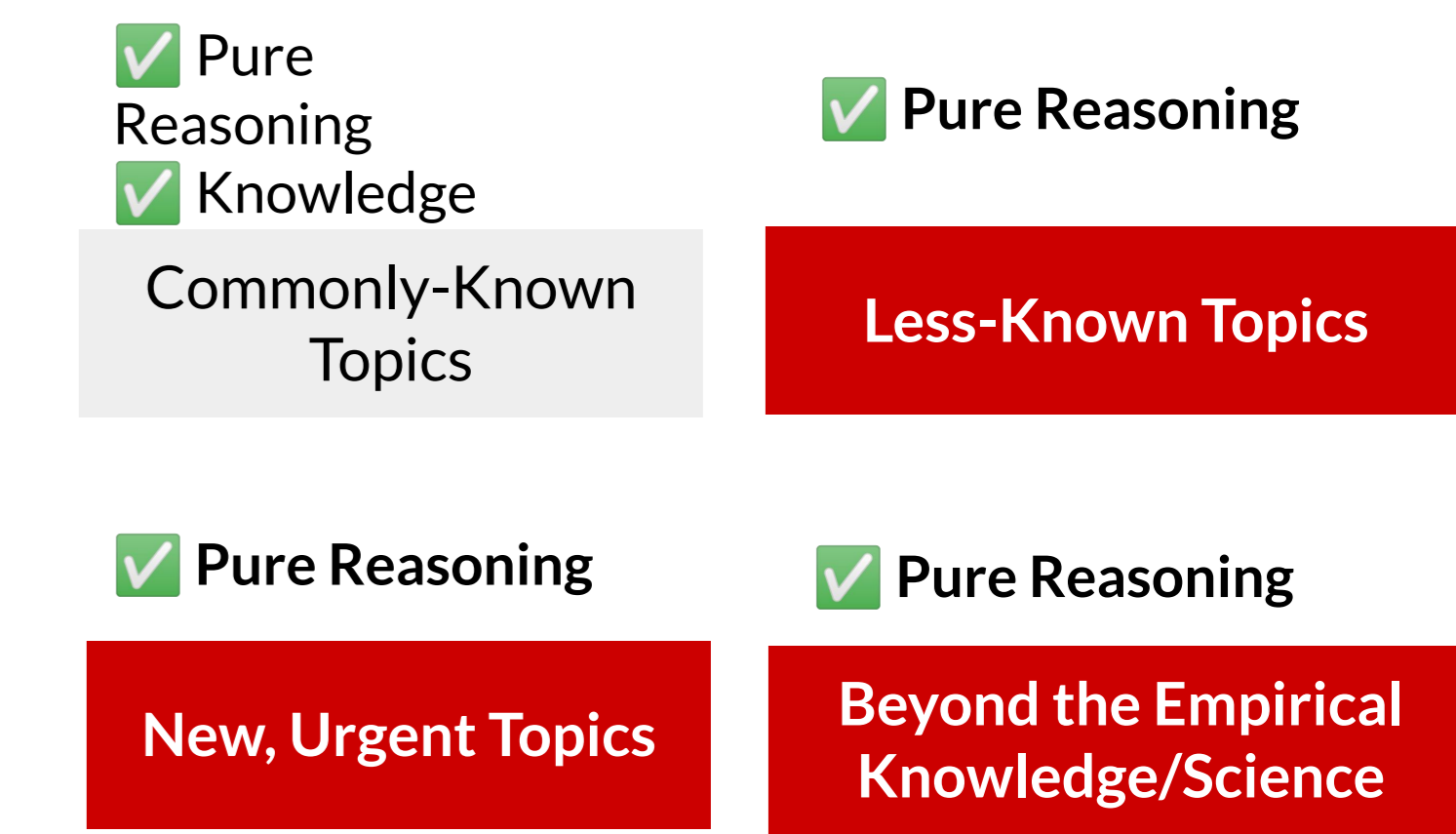
The Problem of Knowledge is largely addressed by LLMs

The Shortcomings of Using Empirical Knowledge



Why do people celebrate holidays? Answer 1-2 sentences

People celebrate holidays for various reasons, including religious or cultural significance, as well as to take a break from routine and spend quality time with family and friends. Holidays often serve as a way to commemorate important events, honor traditions, and foster a sense of community.



The Problem of Reasoning is still challenging

Causal Reasoning

Logical Fallacy Detection

Corr2Cause Inference

arxiv.org/abs/2306.05836

Training Corpus: Ice cream sales, Drowning cases, Hot weather, Vaccination, Fatality rate, Autism, Tech company CEOs, College dropout

How can LLMs process such information? Correlation? Causation? What causes what?

This requires the skill of inferring causation from correlation

We propose a new task: **Corr2Cause Inference**

Suppose we know that A correlates with B. Can we infer that A causes B?

A correlates with B. C correlates with B. However, A is independent of C. Can we infer that A causes B?

Previous tasks: Touching a hot stove Causes Getting burned

Skill being tested in previous work: Empirical knowledge instead of pure causal inference.

CLadder (NeurIPS 2023)

If LLMs are not equipped with strong reasoning ability, it will be easily tricked by incorrect causal reasoning.

- The US has no universal healthcare. Correct Fact
- Italy has universal healthcare. Correct Fact
- We see the US economy much stronger. Correct Fact

So universal healthcare is bad. Wrong Reasoning

(If the audience seeing this reasoning is a US voter, they might vote against establishing universal healthcare.)

Simpson's Paradox

(EMNLP 2022 Findings)
arxiv.org/abs/2202.13758

Our Dataset: LOGIC

Example of Circular Reasoning: She is the best because she is better than anyone else.

Example of False Causality: Every time I wash my car, it rains. So me washing my car has a definite effect on weather.

With a Challenge Set: LOGICCLIMATE

Example of False Causality: Extreme weather-related deaths in the U.S. have decreased by more than 98% over the last 100 years. ... Global warming saves lives.

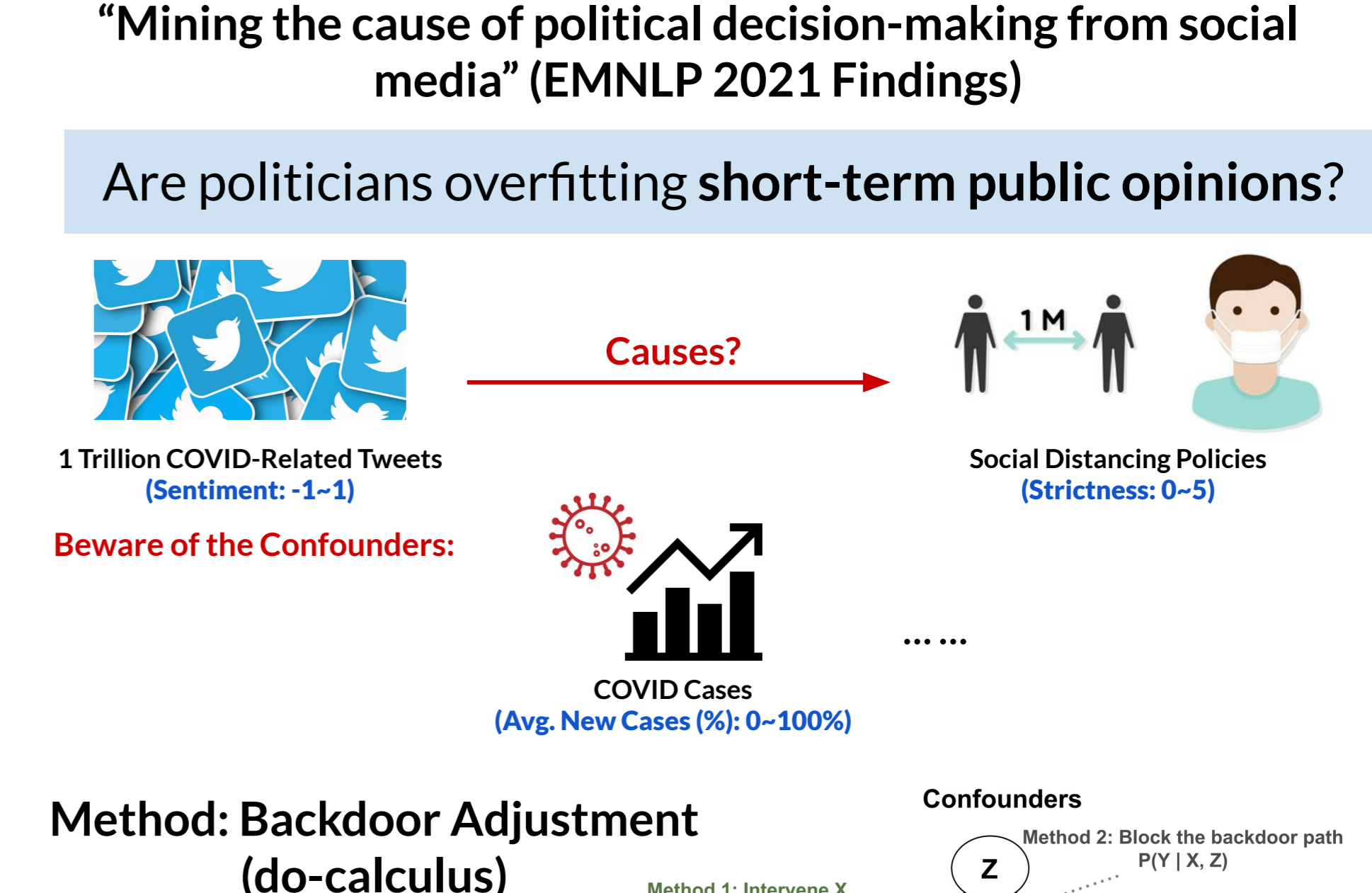
From the article: "There Is No Climate Emergency" (washingtontimes.com)

3. Responsible NLP for Social Good

Promoting NLP for Social Good (NLP4SG)

NLP4SG Paper Analyzer

Case Study: Causal Policy Analysis



12% NLP research is related to social topics

Paper: arxiv.org/abs/2305.05471
Website: nlp4sg.vercel.app/sankey

Paper: aclanthology.org/2021.findings-emnlp.27/