IDSIA SUPSI

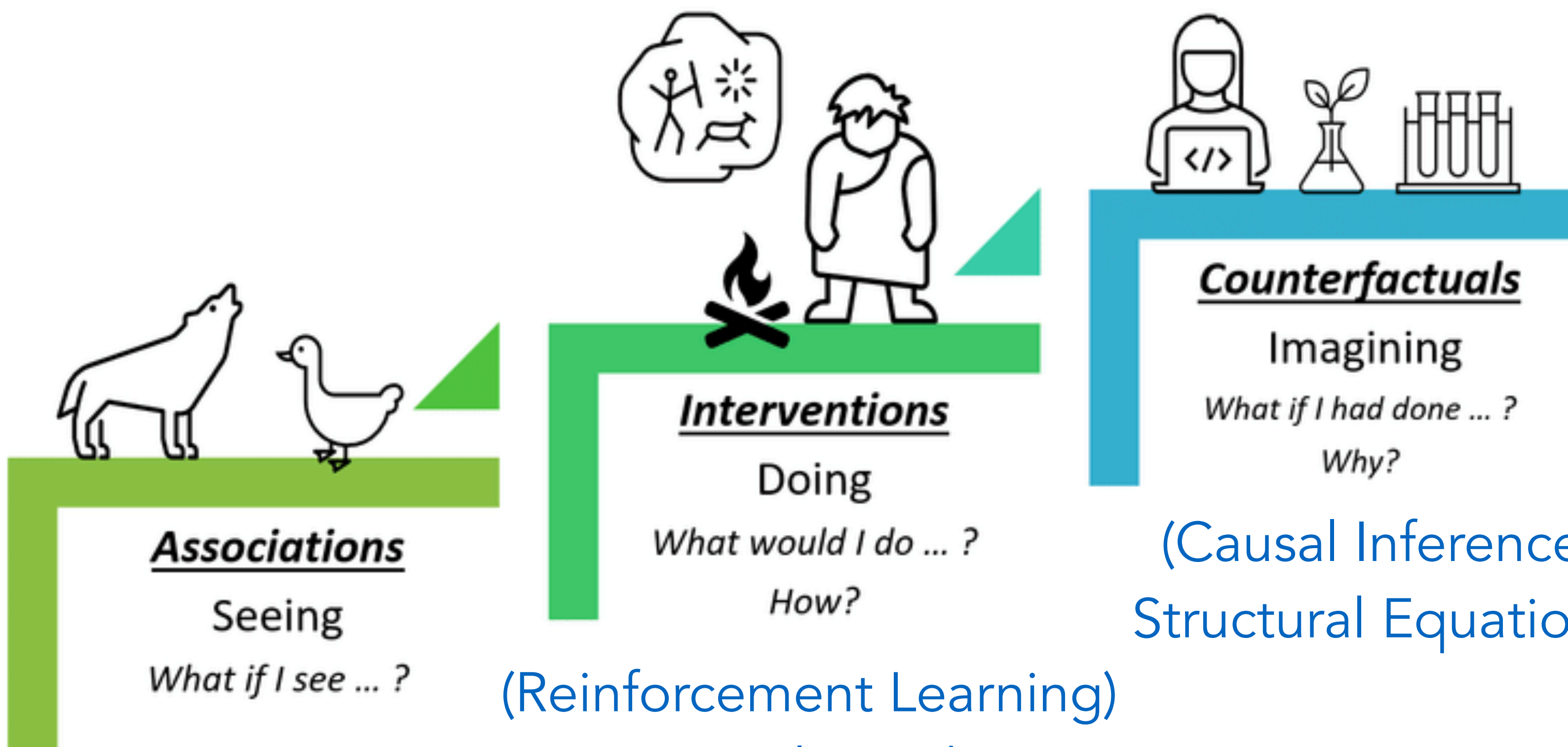
On the Limitations of Zero-shot Classification of Causal Relations by LLMs

Vani Kanjirangat, **Alessandro Antonucci**, Marco Zaffalon (IDSIA)

Seventh International Workshop on Narrative Extraction from Texts

(Text2Story 2024 @ ECIR 2024) Glasgow - March 24th, 2024

Motivation: LLMs climbing Pearl's Ladder of Causation?



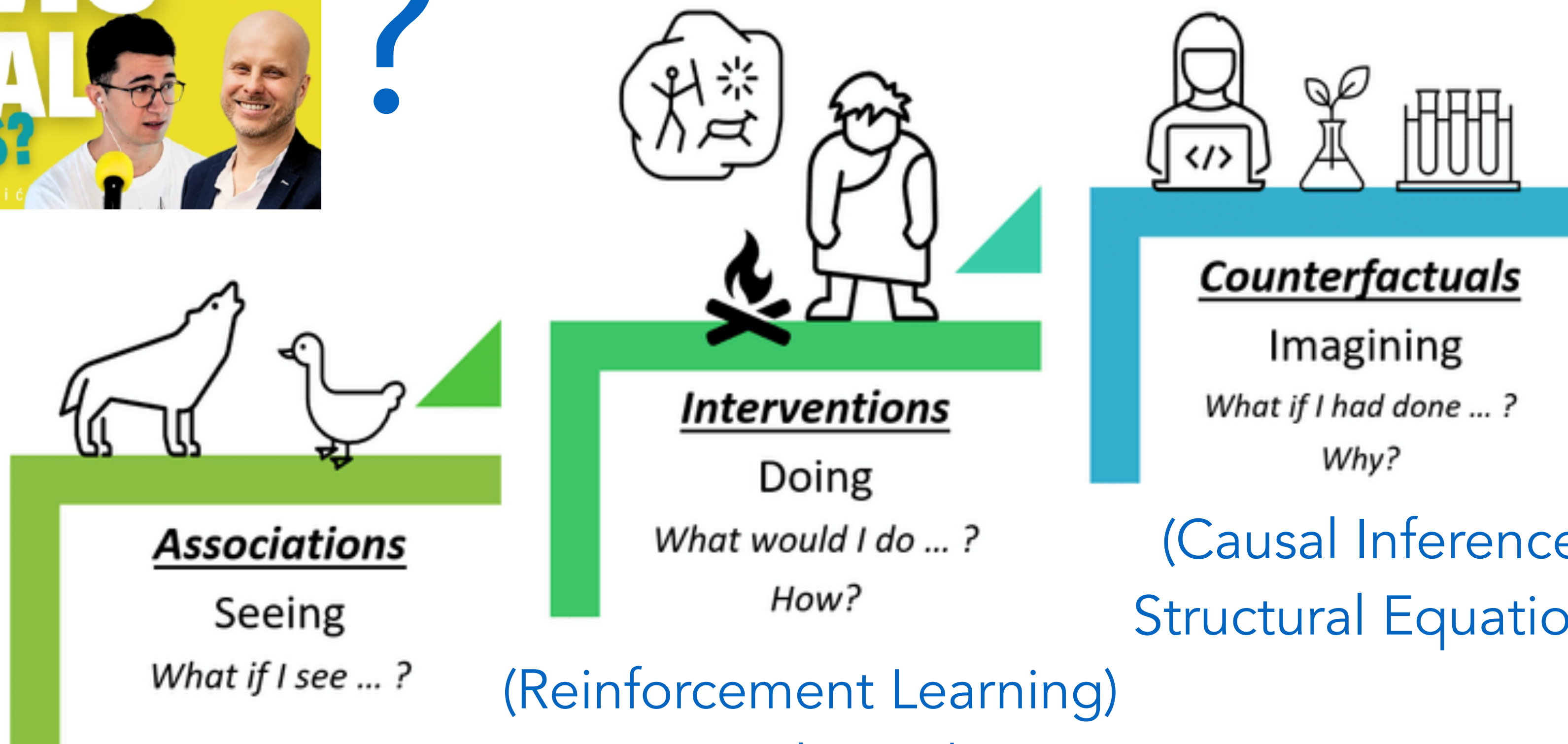
(Machine Learning)
Observational Data

(Reinforcement Learning)
Causal Graph

(Causal Inference)
Structural Equations

Motivation: LLMs climbing Pearl's Ladder of Causation?

?



(Machine Learning)
Observational Data

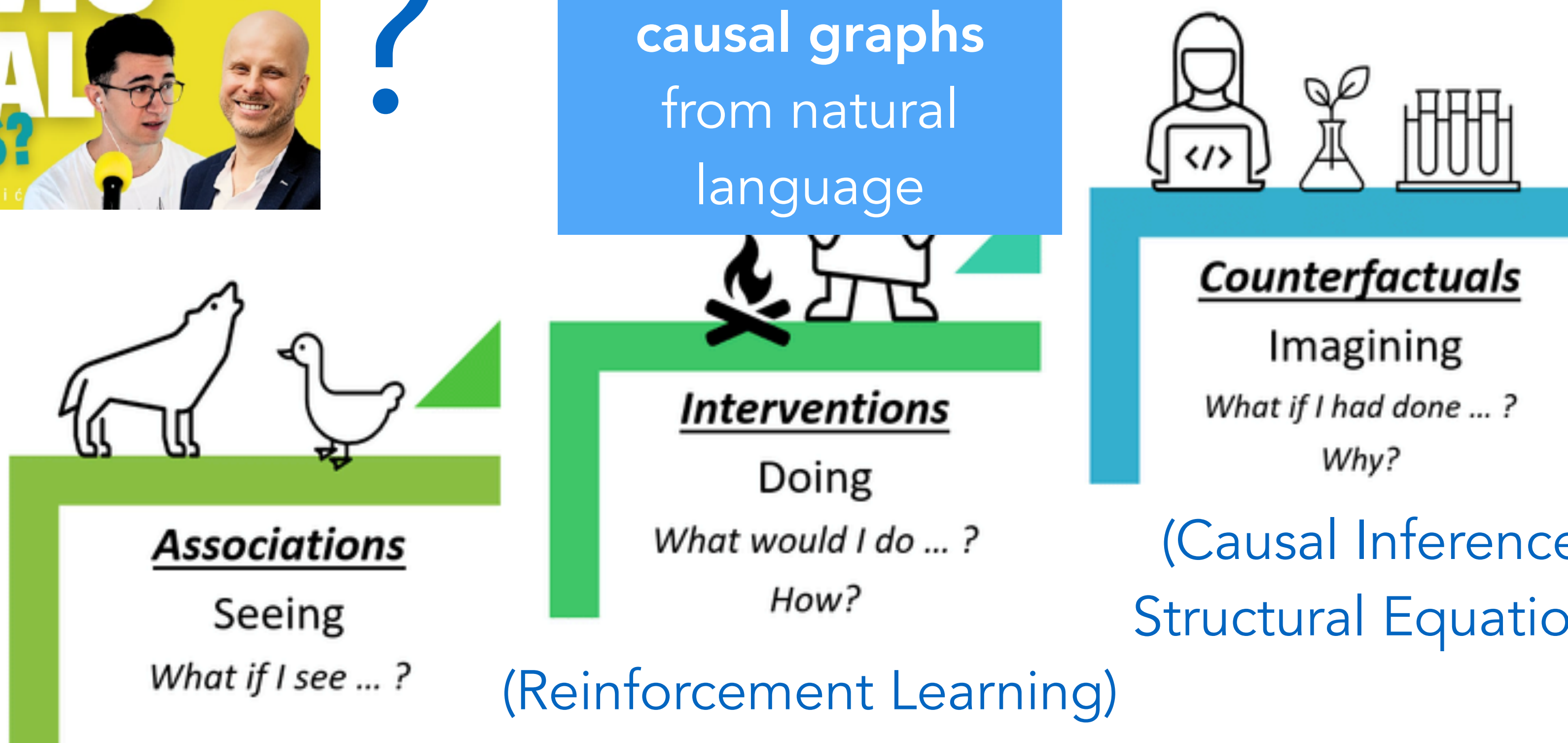
(Reinforcement Learning)
Causal Graph

(Causal Inference)
Structural Equations

Motivation: LLMs climbing Pearl's Ladder of Causation?

?

LLMs to **elicitatate causal graphs** from natural language



(Causal Inference)
Structural Equations

(Reinforcement Learning)
Causal Graph

(Machine Learning)
Observational Data

Our Earlier Work:

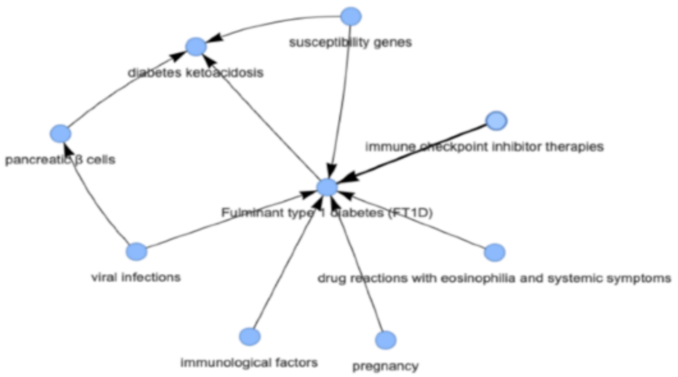
[Submitted on 22 Dec 2023]

Zero-shot Causal Graph Extrapolation from Text via LLMs

Alessandro Antonucci, Gregorio Piqué, Marco Zaffalon

Fulminant type 1 diabetes (FT1D) is a novel type of type 1 diabetes that is caused by extremely rapid destruction of the pancreatic β cells. Early diagnosis or prediction of FT1D is critical for the prevention or timely treatment of diabetes ketoacidosis, which can be life-threatening. Understanding its triggers or promoting factors plays an important role in the prevention and treatment of FT1D. In this review, we summarised the various triggering factors of FT1D, including susceptibility genes, immunological factors (cellular and humoural immunity), immune checkpoint inhibitor therapies, drug reactions with eosinophilia and systemic symptoms or drug-induced hypersensitivity syndrome, pregnancy, viral infections, and vaccine inoculation. This review provides the basis for future research into the pathogenetic mechanisms that regulate FT1D development and progression to further improve the prognosis and clinical management of patients with FT1D.

LLMs



Our Earlier Work:

[Submitted on 22 Dec 2023]

Zero-shot Causal Graph Extrapolation from Text via LLMs

Alessandro Antonucci, Gregorio Piqué, Marco Zaffalon

Fulminant type 1 diabetes (FT1D) is a novel type of type 1 diabetes that is caused by extremely rapid destruction of the pancreatic β cells. Early diagnosis or prediction of FT1D is critical for the prevention or timely treatment of diabetes ketoacidosis, which can be life-threatening. Understanding its triggers or promoting factors plays an important role in the prevention and treatment of FT1D. In this review, we summarised the various triggering factors of FT1D, including susceptibility genes, immunological factors (cellular and humoural immunity), immune checkpoint inhibitor therapies, drug reactions with eosinophilia and systemic symptoms or drug-induced hypersensitivity syndrome, pregnancy, viral infections, and vaccine inoculation. This review provides the basis for future research into the pathogenetic mechanisms that regulate FT1D development and progression to further improve the prognosis and clinical management of patients with FT1D.

PubMed

LLMs

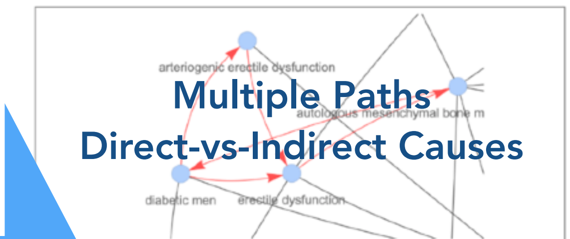

```
You will be provided with a text delimited by the <Text></Text>
xml tags, and a pair of entities delimited by the <Entity></Entity>
xml tags representing entities extracted from the given text.

  Text:
   <Text>Cobalt metal fume and dust cause upper respiratory
   tract irritation, chronic interstitial pneumonitis,
   and skin sensitization.</Text>

  Entities:
   <Entity>fume</Entity>
   <Entity>sensitization</Entity>

Read the provided text carefully to comprehend the context
and content. Examine the roles, interactions, and details
surrounding the entities within the text.

Based only on the information in the text, determine the
most likely cause-and-effect relationship between the entities
from the following listed options (A, B, C):

  Options:
   A: "fume" causes "sensitization";
   B: "sensitization" causes "fume";
   C: "fume" and "sensitization" are not directly causally related;

Your response should analyze the situation in a step-by-step manner
ensuring the correctness of the ultimate conclusion,
which should accurately reflect the likely causal connection
between the two entities, based on the information
presented in the text. If no clear causal relationship is apparent,
select the appropriate option accordingly.

Then provide your final answer within the tags
<Answer>[answer]</Answer>, (e.g. <Answer>C</Answer>).
```

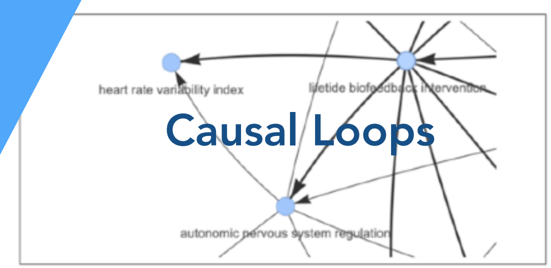
Prompt Engineering

Sentence	Orientation
Zinc is essential for growth and cell division.	A → B
The infection came from a wound.	A ← B
As we saw earlier, helicobacter is responsible for causing stomach ulcer.	A → B
The pseudolesion was caused by drainage of the paraumbilical vein.	A ← B

Good Causal Relation Identification/Orientation

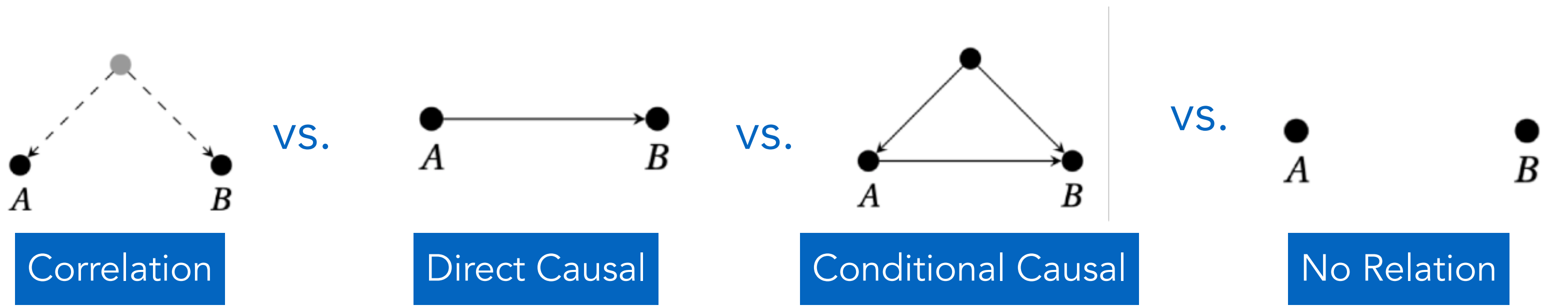
		Ground Truth	
		A → B	A ← B
GPT	A → B	335	7
	A ← B	6	650



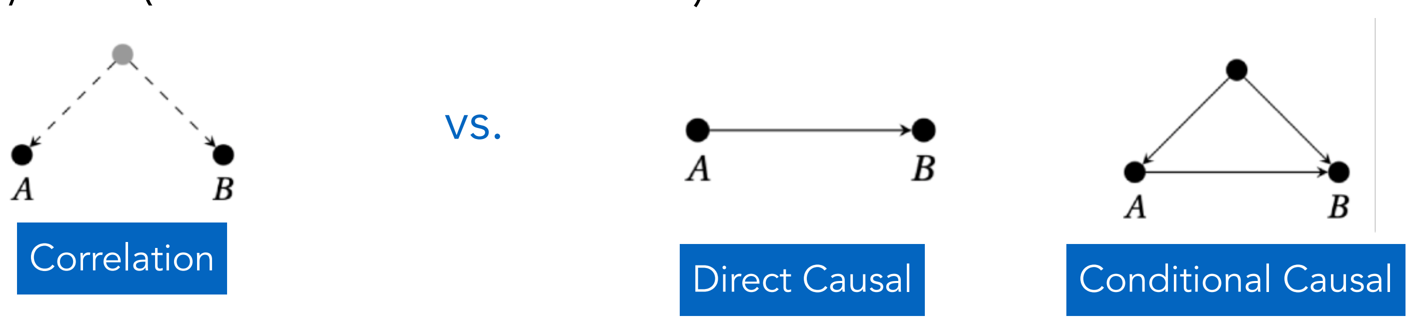


A More Basic Problem: Deciding Causal Nature of a Sentence

- Benchmark from *Detecting causal language use in science findings* (Yu et al., 2019)
- 3061 sentences from PubMed medical abstracts
- Four classes:



- Models? LLMs (GPT3.5 and Falcon) vs. (fine-tuned) BERT
- Binary task (correlation vs. causation):



Zero-Shot Prompt

You are a helpful assistant for causal reasoning and cause-and-effect relationship discovery. Your aim is to identify the entities and to categorize the input sentences into either direct causal relation or conditional causal relation or correlational relation or no relationship exist

intro_msg = You will be provided with a text. Text: <Text>{text}</Text>
instructions_msg = Please read the provided text carefully to comprehend the context and content.

Examine the roles, interactions, and details surrounding the entities within the text. Based only on the information in the text, categorize the causal relation as

0. no relation
1. direct causal
2. conditional causal 3. correlational

Your response should analyze the situation in a step-by-step manner, ensuring the correctness of the ultimate conclusion, which should accurately reflect the likely causal connection based on the information presented in the text.
If no clear causal relationship is apparent,
select the appropriate option accordingly, i.e., 'no relation'.

option_choice_msg = Your response should analyze the situation in a step-by-step manner, ensuring the correctness of the ultimate conclusion,
which should accurately reflect the likely causal connection between the two entities based on the information presented in the text.

If no clear causal relationship is apparent, select the appropriate option accordingly.
Then provide your final answer within the tags <Answer>[answer]</Answer>, (e.g. <Answer>1</Answer>).

Results (Zero-Shot, F1 score)

- GPT $\gg$ Falcon

Model	Approach	Binary class	Multi-class
GPT 3.5 turbo	Zero shot (ZS)	0.59	0.37
Falcon-7b-instruct	Zero shot (ZS)	0.19	0.27
Falcon-40b-instruct	Zero shot (ZS)	0.26	0.38

Expected result

(GPT has more training data and parameters)

Results by Class (Zero-Shot, GPT, F1 score)

	Multi-class				Binary	
	No Rel.	Causal	Cond. Causal	Corr.	No Rel.	Causal
F1-score	0.45	0.39	0.12	0.54	0.59	0.58
Precision	0.68	0.27	0.10	0.60	0.85	0.45
Recall	0.34	0.70	0.14	0.48	0.45	0.85

- Conditional Causal is challenging
- Good Recall for "Direct Causal" (direct causes properly recognised)
- Good Precision for "No Relation" (rarely a true relation as non-relation)

- Bert $\gg$ (ZS) LLM

Model	Approach	Binary class	Multi-class
GPT 3.5 turbo	Zero shot (ZS)	0.59	0.37
BERT-base-cased	Full Fine Tuning (FFT)	0.92	0.87

Supervision (still) makes a (lot of) difference!

Helping Zero-Shot with **Language Cues** ?

You are a helpful assistant for causal reasoning and cause-and-effect relationship discovery. Your aim is to identify the entities and to categorize the input sentences into either direct causal relation or conditional causal relation or correlational relation or no relationship exist

intro_msg = You will be provided with a text. Text: <Text>{text}</Text>
instructions_msg = Please read the provided text carefully to comprehend the context and content.

Examine the roles, interactions, and details surrounding the entities within the text. Based only on the information in the text, categorize the causal relation as

0. no relation
1. direct causal
2. conditional causal 3. correlational

Your response should analyze the situation in a step-by-step manner, ensuring the correctness of the ultimate conclusion, which should accurately reflect the likely causal connection based on the information presented in the text.
If no clear causal relationship is apparent,
select the appropriate option accordingly, i.e., 'no relation'.

option_choice_msg = Your response should analyze the situation in a step-by-step manner, ensuring the correctness of the ultimate conclusion,
which should accurately reflect the likely causal connection between the two entities based on the information presented in the text.

If no clear causal relationship is apparent, select the appropriate option accordingly.
Then provide your final answer within the tags <Answer>[answer]</Answer>, (e.g. <Answer>1</Answer>).

+

You may see the language cues for predicting correct answer.

direct causal may contain cues such as "increase, decrease, lead to, effective in, contribute to, reduce",

conditional causal with "increase, decrease, lead to, effect on, contribute to, result in" along with Cues indicating doubt:may, might, appear to, probably

and correlational with cues such as "association, associated with, predictor, at high risk of"

Results (LLM vs. Fine-Tuning, F1 score)

- Bert $\gg$ LLM with cues

Model	Approach	Binary class	Multi-class
GPT 3.5 turbo	Zero shot (ZS)	0.59	0.37
GPT 3.5 turbo	Zero shot with Cues (ZS-Cues)	0.66	0.51
BERT-base-cased	Full Fine Tuning (FFT)	0.92	0.87

Some improvements,
but still poor wrt fine-tuning

Few-Shot?

You are a helpful assistant for causal reasoning and cause-and-effect relationship discovery. Your aim is to identify the entities and to categorize the input sentences into either direct causal relation or conditional causal relation or correlational relation or no relationship exist

intro_msg = You will be provided with a text. Text: <Text>{text}</Text>
instructions_msg = Please read the provided text carefully to comprehend the context and content.

Examine the roles, interactions, and details surrounding the entities within the text. Based only on the information in the text, categorize the causal relation as

0. no relation
1. direct causal
2. conditional causal 3. correlational

Your response should analyze the situation in a step-by-step manner, ensuring the correctness of the ultimate conclusion, which should accurately reflect the likely causal connection based on the information presented in the text.
If no clear causal relationship is apparent,
select the appropriate option accordingly, i.e., 'no relation'.

option_choice_msg = Your response should analyze the situation in a step-by-step manner, ensuring the correctness of the ultimate conclusion,
which should accurately reflect the likely causal connection between the two entities based on the information presented in the text.

If no clear causal relationship is apparent, select the appropriate option accordingly.
Then provide your final answer within the tags <Answer>[answer]</Answer>, (e.g. <Answer>1</Answer>).

+

[

('Together with mean HbA1c, baseline urine albumin-to-creatinine ratio and presence of hypertension, accurate 3-year new-onset albuminuria prediction may be possible.', 3),

('These findings show that patients with atherosclerotic cardiovascular disease benefit from lowering of LDL cholesterol levels below current targets.', 1),

('Further research should focus on prevention of attrition in families with a lower educational background.', 0),

('At short term, unfavourable effects may occur.', 2)

]

Results (LLM vs. Fine-Tuning, F1 score)

- Bert $\gg$ (FS) LLM

Model	Approach	Binary class	Multi-class
GPT 3.5 turbo	Zero shot (ZS)	0.59	0.37
GPT 3.5 turbo	Zero shot with Cues (ZS-Cues)	0.66	0.51
GPT 3.5 turbo	Few shot with Cues (FS-Cues)	0.62	0.50
BERT-base-cased	Full Fine Tuning (FFT)	0.92	0.87

With 10-shot small improvement of ZS,
but no wrt Cues ...

From Few- to Many-Shot?

- To compete with Bert we might add more examples to the prompt
- Focus on prompting techniques (= no GPT fine-tuning)
- The same examples are used to train BERT
- With 500 example BERT has similar performance as in CV
- GPT? Better than ZS, but still not competitive with Bert
- Sometimes (small) hallucinations: different labels or multiple labels.

Model	No Rel.	Causal	Cond. Causal	Corr.	Avg.
BERT-base-cased	0.86	0.77	0.74	0.86	0.81
GPT 3.5 turbo	0.61	0.42	0.12	0.55	0.43

Bert $\gg$ (MS) LLM

machine learning $\gg$ case-based reasoning

Conclusions and Outlooks

- LLMs as poor zero-shot learners for complex natural language understanting such as in the causal domain
- But, they can be powerful pre-processing tools, with a good potential for the elicitation of causal graphs
- Good for arc orientation, less for arc recognition
- We need (and are working) on the creation of a larger and carefully annotated benchmark
- (Necessary) Future work?
  - Soft Prompting (= prompt tuning)
  - PEFT (Parameter Efficient Tuning)

Conclusions and Outlooks

- LLMs as poor zero-shot learners for complex natural language understanting such as in the causal domain
- But, they can be powerful pre-processing tools, with a good potential fo
- Good for a

alessandro@idsia.ch

- We need (and are working) on the creation of a larger and carefully annotated benchmark
- (Necessary) Future work?
  - Soft Prompting (= prompt tuning)
  - PEFT (Parameter Efficient Tuning)