Can Foundation Models Talk Causality?

MEMBERS:
HIRA TAHIR
FARHAN AHMED
KANWAR MUZAMMIL ROHAIL

Executive Summary

- Investigates foundation models in NLP, focusing on causal understanding.
- Explores BERT and GPT's ability to comprehend causality.
- Conducts experiments to test models on cause-effect tasks.
- Provides insights into strengths and limitations of foundation models.
- Highlights impressive performance on certain tasks but identifi gaps in understanding.
- Discusses implications for NLP advancement and the need for further research.
- Contributes to understanding foundation models' processing o causal information.

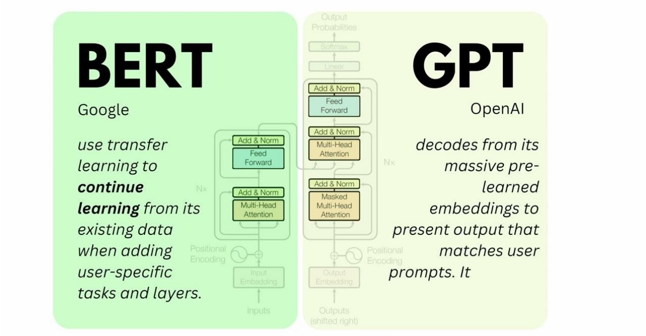
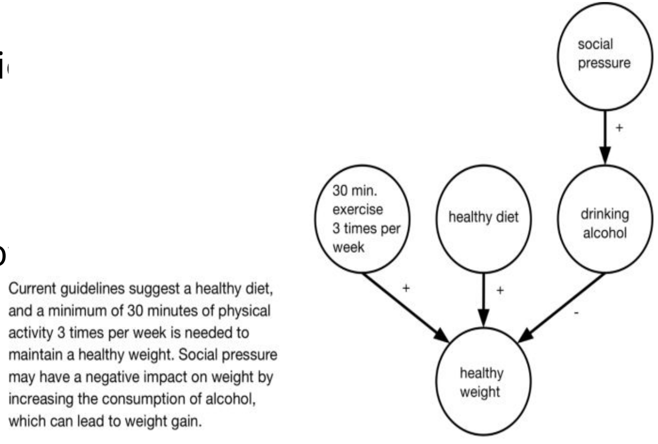


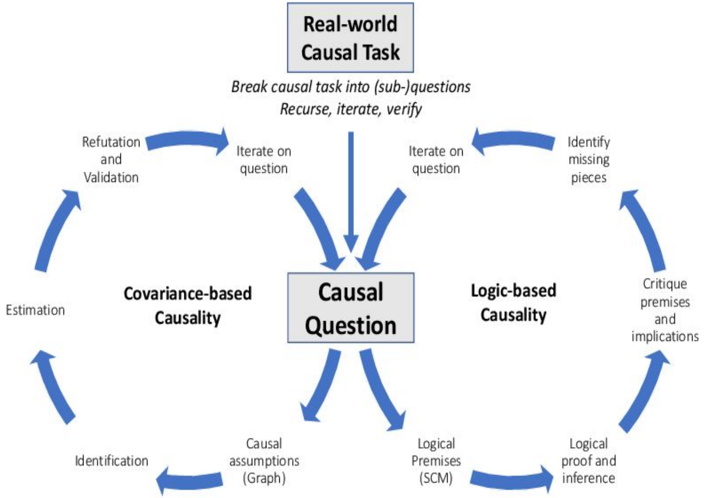
Figure 1: The Transformer - model architecture.

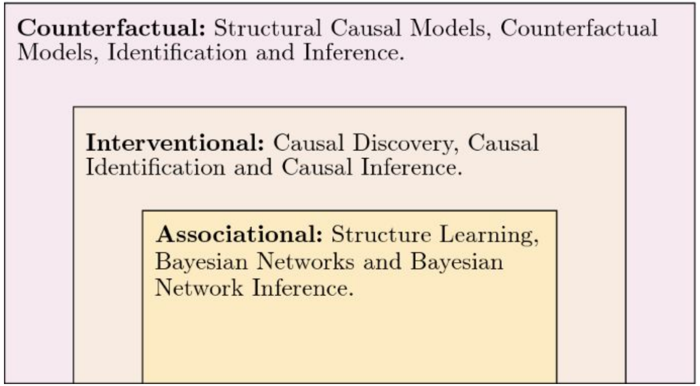


(a) Text-based causal information (b) Causal diagram

Background

- Rise of advanced language models like BERT and GPT for language processing tasks.
- Models excel in tasks such as translation and sentiment analysis.
- Challenge lies in understanding cause-and-effect relationships in text.
- Humans intuitively grasp causal connections, but it's challenging for machines.
- Causal reasoning is complex and involves understanding the reasons behind events.
- Traditional methods manually feed machines rules or knowledge bases for causal reasoning.
- Large-scale models offer an opportunity for machines to learn causal reasoning from data.
- Research aims to evaluate models' capabilities and limitations in understanding cause-and-effect relationships in text.





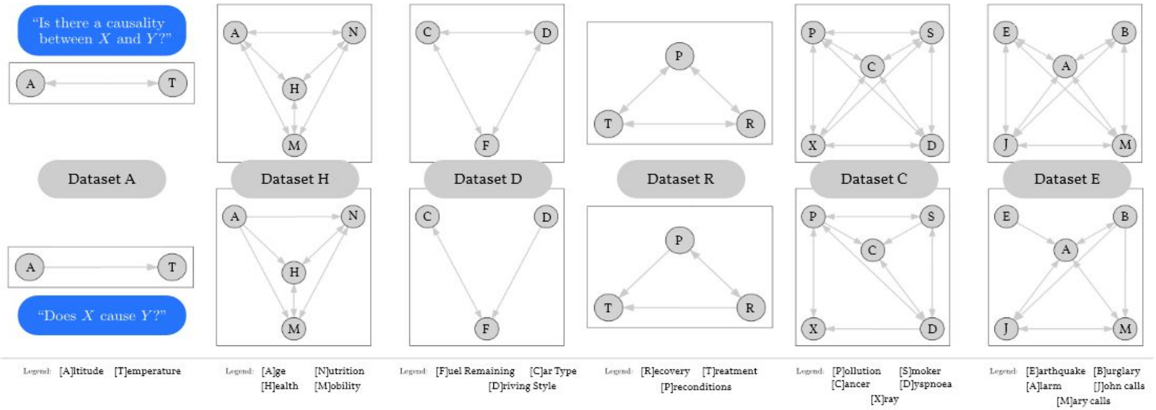
Experiments and Results

- Design of experiments testing models' understanding of causal relationships in text.
- Examples of experimental scenarios and tasks presented to the models.
- Mixed results indicating strengths and weaknesses in models' causal reasoning abilities.
- Implications for advancing language comprehension in AI systems.

Q1 How do the FM graph predictions compare to settings where the causal graph is (partially) known?

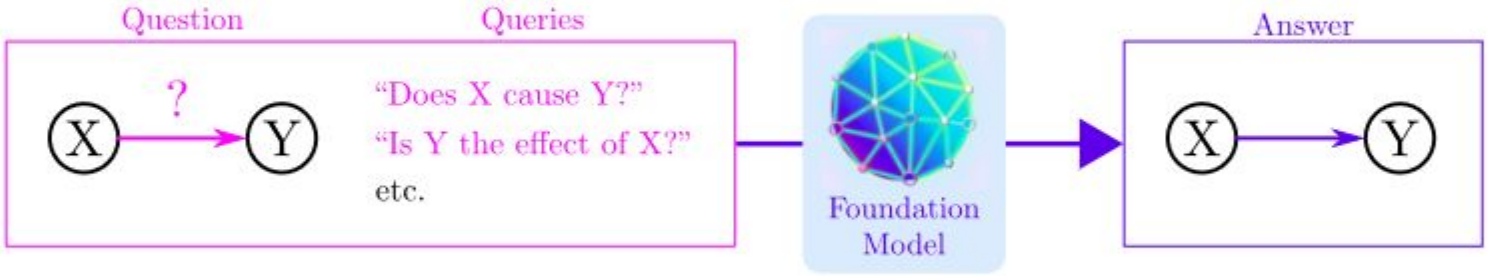
Q2 How do the FM graph predictions perform in "common sense" settings that involve abstract reasoning and intuitive physics?

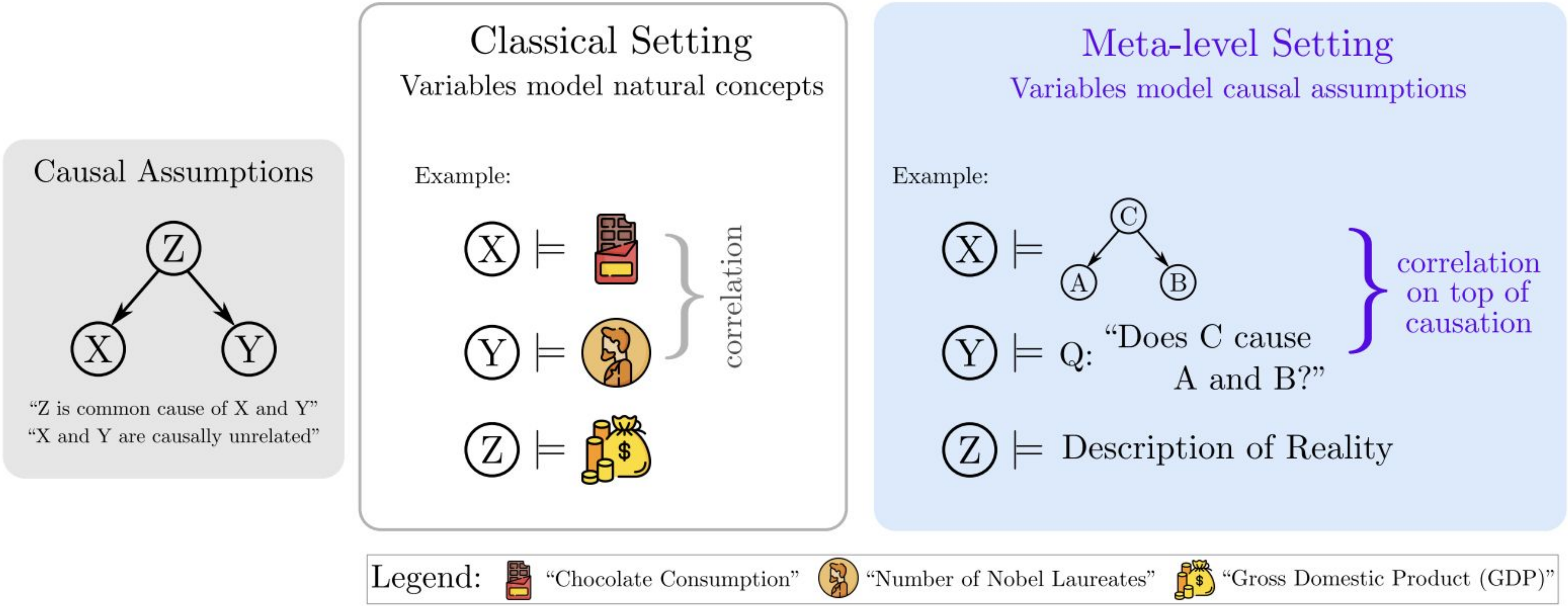
Q3 How do synonyms or more general variable name altercations affect the FM graph prediction?



Methodology

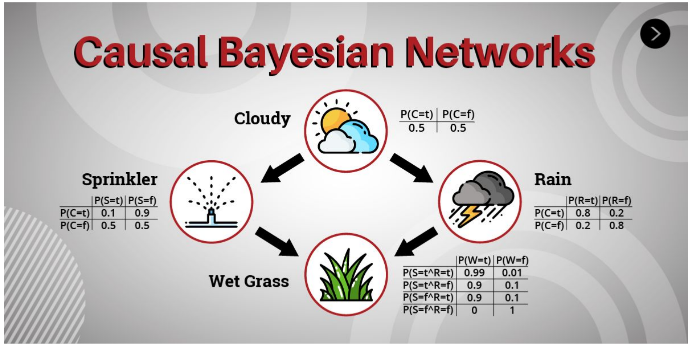
- Systematic approach to investigating causal reasoning in language models.
- Curation of diverse dataset and crafting of experimental tasks.
- Rigorous evaluation metrics and procedures ensuring validity and reliability.
- Advanced statistical analysis techniques applied for insights.

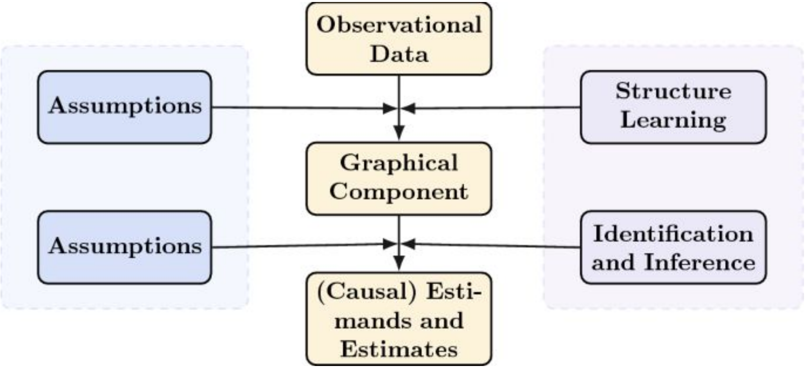



Causal Assumptions
Z
X
Y
"Z is common cause of X and Y"
"X and Y are causally unrelated"
Classical Setting
Variables model natural concepts
Example:
X ⊨
Y ⊨
Z ⊨
correlation
Meta-level Setting
Variables model causal assumptions
Example:
X ⊨
C
A
B
Y ⊨ Q: "Does C cause A and B?"
Z ⊨ Description of Reality
correlation on top of causation
Legend: "Chocolate Consumption" "Number of Nobel Laureates" "Gross Domestic Product (GDP)"

Key Findings

- Nuanced understanding of causal relationships demonstrated by models.
- Contextual sensitivity aiding accurate causal inference.
- Generalization of causal reasoning abilities across domains.
- Limitations observed in counterfactual reasoning.
- Implications for natural language understanding and future research directions.





Discussion Points

- Depth of interpretative understanding versus surface-level associations.
- Role of context in facilitating accurate causal inference.
- Challenges in counterfactual reasoning and manipulation of causal variables.
- Generalizability of causal reasoning abilities across diverse domains.
- Broader implications for AI development and ethical considerations.

Limitations and Open Points

- Scope limitations concerning textual genres and domains analyzed.
- Challenges related to data availability and quality for training causal reasoning models.
- Need for robust evaluation metrics capturing complexities of causal inference.
- Importance of interpretable model architectures for transparent causal reasoning.
- Ethical and societal implications of deploying advanced causal reasoning models in real-world applications.

Any Questions?