# From Correlation to Causation: Understanding Climate Change through Causal Analysis and LLM Interpretations

**Shan Shan**
Zhejiang University
shshan@zju.edu.cn

## Abstract

This research presents a three-step causal inference framework that integrates correlation analysis, machine learning-based causality discovery, and LLM-driven interpretations to identify socioeconomic factors influencing carbon emissions and contributing to climate change. The approach begins with identifying correlations, progresses to causal analysis, and enhances decision making through LLM-generated inquiries about the context of climate change. The proposed framework offers adaptable solutions that support data-driven policy-making and strategic decision-making in climate-related contexts, uncovering causal relationships within the climate change domain.

## 1 Introduction

Why do we seek to precisely understand causality? In the realm of large-scale, high-dimensional datasets, is mere knowledge of correlations sufficient? Does our pursuit of causality stem from mere curiosity, or does it offer substantial practical benefits? Could our perceived understanding of causality, much like Plato's allegory of the shadows in the cave, actually obscure the true nature of reality? As Ludwig Wittgenstein notes in his *Philosophical Investigations*, "The precise and explicit rules governing the logical structure of propositions often serve as a concealed backdrop within our medium of understanding." He further discusses the "crystalline purity of logic," highlighting its indispensable role not merely as an outcome of inquiry, but as a foundational necessity (Wittgenstein, 1967, 4f,c);(Sluga and Stern, 1996, 49-50). Expanding on this framework of logic structure in understanding, Judea Pearl fosters the concept of "understanding" as a means to the sensation of control, specifically through causal inference which he associates with decision-making in intelligent systems Pearl (2014). He posits that a robust understanding of causality is crucial for effective decision-making in intelligent systems, emphasizing that such understanding goes beyond mere data correlation and involves the ability to manipulate and control outcomes[1].

These perspectives are particularly relevant to the development of Large Language Models (LLMs), where design elements like prompts are tailored to reflect human-machine interaction. The architecture of LLMs not only displays the capacity of machines to emulate complex human logical processes but also enables further exploration of causal relationships (Jin et al., 2024; Kiciman et al., 2024; Ceraolo et al., 2024). LLMs' prompted and related design highlights the potential of machines to emulate human logical processes and probe into causal relationships of deep understanding.

However, recent studies have indicated that LLMs are "weak causal parrots", merely reciting the causal knowledge from the training data(Zečević et al., 2023), parroting unintentional remarks. The primary challenges in causality research arise from the lack of a clear definition of causal statements

[1]The Science of Cause and Effect: From Deep Learning to Deep Understanding, https://simons.berkeley.edu/

and the absence of adequate mathematical tools to address these complex questions (Pearl, 2000). This research confronts these challenges by comparing and integrating three approaches: correlation analysis, machine learning-based causality discovery, and LLM-inquiry-driven interpretations, within a human-machine interaction framework focused on the social context of climate change. The progression emphasizes not only the identification of correlations and causal factors but also leverages earlier stages and prior knowledge as benchmarks for advanced logical inquiries and interpretations.

The study aims to delineate the effectiveness of above three approaches within the social science context of climate change, striving to deepen our understanding of causality through a comparative framework. By exploring and progressing through three stages, this research seeks to uncover insights that could lead to more informed decisions and strategies for a deeper understanding of social factors that influence climate change. In summary, the study's contributions are as follows:

- **Comparative Methodological Framework:** The research uses a three-step comparative framework that combines correlation analysis, machine learning-based causality discovery, and LLM-prompt-driven inquiries (See Appendix Figure 1) . This approach improves the study of causality and provides a structured way to assess different methods, offering a framework that could be replicated in future research in a similar climate change context.
- **Application to Climate Change Social Contexts:** The study focuses on the social science aspects of climate change by exploring correlations and causations of social influences on climate change. The research continues to explore how LLMs can understand and help address these critical social issues, providing valuable insights that could shape environmental policies and strategies, thus improving decision making.

# 2 Preliminaries: Causal Relations

It is already known that correlation does not imply causation(Ksir and Hart, 2016; Rohrer, 2018). However, causation is a subset of correlation because a causal relationship inherently implies correlation (but with a cause-and-effect dynamic). Literarily and technically, when exploring causation, correlation represents a closer relationship between two variables than non-relation. Causation cannot exist without correlation, even though correlation alone is not sufficient to establish causation. Following this logical sequence, this research begins by understanding the desired outcome and then determining the necessary steps to achieve it. The study assumes that correlation analysis does not need to be excluded; instead, it could serve as a foundation to narrow down the selection of relationships.

In complex real-world scenarios, identifying the associations between events and variables helps predict outcomes, make informed decisions, and understand the underlying mechanisms of systems (Pearl, 2000). Causal discovery involves identifying the dependent variables of an event of interest and understanding the physical influence relationships between events or variables. Causal structures imply both statistical (conditional) independence and independences to other (non-statistical) information measures (Peters et al., 2017), which is a common task in causal inference. In the domain of causal discovery, machine learning algorithms can primarily be divided into two main categories:

- **Constraint-based Algorithms**: These algorithms rely on tests of statistical independence within the data to uncover potential causal relationships between variables. They attempt to construct causal graphs by analyzing the conditional independencies among variables. A classic method is the PC algorithm (Peter-Clark Algorithm), which iteratively examines and eliminates edges that do not satisfy conditional independencies, thereby inferring the causal structure between variables Spirtes et al. (2001).
- **Score-based Algorithms:** This type of algorithm identifies the best causal graph by assigning scores to different causal models. The process typically involves enumerating and scoring various possible causal graphs, selecting the model with the highest score. Scoring criteria may be based on how well the data fits the model, such as the Bayesian Information Criterion (BIC) or the Minimum Description Length (MDL). The core idea behind these algorithms is that the causal model which best explains the observed data is considered optimal (Liu et al., 2012; Nogueira et al., 2022).

Constraint-based algorithms are generally more efficient as they rely on statistical tests to quickly narrow down the search space, but they may be sensitive to noise in the data and sample size. On the other hand, score-based algorithms, though theoretically capable of exploring a broader model space to find the optimal model, can be computationally expensive in practice due to the need to evaluate a large number of models. Overall, the search over causal graphs between variables is challenged by two distinct factors: the sheer volume of causal graphs, which increases super-exponentially with the addition of nodes, and the constraint of maintaining acyclicity(Cheng et al., 2024).

To address this limitation, Rolland et al. (Rolland et al., 2022) design a novel order-based methods to recover causal graphs from the score of the data distribution in non-linear additive noise models and propose a new efficient method for approximating the score's Jacobian, enabling to recover the causal graph. Specifically, they first sequentially identify leaves of the causal graph by analyzing its entailed observational score, and then remove the identified leaf variables. As a result, one can obtain a complete topological order with a time complexity linear in the number of nodes. Since the node in the ordering can be a parent only of the nodes appearing after it in the same ordering, once a topological order is fixed, the acyclicity constraint is naturally enforced, making the pruning step easier to solve.

# 3 Correlation Analysis: Narrowing the Variable Pool

The correlation step involves correlation analysis, using a heatmap and the Anderson-Darling k-sample test (anderson_ksamp) with a threshold value of 0.1 (i.e. 10%) to identify the influence of each variable on the target variable, carbon emissions. Variables with matching distributions between the training and testing data are retained. The most relevant variables are then carried forward to the causation analysis.

- **Correlation Matrix Calculation**: First, a correlation matrix is calculated for the dataset, quantifying the linear relationships between each pair of variables. The matrix visually distinguishes positive from negative correlations, with coefficients ranging from "-1" strong negative correlation to "+1" strong positive correlation. The full lists of correlation map are presented in Appendix Figure 2.
- **Heatmap Visualization**: Second, the heatmap incorporates hierarchical clustering, which groups variables with similar correlation patterns together. This step enhances the interpretability of the heatmap by organizing it into blocks of highly correlated variables, making it easier to spot clusters of factors that behave similarly.

  The resulting correlation map provides a visual summary of how different factors related to carbon emissions interact with each other. It reveals key drivers of carbon emissions and potential areas for further investigation or intervention, identifying the relationships within complex datasets, and facilitating deeper insights into the underlying dynamics of carbon emissions. Appendix Figure 3 shows the sorted correlations of features with the target variable, carbon emissions.

# 4 Causal Effects Estimation

The method employed in this research is adapted from existing approaches to causal modeling, specifically following the framework outlined by Rolland et al. (2022) (Rolland et al., 2022). In this approach, each variable is modeled as a function of its direct causal parents in the causal graph, along with an additive noise term. The data distribution is defined by these causal relationships, and score functions are used to identify leaf nodes within the graph. Leaf nodes are detected based on the variance of partial derivatives of the score function, which helps distinguish parent-child relationships among variables. The nodes in the graph are arranged in order by finding and removing leaf nodes one by one. To do this, the experiment uses the Stein gradient estimator with ridge RBF kernel regression to estimate the score function.

## 4.1 Causal Graph Construction and Score Matching

Take the dataset $\{V^{2000}, Y^{2005}\}$ as an illustration example, which treats as 16 study variables $X_{1\ldots16} = \{V^{2000}, Y^{2005}\}$. The study assumes the data is generated using the following model:

$$X_i = f_i(\mathrm{pa}_i(X)) + \epsilon_i, \quad i \in \{1, 2, \ldots, 16\},$$

where $\text{pa}_i(X)$ selects the coordinates of $X$ that are direct causal variables of node $i$ in the causal graph, and $\epsilon_i$ is an additive noise term, which might include measurement errors. The associated probability distributions are given by:

$$p(x) = \prod_{i=1}^{d} p(x_i \mid \text{pa}_i(x))$$

$$\log p(x) = \sum_{i=1}^{d} \log p(x_i \mid \text{pa}_i(x))$$

The score function is defined as:

$$s(x) \equiv \nabla \log p(x)$$

The necessary and sufficient conditions for the $j$-th variable to be a leaf node are given by:

$$\forall x, \left(\frac{\partial s_j(x)}{\partial x_j}\right) = c, \quad \text{where } c \text{ is a constant value independent of } x,$$

$$\text{Var}_X\left(\frac{\partial s_j(X)}{\partial x_j}\right) = 0$$

If the $j$-th variable is a leaf node, and the $i$-th variable is its parent node, then:

$$\text{Var}_X\left(\frac{\partial s_j(X)}{\partial x_i}\right) \neq 0$$

Based on this finding provided by Rolland et al., 2022, the experiment achieves topological ordering by sequentially identifying the leaf nodes and removing them one by one. The Jacobian of the score can be approximated by Stein gradient estimator with ridge RBF kernel regression (Rolland et al., 2022).

Once a topological order is estimated, the causal graph becomes constrained to be a sub-graph of a certain fully connected graph. However, it is necessary to prune this fully connected graph to remove spurious edges. This study uses the CAM pruning process to complete the step.

### 4.2 CAM Pruning

The above approaches control confounding variables by retaining key confounders during variable selection, removing irrelevant variables through correlation analysis. This section refining the causal graph via CAM pruning to eliminate spurious relationships while preserving causal integrity.

After arranging the nodes, the graph is refined by using the CAM pruning process, which removes unnecessary connections to reveal the actual causal structure, aligning with methods discussed by Rolland et al. 2022(Rolland et al., 2022). Detailed outputs are provided in the Appendix and include the following metrics [2]:

- Structural similarity: Evaluated using SID and SHD.
- Predictive accuracy: Measured through precision, recall, and F1 score.
- Overall deviation: Assessed using L2 distance.

The graph (See Appendix Figure 4) highlights a structured approach to understanding how specific social factors influence carbon emissions and climate change. By focusing on the most influential variables—access to clean fuels in rural and urban areas and managing urban population growth—strategic decisions and policies can be more effectively targeted.

[2] It is noted that "Variable Selection" is to ensure that important confounders are included before pruning begins, as removing key variables early can lead to residual confounding or spurious relationships. The formal analysis of correlation removes unrelated variables—those that have no meaningful relationship with the target variable or the other variables in the system. These variables are unlikely to act as confounders since they do not introduce residual confounding or spurious relationships when removed.

For validation, after CAM pruning, the causal structure is validated using domain expertise to ensure the robustness of the inferred causal graph. CAM pruning is not a substitute for confounding control methods. It is suggested to be used in combination with other techniques to ensure the validity of causal inferences. This is also the rationale for incorporating LLMs with expertise knowledge for further exploration.

# 5 Validation: From Correlations to Causation via LLM Inquiries

In the specific context of climate change, do LLMs offer better causal inference? To address the request involving the exploration of causality factors for carbon emissions using the World Bank variables ("EG.CFT.ACCS.RU.ZS", "EG.CFT.ACCS.UR.ZS", "SP.URB.TOTL.IN.ZS") as the piror benchmark, the study categorizes questions into five main types for LLMs prompts (See Taxonomy of Causality).

This study follows (Ceraolo et al., 2024)'s CausalQuest database focusing more on the economy and climate change background. The study follows (Ceraolo et al., 2024)'s CausalQuest database, but focuses more on the economic and climate change context. Similarly, the study adopts Pearl's Causal Hierarchy (PCH) framework ((Pearl and Mackenzie, 2018; Bareinboim et al., 2022)), and defines a causal question as one that meets the following criteria: a question is considered causal if it involves, or if its solution process includes, any inquiry into the effects given a specific cause, and the causes given a specific effect, or the causal relationship between the given variables.

## 5.1 Taxonomy of Causality

The causal taxonomy-"Direct, Preventative, Facilitative, Resultative, and Influential"-describes various types of causal relationships that verbs can imply. This approach controls for confounding variables during LLM inquiries by leveraging a structured causal taxonomy to identify, classify, and account for different types of causal relationships(Liang et al., 2023; Cui et al., 2024).

The taxonomy classifies causal relationships into "Direct, Preventative, Facilitative, Resultative, and Influential" categories, ensuring that the LLMs recognize the nature of relationships between variables. By explicitly categorizing verbs that describe causal interactions, it helps avoid misinterpretation of ambiguous or indirect relationships, which could otherwise lead to confounding. In the case of carbon emissions: A variable like access to clean technology might be classified as Facilitative (like "facilitates a reduction in emissions") rather than Direct (like "directly reduces emissions"), ensuring proper distinction. Urbanization, classified as Influential (like "influences emissions through energy use patterns"), ensures its role as a broader contextual factor is not conflated with a direct cause.

"The direct" refers to actions or driven forces that have a straightforward and immediate impact on an outcome. In this condition, the cause directly influences the effect without intermediary steps. For example, "increase" or "trigger" are direct causal verbs because they indicate a direct cause-effect relationship(Girju and Moldovan, 2002; Kozareva, 2012; Riaz and Girju, 2014; Nazaruka, 2020). Example: "Urban access to clean fuels directly reduces carbon emissions."

"The preventative" involves actions or causes that prevent or reduce the likelihood of a particular outcome. These verbs imply that the cause acts as a barrier to a negative effect. Common verbs include "prevent", "reduce" and "inhibit"(Allen, 2005). Example: "Improved access to clean technologies prevents an increase in carbon emissions."

"The facilitative" includes causes that make it easier or more likely for an effect to occur but do not directly cause the effect themselves. Facilitative causes provide support or create conditions that enable the outcome. Verbs like "enable", "allow", or "support" are examples(Harvey et al., 2002; Wolff, 2003). Example: "Access to urban clean fuels facilitates a reduction in carbon emissions."

"The resultative" describes causes that lead to specific outcomes, often emphasizing the result or consequence of an action. These verbs focus on the outcome rather than the action itself. Verbs like "lead to", "result in" or "cause" fit into this category(Boas, 2000; Pena Cervel, 2015). Example: "The urban population increase results in higher carbon emissions."

"The influential" includes actions or factors that exert an influence on the effect but might not completely determine it. These causes often affect the likelihood or intensity of the effect indirectly. Verbs like "influence", "impact" or "affect" are often used(Yee, 1996; Slovic et al., 2007, 2013). Example: "Urbanization influences carbon emissions through changes in energy use patterns."

# 6 Results of Causal Relationship and Interpretations

Based on data of carbon emissions and social impacts (266 countries/regions, 70 socio-economic indicators, 20 years), the analysis identifies that Access to Clean Fuels and Technologies for Cooking (percent of Rural Population)("EG.CFT.ACCS.RU.ZS"), Access to Clean Fuels and Technologies for Cooking (percent of Urban Population)("EG.CFT.ACCS.UR.ZS"), and Urban Population (percent of Total Population)("SP.URB.TOTL.IN.ZS") have strong causal effects on carbon emissions per capita. These findings are consistent with previous research, which has verified the significant influence of these variables in the context of climate change and energy transitions

Access to clean fuels and technologies for cooking, rural("EG.CFT.ACCS.RU.ZS"): This variable is highly influential in the context of climate change as it directly affects carbon emissions through the use of clean versus polluting energy sources in rural areas. A higher score indicates that rural access to clean fuels significantly reduces carbon emissions, highlighting its critical role in mitigating climate change impacts in less urbanized social context (Nathaniel and Adeleye, 2021; Verma et al., 2021).

Access to clean fuels and technologies for cooking, urban("EG.CFT.ACCS.UR.ZS"): Similar to the rural access variable, this factor measures the availability of clean cooking technologies in urban areas. Urban access is crucial since densely populated regions can contribute substantially to emissions. Improving clean energy access in urban areas can lead to a significant reduction in overall emissions, making it a key target for policy interventions (Naeem et al., 2023).

Urban population as a percentage of total population("SP.URB.TOTL.IN.ZS"): This variable captures the influence of urbanization on carbon emissions. As urban populations grow, the demand for energy, transportation, and industrial activity increases, contributing to higher emissions(Hankey and Marshall, 2010; Madlener and Sunak, 2011; Li and Lin, 2015). The score associated with this variable indicates that urbanization plays a major role in driving climate change, necessitating targeted strategies to manage urban growth sustainably.

# 7 Conclusion: Evaluations and Integrations

The three-step causal inference framework for data-driven decision-making in climate change context integrates correlation analysis, machine Learning, and LLM-interpretations. In this framework, correlation analysis helps narrow down and identify connections, causality provides a stricter and more precise understanding of these relationships, and LLMs interpret the results within specific scenarios.

Correlation provides a preliminary view of the relationships by highlighting mutual associations among variables and measures that indicate the extent to which two or more variables paired with each other. It narrows the scope of investigation by identifying potential connections between variables, but they do not provide insights into the nature of these connections.

Causality involves understanding the directional influence from one variable to another. In this research, causality goes a step beyond correlation by aiming to establish a cause-and-effect relationship between variables. Exploring and understanding causality is more stringent and complex because it requires not just observing that two variables occur together but also demonstrating that one variable produces an effect on another.

LLMs could not handle causality explicitly and could not differentiate between mere correlations and true causal relationships ((Zečević et al., 2023; Liu et al., 2024). However, in interpreting results, LLMs can offer insights that are conditioned on their training data and the scenarios they are designed to understand. This means that while LLMs can be adept at identifying patterns and generating responses based on correlations, their ability to correctly interpret causal relationships could be improved by leveraging a structured causal taxonomy.

# References


Allen, C. (2005). *A local grammar of cause and effect: A corpus-driven study*. Ph. D. thesis, University of Birmingham.

Bareinboim, E., J. D. Correa, D. Ibeling, and T. Icard (2022). On pearl's hierarchy and the foundations of causal inference. In H. Geffner, R. Dechter, and J. Y. Halpern (Eds.), *Probabilistic and Causal Inference: The Works of Judea Pearl*, Volume 36 of *ACM Books*, pp. 507–556. ACM.

Boas, H. C. (2000). *Resultative constructions in English and German*. The University of North Carolina at Chapel Hill.

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, Volume 33, pp. 1877–1901.

Ceraolo, R., D. Kharlapenko, A. Reymond, R. Mihalcea, M. Sachan, B. Schölkopf, and Z. Jin (2024). Causalquest: Collecting natural causal questions for ai agents.

Cheng, D., J. Li, L. Liu, J. Liu, and T. D. Le (2024). Data-driven causal effect estimation based on graphical causal modelling: A survey. *ACM Computing Surveys 56*(5), 1–37.

Cui, S., Z. Jin, B. Schölkopf, and B. Faltings (2024). The odyssey of commonsense causality: From foundational benchmarks to cutting-edge reasoning.

Forbes, M., J. D. Hwang, V. Shwartz, M. Sap, and Y. Choi (2020, November). Social chemistry 101: Learning to reason about social and moral norms. In B. Webber, T. Cohn, Y. He, and Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online, pp. 653–670. Association for Computational Linguistics.

Girju, R. and D. Moldovan (2002). Mining answers for causation questions. In *Proc. The AAAI Spring Symposium on Mining Answers from Texts and Knowledge Bases*, pp. 67–82.

Hankey, S. and J. D. Marshall (2010). Impacts of urban form on future us passenger-vehicle greenhouse gas emissions. *Energy Policy 38*(9), 4880–4887.

Harvey, G., A. Loftus-Hills, J. Rycroft-Malone, A. Titchen, A. Kitson, B. McCormack, and K. Seers (2002). Getting evidence into practice: the role and function of facilitation. *Journal of advanced nursing 37*(6), 577–588.

Jin, Z., J. Liu, Z. Lyu, S. Poff, M. Sachan, R. Mihalcea, M. Diab, and B. Schölkopf (2024). Can large language models infer causation from correlation?

Kiciman, E., R. Ness, A. Sharma, and C. Tan (2024). Causal reasoning and large language models: Opening a new frontier for causality.

Kozareva, Z. (2012). Cause-effect relation learning. In *Workshop Proceedings of TextGraphs-7: Graph-based Methods for Natural Language Processing*, pp. 39–43.

Ksir, C. and C. L. Hart (2016). Correlation still does not imply causation. *The Lancet Psychiatry 3*(5), 401.

Kwiatkowski, T., J. Palomaki, O. Redfield, M. Collins, A. Parikh, C. Alberti, D. Epstein, I. Polosukhin, J. Devlin, K. Lee, K. Toutanova, L. Jones, M. Kelcey, M.-W. Chang, A. M. Dai, J. Uszkoreit, Q. Le, and S. Petrov (2019). Natural questions: A benchmark for question answering research. *Transactions of the Association for Computational Linguistics 7*, 452–466.

Li, K. and B. Lin (2015). Impacts of urbanization and industrialization on energy consumption/co2 emissions: does the level of development matter? *Renewable and Sustainable Energy Reviews 52*, 1107–1122.

Liang, P., R. Bommasani, T. Lee, D. Tsipras, D. Soylu, M. Yasunaga, Y. Zhang, D. Narayanan, Y. Wu, A. Kumar, B. Newman, B. Yuan, B. Yan, C. Zhang, C. Cosgrove, C. D. Manning, C. Ré, D. Acosta-Navas, D. A. Hudson, E. Zelikman, E. Durmus, F. Ladhak, F. Rong, H. Ren, H. Yao, J. Wang, K. Santhanam, L. Orr, L. Zheng, M. Yuksekgonul, M. Suzgun, N. Kim, N. Guha, N. Chatterji, O. Khattab, P. Henderson, Q. Huang, R. Chi, S. M. Xie, S. Santurkar, S. Ganguli, T. Hashimoto, T. Icard, T. Zhang, V. Chaudhary, W. Wang, X. Li, Y. Mai, Y. Zhang, and Y. Koreeda (2023). Holistic evaluation of language models.

Liu, X., P. Xu, J. Wu, J. Yuan, Y. Yang, Y. Zhou, F. Liu, T. Guan, H. Wang, T. Yu, et al. (2024). Large language models and causal inference in collaboration: A comprehensive survey. *arXiv preprint arXiv:2403.09606*.

Liu, Z., B. Malone, and C. Yuan (2012). Empirical evaluation of scoring functions for bayesian network model selection. *BMC bioinformatics 13*(Suppl 15), S14.

Madlener, R. and Y. Sunak (2011). Impacts of urbanization on urban structures and energy demand: What can we learn for urban energy planning and urbanization management? *Sustainable Cities and Society 1*(1), 45–53.

Naeem, M. A., M. Appiah, J. Taden, R. Amoasi, and B. A. Gyamfi (2023). Transitioning to clean energy: Assessing the impact of renewable energy, bio-capacity and access to clean fuel on carbon emissions in oecd economies. *Energy Economics 127*, 107091.

Nathaniel, S. P. and N. Adeleye (2021). Environmental preservation amidst carbon emissions, energy consumption, and urbanization in selected african countries: implication for sustainability. *Journal of Cleaner Production 285*, 125409.

Nazaruka, E. (2020). An overview of ways of discovering cause-effect relations in text by using natural language processing. In *Evaluation of Novel Approaches to Software Engineering: 14th International Conference, ENASE 2019, Heraklion, Crete, Greece, May 4–5, 2019, Revised Selected Papers 14*, pp. 22–38. Springer.

Nguyen, T., M. Rosenberg, X. Song, J. Gao, S. Tiwary, R. Majumder, and L. Deng (2016). Ms marco: A human generated machine reading comprehension dataset. In T. R. Besold, A. Bordes, A. S. d'Avila Garcez, and G. Wayne (Eds.), *Proceedings of the Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches 2016 Co-located with the 30th Annual Conference on Neural Information Processing Systems (NIPS 2016)*, Volume 1773 of *CEUR Workshop Proceedings*.

Nogueira, A. R., A. Pugnana, S. Ruggieri, D. Pedreschi, and Gama (2022). Methods and tools for causal discovery and causal inference. *Wiley interdisciplinary reviews: data mining and knowledge discovery 12*(2), e1449.

Ouyang, L., J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. W. Askell, P. F. Christiano, J. Leike, and R. Lowe (2022). Training language models to follow instructions with human feedback. *CoRR abs/2203.02155*.

Pearl, J. (2000). Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress 19*(2), 3.

Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.

Pearl, J. and D. Mackenzie (2018, May). *The Book of Why: The New Science of Cause and Effect* (First edition ed.). New York, NY: Basic Books.

Pena Cervel, M. S. (2015). A constructionist approach to causative frighten verbs. *Linguistics 53*(6), 1247–1302.

Peters, J., D. Janzing, and B. Schölkopf (2017). *Elements of causal inference: foundations and learning algorithms*. The MIT Press.

Riaz, M. and R. Girju (2014). Recognizing causality in verb-noun pairs via noun and verb semantics. In *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*, pp. 48–57.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in methods and practices in psychological science 1*(1), 27–42.

Rolland, P., V. Cevher, M. Kleindessner, C. Russell, D. Janzing, B. Schölkopf, and F. Locatello (2022). Score matching enables causal discovery of nonlinear additive noise models. In *International Conference on Machine Learning*, pp. 18741–18753. PMLR.

Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor (2007). The affect heuristic. *European journal of operational research 177*(3), 1333–1352.

Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor (2013). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk and rationality. In *The feeling of risk*, pp. 21–36. Routledge.

Sluga, H. D. and D. G. Stern (1996). *the Cambridge companion to Wittgenstein*. Cambridge University Press.

Spirtes, P., C. Glymour, and R. Scheines (2001). *Causation, Prediction, and Search* (2 ed.). Cambridge, MA: MIT Press.

Srivastava, A., A. Rastogi, A. Rao, A. A. M. Shoeb, A. Abid, A. Fisch, A. R. Brown, A. Santoro, A. Gupta, A. Garriga-Alonso, et al. (2022). Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *arXiv preprint arXiv:2206.04615*.

Van Hee, C., E. Lefever, B. Verhoeven, J. Mennes, B. Desmet, G. De Pauw, W. Daelemans, and V. Hoste (2015). Detection and fine-grained classification of cyberbullying events. In *Proceedings of the international conference recent advances in natural language processing*, pp. 672–680.

Verma, P., T. Kumari, and A. S. Raghubanshi (2021). Energy emissions, consumption and impact of urban households: A review. *Renewable and Sustainable Energy Reviews 147*, 111210.

Wang, S., G. Durrett, and K. Erk (2018, June). Modeling semantic plausibility by injecting world knowledge. In M. Walker, H. Ji, and A. Stent (Eds.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, New Orleans, Louisiana, pp. 303–308. Association for

Computational Linguistics.
Wittgenstein, L. (1967). *Philosophical Investigations* (3 ed.). Oxford, UK: Basil Blackwell Ltd.
Wolff, P. (2003). Direct causation in the linguistic coding and individuation of causal events. *Cognition 88*(1), 1–48.
Yee, A. S. (1996). The causal effects of ideas on policies. *International organization 50*(1), 69–108.
Zečević, M., M. Willig, D. S. Dhami, and K. Kersting (2023). Causal parrots: Large language models may talk causality but are not causal.
Zhao, W., X. Ren, J. Hessel, C. Cardie, Y. Choi, and Y. Deng (2024). (inthe)wildchat: 570k chatgpt interaction logs in the wild. In *The Twelfth International Conference on Learning Representations*.
Zhou, Y., A. I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, and J. Ba (2022). Large language models are human-level prompt engineers. *arXiv preprint arXiv:2211.01910*.

# A Three-step Framework for Causal Analysis

This research proposes a three-step framework for causal inference that progresses from understanding correlations to establishing causality, and finally to interpreting these relationships via LLMs. This approach leverages the different methods and exploits their distinctive advances to align with the understanding of climate change issues. This structured approach helps to systematically explore and analyze the causality factors associated with carbon emissions, translates data patterns into LLM-inquiry-driven interpretations, which aids in gaining deeper insights and more interpretable policymaking.

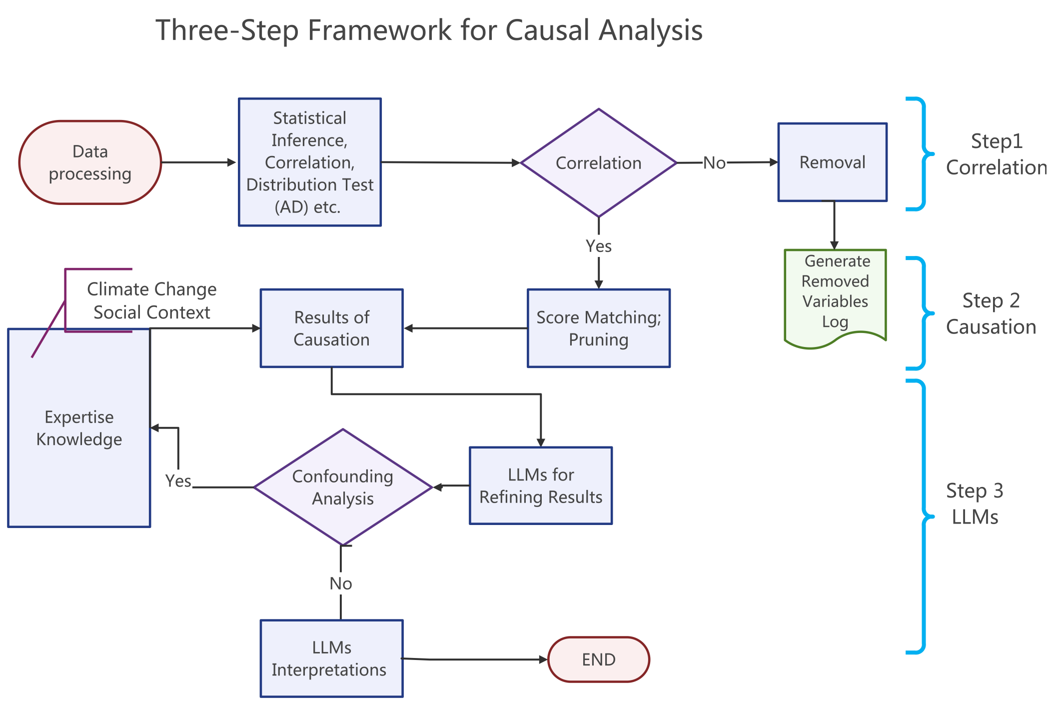


Figure 1: Three-steps Framework for Causal Analysis

# B Setup and Main Results

- Data Availability The data and code that support the findings of this study are available at https://github.com/shanshanfy/climate-change
- Developer Environment Availability The author’s environment ‘ClimateChangePackages.yaml’ file for Conda open-source package management system is provided through:

https://github.com/shanshanfy/climate-change. It allows for isolated environments to manage packages without interference. The file contains the configuration of the project's Python environment, including channels, dependencies, and library versions.

### B.1 Data Processing

The research identifies the socioeconomic factors that influence and contribute to carbon emissions and climate change. The data is available at `https://www.climatewatchdata.org/ghg-emissions`. Total carbon emissions are measured as carbon emissions per capita. The complete carbon emission dataset are collected from 265 countries and includes 100 variables related to carbon emissions for the years 2000, 2005, 2010, 2015, and 2020. Emissions data are sourced from Climate Watch Historical GHG Emissions (1990-2020). 2023. Washington, DC: World Resources Institute.

# C Correlation Map

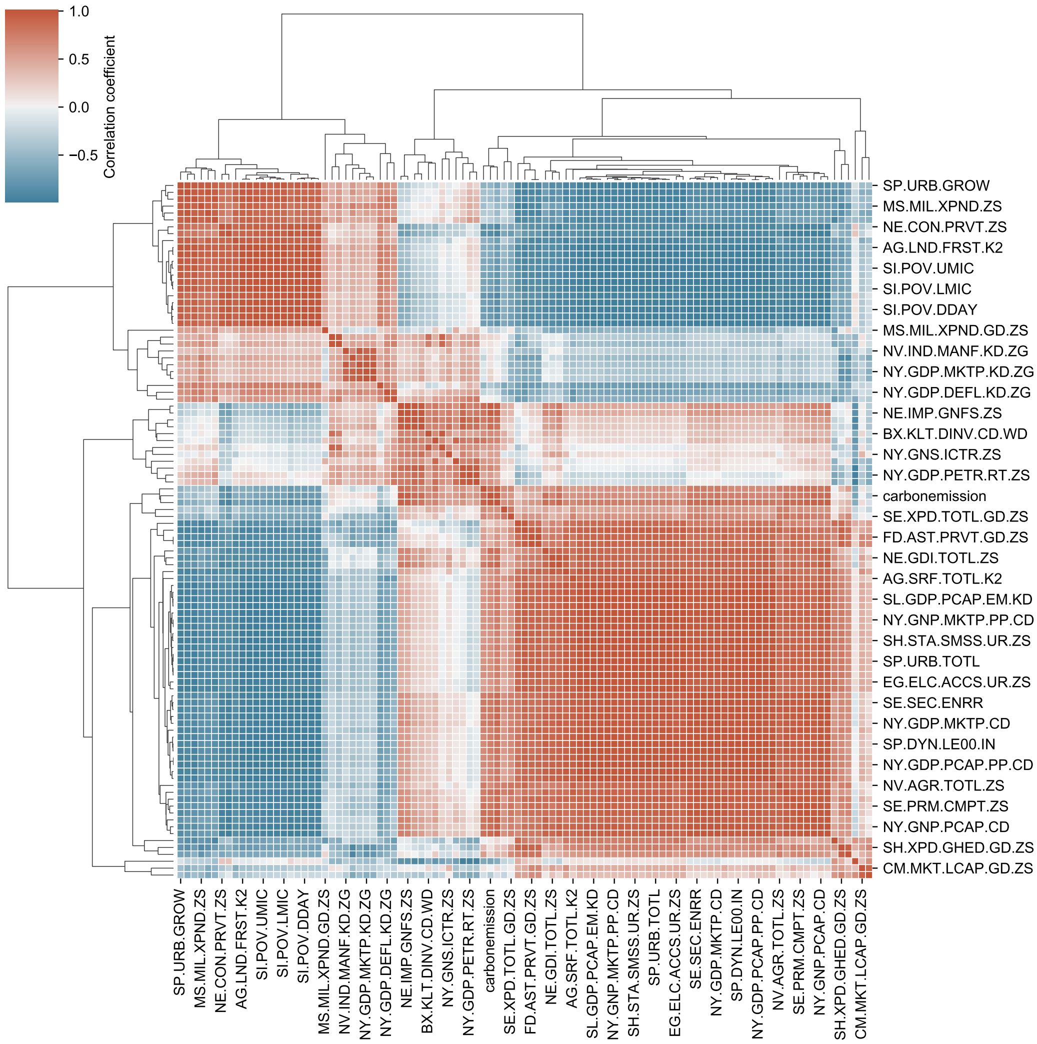


Figure 2: **Clustered Correlation Heatmap of Social Factors Influencing Carbon Emissions** This heatmap illustrates the correlations between various social factors and carbon emissions, highlighting key relationships. The clustering visually groups factors with similar correlation patterns, aiding in identifying which socioeconomic indicators most strongly influence carbon emissions, thereby providing insights into the complex interplay between social behavior and climate impact. The dendrogram, shown as lines on the top and left of the heatmap, represents hierarchical clustering. It groups variables based on similarity of correlation or distance, with shorter line heights indicating higher similarity. Variables in the rows and columns are grouped to identify clusters with closely related pairwise relationships.

### C.1 Ordered Feature Correlation with Carbon Emissions

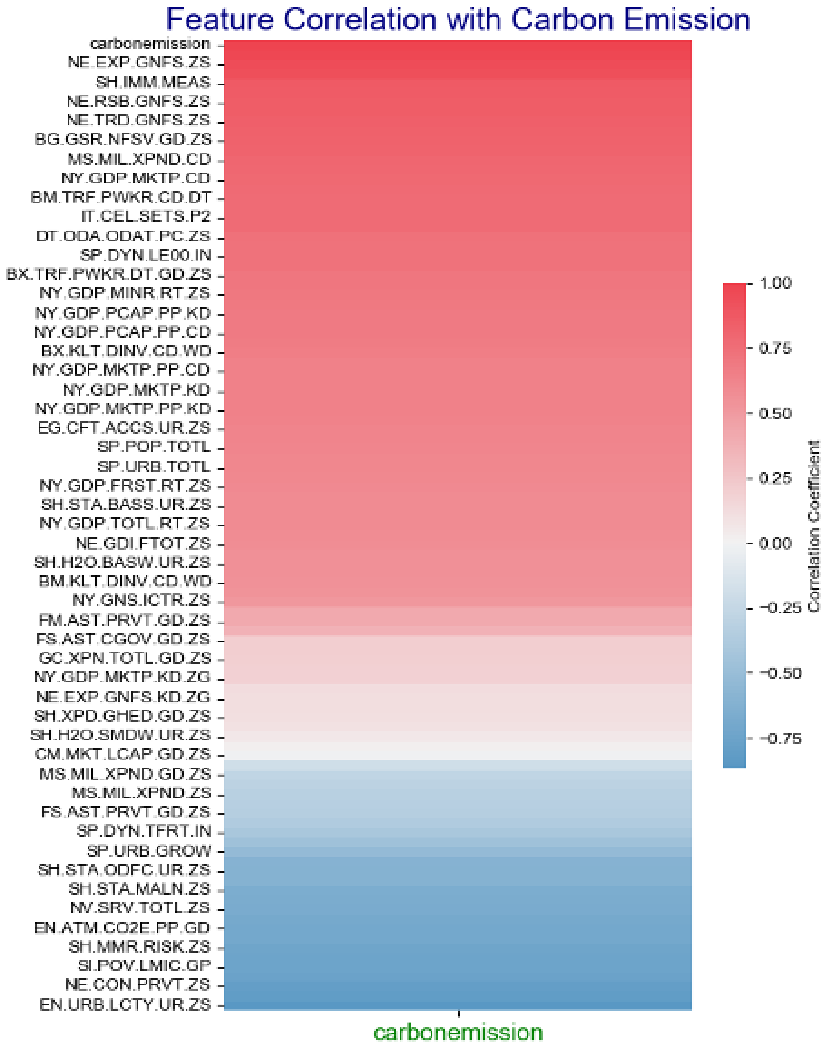


Figure 3: **Ordered Feature Correlation with Carbon Emissions.** This figure shows the correlation between various social and economic factors and carbon emissions per capita, highlighting key influences such as energy use, GDP, urban population, and access to clean technologies.

# D Causal Analysis

## D.1 Data Descriptions

Given the substantial amount of missing data, the experiment eliminated columns with a missing ratio greater than 40%, as well as some columns with data that are difficult to observe. The results have identified 15 studied variables related to the carbon emission variable, including "EG.CFT.ACCS.ZS", "EG.CFT.ACCS.RU.ZS", "EG.CFT.ACCS.UR.ZS", "EG.ELC.ACCS.UR.ZS", "EG.ELC.ACCS.ZS", "SP.URB.TOTL", "SP.URB.TOTL.IN.ZS", "SP.URB.GROW", "SE.SEC.DURS", "EG.FEC.RNEW.ZS", "SP.RUR.TOTL.ZS", "SP.RUR.TOTL", "AG.LND.FRST.ZS", "ER.FSH.CAPT.MT".The description of these variables is shown in Table 1.

Subsequently, the study treats the data into a numerical matrix $V^t \in \mathbb{R}^{134\times 15}$ and denote 'EN.ATM.CO2E.PC' by $Y^t \in \mathbb{R}^{134\times 1}$, where $t \in \{2000, 2005, 2010, 2015, 2020\}$ denotes the year in which the data was collected. We then normalize and standardize each column of data. Finally, using a five-year interval as a step, the study investigates the causal relationship between $X$ and $Y$. The integrated observational data is as follows:

$$D = \{(V^{2000}, Y^{2005}), (V^{2005}, Y^{2010}), (V^{2010}, Y^{2015}), (V^{2015}, Y^{2020})\},$$

Table 1: Description of the Studied Variables for Causal Analysis

| Variable | Description |
|---|---|
| EG.CFT.ACCS.ZS | Access to clean fuels and technologies for cooking (% of population) |
| EG.CFT.ACCS.RU.ZS | Access to clean fuels and technologies for cooking, rural (% of rural population) |
| EG.CFT.ACCS.UR.ZS | Access to clean fuels and technologies for cooking, urban (% of urban population) |
| EG.ELC.ACCS.UR.ZS | Access to electricity, urban (% of urban population) |
| EG.ELC.ACCS.ZS | Access to electricity (% of population) |
| SP.URB.TOTL | Urban population |
| SP.URB.TOTL.IN.ZS | Urban population (% of total population) |
| SP.URB.GROW | Urban population growth (annual %) |
| SE.SEC.DURS | Secondary education, duration (years) |
| EG.FEC.RNEW.ZS | Renewable energy consumption (% of total final energy consumption) |
| SP.RUR.TOTL.ZS | Rural population (% of total population) |
| SP.RUR.TOTL | Rural population |
| AG.LND.FRST.ZS | Forest area (% of land area) |
| ER.FSH.CAPT.MT | Capture fisheries production (metric tons) |
| EN.ATM.CO2E.PC | CO2 emissions (metric tons per capita) |

## D.2 Scoring Matching Output: Causal Relationships Among Social Factors and Carbon Emissions

| | EG.CFT.ACCS.ZS | EG.CFT.ACCS.RU.ZS | EG.CFT.ACCS.UR.ZS | EG.ELC.ACCS.UR.ZS | EG.ELC.ACCS.ZS | SP.URB.TOTL | SP.URB.TOTL.IN.ZS | SP.URB.GROW | SE.SEC.DURS | EG.FEC.RNEW.ZS | SP.RUR.TOTL.ZS | SP.RUR.TOTL | AG.LND.FRST.ZS | ER.FSH.CAPT.MT | CO2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EG.CFT.ACCS.ZS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 4 | 0 | 0 | 0 |
| EG.CFT.ACCS.RU.ZS | 0 | 0 | 4 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 2 |
| EG.CFT.ACCS.UR.ZS | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 4 | 0 | 0 | 1 |
| EG.ELC.ACCS.UR.ZS | 0 | 4 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 4 | 0 | 0 | 0 |
| EG.ELC.ACCS.ZS | 0 | 1 | 0 | 3 | 0 | 4 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| SP.URB.TOTL | 0 | 2 | 0 | 1 | 0 | 0 | 1 | 0 | 4 | 0 | 4 | 1 | 0 | 0 | 0 |
| SP.URB.TOTL.IN.ZS | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| SP.URB.GROW | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SE.SEC.DURS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EG.FEC.RNEW.ZS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SP.RUR.TOTL.ZS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 |
| SP.RUR.TOTL | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| AG.LND.FRST.ZS | 0 | 2 | 0 | 0 | 1 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ER.FSH.CAPT.MT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CO2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Figure 4: Causal Relationships Among Social Factors and Carbon Emissions. This scoring map illustrates the causal relationships between various social factors and carbon emissions, highlighting key variables: access to clean fuels and technologies for cooking in rural and urban areas (EG.CFT.ACCS.RU.ZS and EG.CFT.ACCS.UR.ZS) and urban population percentage (SP.URB.TOTL.IN.ZS). These factors show strong causal effects on carbon emissions per capita, emphasizing the interconnectedness of urbanization, energy use, and climate change.

## D.3 CAM Pruning

• Structural similarity (via SID and SHD). • Predictive accuracy (via precision, recall, and F1 score). • Overall deviation (via L2 distance).

### D.3.1 Function Definition: BackRE

The `backRE` function calculates several metrics to evaluate the accuracy of a predicted Directed Acyclic Graph (DAG) against the target DAG. Below is the Python implementation:

```
def backRE(tar_DAG, P_KCI):
    sid_val = SID(tar_DAG, P_KCI)
    shd_val = SHD(tar_DAG, P_KCI)
    precision, recall, f1 = f1_score(tar_DAG, P_KCI)
    distance = l2_distance(tar_DAG, P_KCI)
    return [sid_val, shd_val, precision, recall, f1, distance]
```

The function computes:

- `SID`: Structural Intervention Distance.
- `SHD`: Structural Hamming Distance.
- Precision, Recall, and F1-score to evaluate edge predictions.
- L2 Distance to measure overall deviation between the target DAG and the predicted DAG.

#### D.3.2 L2 Distance Formula

The L2 Distance is calculated using the formula:

$$\text{L2 Distance} = \sqrt{\sum_{i,j} \left(A_{\text{true}}[i,j] - A_{\text{pred}}[i,j]\right)^2}$$

where:

- $A_{\text{true}}[i,j]$: Entry in the adjacency matrix of the true DAG.
- $A_{\text{pred}}[i,j]$: Entry in the adjacency matrix of the predicted DAG.

This metric provides a scalar measure of the overall deviation between the true and predicted graphs.

# E LLM Inquires

## E.1 Related Works

As public use of LLMs for tasks, various resources and tools have emerged to aid in prompt engineering and discovery[3]. Instruction Categories provide different strategies for prompt engineering, and this study employs the following methods. Zero-shot Evaluation Instruction and Zero-shot-CoT Instruction are similar, with the latter explicitly incorporating "chain of thought" reasoning Brown et al. (2020); Zhou et al. (2022); Srivastava et al. (2022). Both approaches assess the model's ability to apply its training to new, unseen tasks without prior specific examples. Few-shot Evaluation Instruction and Resample Instruction involve adaptive learning from a small set of examples or feedback, iteratively refining the prompts. Forward Generation predicts subsequent content based on the preceding context, commonly used in natural language generation.

However, there is a lack of a comprehensive collection of causal questions in previous research Ceraolo et al. (2024). While related databases such as Google (Kwiatkowski et al., 2019), Bing (Nguyen et al., 2016), and questions posed to LLMs (H-to-LLM) from sources like ShareGPT and WildChat (Zhao et al., 2024) exist, none specifically focus on new sources of natural-causal questions, particularly causal questions directly asked to LLMs (Ouyang et al., 2022; Jin et al., 2024). Moreover, there is no dedicated database that addresses the context of climate change.

## E.2 LLM-Generated Mixed Questions

***Understanding Variables***

- **Direct:** What does `EG.CFT.ACCS.RU.ZS` represent in the context of global carbon emissions?
- **Influential:** How might urban access to clean fuels (`EG.CFT.ACCS.UR.ZS`) impact carbon emissions?
- **Facilitative:** What is the significance of `SP.URB.TOTL.IN.ZS` in studying urbanization effects on the environment?
- **Influential:** How do these variables interact to influence overall carbon emissions?

***Historical Data Analysis***

- **Resultative:** What trends are observable in `EG.CFT.ACCS.RU.ZS` over the last decade?
- **Resultative:** Has there been a significant change in `EG.CFT.ACCS.UR.ZS` data in major industrial countries?
- **Resultative:** How has the urban population percentage (`SP.URB.TOTL.IN.ZS`) changed in emerging economies?

[3] `https://platform.openai.com/docs/guides/prompt-engineering`
`https://huggingface.co/spaces/Gustavosta/MagicPrompt-Stable-Diffusion`
`https://promptomania.com/stable-diffusion-prompt-builder/`

- **Influential:** What historical events have significantly impacted these variables?

*Predictive Modeling*

- **Resultative:** Can we predict future trends in `EG.CFT.ACCS.RU.ZS` using past data?
- **Influential:** How might changes in `EG.CFT.ACCS.UR.ZS` predict shifts in urban carbon emissions?
- **Facilitative:** What models can forecast the growth of urban populations (`SP.URB.TOTL.IN.ZS`)?
- **Preventative:** What are the potential future scenarios for these variables under different policy implementations?

*Policy Impact Evaluation*

- **Influential:** How have recent policies affected rural access to clean technologies (`EG.CFT.ACCS.RU.ZS`)?
- **Resultative:** What are the environmental impacts of improved urban access to clean fuels (`EG.CFT.ACCS.UR.ZS`)?
- **Influential:** How does urbanization measured by `SP.URB.TOTL.IN.ZS` correlate with national carbon emission policies?
- **Preventative:** What policy measures could potentially alter the trends in these variables most effectively?

### E.3 LLM-Generated Why Questions

*Understanding Variables*

- **Direct:** Why does `EG.CFT.ACCS.RU.ZS` matter in the context of global carbon emissions?
- **Influential:** Why might urban access to clean fuels (`EG.CFT.ACCS.UR.ZS`) influence carbon emissions?
- **Facilitative:** Why is `SP.URB.TOTL.IN.ZS` significant when studying the effects of urbanization on the environment?
- **Influential:** Why do these variables interact to influence overall carbon emissions?

*Historical Data Analysis*

- **Resultative:** Why are there observable trends in `EG.CFT.ACCS.RU.ZS` over the last decade?
- **Resultative:** Why has there been a significant change in `EG.CFT.ACCS.UR.ZS` data in major industrial countries?
- **Resultative:** Why has the urban population percentage (`SP.URB.TOTL.IN.ZS`) changed in emerging economies?
- **Influential:** Why have certain historical events significantly impacted these variables?

*Predictive Modeling*

- **Resultative:** Why can past data on `EG.CFT.ACCS.RU.ZS` be used to predict future trends?
- **Influential:** Why might changes in `EG.CFT.ACCS.UR.ZS` predict shifts in urban carbon emissions?
- **Facilitative:** Why are certain models effective at forecasting the growth of urban populations (`SP.URB.TOTL.IN.ZS`)?
- **Preventative:** Why could potential future scenarios for these variables differ under various policy implementations?

*Policy Impact Evaluation*

- **Influential:** Why have recent policies affected rural access to clean technologies (EG.CFT.ACCS.RU.ZS)?
- **Resultative:** Why do improved urban access to clean fuels (EG.CFT.ACCS.UR.ZS) have environmental impacts?
- **Influential:** Why does urbanization, as measured by SP.URB.TOTL.IN.ZS, correlate with national carbon emission policies?
- **Preventative:** Why might certain policy measures most effectively alter the trends in these variables?

# F Limitations and Discussions

The study knocks on the door of causal and inference and evaluates the LLM-inquiry performance. However, understanding how to question causality within LLMs also involves recognizing the social norms embedded in human-machine interactions, as well as the social and moral dynamics present in language (Van Hee et al., 2015; Wang et al., 2018; Forbes et al., 2020; Cui et al., 2024). These aspects are far more complex than simple data patterns.

Technically, three limitations exist. First, for data dependency, the accuracy and reliability of the causal inferences drawn from this framework depend heavily on the quality and completeness of the underlying data. Poor data quality or gaps can lead to incorrect conclusions, potentially misguiding important policy decisions. Second, for model assumptions, the three-step framework relies on the global carbon emissions and climate change assumptions that may need to be supported across different scenarios or contexts, particularly in the complex, multifactorial climate change domain. Third, for generalizability, findings derived from this framework are context-specific and may not apply to different settings or scenarios without adjustments.

In future research, with more data and a deeper grounding in real-world societal settings, studies on vertical domains could be expanded on a larger scale and have a more profound impact on policymaking.

# G Acknowledgment

The author thanks Anpeng Wu for his contributions to the analytical work on causal methodology and analysis. The author thanks Prof. Fei Wu for his comments and suggestions on causal analysis and Dr. Xin Qiu for his comments on the article. The author thanks the NeurIPS reviewers for the valuable comments.

# H Ethical Considerations

The research does not use any privacy-sensitive data. It utilizes a publicly available World Bank dataset containing no information about ethical conflicts.