

A cautionary (Canterbury) tale of Causal AI

Chris Hand¹ and Elena Fitkov-Norris²

1. Department of Strategy Marketing and Innovation, Kingston Business School

2. Department of Accounting Finance and Informatics, Kingston Business School

Track: Research Methods

Paper type: Development paper

A cautionary (Canterbury) tale of Causal AI

Introduction and Background

Artificial Intelligence has attracted a lot of attention as a tool for research and for its implications for learning and teaching (e.g. Stoken-Walker and Van Noorden, 2023). However, most recent attention seems to have been focused on Generative AI (GenAI). GenAI is typified by platforms like ChatGPT and generates a response to a prompt based on the data it has been trained on. There is however a second type of AI which will also have an impact on how research is carried out: Causal AI (Fitkov-Norris and Kocheva, 2025). Causal AI utilises Directed Acyclic Graphs (DAGs) and Structural Causal Models (SCMs) to identify and quantify causal relationships in data models by controlling for the impact of confounders (Pearl and Mackenzie, 2018). In addition, through the use of causal discovery algorithms, Causal AI has the potential to enable researchers to identify sets of unbiased causal relationships between variables from observational data, without having to specify in advance if and how the variables may be related (Malinsky and Danks, 2018). Both forms of AI are data driven methods and as such are largely atheoretical. This is in stark contrast to standard approaches to quantitative research in Business and Management which tends to be more deductive, testing sets of proposed hypotheses or relationships between a set variables or constructs (although data driven approaches are increasingly used e.g. Valizade et al., 2024). Such AI methods (and other machine learning methods) adopt an approach which is arguably purely inductive, in that the results are allowed to emerge from the data.

Advocates of Causal AI over predictive AI methods (e.g. Sgaier et al., 2020) suggest that Causal AI methods avoid the potential pitfalls of predictive AI, in particular confusing correlation with causation by identifying underlying causal relationships. However, such results are still purely data-driven. Considering AI more generally (and in particular Large Language Models) Felin and Holweg (2024) argue that AI methods are largely backward-looking (in that they are dependent on the data they are trained on) whereas human cognition is more forward-looking, characterised as theory-based causal reasoning. Bishop (2021) is more forthright, arguing that AI is fundamentally limited by the way it learns. AI tools do not have a conceptual understanding of relationships between variables which would provide a way of sense-checking the results.

Causal AI models are increasingly available, not just through languages such as R and Python, but through more user-friendly interfaces such as Microsoft's ShowWhy (<https://github.com/microsoft/showwhy>). Given this availability it is useful for researchers to be aware of any limitations of the methods. Our aim in this paper is to explore Causal AI's ability to identify a set of underlying relationships in a dataset and to investigate its ability to distinguish between correlation and causation in a single cross-section data set.

Study Aim

The aim of the study is to test Causal AI algorithms' ability to discern underlying relationships between a set of variables. Prior studies have compared causal maps produced by an algorithm to those produced by (human) experts (e.g. Prosperi et al., 2021). In this study we adopt different approach. A dataset was created for this study, so the underlying relationships between variables are known. The data was created to test the algorithms' ability to determine underlying relationships and to distinguish between correlations and causal relationships.

Overview of Causal Discovery Algorithms

Causal discovery algorithms seek to infer causal structures from observational data by identifying the direction and strength of causal relationships between variables. Unlike

traditional predictive models, which primarily capture statistical associations, causal discovery methods aim to distinguish between correlation and causation. This section examines three prominent causal discovery algorithms: the PC algorithm, Direct LiNGAM, and Deep End-to-End Causal Inference (DECI), each of which have their own methodological assumptions.

The PC Algorithm

The PC (Peter-Clark) algorithm is a constraint-based causal discovery method that identifies causal structures through conditional independence testing (Glymour et al, 2019). It begins by constructing a fully connected undirected graph, progressively eliminating edges between variables based on statistical tests for conditional independence. Once a skeletal graph is obtained, the algorithm applies a series of orientation rules to direct causal relationships while ensuring the resulting structure remains acyclic.

A principal advantage of the PC algorithm is its ability to handle high-dimensional data efficiently, making it particularly suitable for large-scale causal discovery tasks. Furthermore, its reliance on independence testing allows it to infer causal structures without requiring predefined functional forms for relationships between variables. However, its performance is highly dependent on the reliability of the conditional independence tests, which can be compromised by small sample sizes or measurement noise. Additionally, the algorithm assumes causal sufficiency, meaning that all common causes of the observed variables are included in the dataset. This assumption may not always hold in practice, such as in the case of unobserved confounders for example, potentially leading to biased causal estimates.

Direct LiNGAM

The Linear Non-Gaussian Acyclic Model (LiNGAM) is a model-based causal discovery algorithm that assumes causal relationships follow a linear structural equation model (SEM) with non-Gaussian noise (Shimizu et al., 2011). The Direct LiNGAM variant refines this approach by directly estimating causal directions without the need for iterative resampling, thus improving computational efficiency.

A key strength of Direct LiNGAM is its ability to infer causal relationships without the need for prior interventions, making it particularly useful for observational datasets. By leveraging Independent Component Analysis (ICA), the algorithm exploits statistical properties of non-Gaussian noise to infer causal ordering among variables. This allows it to distinguish cause from effect, even in cases where traditional correlation-based methods fail.

Despite its advantages, Direct LiNGAM is constrained by its reliance on the linearity assumption, which limits its applicability to non-linear causal relationships. Furthermore, while the assumption of non-Gaussian noise aids causal inference, it may not always be a realistic condition in real-world datasets. The algorithm also struggles when dealing with latent confounding, that is, unobserved variables that influence multiple observed variables simultaneously, potentially leading to incorrect causal structures.

Deep End-to-End Causal Inference (DECI)

The Deep End-to-End Causal Inference (DECI) framework represents a more recent approach to causal discovery, integrating deep learning techniques with causal inference principles (Geffner et al, 2022). Unlike classical methods such as PC and LiNGAM, which rely on statistical independence assumptions, DECI employs a variational inference framework to learn causal structures in a data-driven manner. By parameterising structural equation models (SEMs) with neural networks, it is capable of modelling complex causal relationships, including non-linear interactions and latent confounding.

A primary advantage of DECI is its flexibility in handling large and complex datasets, particularly those with highly non-linear dependencies or hidden common causes. This makes it particularly well-suited for applications in fields such as biomedical sciences and finance, where traditional methods struggle due to intricate causal relationships. Additionally, DECI does not require strong assumptions about the data-generating process, enabling it to uncover causal relationships where classical algorithms may fail. However, DECI is not without limitations. As a deep learning-based method, it needs substantial computational resources and large training datasets to achieve robust results. Moreover, it operates as a black-box model, meaning that its inferred causal structures are often less interpretable than those produced by traditional approaches.

Data and Results

A dataset containing 11 variables was created, each containing 2,000 observations. The values for the exogenous variables were created from randomly selected values between 1 and 10. The endogenous variables, including the overall target (or dependent variable) were generated from values of the exogenous variables multiplied by a path coefficient plus a random noise component. The target variable was calculated using a logistic function, to give a dummy variable as the target. The relationships and path coefficients used are shown in figure 1. Variables kilo and juliet are strongly correlated, but not causally related ($r = 0.7$).

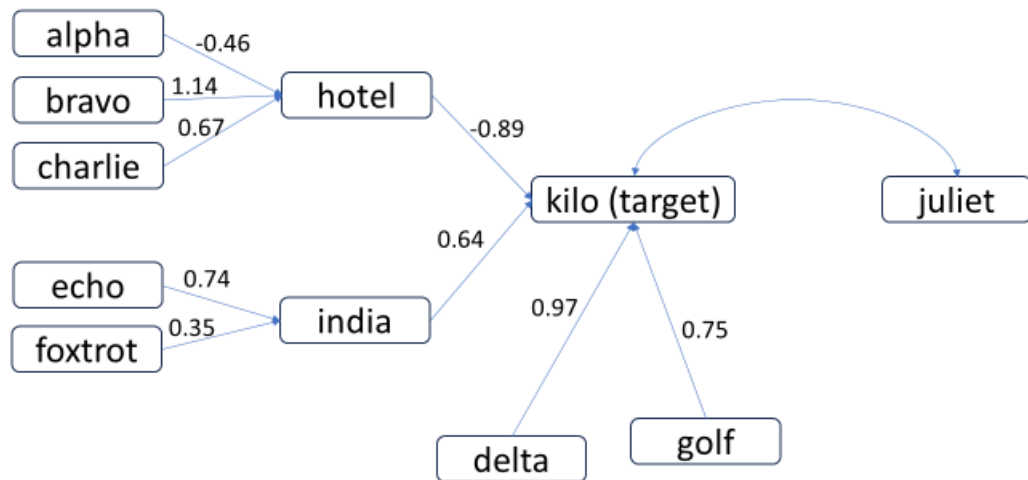


Figure 1. Underlying relationships between variables

The variable names were chosen deliberately to only act as labels for each variable and not to give a hint as to how the variables related to each other. The underlying model was set up so the target variable kilo is predicted by variables alpha, bravo and charlie mediated via hotel; echo and foxtrot mediated via india and by delta and golf directly. Variable juliet is correlated with kilo but was not used to generate it. Each variable had a random noise component added to it so that the endogenous variables (hotel, india and kilo) were not perfectly predicted by the antecedent variables. The challenges for the Causal AI algorithm were to a) identify this underlying structure and b) not suggest that variable juliet is causally linked to variable kilo. To make this a true test of the Causal AI algorithm's ability

the data was created by one author whilst the Causal AI algorithm was run by the other author without knowledge of the underlying set of relationships.

Results

The results derived from the PC, DirectLINGAM and DECI algorithms derived from the data are shown in Figures 2, Figure 3 and Figure 4 respectively. In the figures below the arrows indicate the relationships identified by the algorithms.

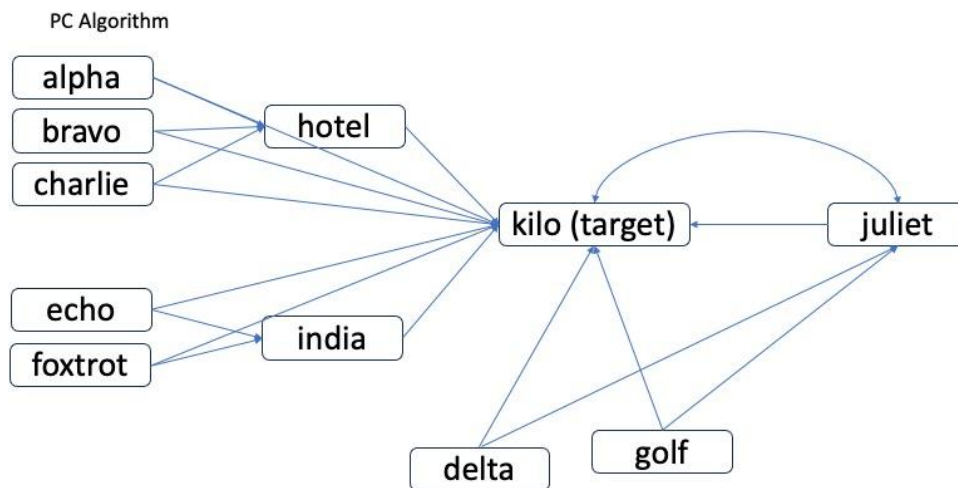


Figure 2 Dependencies Identified using the PC Algorithm

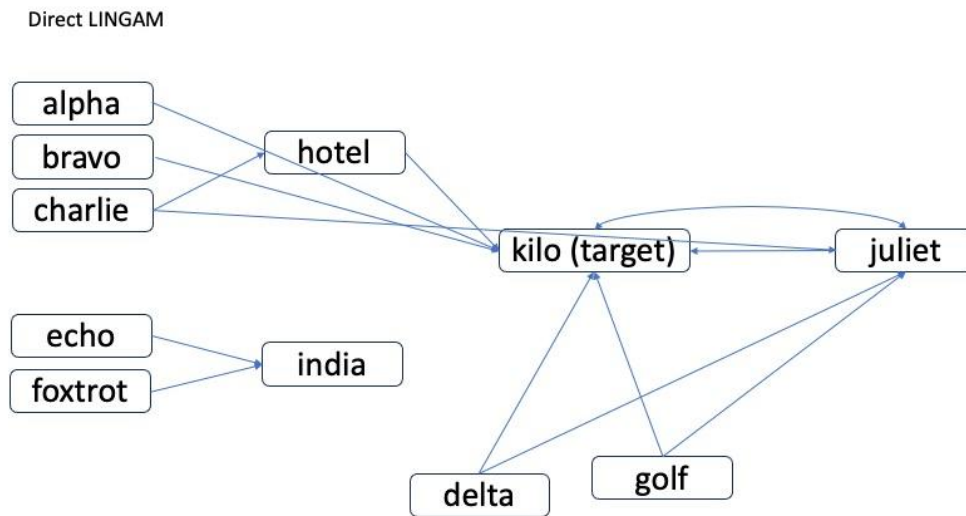


Figure 3 Dependencies Identified using the DirectLINGAM Algorithm

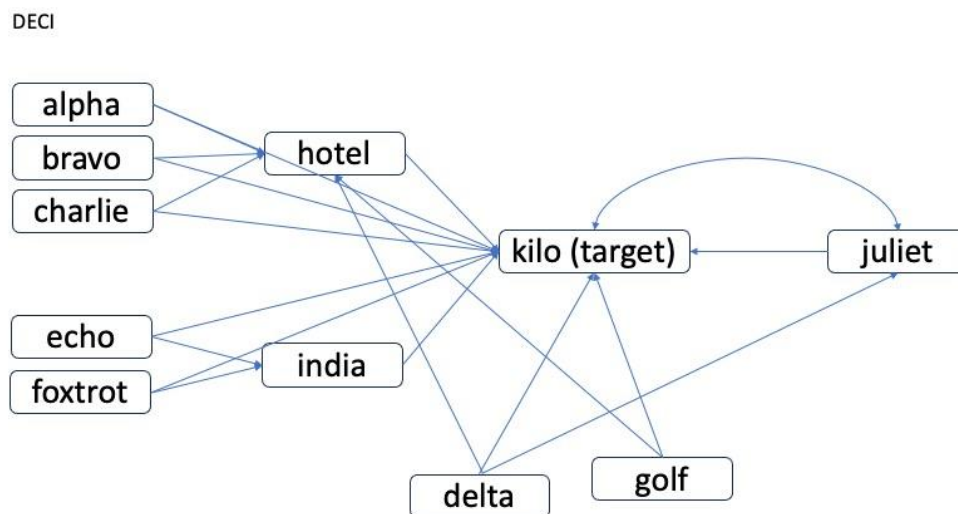


Figure 4 Dependencies Identified using the DECI Framework

All three algorithms incorrectly classified juliet as having a direct causal effect on kilo, failing to recognise that its association with the target variable was due to correlation. While this result was expected in the case of the PC algorithm, which assumes causal sufficiency, the failure of Direct LiNGAM and DECI, both of which are more sophisticated algorithms, is notable. This seems to undermine the claims made for Causal Ai. The inability to detect the hidden confounder highlights a critical challenge: even powerful causal discovery methods can lead to erroneous conclusions when applied by inexperienced researchers, and highlights the need for a conceptual understanding of relationships between variables to provide a sense-check on the findings.

Beyond this limitation, the PC algorithm demonstrated the most successful causal discovery of the underlying data generation mechanism. It correctly identified all direct relationships and indirect pathways, outperforming both Direct LiNGAM and DECI. However, it produced an overly complex model, as it inferred causal links in addition to indirect effects, rather than recognising them purely as indirect effects. Despite this, the PC algorithm was the only method to correctly identify both hotel and india as moderator variables, further demonstrating its relative robustness in structure discovery.

Direct LiNGAM, and DECI, successfully inferred some direct causal effects but struggled to detect indirect relationships, such as the moderating role of hotel and india. This suggests that while Direct LiNGAM and DECI are effective in special cases such as non-linear relationships or data with non-Gaussian noise may not be well-suited for cases where variables exert mediated influences.

Conclusions

The findings from this study highlight several key considerations for researchers using causal discovery algorithms. First, the incorrect classification of juliet as a direct cause of kilo across all three methods underscores the risk of failing to account for unmeasured confounders. This suggests that no causal discovery algorithm should be applied without careful domain knowledge and, where possible, validation through other methods. Whilst more sophisticated algorithms like Direct LiNGAM and DECI are often assumed to outperform other methods, their failure to detect hidden confounding demonstrates that they are not infallible.

Although the PC algorithm outperformed the other methods in identifying causal relationships, its tendency to generate overly complex models suggests that further refinement is needed to distinguish between direct and indirect effects more effectively. This highlights the need for complementary techniques, such as domain expertise, sensitivity analyses, and robustness checks, to ensure the validity of causal conclusions.

Overall, this study highlights the importance of caution in applying causal discovery methods. While these algorithms can provide valuable insights, their limitations must be carefully considered, and researchers should interpret their outputs critically to avoid drawing misleading causal inferences.

References

- Bishop, J.M. (2021) “Artificial Intelligence is Stupid and Causal Reasoning Will Not Fix It”, *Frontiers in Psychology*, Volume 11, 513474
- Felin, T. and Holweg, M. (2024) “Theory is all you need: AI, Human Cognition, and Causal Reasoning”, *Strategy Science*, 9(4), 346-371.
- Fitkov-Norris, E. and Kocheva, N. (2025). Leveraging AI for strategic foresight: Unveiling future horizons. In M. M. Crawford & G. Wright (Eds.), *Improving and enhancing scenario planning: Futures thinking* (pp. 285–304). Edward Elgar.
- Geffner, T., Antoran, J., Foster, A., Gong, W., Ma, C., Kiciman, E., Sharma, A., Lamb, A., Kukla, M., Pawlowski, N., Allamanis, M. & Zhang, C. (2022) “Deep End-to-end Causal Inference”, *arXiv*. <https://doi.org/10.48550/arxiv.2202.02195>

Glymour, C., Zhang, K. & Spirtes, P. (2019) “Review of Causal Discovery Methods Based on Graphical Models”, *Frontiers in Genetics*, 10, 524. <https://doi.org/10.3389/fgene.2019.00524>

Malinsky, D. and Danks, D. (2018). Causal discovery algorithms: A practical guide. *Philosophy Compass*, 13(1). <https://doi.org/10.1111/phc3.12470>

Pearl, J. and Mackenzie, D. (2018) *The Book of Why: the new science of cause and effect*, London, Penguin.

Prosperi M, Salemi M, Ghosh S, Lyu T, Bian J, Chen Z, Zhao J. (2021) “Causal AI with Real World Data: Do Statins Protect from Alzheimer's Disease Onset?” *Proceedings of the 5th International Conference on Medical and Health Informatics*, 296-303.

Sgaier, S.K., Huang, V. and Charles, G. (2020) “The Case for Causal AI”, *Stanford Social Innovation Review*, Summer, 50-55.

Shimizu, S., Hoyer, P. O., Hyvarinen, A. & Kerminen, A. (2006) “Linear Non-Gaussian Acyclic Model for Causal Discovery”, *Journal of Machine Learning Research*, 7, 2003–2030.

Stokel-Walker, C. and Van Noorden, R. (2023) “What ChatGPT and generative AI mean for Science”, *Nature*, 614, 214-216.

Valizade, D., Schulz, F., Nicoara, C. (2024) “Towards a Paradigm Shift: How Can Machine Learning Extend the Boundaries of Quantitative Management Scholarship?”, *British Journal of Management*, 35(1), 99-114.