ARL-TR-9599 ● Oct 2022

Causality and Machine Learning Review

by Atul Rawal, Adrienne Raglin, Danda B Rawat, and Brian M Sadler

NOTICES

Disclaimers

ARL-TR-9599 ● Oct 2022

Causality and Machine Learning Review

Adrienne Raglin and Brian M Sadler
DEVCOM Army Research Laboratory

Atul Rawal and Danda B Rawat
Howard University

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED (From - To)**
October 2022	Technical Report	January–July 2022

4. TITLE AND SUBTITLE	
Causality and Machine Learning Review	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Atul Rawal, Adrienne Raglin, Danda B Rawat, and Brian M Sadler	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
DEVCOM Army Research Laboratory ATTN: FCDD-RLC-IT Adelphi, MD 20783-1138	ARL-TR-9599
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Causal inference has likely been a part of science as long as science itself, with the idea of cause and effect having defined the fundamental sciences from Newton's laws to the devastating COVID-19 pandemic. The cause explains the "why," whereas the effect describes the "what." The domain itself encompasses a plethora of disciplines from statistics and computer science to economics and philosophy. Recent advancements in machine learning and artificial intelligence systems have nourished a renewed interest in identifying and estimating the cause-and-effect relationship from the substantial amount of available observational data. This has resulted in various new studies aimed at providing novel methods for identifying and estimating causal inference. We include a detailed taxonomy of causal inference frameworks, methods, and evaluation. An overview of causality for security is also provided. Open challenges are delineated, and measures for evaluating robustness of causal inference methods are described. This report aims to provide a comprehensive survey on such studies of causality. We provide an in-depth review of causality frameworks and describe the different methods.

15. SUBJECT TERMS

Military Information Sciences, causal reasoning, machine learning, artificial intelligence, causality

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Adrienne Raglin
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	40	**19b. TELEPHONE NUMBER (Include area code)**
Unclassified	Unclassified	Unclassified			(301) 394-0210

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

The concept of causality, also referred to as cause and effect, has defined fundamental science since the birth of science itself. Cause and effect remain at the core of any scientific discovery, where the cause explains the “why” and the effect describes the “what.” Causality is often misused interchangeably with correlation even though correlation does not imply causation. Even though correlation is critical for science, misrepresenting correlation as causality can have adverse effects. For example, the correlation between Covid-19 and some medications caused unproven theories to spread about possible treatments for disease. Correlation refers to the relationship between two variables with a specific trend, whereas causality is the cause-and-effect relationship where the cause is responsible for the effect and the effect is somewhat reliant on the cause.[1,2] Thus, causal learning is the process of generating causal connections from the data.[3–5] Causality plays a vital and omnipresent role in our daily lives as well. Every decision we make has a cause-and-effect variable that dictates how we live our lives. Therefore, it is crucial to assume that causal learning is a critical component of any artificial intelligence (AI) or machine learning (ML) system, regardless of its use in both commercial and military applications.[2] Causal learning has seen an increase in research activity within the past two decades with yearly publications reflecting the rapid rise in causal research (Fig. 1).

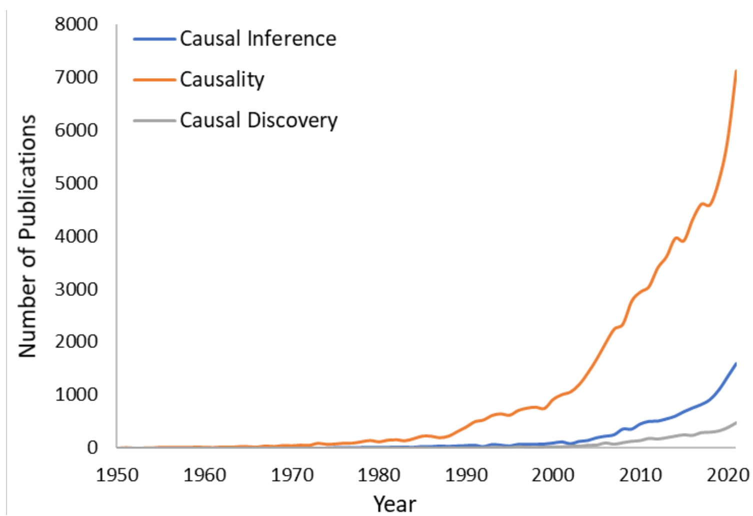

Fig. 1 Yearly publications for causal inference and causality (data derived from Scopus)

Recent advances in AI/ML systems in the past decade has put artificial reasoning systems in the forefront of many industries. With AI/ML systems expected to act autonomously and display human-like intelligence, there are still some fundamental challenges remaining, such as robustness, transferability, explainability, and causality. While AI/ML systems have made tremendous achievements in prediction accuracy and precision, they are still inherently black box models and thus lack the explanations for how the system came to the prediction that it did. This has caused unwarranted issues with the use of such systems, where biased predictions were made affecting human lives. This gave rise to Explainable AI (XAI), which has been seen as the solution for the black box problem where AI/ML systems are able to explain their decision-making process to the end users. One of the goals for the development of XAI systems is to mitigate bias from not only the model itself, but also from the incoming data used to make the predictions.[6] Algorithmic/model bias can be recognized and mitigated using various techniques, but inherent bias within the data itself is harder to mitigate. Therefore, causality is critical for identifying and mitigating bias from the data for AI systems. According to Dr Judea Pearl, causality can allow AI/ML systems to "choreograph a parsimonious and modular representation of their environment, interrogate that representation, distort it through acts of imagination, and finally answer 'What if?' type questions."[7] For further reading on XAI, we suggest readers review detailed surveys such as Rawal et al.,[6] Gunning et al.,[8] Xu et al.,[9] and Arrieta et al.[10]

Even though there are related surveys and fundamental studies on causality, such as the ones from Judea Pearl,[3,7,12] Morgan et al.,[4] Yao et al.,[5] and Gianicolo et al.[11] that provide great overviews, an up-to-date survey that provides a more comprehensive look at not just causality, but its goals and evaluation metrics as relating to AI/ML, is also needed. This survey report aims to fill the gaps in literature by providing a comprehensive survey that looks at all aspects of causality from development to evaluation and highlights some of the more recent breakthroughs and advances made toward causal AI/ML systems. The main contributions of this survey include the following:

- We present a detailed overview of causality by focusing on all aspects of the field from design and development to evaluation.
- We summarize a comprehensive taxonomy for design/development and evaluation of causality (page 17).
- We provide a comparison of the causal learning methods.
- We provide insights into the use of causality for cybersecurity and highlight some recent advances toward causal security.
- We present an open discussion of remaining challenges in the field and perspectives on recommendations for addressing them.

This report is organized as follows. Section 2 presents a taxonomy and insight into the levels of causal inference. Section 3 provides a brief survey of the design and development methods for utilization of causal inference and causal discovery. Section 4 describes techniques that are used for measuring the effectiveness of causal AI/ML systems. Section 5 provides a brief overview of causality for cybersecurity. Section 6 discusses open challenges and current trends in causality research. Concluding remarks follow in Section 7.

2. Overview of Causality

As mentioned before, causality in layman's term is the relationship between a cause and an effect. However, causation must be differentiated from statistical association. From an ML perspective, causal learning translates to the estimation of the changes in a variable's prediction/decision made by the model if a different variable has been modified or manipulated. Here, the variable that is modified/manipulated is referred to as the treatment, whereas the variable whose change is being investigated is referred to as the outcome. Covariates are the background features or variables, whereas the confounders are variables that causally affect both the treatment and the outcome.

The causal relations derived from data can be classified into the three general categories of Association, Intervention, and Counterfactuals (Fig. 2). These categories form the basis of the three-level causal hierarchy presented by Judea Pearl where causal questions at each level can only be answered when the available information corresponds to that specific level or higher.[7]

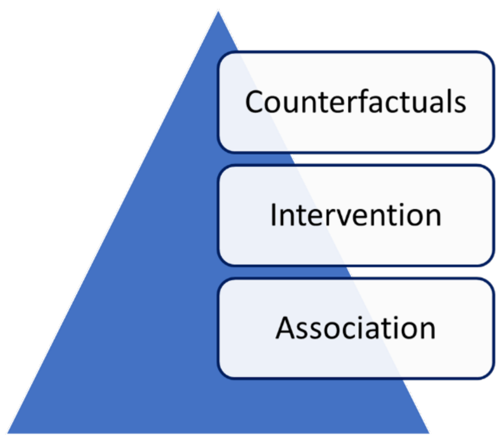

Fig. 2 Causal hierarchy as presented by Judea Pearl[7]

Association refers to the statistical relations within the raw data. Associative questions can be deduced directly from the raw data using traditional statistical techniques. This level of relations or questions are the foundation of ML, where decisions/predictions are made from statistical relations from the raw data without the need for causal information.

Intervention involves estimating the effect of an action. Questions at this level involve deductions from the raw data as well as modifying/manipulating the variables (treatment). They cannot be answered solely from the statistical relations within the raw data. The causal structure of the variables within the system must be comprehended.

Counterfactuals are the highest level of causal information, as they incorporate associative and interventive questions within them. They can readily answer questions at both Association and Intervention levels and are related to causal reasoning/inference where estimates from unobserved outcomes can be predicted.

Causality with ML can help in answering a plethora of questions for various applications (Table 1). These questions can mainly be grouped into two types:

1) How much does changing one variable (treatment) affect the target variable (outcome)?

2) Which variable needs to be modified/manipulated (treatment) to see a change in the target variable (outcome)?

These two questions govern the study of causality with ML and can be referred to as causal inference and causal discovery, respectively.[13,14] With causal inference, causal effects can be investigated by studying the extent to which a cause can be manipulated to influence an effect. With causal discovery, causal relations can be established from data between the variables.

Causality can be investigated via either causal inference or causal discovery through two formal frameworks, structural causal models (SCMs) and potential outcome framework. SCMs consist of a causal graph and structural equations. They provide a comprehensive theory for causality.[3,5,15,16] Causal graphs are directed graphs that describe the causal effects between the variables, where each node represents a random variable, including the outcome treatment and other observed/unobserved variables. A directed line $x \rightarrow y$ indicates a causal effect of x on y (Fig. 3).[15] These graphs inherit the conditional independence criteria as they form a class of Bayesian networks with causal effects represented by the edges.

Table 1 Causal hierarchy as presented by Judea Pearl[7]

Level of causality	Activity	Questions	Examples
Association	Seeing	What is?	What do the polls say about the midterm election?
Intervention	Doing Intervening	What if?	What happens when a candidate is endorsed by the president?
Counterfactuals	Reasoning Imagining	Why? What caused the effect? What happens when a variable is modified?	Was it the presidential endorsement that caused the rise in the polls? What if the candidate changed their stance on a major issue?

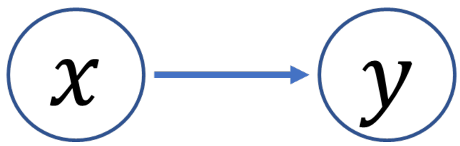

Fig. 3 Causal graph indicating the causal effect of *x* on *y*

Structural equations represent the effect of the treatment that directly causes an outcome. Given a structural equation along with the causal graph, the causal effects can be defined by the directed edges within the graph. Further details on structural equations can be found in the surveys mentioned previously.

The potential outcome framework proposed by Rubin[17] states that causality is tied to the treatment (modification/manipulation) applied to a unit.[18] A unit is defined by Yao et al.[5] as the atomic research object in the treatment effect study. The comparison of the units' potential outcomes of treatments yields the treatment effect. It is widely used for learning causal effects with reference to a treatment–outcome pair. Given a treatment and outcome, the potential outcome of an instance is the outcome that would have been observed if the instance had received treatment. This allows for the simpler expression of the causal challenge as presented by Hoyer et al., where only a single outcome is possible for a given instance.[19,20] The individual treatment effect (ITE) can be defined via the potential outcomes as the difference between the outcomes of a given instance with two different treatments. A binary treatment is often assumed for ITEs. Based on the ITE, the average treatment effect (ATE) can be estimated when extended on arbitrary populations. For subpopulations, the conditional average treatment effect (CATE) can be defined. For formal mathematical and detailed definitions of the terms mentioned in this section, readers are encouraged to read detailed surveys in Pearl,[3,16] Yao et al.,[5] and Guo et al.[15]

Between the two frameworks, an inference made in one framework can be readily translated into the other because as they are logically equivalent.[3,5,15] However, there are distinctions between the two frameworks. In structural causal models, causal effect of any variable can be investigated, making them the preferred framework for learning causal relations between variables.[21] In the potential outcome framework, the causal effects are only known for the treatment and specific variables, which makes the framework capable of modeling causal effects without knowing the complete causal graph.[21] Potential outcome framework also eases the work of estimating treatment effects by developing estimators.[15]

3. Learning Casual Effects and Relations

Causality can be extracted from raw data by either causal inference or causal discovery. It has been used in a plethora of applications ranging from medicine to agricultural sciences (Fig. 4). This section provides an overview of the different methods and techniques used for causal inference and causal discovery. For detailed information on different methods based on the underlying techniques such as stratification, matching, re-weighting, readers are encouraged to read the survey by Yao et al.[5]

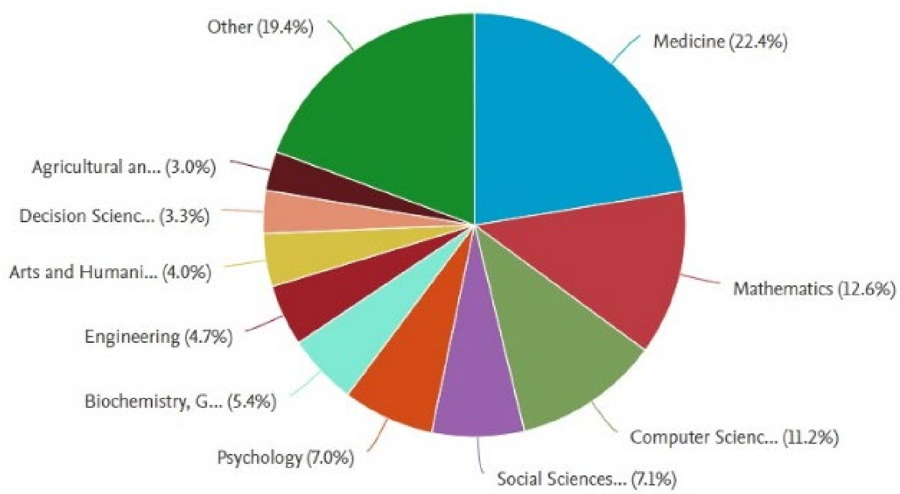

Fig. 4 Applications for causality (data derived from Scopus)

3.1 Methods for Causal Inference

Causal inference can be done with a broad range of methods. Causal inference deals with investigating causal effects of variables within the raw data. Learning causal effects for a given instance refers to quantifying how the outcome is predicted to change when the treatment is modified. This is helpful for various applications as shown in Fig. 3, where the causal effects for different populations are of critical importance. This can be done for both the SCM and potential outcome frameworks.

Propensity Score is defined by Morgan and Winship as the conditional probability of a treatment given the background variables.[1] Given a set of specific observed covariates, the propensity score refers to the probability of a unit being assigned to one specific treatment.[5] These methods extract the ATE by splitting instances into strata and treating each one as a randomized control trial. Methods based on propensity scores are propensity score matching (PSM), propensity score-based stratification, inverse probability of treatment weighting, and propensity-based adjustment.[1,4,22–25] PSM matches treated instances to controlled instances with similar propensity scores. Austin[24] provides an in-depth analysis of the different propensity-based methods mentioned.

Covariate balancing methods learn the sample weights via regression and sample re-weighting.[26] Some of the methods available in literature include entropy balancing (EB),[27] approximate residual balancing (ARB),[28] covariate balancing propensity score (CBPS),[29] and covariate balancing generalized propensity score.[30] EB learns sample weights of instances under control to match the two groups. It prevents data loss by keeping the weights close to the base weight and allows for larger constraints. ARB extracts average treatment effect from data by combining the weights balance and regularized regression adjustment. This is done by learning the sample weights, then fitting the regularized regression adjustment model, and finally estimating the average treatment effect. Compared with EB, ARB can handle the sparseness of high-dimensional data.[15,31] CBPS combines covariate balancing and propensity scores to derive the balancing score from the propensity score to increase the robustness of the model toward misspecification of the propensity score.[5]

Regression adjustment method are based on supervised ML where a function is fitted to predict the probability distribution with features and labels. Here, instead of the probability distribution, we are interested in interventional distribution and counterfactuals. The counterfactual outcome is based on the features x and treatments t, where adjustment can be done two different ways. First is by fitting a single function to estimate the probability distribution with the features and treatment, which can deduce the ITE. The second way is to fit the model for each individual potential outcome and then estimate the ATE. Different regression adjustment methods have been proposed in the literature, such as the doubly robust estimation,[32] targeted maximum likelihood estimator (TMLE),[33] and ARB.[24,33–35]

While the mentioned methods are capable and robust for learning causal effect from raw data, they are not useful for data without observed confounders. In such cases the assumption of unconfoundedness remains unsatisfied, and causal effects cannot be extracted from the data. Instrumental Variable[15] methods allow us to extract causal effects from data with unobserved confounders. The IV influences the causal outcome by directly affecting the treatment.[15] The front-door criterion[36] method is another method for learning causal effects from data with unobserved confounders. Regression discontinuity design (RDD) is also capable of learning causal effects from data with unobserved confounders. RDD has been further expanded into other methods such as the Sharp RDD[37,38] and Fuzzy RDD.[37,39]

3.2 Methods for Causal Discovery

Causal discovery answers the question of finding causal relations within the raw data. Similar to causal inference, there are a handful of methods to extract causal relations from data. Learning causal relation for a given instance refers to determining whether modifying a variable x causes modification for the target variable x'. This can be done by hypothesizing that the causal relation between the variables in the data can be detected by statistical dependencies.[40,41] Here, the causal relations can be learned via three general algorithms: constraint-based (CB), score-based (SB), and functional causal model (FCM)-based algorithms.[42] Constraint- and score-based algorithms are derived from statistical relations to determine causal graphs, while functional causal models estimate structural equation coefficients to learn causal relations.[15]

CB algorithms derive causal relations from causal graphs that satisfy the conditional independence based on the faithfulness assumption via statistical testing.[40] Examples of CB algorithms include the Peter–Clarke algorithm,[40] inferred causation (IC) algorithm and its variants,[16,40,43–48] and fast causal inference (FCI) and its variants.[49–54] The Peter–Clarke algorithm derives an undirected causal graph from the raw data and then predicts the directions for the edges for a complete causal graph.[40] The FCI algorithms were proposed for searches through extended causal graphs.

SB algorithms were proposed to overcome the faithfulness assumption by replacing the conditional independence tests with goodness-of-fit tests. By maximizing the scoring criteria to derive the causal graph's score from the raw data, these algorithms are able to learn causal graphs. The score function along with the structural equations need to be explicitly presented for the goodness-of-fit tests. The structural equations are crucial to describe how a variable is affected causally by its parent variables and noise. Given the parameterized structural equation, the score function maps the candidate causal graphs.[15] Examples of SB algorithms include the Bayesian information criterion (BIC) score,[55] factorized normalized maximum likelihood (NML) universal model,[56] Bayesian Dirichlet score,[57] Greedy Equivalence Search (GES),[58] Fast GES,[59] and the adaptation of the Greedy SP algorithm by Wang et al.[60] Hybrid algorithms that combine CB and SB methods are also present in literature, for example the Max-Min Hill-Climbing (MMHC) algorithm.[61–63]

FCM-based algorithms are capable of differentiating between directed acyclic graphs (DAGs) from the same class. Here, a function is written as the function of its direct causes and a noise term.[15] Examples of FCM based algorithms include the Linear Non-Gaussian Acyclic Model (LiNGAM), Independent Component Analysis LiNGAM (ICA-LiNGAM),[64] DirectLiNGAM,[65] and auto-regressive LiNGAM.[66] Additionally, relaxing the linear restriction on the variable-noise distribution relation via additive noise models have also been proposed to reduce the search space of causal graphs.[14,19,67] Glymour et al. provide an in-depth survey of various causal discovery methods based on graphical models.[68]

In addition to the three categorized methods, other methods have also been proposed in literature for learning causal relations. Yao et al. provides detailed reviews of methods such as re-weighting, stratification, matching, tree-based, representation learning, multi-task learning, and meta-learning methods.[5] Stratification methods adjust confounders based on subclassification or blocking.[18] Matching based methods are capable of reducing the estimation bias due to confounders and estimating the counterfactual. Tree-based models are a predictive modeling method based on decisions trees. These include the algorithms such as Classification and Regression Tree (CART),[69] Bayesian Additive Regression Tree (BART),[70,71] and Random Forest (RF) methods.[72]

Due to the available methods for causal inference and causal discovery, AI/ML tools have been investigated to aid in the search for causality. ML models such as RF, Deep Learning (DL), Reinforcement Learning (RL), and Neural Networks (NNs) have all been used to learn causality. NN models such as Global Vectors (Glove)[73] and Recurrent Neural Network (RNN)[74] have been proposed to study the causal effects of group formation on loans for a financing application. Pham et al. applied NNs to estimate causal distributions.[75] DL models for causality include the Causal Effect Variational Autoencoder (CEVAE),[76] TARnet,[77] and Balancing Counterfactual Regression.[78] Tables 2 and 3 provide an overview of the available ML toolboxes and algorithms for causal discovery and inference.[*]

[*] For an updated list visit the Git repository for Dr Huan Liu's research group from Arizona State University: https://github.com/rguo12/awesome-causality-algorithms.

Table 2 Available codes for causal learning

Causal learning	Name	Language	Paper and link
ITE	PSM	Python	https://github.com/akelleh/causality/tree/master/causality/estimation[22]
	Counterfactual Regression	Python	https://github.com/oddrose/cfrnet[77]
	CEVAE	Python	https://github.com/AMLab-Amsterdam/CEVAE[76]
	Causal Forest	R Python	https://github.com/kjung/scikit-learn https://github.com/grf-labs/grf[81]
ATE	EB	R	https://github.com/cran/ebal[27]
	TMLE (Regression adjustment)	R	https://cran.r-project.org/web/packages/tmle/index.html[82]
	Inverse probability re-weighting	R	https://github.com/cran/ipw[22]
	DRE (Regression adjustment)	R	https://github.com/gregridgeway/fastDR[83]
Causal effect	Dose response networks (DRNets)	Python	https://github.com/d909b/drnet[84]
	Causal Impact	R Python	https://github.com/synth-inference/synthdid https://github.com/MasaAsami/pysynthdid[85]
	Linked Causal Variational Autoencoder (LCVA)	Python	https://github.com/rguo12/CIKM18-LCVA[86]
	Deconfounded RL	Python	https://github.com/CausalRL/DRL[87]
Causal relations	Causal PSL	Java	https://bitbucket.org/linqs/causpsl/src/master/[88]
	Temporal Causal Discovery Framework (TCDF)	Python	https://github.com/M-Nauta/TCDF[89]
	PC algorithm	Python R Julia	https://github.com/keiichishima/pcalg https://github.com/cran/pcalg https://github.com/mschauer/CausalInference.jl[90]
	Scalable and hybrid ensemble-Based causality discovery	Python	https://github.com/big-data-lab-umbc/ensemble_causality_learning[91]

Table 3 Available toolboxes for causal learning

Toolbox	Language	Paper and link
Uber CausalML	Python	https://github.com/uber/causalml[92]
CausalNex	Python	https://github.com/quantumblacklabs/causalnex
Causal DiscoveryToolbox	Python	https://github.com/FenTechSolutions/CausalDiscoveryToolbox[93]
CausalToolbox	R	https://github.com/forestry-labs/causalToolbox[94]
CausalVAE	Python	https://github.com/huawei-noah/trustworthyAI[95]
Econ ML	Python	https://econml.azurewebsites.net/spec/spec.html[96]
DoWhy	Python	https://github.com/microsoft/dowhy[97]
gCastle	Python	https://github.com/huawei-noah/trustworthyAI[98]
TETRAD Toolbox	R	https://github.com/bd2kccd/r-causal[99]
JustCause	Python	https://github.com/inovex/justcause

These methods are useful for extracting causal information derived from experimental data. However, for deriving causality from observational data, the following assumptions must be made for consistent and accurate causal estimates.[5,15,79,80]

Assumption 1: Stable Unit Treatment Value Assumption (SUTVA) – The potential outcomes for any units do not vary with the treatment assigned to others. For each unit, there is only one version of treatment level that leads to one potential outcome. Here, the independence of each unit and well-defined treatment levels are emphasized.

Assumption 2: Ignorability – For any given background variable, the treatment is independent of the potential outcomes.

Assumption 3: Positivity – The probability of receiving every value of treatment conditional on some measured covariates X is greater than zero.[80]

Assumption 4: Consistency – For any assigned treatment, the potential outcome is independent of the treatment.

Assumption 5: Causal Sufficiency – A set of variables is causally sufficient for a process, if and only if it includes all common causes of every two pairs in the set.[80]

Assumption 6: Faithfulness – The independence explained from any data generating causal graphs may not be violated by the statistical relations between variables within the data.

4. Evaluation Metrics

Causality in many cases is estimated/predicted based on causal experiments. For example, in the medical field, Randomized Control Trials are performed for this reason. However, experimental trials are not always feasible, and investigators must rely on observational studies to conduct causal studies. This can cause issues to arise such as the lack of randomization within the observational data. Thus, the issue of effect of confounders must be accounted for in the causal studies. Traditional methods for dealing with the confounders in causal studies include stratification, matching-based methods, and PSM.[100–102] While these methods are able to mitigate the issues arising from the presence of observed and unobserved confounders, they still have to be evaluated for causal accuracy. This section provides an overview of some methods used to evaluate the robustness of causal methods from either experimental or observational studies.

Some of the first evaluation metrics for causality come from Sir Bradford Hill, who proposed the aspects of causality for studies where experimental data was not readily available and investigators had to rely on observational data. He proposed the following the criteria for causality[103–105]:

- *Strength*: Strong association between the variables, decreasing the probability for the randomized correlation.
- *Consistency*: Association observed in multiple studies under varying conditions.
- *Specificity*: Causation between specific variables.
- *Temporality*: Effect is temporally subsequent to cause.
- *Plausibility*: Plausible link between the cause and effect.
- *Coherence*: Non-conflicting causal interpretation.
- *Analogy*: Similar causes with similar effects.
- *Experimental evidence*: Experimental evidence for observational data will be the best evidence for any causal reasoning (if possible).
- *Biological gradient*: For medical studies, a dose-response effect is displayed.

Evaluating causality is often done via a comparison of the estimated/predicted causal graphs to the ground truth via the equivalence class condition. The condition of equivalence class states that "two causal graphs can belong to the same class if and only if each conditional independence that the one graph is also implied by the other."[15] de Jongh et al. present an excellent review of evaluation metrics for comparing the distances between the predicted causal graphs and the ground truth.[106] They proposed the following metrics and measures:

- *Missing edges*: Edges present in the ground truth but not in the learned graph. (Lower is better.)
- *Extra edges*: Edges present in the learned graph but not in the ground truth. (Lower is better.)
- *Correct edges*: Edges present in both learned graph and the ground truth. (Higher is better.)
- *Correct edge direction*: Correctly oriented edges in the learned graph. (Higher is better.)
- *Incorrect edge direction*: Incorrectly oriented edges in the learned graph. (Lower is better.)
- *Topology*: Normalized and weighted combination of the missing, extra, and correct edges. (Higher is better.)
- *Hamming Distance*: Number of edits needed for the learned graph to imitate the ground truth. (Higher is better.)

The structural hamming distance (SHD) has been most commonly used to compare the causal graphs with the ground truths.[62,65,107] It is defined as the number of edits needed for the predicted causal graph to become the same as the ground truth. Beyond the relation between the predicted and ground truth causal graphs, other metrics used for measuring the accuracy of causality include the following[80]:

- *True/False Positive Rates*: True positive rate (TPR) is the number of common edges found in the learned causal graph plus the edges in the ground truth divided by the edges in the ground truth. Similarly, the false positive rate (FPR) is the number of common edges found in the learned causal graph plus the edges in the ground truth divided by the absolute difference between the number of edges in the ground truth and the learned causal graph.[108–111]
- *Receiver Operator Characteristic (ROC) curve*: The ROC curve is a common metric for the accuracy of statistical and machine learning models. It is simply the ratio of the TPR and the FPR.[112–116]
- *Precision, Recall, and Precision Recall Curve*: Precision is the proportion of correct positive identifiers. It is ratio of total correct positive predictions to the total positive predictions. Recall is the proportion of the actual

positives predicted accurately. It is the ratio of correct positive predictions to all the predictions. The precision recall curve is the ratio of the precision and recall.[116,117]

- *F1-Score and F-test*: F1 Score is weighted average of Precision and Recall. F-test is a statistical test. To perform the F-test, a null and alternative hypothesis is used to derive the F-value and F-Statistic.[118]
- *Mean Squared Error (MSE)*: MSE is the average of the square of the difference between the correct and predicted values.[119,120]

Figure 5 provides a taxonomy for both methods and evaluation metrices for causal learning as provided in the preceding sections.

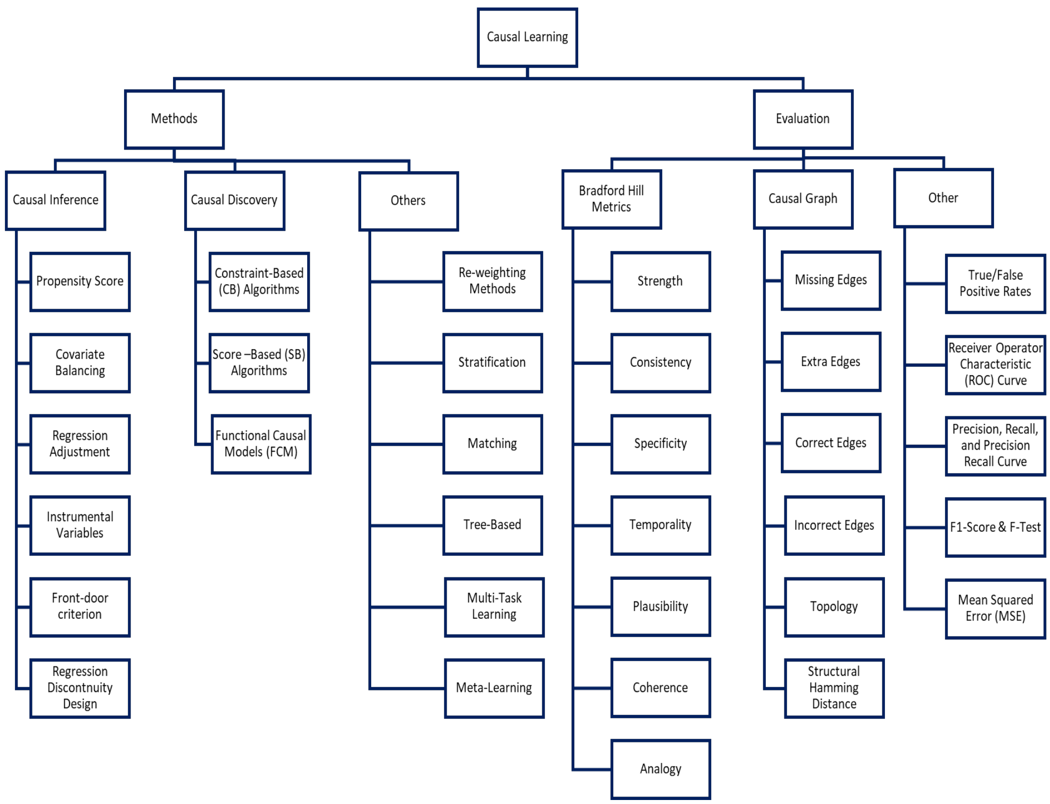

Fig. 5 Taxonomy for causal learning methods and evaluation

5. Causal Learning and Cybersecurity

In the field of cybersecurity, the comprehension of a cyberattack, including how it unfolds in real-time, is critical for developing techniques to defend against them. Here, causal learning can play a crucial role in understanding and mitigating these cyber threats. However, the application of causality with cybersecurity is still a very novel concept, as there is a scarcity of studies and reports in literature, as shown in

Fig. 6. Even though the field is still in its infancy, causal cybersecurity is poised to be vital in securing the next generation of AI/ML systems.

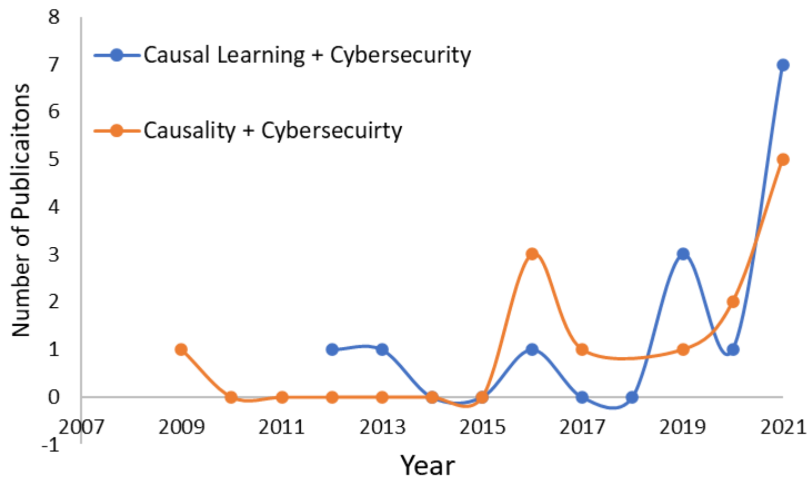

Fig. 6 Studies for causal learning and causality with cybersecurity (data derived from Scopus)

The application of causal learning for threat detection and mitigation via causal effect of actions from observational data is being pursued by researchers. For example, at the world-renowned Alan Turing Institute, causal inference is being used for improved cybersecurity threat detection.[121] Research applications such as cyberattack indicator identification, the sequence for an optimal attack and counter-defense, addition of a causal dimension to network threat detection, and the combination of causal queries with cyber data sets are being studied at the institute. Dhir et al. provided some benefits of using causality with cybersecurity[121]:

- Causal approach allows for the comprehension of the cyberattack tactic, and determination of their consistency.
- Potential for a defensive strategy for cyberattacks if the causal structure of attacks can be learned from data.

Dhir et al. described the use of causality for active cyber defense by applying the causal inference on the MITRE ATTACK framework to achieve threat detection and mitigation.[121]

Qin et al. provided another use-case example for using causal inference to analyze and predict attack scenarios using the US Defense Advanced Research Projects Agency (DARPA) Grand Challenge Problem data set.[122] The authors built an attack-tree-based causal network with characteristics such as the final goal of the attack, subgoals, and evidence. They identified and predicted upcoming attacks via the causal network based on the attack trees (Fig. 7).

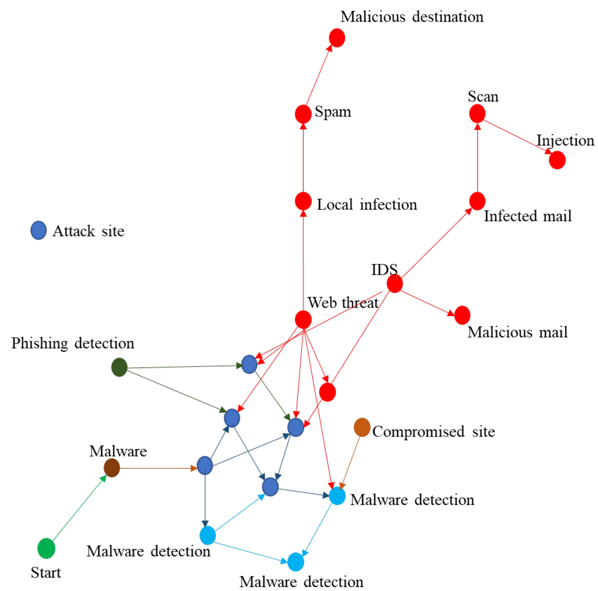

Fig. 7 Notional example of a causal graph for cyberattack

Causal discovery algorithms have also been used in cybersecurity. Mueller et al. used causal discovery for cybersecurity research and provided assessment of cyberattack databases with the Cyber Kill Chain framework.[123] Cause–effect relationships for attack types were discussed for the characterization of attack types from real cyberattack data sets. Evidence for several phases of the Cyber Kill Chain framework was presented with the use of causal discovery algorithms such as the Fast Greedy Equivalent Search (FGES).

Tople et al. presented the use of causal learning for privacy guarantee using causal structures.[124] The study showed better performance for causal structure models for data generalization than trained distribution models. This was used to theorize the relation between causality and privacy: Causal models are more robust to membership inference attacks and offer much better privacy guarantees. The study provided an outlook for the use of causal learning for training models to achieve better robustness against privacy attacks.

6. Open Challenges and Perspectives

Even though causal inference and causal discovery have made remarkable strides, numerous challenges still remain. These include the lack of experimental data sets for studies, lack of ground truth for current relevant data sets, lack of universally adopted definition, standards, and measures for the causality of AI/ML systems, the need for better evaluation metrics for causality-based ML models, the balance between causality and performance, and the challenges of making DL models causal.

The lack of experimental data for today's relevant applications remains a major challenge. While learning causality from observational data has achieved major advances within the past decade, there is no substitute for data from experimental studies for causality. Since experimental studies are not always feasible and practical (both cost and timewise), observational studies are performed for the ease of data compilation. To solve this problem, a repository with experimental data sets for applications relevant to today's challenges, such as the Internet of Things and the Internet of Battlefield Things, is needed for the practical and robust integration of causality with AI/ML systems. Synthetic data emulating real-world experimental conditions can be generated and used where possible. Learning causality from observational data presents its own challenges such as the handling of anomalies. Toward this end, Kaize et al. proposed a DL model that models the topological structure via an autoencoder that uses learned embeddings to reconstruct the original data.[125] Additionally, other challenges for observational data include the complexity in handling data entanglement, complex treatments, and temporal observations.[126–129]

The absence of ground truth for observational data. When observational data is available for studies, the unavailability of ground truth is a hindrance for studies involving current state of the art applications and data sets for both commercial and military applications. Thus, there needs to be an alternative for studies when it is not possible to obtain ground truth for the data. Feature labels can be used as a substitute for some observational data from modalities such as audio, text, and images. For example, images and audio labeled as "beach" can be set to the ground truth with the image of sand and water along with the audio of waves at a beach.

The lack of a universal standard. Another vital challenge within the field of causality is the ambiguity of definitions for the available terminology. As mentioned previously, correlation is often implied to be causation. Furthermore, terms like causal inference and causal discovery are used interchangeably as synonyms and have only in the past decade taken on strict distinct definitions. This can also be due to the lack of a standard unified theory of causality even though fundamental works from Pearl and Spirtes et al. have been monumental in the effort to understand causality.[3,12,16,36,40,90,130] A unifying framework for causality provides common ground for researchers to contribute to distinctive needs and challenges of the field. Additionally, the lack of evaluation metrics for causal models in the absence of ground truth is also a hindrance. Even though the works presented in the previous sections are capable of evaluating and measuring the models' accuracy, they rely heavily on the presence of ground truth.

Model selection. When learning causality from observational data, model selection plays a vital role in the accuracy of causal learning. Since there are no readily available ML models that are explicitly used for causal learning, model selection will depend on the type of application. Using causality to improve ML and using ML to improve causality are both open-ended challenges that need to be addressed by the field. Potential research directions for the field include the use of causality for XAI bias detection, and mitigation.[131,132] Moraffah et al. present an excellent survey of causal interpretable models with insights into the methods and evaluation metrics for causal interpretability.[79]

Causality for time series data. Even though causality is generally seen as a time-dependent concept, learning causality from time-series data remains more challenging than with time-independent data. Nonstationary time series, noise presence, and data scarcity are some of the challenges arising from time-series data. To this end, Moraffah et al. present an in-depth survey of existing methods for causal inference from time-series data.[80] The survey categorizes time-series causality into three types: time-invariant treatment effect, time-varying treatment effect, and dynamic regimes. It also provides examples of causal discovery for various scientific applications such as identification of stock indicators[133] and variables for climate change.[134,135]

Imperfections and bias in data. Bias within data, whether experimental or observational, is a major challenge faced by not just causal research, but the AI/ML research field as a whole. Fairness and bias detection are critical for the success of any causal AI/ML system regardless of the application. Accountable AI, including laws within both the United States and Europe, demand fair and unbiased decision making from causal models that affect human lives. For experimental studies, the quality of randomization dictates the validity of randomized experiments where false relations between variables could be detected due to sampling. Imperfections within the data such as bias, missing/incorrect values, errors with measurements, and noise can lead to wrongful and errored causal relations. Causal models, unlike statistical models, might not be robust against such imperfections and can produce causal relations within variables that have no causal relations whatsoever. Sample bias from observational data, especially big data, is also a major concern because these data are collected freely without any set of rules. There is a need for methods to detect and mitigate such imperfections within data.[136]

Hybrid observational and experimental data. Even with all the advances in computational studies, they are still looked upon as being second-class to experimental studies. Due to the lack of experimental data/studies, observational studies are conducted in substitution. However, these studies are sometimes considered to be unreliable compared with experimental studies with controlled trials. Therefore, studies that combine both observational and experimental portions should be developed to overcome the shortcomings of the two types of studies. Post hoc methods to combine experimental and observational data for causal studies would be immensely helpful for a variety of applications.[136]

Uncertainty quantification. Uncertainty of information is another challenge faced by the data industry. This problem is very relevant for causal studies done from observational data where the data is finite. Quantifying uncertainty is needed for handling prediction errors resulting from model complexity and data quantity/quality.[136]

7. Conclusion

Causal learning will play a vital role in the development and application of artificial reasoning systems. In this report we presented a taxonomy and literature survey of causal learning. Different terms associated with causality were defined, and goals and methods for the design and development causal learning models were presented. Various challenges were also described.

8. References

1. Morgan SL, Winship C. Counterfactuals and causal inference. Cambridge University Press; 2015.

2. Pearl J. Theoretical impediments to machine learning with seven sparks from the causal revolution. arXiv preprint; 2018. arXiv:1801.04016.

3. Pearl J. Causal inference in statistics: an overview. Statistics Surveys. 2009;3:96–146.

4. Morgan SL, Winship C. Counterfactuals and causal inference: methods and principles for social research. 2nd ed. (Analytical Methods for Social Research). Cambridge University Press; 2014.

5. Yao L, Chu Z, Li S, Li Y, Gao J, Zhang A. A survey on causal inference. ACM Trans Knowl Discov Data. 2021;15(5). Article 74. doi: 10.1145/3444944.

6. Rawal A, McCoy J, Rawat D, Sadler B, Amant R. Recent advances in trustworthy explainable artificial intelligence: status, challenges and perspectives. Tech Rxiv preprint; 2021. doi: 10.36227/techrxiv .17054396.v1.

7. Pearl J. The seven tools of causal inference, with reflections on machine learning. Communications of the ACM. 2019;62(3):54–60.

8. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z. XAI—explainable artificial intelligence. Science Robotics. 2019;4(37):eaay7120.

9. Xu F, Uszkoreit H, Du Y, Fan W, Zhao D, Zhu J. Explainable AI: a brief survey on history, research areas, approaches and challenges. Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing. Springer; 2019. p. 563–574.

10. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion. 2020;58:82–115.

11. Gianicolo EAL, Eichler M, Muensterer O, Strauch K, Blettner M. Methods for evaluating causality in observational studies. Dtsch Arztebl Int. 2020 Feb 14;116(7):101–107. doi: 10.3238/arztebl.2020.0101.

12. Pearl J. Statistics and causal inference: a review. Test. 2003;12(2):281–345.

13. Gelman A. Causality and statistical learning. American Journal of Sociology. 2011;117:955–966.

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

15. Guo R, Cheng L, Li J, Hahn PR, Liu H. A survey of learning causality with data. ACM Computing Surveys. 2021;53(4):1–37. doi: 10.1145/3397269.

16. Pearl J. Causality. Cambridge University Press; 2009.

17. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology. 1974;66(5):688.

18. Imbens GW, Rubin DB. Causal inference in statistics, social, and biomedical sciences. Cambridge University Press; 2015.

19. Hoyer P, Janzing D, Mooij JM, Peters J, Schölkopf B. Nonlinear causal discovery with additive noise models. Advances in Neural Information Processing Systems. 2008;21.

20. Holland PW. Statistics and causal inference. Journal of the American Statistical Association. 1986;81(396):945–960.

21. Aliprantis D. A distinction between causal effects in structural and Rubin causal models. Federal Reserve Bank of Cleveland (US); 2015. Paper No.: 15-05.

22. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.

23. Rosenbaum PR. Model-based direct adjustment. Journal of the American Statistical Association. 1987;82(398):387–394.

24. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research. 2011;46(3):399–424.

25. Hirano K, Imbens GW, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica. 2003;71(4):1161–1189.

26. Kuang K, Cui P, Li B, Jiang M, Yang S. Estimating treatment effect in the wild via differentiated confounder balancing. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2017. p. 265–274.

27. Hainmueller J. Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies. Political Analysis. 2012;20(1):25–46.

28. Athey S, Imbens GW, Wager S. Approximate residual balancing: debiased inference of average treatment effects in high dimensions. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2018;80(4):597–623.

29. Imai K, Ratkovic M. Covariate balancing propensity score. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2014;76(1):243–263.

30. Fong C, Hazlett C, Imai K. Covariate balancing propensity score for a continuous treatment: application to the efficacy of political advertisements. The Annals of Applied Statistics. 2018;12(1):156–177.

31. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological). 1996;58(1):267–288.

32. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. American Journal of Epidemiology. 2011;173(7):761–767.

33. Van Der Laan MJ, Rubin D. Targeted maximum likelihood learning. The International Journal of Biostatistics. 2006;2(10).

34. Imbens GW. Nonparametric estimation of average treatment effects under exogeneity: a review. Review of Economics and Statistics. 2004;86(1):4–29.

35. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Statistics in Medicine. 2004;23(19):2937–2960.

36. Pearl J. Causal diagrams for empirical research. Biometrika. 1995;82(4):669–688.

37. Campbell DT. Reforms as experiments. American Psychologist. 1969;24(4):409.

38. Anderson M, Magruder J. Learning from the crowd: regression discontinuity estimates of the effects of an online review database. The Economic Journal. 2012;122(563):957–989.

39. Angrist JD, Lavy V. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. The Quarterly Journal of Economics. 1999;114(2):533–575.

40. Spirtes P, Glymour CN, Scheines R, Heckerman D. Causation, prediction, and search. MIT Press; 2000.

41. Schölkopf B, Janzing D, Peters J, Sgouritsa E, Zhang K, Mooij J. On causal and anticausal learning. arXiv preprint; 2012. arXiv:1206.6471.

42. Malinsky D, Danks D. Causal discovery algorithms: a practical guide. Philosophy Compass. 2018;13(1):e12470.

43. Fukumizu K, Gretton A, Sun X, Schölkopf B. Kernel measures of conditional dependence. Advances in Neural Information Processing Systems. 2007;20.

44. Kalisch M, Bühlman P. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. Journal of Machine Learning Research. 2007;8(3).

45. Le TD, Hoang T, Li J, Liu L, Liu H, Hu S. A fast PC algorithm for high dimensional causal discovery with multi-core PCs. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2016;16(5):1483–1495.

46. Ramsey JD. A scalable conditional independence test for nonlinear, non-Gaussian data. arXiv preprint; 2014. arXiv:1401.5031.

47. Sejdinovic D, Sriperumbudur B, Gretton A, Fukumizu K. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. The Annals of Statistics. 2013;2263–2291.

48. Zhang K, Peters J, Janzing D, Schölkopf B. Kernel-based conditional independence test and application in causal discovery. arXiv preprint; 2012. arXiv:1202.3775.

49. Spirtes PL, Meek C, Richardson TS. Causal inference in the presence of latent variables and selection bias. arXiv preprint; 2013. arXiv:1302.4983.

50. Peters J, Janzing D, Schölkopf, B. Causal inference on time series using restricted structural equation models. Advances in Neural Information Processing Systems. 2013;26.

51. Kocaoglu M, Dimakis A, Vishwanath S. Cost-optimal learning of causal graphs. Proceedings of the International Conference on Machine Learning. Proceedings of Machine Learning Research (PMLR); 2017. p. 1875–1884.

52. Entner D, Hoyer PO. On causal discovery from time series data using FCI. Probabilistic Graphical Models. 2010;121–128.

53. Colombo D, Maathuis MH, Kalisch M, Richardson TS. Learning high-dimensional directed acyclic graphs with latent and selection variables. The Annals of Statistics. 2012;40(1):294–321.

54. Chu T, Glymour C, Ridgeway G. Search for additive nonlinear time series causal models. Journal of Machine Learning Research. 2008;9(5).

55. Schwarz G. Estimating the dimension of a model. The Annals of Statistics. 1978;6(2):461–464.

56. Roos T, Silander T, Kontkanen P, Myllymaki P. Bayesian network structure learning using factorized NML universal models. Proceedings of the 2008 Information Theory and Applications Workshop. IEEE; 2008. p. 272–276.

57. Heckerman D, Geiger D, Chickering DM. Learning Bayesian networks: the combination of knowledge and statistical data. Machine Learning. 1995;20(3):197–243.

58. Chickering DM. Optimal structure identification with greedy search. Journal of Machine Learning Research. 2002;3(Nov):507–554.

59. Ramsey J, Glymour M, Sanchez-Romero R, Glymour C. A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. International Journal of Data Science and Analytics. 2017;3(2):121–129.

60. Wang Y, Solus L, Yang K, Uhler C. Permutation-based causal inference algorithms with interventions. Advances in Neural Information Processing Systems. 2017;30.

61. Wong ML, Lee SY, Leung KS. A hybrid approach to discover Bayesian networks from databases using evolutionary programming. Proceedings of the 2002 IEEE International Conference on Data Mining. IEEE; 2002. p. 498–505.

62. Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. Machine Learning. 2006;65(1):31–78.

63. Tsamardinos I, Aliferis CF, Statnikov AR, Statnikov E. Algorithms for large scale Markov blanket discovery. Proceedings of the Florida Artificial Intelligence Research Society. 2003;2:376–380.

64. Shimizu S, Hoyer PO, Hyvärinen A, Kerminen A, Jordan M. A linear non-Gaussian acyclic model for causal discovery. Journal of Machine Learning Research. 2006;7(10).

65. Shimizu S, Inazumi T, Sogawa Y, Hyvärinen A, Kawahara Y, Washio T, Hoyer PO, Bollen K. DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. Journal of Machine Learning Research. 2011;12:1225–1248.

66. Hyvärinen A, Zhang K, Shimizu S, Hoyer PO. Estimation of a structural vector autoregression model using non-Gaussianity. Journal of Machine Learning Research. 2010;11(5).

67. Hoyer PO, Hyvärinen A, Scheines R, Spirtes PL, Ramsey J, Lacerda G, Shimizu S. Causal discovery of linear acyclic models with arbitrary distributions. arXiv preprint; 2012. arXiv:1206.3260.

68. Glymour C, Zhang K, Spirtes P. Review of causal discovery methods based on graphical models. Frontiers in Genetics Review; 2019 June 04. doi: 10.3389/fgene.2019.00524.

69. Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences. 2016;113(27):7353–7360.

70. Chipman H, George E, McCulloch R. Bayesian ensemble learning. Advances in Neural Information Processing Systems. 2006;19.

71. Chipman HA, George EI, McCulloch RE. BART: Bayesian additive regression trees. The Annals of Applied Statistics. 2010;4(1):266–298.

72. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association. 2018;113(523):1228–1242.

73. Pennington J, Socher R, Manning CD. Glove: global vectors for word representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2014. p. 1532–1543.

74. Mikolov T, Karafiát M, Burget L, Cernocký J, Khudanpur S. Recurrent neural network based language model. Interspeech. 2010; 2(3)1045–1048.

75. Pham TT, Shen Y. A deep causal inference approach to measuring the effects of forming group loans in online non-profit microfinance platform. arXiv preprint; 2017. arXiv:1706.02795.

76. Louizos C, Shalit U, Mooij JM, Sontag D, Zemel R, Welling M. Causal effect inference with deep latent-variable models. Advances in Neural Information Processing Systems. 2017;30.

77. Shalit U, Johansson FD, Sontag D. Estimating individual treatment effect: generalization bounds and algorithms. Proceedings of the International Conference on Machine Learning. Proceedings of Machine Learning Research (PMLR). 2017;70:3076–3085.

78. Johansson F, Shalit U, Sontag D. Learning representations for counterfactual inference. Proceedings of Machine Learning Research (PMLR). 2016;48:3020–3029.

79. Moraffah R, Karami M, Guo R, Raglin A, Liu H. Causal interpretability for machine learning – problems, methods and evaluation. ACM SIGKDD Explorations Newsletter. 2020;22(1):18–33.

80. Moraffah R, Sheth P, Karami M, Bhattacharya A, Wang Q, Tahir A, Raglin A, Liu H. Causal inference for time series analysis: problems, methods and evaluation. Knowledge and Information Systems. 2021;1–45.

81. Stefan W, Susan A. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association. 2018;113(523):1228–1242. doi: 10.1080/01621459.2017.1319839.

82. Gruber S, van der Laan M. tmle: an R package for targeted maximum likelihood estimation. Journal of Statistical Software. 2012;51(13)1–35. doi: 10.18637/jss.v051.i13.

83. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics. 2005;61(4):962–973. doi: 10.1111/j.1541-0420.2005.00377.x.

84. Schwab P, Linhardt L, Bauer S, Buhmann JM, Karlen W. Learning counterfactual representations for estimating individual dose-response curves. Proceedings of the AAAI Conference on Artificial Intelligence. 2020;34(04):5612–5619.

85. Brodersen KH, Gallusser F, Koehler J, Remy N, Scott SL. Inferring causal impact using Bayesian structural time-series models. The Annals of Applied Statistics. 2015;9(1):247–274. doi: 10.1214/14-AOAS788.

86. Rakesh V, Guo R, Moraffah R, Agarwal N, Liu H. Linked causal variational autoencoder for inferring paired spillover effects. Proceedings of the 27th ACM International Conference on Information and Knowledge Management; 2018. p. 1679–1682.

87. Lu C, Schölkopf B, Hernández-Lobato JM. Deconfounding reinforcement learning in observational settings. arXiv preprint; 2018. arXiv:1812.10576.

88. Sridhar D, Pujara J, Getoor L. Scalable probabilistic causal structure discovery. Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI); 2018. p. 5112–5118.

89. Nauta M, Bucur D, Seifert C. Causal discovery with attention-based convolutional neural networks. machine learning and knowledge extraction. Mach Learn Knowl Extr. 2019;1(1):312–340. https://www.mdpi.com/2504-4990/1/1/19.

90. Spirtes P, Glymour C, Scheines R. Causation, prediction, and search. Vol. 1. MIT Press; 2001.

91. Guo P, Ofonedu A, Wang J. Scalable and hybrid ensemble-based causality discovery. Big Data Research. 2021;26. https://par.nsf.gov/servlets/purl/10217090.

92. Chen H, Harinen T, Lee J-Y, Yung M, Zhao Z. CausalML: Python package for causal machine learning. arXiv preprint; 2020. arXiv:2002.11631.

93. Kalainathan D, Goudet O. Causal discovery toolbox: uncover causal relationships in Python. arXiv preprint; 2019. arXiv:1903.02278.

94. Künzel SR, Sekhon JS, Bickel PJ, Yu B. Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences. 2019;116(10):4156–4165.

95. Yang M, Liu F, Chen Z, Shen X, Hao J, Wang J. CausalVAE: disentangled representation learning via neural structural causal models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2021. p. 9593–9602.

96. Syrgkanis V, et al. Causal inference and machine learning in practice with EconML and CausalML: industrial use cases at Microsoft, TripAdvisor, Uber. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining; 2021. doi: abs/10.1145 /3447548.3470792.

97. Sharma A, Kiciman E. DoWhy: an end-to-end library for causal inference. arXiv preprint; 2020. arXiv:2011.04216.

98. Zhang K, Ahu S, Kalander M, Ng I, Ye J, Chen Z, Pan L. gCastle: a Python toolbox for causal discovery. arXiv preprint; 2021. arXiv:2111.15155.

99. Ramsey JD, Zhang K, Glymour M, Romero RS, Huang B, Ebert-Uphoff I, Samarasinghe S, Barnes EA, Glymour C. TETRAD – a toolbox for causal discovery. Proceedings of the 8th International Workshop on Climate Informatics; 2018. https://www.atmos.colostate.edu/~/PAPERS/CI2018_paper_35.pdf.

100. Röhrig B, Du Prel J-B, Blettner M. Study design in medical research: part 2 of a series on the evaluation of scientific publications. Deutsches Ärzteblatt International. 2009;106(11):184.

101. Röhrig B, Du Prel J-B, Wachtlin D, Blettner M. Types of study in medical research: part 3 of a series on evaluation of scientific publications. Deutsches Ärzteblatt International. 2009;106(15):262.

102. Kuss O, Blettner M, Börgermann J. Propensity score: an alternative method of analyzing treatment effects: part 23 of a series on evaluation of scientific publications. Deutsches Ärzteblatt International. 2016;113(35–36):597.

103. Olsen J, Jensen UJ. Causal criteria: time has come for a revision. European Journal of Epidemiology. 2019;34(6):537–541.

104. Dekkers OM. The long and winding road to causality. European Journal of Epidemiology. 2019;34(6)533–535.

105. Hill AB. The environment and disease: association or causation? Proc R Soc Med. 1965;58(5): 295–300.

106. de Jongh M, Druzdzel MJ. A comparison of structural distance measures for causal Bayesian network models. In: Recent advances in intelligent information systems, challenging problems of science, computer science series. Academic Publishing House EXIT; 2009. p. 443–456.

107. Peters J, Bühlmann P, Meinshausen N. Causal inference by using invariant prediction: identification and confidence intervals. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2016;78(5):947–1012.

108. Hyttinen A, Plis S, Järvisalo M, Eberhardt F, Danks D. Causal discovery from subsampled time series data by constraint optimization. Proceedings of the Conference on Probabilistic Graphical Models. Proceedings of Machine Learning Research (PMLR). 2016;52:216–227.

109. Runge J, Nowack P, Kretschmer M, Flaxman S, Sejdinovic D. Detecting and quantifying causal associations in large nonlinear time series data sets. Science Advances. 2019;5(11):eaau4996.

110. Runge J. Causal network reconstruction from time series: from theoretical assumptions to practical estimation. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2018;28(7):075310.

111. Schaechtle U, Stathis K, Bromuri S. Multi-dimensional causal discovery. Proceedings of the 23rd International Joint Conference on Artificial Intelligence; 2013. https://www.ijcai.org/Proceedings/13/Papers/245.pdf.

112. Haufe S, Müller K-R, Nolte G, Krämer N. Sparse causal discovery in multivariate time series. Proceedings of Workshop on Causality: Objectives and Assessment. Proceedings of Machine Learning Research (PMLR) 2010;6:97–106.

113. Khanna S, Tan VY. Economy statistical recurrent units for inferring nonlinear granger causality. arXiv preprint; 2019. arXiv:1911.09879.

114. Löwe S, Madras D, Zemel R, Welling M. Amortized causal discovery: learning to infer causal graphs from time-series data. arXiv preprint; 2020. arXiv:2006.10833.

115. Wu T, Breuel T, Skuhersky M, Kautz J. Nonlinear causal discovery with minimum predictive information regularization.

116. Xu C, Huang H, Yoo S. Scalable causal graph learning through a deep neural network. Proceedings of the 28th ACM International Conference on Information and Knowledge Management; 2019. p. 1853–1862.

117. Tank A, Covert I, Foti N, Shojaie A, Fox E. Neural Granger causality. arXiv preprint; 2018. arXiv:1802.05842.

118. Papana A, Kyrtsou C, Kugiumtzis D, Diks C. Simulation study of direct causality measures in multivariate time series. Entropy. 2013;15(7)2635–2661.

119. Gong M, Zhang K, Schölkopf B, Glymour C, Tao D. Causal discovery from temporally aggregated time series. Proceedings of the Conference on Uncertainty in Artificial Intelligence; 2017. https://www.ncbi.nlm.nih.gov /pmc/articles/PMC5995575/.

120. Bica I, Alaa A, Van Der Schaar M. Time series deconfounder: estimating treatment effects over time in the presence of hidden confounders. Proceedings of the International Conference on Machine Learning. Proceedings of Machine Learning Research (PMLR). 2020;119:884–895.

121. Dhir N, Hoeltgebaum H, Adams N, Briers M, Burke A, Jones P. Prospective artificial intelligence approaches for active cyber defense. arXiv preprint; 2021. arXiv:2104.09981.

122. Qin X, Lee W. Attack plan recognition and prediction using causal networks. Proceedings of the 20th Annual Computer Security Applications Conference. IEEE; 2004. p. 370–379.

123. Mueller WG, Memory A, Bartrem K. Causal discovery of cyber attack phases. Proceedings of the 18th IEEE International Conference on Machine Learning and Applications (ICMLA); 2019. p. 1348–1352.

124. Tople S, Sharma A, Nori A. Alleviating privacy attacks via causal learning. Proceedings of the International Conference on Machine Learning. Proceedings of Machine Learning Research (PMLR). 2020;119:9537–9547.

125. Ding K, Li J, Bhanushali R, Liu H. Deep anomaly detection on attributed networks. Proceedings of the 2019 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics (SIAM); 2019. p. 594–602.

126. Li Y, Guo R, Wang W, Liu H. Causal learning in question quality improvement. Proceedings of the International Symposium on Benchmarking, Measuring and Optimization. Springer; 2019. p. 204–214.

127. Marin E, Guo R, Shakarian P. Temporal analysis of influence to predict users' adoption in online social networks. Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. Springer; 2017. p. 254–261.

128. Sarkar S, Guo R, Shakarian P. Using network motifs to characterize temporal network evolution leading to diffusion inhibition. Social Network Analysis and Mining. 2019;9(1):1–24.

129. Toulis P, Volfovsky A, Airoldi EM. Propensity score methodology in the presence of network entanglement between treatments. arXiv preprint; 2018. arXiv:1801.07310.

130. Pearl J. The foundations of causal inference. Sociological Methodology. 2010;40(1):75–149.

131. Kusner MJ, Loftus J, Russell C, Silva R. Counterfactual fairness. Advances in Neural Information Processing Systems. 2017;30.

132. Arjovsky M, Bottou L, Gulrajani I, Lopez-Paz D. Invariant risk minimization. arXiv preprint; 2019. arXiv:1907.02893.

133. Hiemstra C, Jones JD. Testing for linear and nonlinear Granger causality in the stock price-volume relation. The Journal of Finance. 1994;49(5)1639–1664.

134. Stips A, Macias D, Coughlan C, Garcia-Gorriz E, Liang XS. On the causal structure between CO_2 and global temperature. Scientific Reports. 2016;6(1):1–9.

135. Runge J, Bathiany S, Bollt E, Camps-Valls G, Coumou D, Deyle E, Glymour C, Kretschmer M, Mahecha MD, Muñoz-Marí J, et al. Inferring causation from time series in Earth system sciences. Nature Communications. 2019;10(1)1–13.

136. Guyon I, Janzing D, Schölkopf B. Causality: Objectives and assessment. Proceedings of Workshop on Causality: Objectives and Assessment. Machine Learning Research (PMLR). 2010;6:1–42.

List of Symbols, Abbreviations, and Acronyms

AI	artificial intelligence
ARB	approximate residual balancing
ATE	average treatment effect
BART	Bayesian Additive Regression Tree
BIC	Bayesian information criterion
CART	Classification and Regression Tree
CATE	conditional average treatment effect
CB	constraint based
CBPS	covariate balancing propensity score
CEVAE	Causal Effect Variational Autoencoder
DAG	directed acyclic graph
DARPA	US Defense Advanced Research Projects Agency
DL	Deep Learning
EB	entropy balancing
FCI	fast causal inference
FCM	functional causal model
FGES	Fast Greedy Equivalent Search
FPR	false positive rate
GES	Greedy Equivalent Search
Glove	Global Vectors
IC	inferred causation
ICA-LiNGAM	Independent Component Analysis LiNGAM
ITE	individual treatment effect
IV	Instrumental Variable
LCVA	Linked Causal Variational Autoencoder
LiNGAM	Linear Non-Gaussian Acyclic Model

ML	machine learning
MMHC	Max-Min Hill-Climbing
MSE	mean squared error
NML	normalized maximum likelihood
NN	Neural Network
PSL	Partial Bayesian network Structure Learning
PSM	propensity score matching
RDD	regression discontinuity design
RF	Random Forest
RL	Reinforcement Learning
RNN	Recurrent Neural Network
ROC	Receiver Operator Characteristic
SB	score based
SCM	structural causal model
SHD	structural hamming distance
SUTVA	Stable Unit Treatment Value Assumption
TCDF	Temporal Causal Discover Framework
TMLE	Targeted Maximum Likelihood Estimator
TPR	true positive rate
XAI	Explainable AI

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 DEVCOM ARL
(PDF) FCDD RLD DCI
TECH LIB

2 DEVCOM ARL
(PDF) FCDD RLC IT
A RAGLIN
B SADLER

2 HOWARD UNIVERSITY
(PDF) A RAWAL
D RAWAT