# A Causal AI Suite for Decision-Making

**Emre Kıcıman**
Microsoft Research
emrek@microsoft.com

**Eleanor Dillon**
Microsoft Research
eldillon@microsoft.com

**Darren Edge**
Microsoft Research
darren.edge@microsoft.com

**Adam Foster**
Microsoft Research
adam.e.foster@microsoft.com

**Agrin Hilmkil**
Microsoft Research
agrinhilmkil@microsoft.com

**Joel Jennings**
Microsoft Research
joeljennings@microsoft.com

**Chao Ma**
Microsoft Research
chao.ma@microsoft.com

**Robert Ness**
Microsoft Research
robertness@microsoft.com

**Nick Pawlowski**
Microsoft Research
nick.pawlowski@microsoft.com

**Amit Sharma**
Microsoft Research
amshar@microsoft.com

**Cheng Zhang**
Microsoft Research
cheng.zhang@microsoft.com

## Abstract

Critical data science and decision-making questions across a wide variety of domains are fundamentally causal questions. We present a suite of open-source causal tools and libraries that aims to simultaneously provide core causal AI functionality to practitioners and create a platform for research advances to be rapidly deployed. In this paper, we describe our contributions towards such a comprehensive causal AI suite of tools and libraries, its design, and lessons we are learning from its growing adoption. We hope that our work accelerates use-inspired basic research for improvement of causal AI.

## 1 Introduction

Critical data science and decision-making questions across a wide variety of domains are fundamentally causal questions. Whether estimating the impact of a marketing campaign, understanding the reasons for customer churn, predicting the impact of climate change, or identifying which drug may work best for which patient, answering key questions relies on successfully modeling cause-and-effect relationships that underlie a system. As the field of data science and applications of data-driven decision making expand, many people are recognizing the importance of causal methods for providing robust insights. However, valid causal analysis requires a combination of human domain expertise (to make reasonable assumptions) and sophisticated analytical frameworks (to build those assumptions

NeurIPS 2022Workshop on Causal Machine Learning for Real-World Impact (CML4Impact 2022).

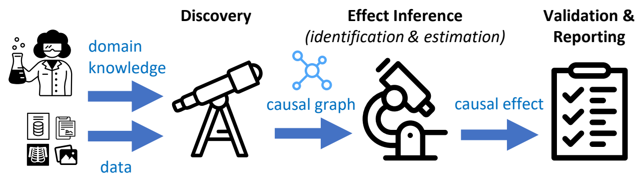



Figure 1: End-to-end causal inference pipeline

into causal models and learning algorithms), which makes it particularly important to develop robust tools that empower analysts and decision-makers to use the latest statistical advances.

The causal AI research area is still early in its development, however, and as with any technology area, will require many more advances and iterative practical deployments to reach its full impact. As argued by Stokes [40], fundamental research is accelerated when it is motivated and informed by considerations of use. To enable and accelerate such "use-inspired basic research" in causality, we should create mechanisms that enable and ease the practical usage of state of the art technology, and collect results and challenges to inform the direction of research. This requires that we broaden the accessibility of causal methods beyond today's causal experts, and make them available to the larger audiences of scientists, decision-makers, and other practitioners, so that we may discover the fundamental challenges of causality that stymie its full adoption. Learning from broad usage of these methods, and in particular the use cases in which current methods are insufficient or fail, will motivate and speed new research directions that our field might not otherwise prioritize or even be aware of.

We present a suite of causal tools and libraries that aims to simultaneously provide core causal AI functionality to practitioners and create a platform for research advances to be rapidly deployed. First, this suite aims to be easy for practitioners, including non-causal experts, to use. Importantly, the tools should provide scaffolding to those familiar with non-causal machine learning and data science processes as they learn causal workflows. Ideally, tools should make it easy for people to make the right assumptions and come to the correct causal conclusions, and make it difficult for people to make mistakes. Second, to enable new research advances to be rapidly deployed, the suite provides an interface that clearly separates the problem (e.g., a given causal task we wish to perform or causal question we wish to answer) from the algorithmic and methodological implementations that may solve the problem. Finally, to encourage cross-pollination of ideas across disparate causal communities of interest, the suite of tools and libraries bridges across multiple causal frameworks and tasks, including graphical approaches to causal reasoning, potential outcomes approaches to effect inference, causal discovery algorithms, and others. Parts of our causal AI suite are already widely used across industry and academia, with over 1.5 million downloads, and newer components are receiving significant interest as well.

In this paper, we describe our contributions towards such a comprehensive causal AI suite tools and libraries, our design decisions, lessons learned from their usage, and what is next.

## 2 A Causal AI Suite

The primary goals of our causal AI suite are (1) to broaden usage of causal methods by data scientists and decision-makers in practice; and (2) act as a flywheel to accelerate fundamental causal research and impact. The initial focus of our suite is the end-to-end causal inference task, as shown in Figure 1. The inputs to the pipeline are data and domain knowledge, capturing key causal assumptions as partial structural knowledge of a causal graph and other non-graphical assumptions. We use causal discovery to aid the exploration and inference process.

We first discuss our core design principles; and then present the major libraries and tools that make up our causal AI suite: DoWhy, EconML, Causica, and ShowWhy.

### 2.1 Design Principles and Decisions

**Scaffolding the causal analysis process:** The right API abstraction can guide users to follow best practices of causal inference and characterize the impact of causal assumptions on the estimated effect. For instance, incorrect identification leads to bias in an estimate that cannot be corrected by simply optimizing across different estimation methods ([38]). In the absence of a global cross-validation evaluation procedure as in supervised machine learning, the steps for stating and validating assumptions are a critical part of causal inference ([35]).

Thus, a key design decision in our causal AI suite is organized around an analysis pipeline that elevates domain knowledge and its assumptions as a first-class entity. These assumptions dynamically guide the identification of algorithmic approaches, e.g., for causal estimation. And, these assumptions are all subject to validation, refutation, or sensitivity analysis in our pipeline.

**Separation of causal questions from algorithmic and methodological implementations:** Given the rapid pace of advancement in causal methods and machine learning more broadly, we expect to see significant changes in algorithms and methodological approaches for causal inference. To enable and encourage such advancement, we build API abstractions that cleanly separate how we represent and ask causal questions from how those questions are answered. The right abstractions also allow easier reuse and flexible combinations of algorithms and core components across related causal tasks.

For example, given a graphical causal model, the identification step of causal effect inference is a causal reasoning problem whereas the estimation step is a statistical problem. Our libraries enforce this boundary by having two separate API calls for identification and estimation. This allows modularity, allowing us to rapidly expand our set of supported identification and estimation methods independently. Furthermore, we can "mix-and-match" identification and estimation methods: multiple estimation methods can be used for a single identified estimand and vice versa.

**Find common abstractions across causal frameworks**. To encourage cross-pollination of ideas, we build abstractions that bridge across multiple causal frameworks and tasks, including graphical approaches to causal reasoning, potential outcomes approaches to effect inference, causal discovery algorithms, and others. The exercise of finding common ways of expressing the causal assumptions used by graphical reasoning methods, panel data methods, and causal discovery encourages us to consider both the similarities and distinctive strengths of methods, algorithms, and scenarios.

All of the components below are available as open-source Python packages [1].

### 2.2 DoWhy

DoWhy is an end-to-end library for causal analysis that builds on the latest research in modeling assumptions and robustness checks ([6, 25]), and provides an easy API interface. To scaffold and guide users through the best practices of causal inference, DoWhy's API is organized around the four fundamental steps for causal analysis: Model, Identify, Estimate, and Refute. **Model** encodes prior knowledge as a formal causal graph, **identify** uses graph-based methods such as the ID algorithm to identify the causal effect, **estimate** uses statistical methods, such as EconML's sophisticated methods for estimating the identified estimand, and finally **refute** tests the validity of the assumed graph and resultant estimate. DoWhy unifies two powerful frameworks: causal graphs ([30]) and potential outcomes ([21]). Specifically, DoWhy uses graph-based criteria and do-calculus for modeling assumptions and identifying a non-parametric causal effect. For estimation, it switches to methods based on potential outcomes, including methods from external libraries like EconML ([28]) and CausalML ([10]).

In 2022, DoWhy became the first library to join the independent community-governed PyWhy organization as part of an effort to encourage the development of an open source ecosystem for causal machine learning.

### 2.3 EconML

The EconML library complements DoWhy with a focus on sophisticated estimation methods based on the latest advances in causal machine learning. EconML applies advanced machine learning techniques to estimate individualized causal responses from observational or experimental data. The suite of estimation methods provided in EconML represents the latest advances in causal machine

[1] DoWhy: `https://github.com/py-why/dowhy/`; EconML: `https://github.com/microsoft/econml/`; Causica: `https://github.com/microsoft/causica/`; ShowWhy: `https://github.com/microsoft/showwhy/`

learning. Supervised machine learning algorithms (such as random forests, boosting, lasso, and neural nets) have become enormously effective at predicting an outcome $Y$ from a set of features. They achieve this good performance, and improve on earlier ad hoc methods, with automated model selection, usually with cross-fitting. The central advance of causal machine learning is to recognize that many individual steps within causal analysis can benefit from machine learning. EconML encodes this causal machine learning approach (e.g., DoubleML, meta-learners, and many others) into a simple API that builds intuitively on other popular ML packages such as scikit-learn. Armed with an appropriate data sample, analysts can estimate reliable causal effects with a few lines of code. In the background, EconML estimates "nuisance" functions using flexible ML models—including the option to use AutoML to choose across multiple algorithms, automatically cross-fits to tune those models—and builds these estimates into carefully constructed final models that use the latest techniques to deliver interpretable causal effects. Each approach in the library includes confidence intervals, many constructed analytically for added efficiency, and enables flexible modeling of *heterogeneous* treatment effects to understand how treatment effects vary across a population and create targeted, personalized plans of action.

### 2.4 Causica

Causica develops deep learning methods for end-to-end causal inference (DECI), using existing data and domain constraints to both perform causal discovery and compute causal inference quantities such as (conditional) average treatment effects. Causica also provides core functionality for a variety of causal tasks, including missing value prediction, best next question, causal discovery, and causal inference. The DECI model is a generative approach that employs an additive noise structural equation model (ANM-SEM) to capture the functional relationships among variables and exogenous noise, while simultaneously learning a variational distribution over causal graphs [15]. DECI estimates causal quantities (ATE, CATE, and ITE) by applying the relevant interventions to its learnt causal graph and then sampling from the generative model.

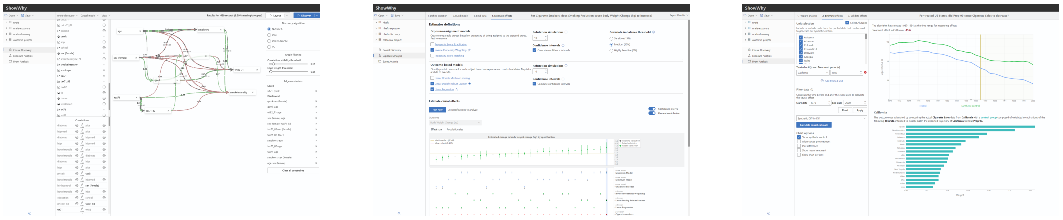
Many real-world application are more complex than this and cannot be modelled with simple ANM-SEMs. A large set of domains, such as finance, healthcare, or sales, deal with time-series data and are interested in predicting the temporal progression of interventions. Rhino [17] extends the DECI framework with an auto-regressive formulation to introduce a novel functional form for temporal SEMs. The functional form models instantaneous effects, complex functional relationships, and history-dependent noise. The Rhino SEM is shown to be structure identifiable under common causal assumptions and can be learned using the same variational inference based framework as DECI.

Another common – but often inaccurate – assumption of causal modelling is the absence of unobserved confounders. For example, many sales scenarios are influenced by a latent confounder describing the global economic situation. In [5], DECI is generalized to handle the existence of latent confounders with the help of Acyclic Directed Mixed Graphs (ADMGs). Theoretically, it is shown that presence of latent confounding is identifiable under the assumptions of bow-free ADMGs with nonlinear additive noise models. This insight enables a straightforward extension of DECI, capable of learning both the causal structure and the posterior values of of latent confounders at the same time. This further boosts the empirical performance of DECI in both causal discovery and causal inference.

### 2.5 ShowWhy

ShowWhy builds atop this ecosystem of Causal AI libraries to develop a new class of no-code user interfaces (GUIs) that empower domain experts to become "decision scientists" who can independently ask a causal question, develop causal estimates, and present and defend causal evidence to an audience of decision makers. We expect that making causal methods available to this broader audience of stakeholders will motivate and accelerate new use-inspired basic research. As new algorithms and methods are implemented underneath the shared APIs of the DoWhy, EconML, and Causica libraries, ShowWhy's no-code user interface can easily leverage them.

Since the broader audience of ShowWhy users may not be familiar with specific machine learning estimators or causal inference more generally, a key design principle for our no-code interfaces is to use a range of estimators, problem specifications, and visualization techniques to communicate the overall balance of evidence regarding the existence and strength of potential causal relationships.



(a) Exploration: Use causal discovery methods to explore causal links among variables.

(b) Exposure Analysis: Use effect inference methods to see whether exposure causes outcomes.

(c) Event analysis: Use panel data analyses to determine whether an event caused an outcome.

Figure 2: The ShowWhy no-code interfaces enable domain experts without coding expertise to use the latest methods to answer key causal questions

ShowWhy's no-code interfaces also recognize the fact that decision making is an ongoing process, often involving rapid iterative feedback loops, that extends beyond any individual decision. We can even view it as a new kind of decision science: first, discovering candidate interventions; second, deciding which interventions to make; and third, evaluating the impact of chosen interventions over time. Figure 2 presents the three major workflows supported by ShowWhy. The first workflow helps users answer exploratory questions of the form "do variables in some dataset have causal links? The next workflow helps users answer questions of the form "does exposure to a specific treatment cause a specific outcome?" The third and final workflow helps users answer questions of the form "did a specific event cause a specific outcome over time?"

# 3 Example usage

Parts of the causal suite are widely used across industry and academia: DoWhy and EconML, for example, have been downloaded over 1.5 million times. Through adoption and usage of these tools, we have found several templates of causal analysis scenarios to be commonly used: (1) identifying causal drivers of key metrics and outcomes; (2) return-on-investment analyses, such as identifying the impact of customer loyalty programs; (3) robust prediction and forecasting, such as modeling future impacts of expected climate change; and (4) segmentation and personalized treatment scenarios.

Here, we present an example usage of these causal analysis libraries for a multi-investment attribution task, identifying the business actions that have a causal effect on revenue from enterprise customers.

## 3.1 Example: Multi-investment Attribution

A software company would like to know whether its multiple outreach efforts to their business customers are successful in boosting sales. They would also like to learn how to better target different incentives to different customers. In other words, they would like to learn the treatment effect of each investment on customers' total expenditure on the company's products—particularly the heterogeneous treatment effect.

Here, we show how tools from the Causica, EconML and DoWhy libraries can use historical investment data to learn the effects of multiple investments. We then use Causica to discover the causal graph: the relationship between each variable in the simulated data. With this generated graph, we use DoWhy to identify an appropriate strategy to estimate the causal effect. We pass this recommendation to EconML to estimate the personalized treatment effects for each customer. We also show an alternative effect estimation using Causica. Finally, we use DoWhy to test the assumptions underlying the causal estimation. The full walkthrough is available as a Jupyter notebook at `https://aka.ms/causal_suite_notebook`.

### 3.1.1 Data

We create a simulated dataset of 10,000 customers. Each customer is associated with annual revenue from the customer and whether they received technical support, a discount, or were targeted by new outreach strategies. Each customer is annotated with 10 additional characteristics that might affect revenue, such as the company's employee count, whether the customer is a commercial or public sector, a global company, etc. Data is simulated to mimic realistic correlations between features.

#### 3.1.2 Discovering the Causal Graph

The first step in our analysis, shown in Listing 1, is to extract the causal graph using a causal discovery algorithm, here using the `DECI` algorithm from Causica [15]. To inform the causal discovery procedure, we provide several simple constraints in the form of a `constraint_matrix`. In this case study, we declare that the *revenue* outcome variable cannot be the cause of other nodes; that certain attributes of companies cannot be changed by others; and that treatments do not cause each other. We omit this code for brevity.

Listing 1: Causal Discovery Code Example

```
## Causal Discovery step
model = DECI.create("mymodel", modeldir, dataset.variables,
                    cfg, device="gpu")
model.set_graph_constraint(constraint_matrix) # provide domain knowledge
model.run_train(dataset, training_params)
discovered_graph = model.networkx_graph()
```

#### 3.1.3 Treatment effect identification and estimation

The causal graph identifies the likely paths of connection between features. This next step will quantify the strength of those relationships between treatment and outcomes. Our tools estimate both the average treatment effect across all customers (ATE) and how these treatment effects vary across customer features.

DoWhy tools can help users identify an appropriate causal model for their question and understand which confounders, or conditioning variables, they should include in their estimation. To identify an estimation approach, DoWhy needs to work with a causal graph that describes the problem space. DoWhy can work with a causal graph generated through discovery, or input directly by users based on their domain knowledge, including perhaps identifying features that may affect the outcome and treatment but are unobservable.

Listing 2 runs DoWhy's identification algorithm. In this scenario, DoWhy recommends a backdoor estimation approach, which relies on the unconfoundedness assumption, and selects all of the potential confounding variables except for one which was not identified as a confounder in the graph. We then use DoWhy to call the Linear Double Machine Learning model from the EconML library for the estimation step.

Listing 2: Identification and Estimation Code Example

```
## Identification
dw = dowhy.CausalModel(data=dataset,graph=discovered_graph,
                       treatment="Tech Support", outcome="Revenue",
                       effect_modifiers = ['Global Flag', 'Size'])
estimand = dw.identify_effect(method_name="maximal-adjustment",
                              estimand_type='nonparametric-cde')

# Estimate
estimate = dw.estimate_effect(estimand,
                              confidence_intervals=True,
                              method_name="backdoor.econml.dml.LinearDML")
```

#### 3.1.4 Validation and Reporting

If causal assumptions are incorrect or otherwise violated, then identified causal estimands will be biased in ways that cannot be corrected through improved estimation methods ([37]). As validating assumptions is a critical part of causal inference ([35]), we briefly describe the refute step here. We leave code examples of refutation and validation out of the paper due to space considerations.

To aid data scientists' inspection of the validity of causal assumptions, DoWhy provides two kinds of tests: 1) *Refutations* are necessary—but not sufficient—tests of the validity of a causal analysis must pass; and 2) *Sensitivity analyses* that measure the robustness of a causal estimate to violations of assumptions [34]. Tests supported by DoWhy include addition of random common causes, placebo treatments, data subset validation, bootstrap validation, and testing of independence constraints entailed by the causal graph.

## 4 Learning from practical deployment and next steps

These tools have been used in a wide variety of domains, including health and medicine [42, 13, 4, 3, 12, 11, 1, 39, 29, 7], logistics and travel [43, 9, 31, 41], marketing and sales [2, 19, 36], environment and agriculture [22, 14, 16], manufacturing [26], law [20], education [23, 8], and others. From the questions and feedback we receive, as well as from our use of these packages in our own application scenarios, we see several areas of both fundamental interest and with potential for practical impact.

**Improved elicitation of domain knowledge**. First, we see that data scientists and domain experts often have difficulty encoding domain knowledge correctly–whether because domain experts are unavailable or for lack of causal expertise. As a result, we are seeking new ways to ease the elicitation of domain knowledge, including tighter integration of causal discovery and interactive tools for exploration and bootstrapping. Other promising directions include seeking to extract structural domain knowledge from other sources, such as the academic literature [32] and trusted simulators [27].

**Improved validation, refutation, and sensitivity analyses.** Continual improvement of validation and refutation methods, including integration with A/B experimentation methodologies when available is an important need for many of the scenarios we observe.

**Improved support for unstructured and high-dimensional text and image data**. In many application areas today, the richest source of data about a domain is text or image data, such as clinical notes in electronic health records, or satellite imagery capturing climate impacts. While methods for causal inference in the context of high-dimensional data are being explored [33, 18], further work is needed to create approaches for analysis of *unstructured* data, e.g., building on causal representation methods [24].

Our next steps include continuing to grow the open-source community and ecosystem for Causal AI tools and libraries; expanding capabilities to new tasks and APIs, such as causal representation learning, and robust prediction. We encourage others to join us in creating a vibrant ecosystem that serves both researchers and practitioners.

## 5 Acknowledgments

We would like to thank Keith Battocchi, Mónica Carvajal, Denise Chen, Nathan Evans, Andrés Morales Esquivel, Jonathan Larson, Friederike Niedtner, Dayenne de Souza, Christopher Trevino, Ha Trinh, Fabio Vera, Robert King, Ahmed Mostafa, and Thomas Kapler, our collaborators in the PyWhy organization, Peter Götz, Patrick Blöbaum, Kailash Budhathoki and all our open-source contributors for discussions and contributions building the Causal AI Suite.

## References

[1] Rohan Aanegola, Shinpei Nakamura Sakai, and Navin Kumar. Longitudinal analysis of the determinants of life expectancy and healthy life expectancy: A causal approach. *Healthcare Analytics*, 2:100028, 2022.

[2] Carlos Aguilar Palacios et al. Interpretable machine learning for promotional sales forecasting. 2021.

[3] Alejandro Almodóvar Espeso. *Implementation and analysis of Causal Inference techniques for treatment effects estimation applied to Covid-19 cases*. PhD thesis, Telecomunicacion, 2022.

[4] AR Anness, A Clark, K Melhuish, FMT Leone, MW Osman, D Webb, T Robinson, N Walkinshaw, A Khalil, and HA Mousa. Maternal hemodynamics and neonatal birth weight in preg-

nancies complicated by gestational diabetes: new insights from novel causal inference analysis modeling. *Ultrasound in Obstetrics & Gynecology*, 2022.

[5] Matthew Ashman, Chao Ma, Agrin Hilmkil, Joel Jennings, and Cheng Zhang. Causal reasoning in the presence of latent confounders via neural admg learning. In *NeurIPS 2022 Workshop on Causality for Real-world Impact*.

[6] Susan Athey and Guido W Imbens. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32, 2017.

[7] Orel Babayoff, Onn Shehory, Meishar Shahoha, Ruth Sasportas, and Ahuva Weiss-Meilik. Surgery duration: Optimized prediction and causality analysis. *Plos one*, 17(8):e0273831, 2022.

[8] Diana Cárdenas, Finnian Lattimore, Daniel Steinberg, and Katherine J Reynolds. Youth well-being predicts later academic success. *Scientific reports*, 12(1):1–13, 2022.

[9] Rishabh Singh Chauhan, Christoffer Riis, Shishir Adhikari, Sybil Derrible, Elena Zheleva, Charisma F Choudhury, and Francisco Camara Pereira. Causality in travel mode choice modeling: A novel methodology that combines causal discovery and structural equation modeling. *arXiv preprint arXiv:2208.05624*, 2022.

[10] Huigang Chen, Totte Harinen, Jeong-Yoon Lee, Mike Yung, and Zhenyu Zhao. Causalml: Python package for causal machine learning, 2020.

[11] Thomas Alexander Christoph, Onur Galoglu, Jiaqi Lu, Sarah Lumpp, Daniel Sturm, Niki Kilbertus, Co-Mentor M Sc Konstantin Goebler, and M Sc David Strieder. A robust comparison of causal effects from observational data in healthcare. 2021.

[12] Diego F Cuadros, Juan D Gutierrez, Claudia M Moreno, Santiago Escobar, DeWolfe Miller, Godfrey Musuka, Ryosuke Omori, Phillip Coule, and Neil J MacKinnon. Impact of healthcare capacity disparities on the covid-19 vaccination coverage in the united states.

[13] Diego F Cuadros, Juan D Gutierrez, Claudia M Moreno, Santiago Escovar, F DeWolfe Miller, Godfrey Musuka, Ryosuke Omori, Phillip Coule, and Neil J MacKinnon. Impact of healthcare capacity disparities on the covid-19 vaccination coverage in the united states: A cross-sectional study. *medRxiv*, 2022.

[14] Limin Feng, Ting Yang, and Zifa Wang. Performance evaluation of photographic measurement in the machine-learning prediction of ground pm2. 5 concentration. *Atmospheric Environment*, 262:118623, 2021.

[15] Tomas Geffner, Javier Antoran, Adam Foster, Wenbo Gong, Chao Ma, Emre Kiciman, Amit Sharma, Angus Lamb, Martin Kukla, Nick Pawlowski, Miltiadis Allamanis, and Cheng Zhang. Deep end-to-end causal inference, 2022.

[16] Georgios Giannarakis, Vasileios Sitokonstantinou, Roxanne Suzette Lorilla, and Charalampos Kontoes. Towards assessing agricultural land suitability with causal machine learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1442–1452, 2022.

[17] Wenbo Gong, Joel Jennings, Cheng Zhang, and Nick Pawlowski. Rhino: Deep causal temporal relationship learning with history-dependent noise. In *NeurIPS 2022 Workshop on Causality for Real-world Impact*, 2022.

[18] Jason Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. Deep IV: A flexible approach for counterfactual prediction. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1414–1423. PMLR, 06–11 Aug 2017.

[19] David Hason Rudd, Huan Huo, and Guandong Xu. Improved churn causal analysis through restrained high-dimensional feature space effects in financial institutions. *Human-Centric Intelligent Systems*, pages 1–11, 2022.

[20] Zhe Huang, Xinyue Zhang, Ruofan Wang, and Daniel L Chen. Deep iv in law: Appellate decisions and texts impact sentencing in trial courts. *Elements in Law, Economics and Politics*, 2022.

[21] Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

[22] Qiao Kang, Xing Song, Xiaying Xin, Bing Chen, Yuanzhu Chen, Xudong Ye, and Baiyu Zhang. Machine learning-aided causal inference framework for environmental data analysis: a covid-19 case study. *Environmental Science & Technology*, 55(19):13400–13410, 2021.

[23] Marzieh Karimi-Haghighi, Carlos Castillo, and Davinia Hernández-Leo. A causal inference study on the effects of first year workload on the dropout rate of undergraduates. In *International Conference on Artificial Intelligence in Education*, pages 15–27. Springer, 2022.

[24] Jivat Neet Kaur, Emre Kiciman, and Amit Sharma. Modeling the data-generating process is necessary for out-of-distribution generalization, 2022.

[25] Emre Kıcıman and Amit Sharma. Tutorial on causal inference and counterfactual reasoning, 2018.

[26] Rui Liu and Siddharth Misra. Monitoring the propagation of mechanical discontinuity using data-driven causal discovery and supervised learning. *Mechanical Systems and Signal Processing*, 170:108791, 2022.

[27] Daniel McDuff, Yale Song, Jiyoung Lee, Vibhav Vineet, Sai Vemprala, Nicholas Alexander Gyde, Hadi Salman, Shuang Ma, Kwanghoon Sohn, and Ashish Kapoor. Causalcity: Complex simulations with agency for causal discovery and reasoning. In *Conference on Causal Learning and Reasoning*, pages 559–575. PMLR, 2022.

[28] Microsoft-Research. EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. https://github.com/microsoft/EconML, 2019. Version 0.6.

[29] Beatriz Muñoz-Ospina, Daniela Alvarez-Garcia, Hugo Juan Camilo Clavijo-Moran, Jaime Andrés Valderrama-Chaparro, Melisa García-Peña, Carlos Alfonso Herrán, Christian Camilo Urcuqui, Andrés Navarro-Cadavid, and Jorge Orozco. Machine learning classifiers to evaluate data from gait analysis with depth cameras in patients with parkinson's disease. *Frontiers in Human Neuroscience*, 16, 2022.

[30] Judea Pearl. *Causality*. Cambridge university press, 2009.

[31] Huy Pham, Xuan Jiang, and Cong Zhang. Causality and advanced models in trip mode prediction: Interest in choosing swissmetro. 2022.

[32] Hoifung Poon, Chris Quirk, Charlie DeZiel, and David Heckerman. Literome: Pubmed-scale genomic knowledge base in the cloud. *Bioinformatics*, 30(19):2840–2842, 2014.

[33] Scott Powers, Junyang Qian, Kenneth Jung, Alejandro Schuler, Nigam H Shah, Trevor Hastie, and Robert Tibshirani. Some methods for heterogeneous treatment effect estimation in high dimensions. *Statistics in medicine*, 37(11):1767–1787, 2018.

[34] James M Robins, Andrea Rotnitzky, and Daniel O Scharfstein. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 1–94. Springer, 2000.

[35] Paul R Rosenbaum, PR Rosenbaum, and Briskman. *Design of observational studies*, volume 10. Springer, 2010.

[36] David Hason Rudd, Huan Huo, and Guandong Xu. Causal analysis of customer churn using deep learning. In *2021 International Conference on Digital Society and Intelligent Systems (DSInS)*, pages 319–324. IEEE, 2021.

[37] Amit Sharma and Emre Kiciman. Dowhy: An end-to-end library for causal inference. Causal Data Science Meeting (https://causalscience.org/), November 2020.

[38] Amit Sharma, Vasilis Syrgkanis, Cheng Zhang, and Emre Kıcıman. Dowhy: Addressing challenges in expressing and validating causal assumptions. *arXiv preprint arXiv:2108.13518*, 2021.

[39] Paul Stephan, Felix Wortmann, Kevin Koch, et al. Understanding the interactions between driving behavior and well-being in daily driving: Causal analysis of a field study. *Journal of medical Internet research*, 24(8):e36314, 2022.

[40] Donald E Stokes. *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press, 2011.

[41] Vasilis Syrgkanis, Greg Lewis, Miruna Oprescu, Maggie Hei, Keith Battocchi, Eleanor Dillon, Jing Pan, Yifeng Wu, Paul Lo, Huigang Chen, et al. Causal inference and machine learning in practice with econml and causalml: Industrial use cases at microsoft, tripadvisor, uber. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pages 4072–4073, 2021.

[42] Zheyu Wang, Jiye An, Hui Lin, Jiaqiang Zhou, Fang Liu, Juan Chen, Huilong Duan, Ning Deng, et al. Pathway-driven coordinated telehealth system for management of patients with single or multiple chronic diseases in china: System development and retrospective study. *JMIR Medical Informatics*, 9(5):e27228, 2021.

[43] Xinlan Xia, Zhiyuan Luo, and Qiuran Lyu. Analysis of aircraft delay with deep belief network and causal graph methodology. In *2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, pages 525–531. IEEE, 2022.