Bus Inf Syst Eng
https://doi.org/10.1007/s12599-026-00999-x

EDITORIAL

Causal Machine Learning in Information Systems Research

Moritz von Zahn · Arda Güler · Jella Pfeiffer · Hajo A. Reijers · Oliver Hinz

1 Introduction

Researchers aim to identify causal relationships to understand the world and develop theories, while practitioners and especially managers focus on how their actions produce effects in order to guide decisions. With the recent advancements in machine learning, a new and promising class of methods has emerged for causal inference: causal machine learning (causal ML).

While traditional statistical methods have long addressed causal inference, they are often constrained by functional forms and limited scalability in high-dimensional settings. Causal ML builds on the foundational strengths of traditional ML, that is, the ability to flexibly learn complex patterns from data (Jordan and Mitchell 2015), and leverages these strengths to estimate causal effects. Importantly, causal ML can estimate causal effects at the subgroup or even individual level in settings where traditional methods reach their limits (Feuerriegel et al. 2024). Given the crucial role of causal inference, causal ML thus holds immense potential for domains such as public policy (Strittmatter 2023), healthcare (Feuerriegel et al. 2024), and marketing (von Zahn et al. 2025). As we argue in this editorial, it offers similar potential for Information Systems (IS) research.

Building on the recent call for machine learning in IS research (Padmanabhan et al. 2022), we argue that the emergence of causal ML positions IS researchers to contribute along two main paths. First, researchers can leverage causal ML as a *method* to uncover causal relationships underlying phenomena relevant to the field, thereby enhancing IS theory or making methodological contributions by adapting causal ML algorithms to the unique requirements of sociotechnical systems. Second, researchers can study causal ML as an emerging *phenomenon*, including its development, application, and effects on organizational decision-making. To do so, they can leverage the full spectrum of IS research methodologies. While other avenues may exist, we consider these two paths to be the primary opportunities for IS scholars engaging with causal ML.

The following sections provide an overview of causal ML methods and broad guidelines for their application, along with the intuition behind them. We also outline ideas on how IS researchers engaging with causal ML can make meaningful contributions to the field.

2 Foundations of Causal Machine Learning

Causal ML is designed to answer *what-if* questions: What would have happened had an alternative decision been made? For instance, a ride-sharing platform may want to understand how a specific user's weekly ride count would have changed if certain features of the app had been modified. While traditional methods for causal inference

M. von Zahn (✉) · A. Güler · O. Hinz
Information Systems and Information Management, Goethe University Frankfurt, Frankfurt am Main, Germany
e-mail: vzahn@wiwi.uni-frankfurt.de

J. Pfeiffer
Institute for Information Systems (WIN), Karlsruhe Institute of Technology, Karlsruhe, Germany

H. A. Reijers
Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

(e.g., from econometrics) can address such questions, they typically rely on strong assumptions about the parametric functional form of the relationship between covariates (in this case, user characteristics), the treatment (a feature change in the app), and the outcome (the number of rides taken), often assuming linear and additive effects that may not reflect reality. In practice, researchers often lack both intuition and theoretical guidance to select appropriate functional forms, which may lead to model misspecification and biased inference (Shi et al. 2025).

The challenges of traditional methods for causal inference become particularly pronounced in high-dimensional settings, where the number of potentially relevant covariates is large, and their effect on both the treatment and outcomes may be complex and nonlinear. Following our previous example, a new app feature's causal effect on a user's ride frequency may depend on a combination of ride history, spatiotemporal factors, and user characteristics such as technological literacy and demographics, interacting in ways that are difficult to specify ex-ante.

Traditional ML can capture such complex, nonlinear patterns in high-dimensional data (Jordan and Mitchell 2015). In the ride-sharing example, traditional ML can be trained on rich behavioral data such as a user's ride history to predict a user's weekly ride count given a feature change. However, traditional ML methods optimize for predictive accuracy with regard to the outcome, such as the ride count; they cannot predict how the outcome changes *because* of a treatment. To estimate the effect of a treatment, such as that of a new app feature on the ride count, we can rely on an emerging class of methods that combine the flexibility of ML with the principles of causal inference, namely, causal ML methods. In this section, we provide a foundation for causal ML by outlining the fundamental problem and core assumptions of causal inference, differences between experimental and observational data, and finally an overview on the variety of causal ML methods.

2.1 The Fundamental Problem of Causal Inference

To understand the challenges that come with embedding ML in a causal framework for estimating causal effects, we first introduce the *potential outcomes* framework (Rubin 1974) for causal inference.[1] In this framework, each unit (e.g., a user on a ride-sharing platform) has a set of potential outcomes corresponding to each possible treatment condition, that is, interventions or actions that may affect an outcome. For simplicity, we focus on the case of binary treatments, where each unit either receives the treatment or does not. However, for any given unit, we can only observe one of the potential outcomes, namely the one corresponding to the actual treatment, while the counterfactual potential outcome under the alternative treatment remains unobserved. In the ride-sharing example, the treatment would be changing a certain app feature, with potential outcomes representing a user's weekly ride count with and without that feature change, respectively, only one of which is ever observable. This inherent limitation is known as the fundamental problem of causal inference (Holland 1986).

Because we cannot observe both potential outcomes for the same individual, we cannot directly estimate the individual-level causal effect, defined as the difference between the potential outcomes. Instead, researchers typically focus on population-level estimands such as the *Average Treatment Effect* (ATE), i.e., the mean causal effect for the population. Continuing the ride-sharing example, researchers may perform an experiment in which users are offered a new app feature at random to then compare the average ride count, thereby yielding the ATE that captures the average increase (or decrease) in weekly rides due to the new app feature. However, treatment effects often vary substantially across subpopulations. For example, the effect of feature changes on ride frequency may differ by age, location, and prior usage. To capture such heterogeneity, researchers may consider the Conditional Average Treatment Effect (CATE), an estimand that characterizes how the treatment effect varies with observed covariates. Incorporating a richer set of covariates allows estimation of treatment effects at finer levels of granularity, including subgroups or individuals.

Estimating the CATE from data is inherently challenging. Not only are counterfactual outcomes unobservable, but treatment assignment is often non-random and may depend on individual characteristics, introducing systematic bias (Gordon et al. 2023). For instance, in the ride-sharing example, low-frequency users may be less likely to receive the new app feature. As a result, simply comparing outcomes between treated and untreated groups may over- or underestimate the effect of the feature due to selection bias, as both groups differ in their ride behavior prior to the app change. Traditional ML methods typically do not consider such biases, as they solely focus on exploiting

---

[1] Alternative frameworks include *Structural Causal Models (SCM)* and *Directed Acyclic Graphs (DAGs)*. DAGs provide a graphical representation of causal relationships (cause-to-effect) between variables, whereas SCM quantify these relationship through structural equations that incorporate both causal mechanisms and random variation (Pearl 2009). While we focus on the potential outcomes framework, which is the most widely used foundation for modern causal ML methods (e.g., Wager and Athey 2018; Chernozhukov et al. 2018; Künzel et al. 2019), these frameworks are not mutually exclusive but rather complementary (Athey and Imbens 2017).

associative patterns in observed data to accurately predict outcomes. By contrast, causal inference requires careful consideration of how the data were generated and how a treatment was assigned. To draw causal conclusions from observed data, researchers must therefore make explicit assumptions about the data-generating process (Pearl 2009), which we detail in the following.

2.2 Causal Machine Learning is not Automatically Causal: Assumptions for Valid Causal Estimation

For valid ATE and CATE estimation, there are three standard assumptions in causal inference, in addition to the assumption of independent and identically distributed data (Imbens and Rubin 2015). Applying causal ML methods alone does not guarantee valid causal estimates. If any of the following assumptions are violated, the resulting estimated causal relationship may be biased. Following our previous example of a ride-sharing platform, we explain the three assumptions and illustrate how they may be violated in practice.

2.2.1 Stable Unit Treatment Value Assumption (SUTVA)

The *Stable Unit Treatment Value Assumption (SUTVA)* assumes no interference between units, that is, the treatment received by one unit does not affect the potential outcomes of other units. In our example this would imply that introducing the new app feature to one user has no impact on how frequently other users take rides. In practice, this assumption may be violated if the new app feature leads to a strong surge in ride demand, resulting in shortages or longer wait times that impact the experience of non-treated users. Additionally, SUTVA requires treatment consistency, meaning that the observed outcome for a user corresponds exactly to the treatment they received, with no hidden variation of the treatment. Consistency may be violated if some app users receive the new app feature in a different version, e.g., when the treated group see different variants of interface elements that could affect outcomes.

2.2.2 Overlap (Positivity)

The *overlap* (or *positivity*) assumption requires that every unit has a non-zero probability of receiving each treatment. In other words, for every combination of observed characteristics, there must be both treated and untreated individuals in the data. This ensures that we can make meaningful comparisons across groups. Violations occur when certain subpopulations always receive or never receive the treatment, often due to deterministic assignment rules or strict eligibility criteria. In the ride-sharing example, overlap is violated if certain user segments, say, new users who just signed up, are always given the new app feature, while long-term frequent riders are never eligible. In this case, we cannot estimate how the modification of the app would affect these groups, as we never observe the untreated or treated condition, respectively.

2.2.3 Unconfoundedness (Ignorability)

The last assumption is *unconfoundedness* (or also *ignorability*): the treatment assignment is independent of the potential outcomes, conditional on observed characteristics. This means that, once we account for the observed characteristics, there are no unmeasured variables (confounders) that jointly influence both the treatment assignment and the outcome. This assumption may be violated if the platform assigns the new app feature based on internal scores, such as predicted risk of churning, that are based on covariates not available in the dataset. In that case, even after adjusting for observable characteristics, treated and untreated users may differ systematically in motivation or usage intent, which affects the outcome. Such unmeasured confounding leads to biased treatment effect estimates.

2.3 Experimental versus Observational Data

Researchers can address these assumptions by conducting a careful and thought-through experiment or by leveraging properties of observational data. In randomized experiments, such as A/B tests commonly used in business practice, treatment assignment is independent of individual characteristics by design. As a consequence, unconfoundedness and overlap typically hold, which makes randomized experiments the gold standard for identifying treatment effects, and thus, for causal inference. While randomized experiments yield treatment effects with high internal validity, they are often conducted on smaller sample sizes or in a short time frame (e.g., due to opportunity costs or operational constraints), which may limit the generalizability of the findings to the broader population or longer time periods. Conversely, observational data accumulate larger and potentially richer datasets reflecting real-world behavior over a longer period. However, they potentially lack internal validity because treatment assignment depends on the unit's covariates and a (potentially unknown) deterministic mechanism (Van Goffrier et al. 2023; Cinelli et al. 2025). In particular, key assumptions such as unconfoundedness are more likely to be violated, and researchers must carefully assess the credibility of their identification strategy using domain knowledge and, where possible, tools such as causal diagrams or sensitivity analyses (Feuerriegel et al. 2024).

2.4 Causal ML Methods

There is a growing body of research on distinct methods for causal ML estimating causal estimands such as the ATE and CATE in a data-driven, flexible manner. While a comprehensive review of all methods is beyond the scope of this editorial, we provide an overview of widely adopted state-of-the-art methods supported by established software implementations to offer a practical starting point for IS researchers.

Causal ML algorithms for CATE estimation can be broadly grouped into two categories.[2] The first includes *model-based* approaches, which adapt existing ML algorithms to estimate the CATE directly by embedding them within a causal inference framework (see Table 1, upper panel). Most prominent examples include the causal forest (Wager and Athey 2018) and the Bayesian causal forest (Hahn et al. 2020), both of which extend the random forest to estimate heterogeneous effects. Similarly, TARNet (Shalit et al. 2017) uses neural networks to learn representations tailored for individual treatment effect estimation. Model-based approaches are typically accessible through off-the-shelf implementations, allowing researchers to estimate heterogeneous treatment effects in large-scale, high-dimensional settings with relatively little implementation effort. In addition, tree-based methods such as the causal forest offer a degree of interpretability by recursively partitioning the data into subgroups, thereby identifying the specific covariates that drive treatment heterogeneity.

The second category comprises *model-agnostic* algorithms, referred to as *meta-learners*, which decompose the CATE estimation task into several estimation steps that can be addressed using arbitrary ML methods (Künzel et al. 2019). These algorithms include plug-in learners such as the *S-learner* and *T-learner*, and more sophisticated two-step learners such as the *R-, X-* and *DR-learner* (Künzel et al. 2019; Nie and Wager 2021; Kennedy 2023). A complementary approach is provided by *Double/Debiased Machine Learning* (DML), which estimates average treatment effects (ATE) by combining flexible machine learning with semiparametric theory and orthogonalization techniques (Chernozhukov et al. 2018) and its core ideas also extend to CATE estimation (Shi et al. 2025). By separating models for outcome and treatment, respectively, these methods offer greater flexibility and transparency at the cost of additional modeling complexity. The key distinction among these learners lies in how they structure the estimation process (see Table 1, lower panel). Since no single meta-learner dominates across all settings, we refer readers to comparative studies and methodological tutorials for guidance on their performance under different data settings (see e.g., Shalit et al. 2017; Künzel et al. 2019; Curth and Van der Schaar 2021; Okasa 2022; Hu et al. 2024).

Several notable open-source software packages support the implementation of modern causal ML methods. In Python, the EconML library developed by Microsoft Research provides a wide range of meta-learners, DML estimators, forest-based models (including functions for interpretability), and more advanced methods relying on instrumental variables. The CausalML package offers additional functionality for uplift modeling and tree-based learners. In R, the grf package implements Causal Forests, while bartCause supports Bayesian Causal Forests. The DoubleML package is available in both R and Python and enables flexible DML estimation with support for cross-fitting and orthogonalization (Chernozhukov et al. 2018). These libraries collectively enhance the accessibility and reproducibility of ML-based causal inference for researchers.

3 Opportunities for Contributions in IS

As causal ML is constantly evolving and offers exciting opportunities for IS researchers, we argue that there are two main paths for the IS community to make significant contributions. While there may be other ways to engage with causal ML, we believe these two paths are particularly important: In Path I, scholars can leverage this novel method to both investigate causal questions of interest to IS theory and potentially make methodological contributions by adapting causal ML for the specific requirements of our field. In Path II, which addresses the increasing relevance of causal ML in practice, researchers can study its development, application, and impact within sociotechnical systems, utilizing the full spectrum of methodologies common in IS research. In the following sections, we provide guidance, inspiration (hopefully), and specific ideas for both paths.

3.1 Path I: Leveraging Causal ML for Theory Building and Methodological Advancement

The IS field is deeply engaged in theory building, a process for which identifying causal relationships is essential (Gregor 2006). IS researchers typically investigate these relationships to test existing theories, explore the boundary conditions of their failure, or generate new theoretical insights (Gupta et al. 2018). While randomized experiments are often considered the gold standard for such inquiries (Maruping et al. 2025), executing experiments

---

[2] The average treatment effect (ATE) can be obtained by aggregating individual- or group-level CATE estimates across the population.

Table 1 Overview of causal machine learning methods for estimating CATE

Model type	Method	Estimation strategy
Model-based	Tree-based methods (*Causal Forest, Bayesian Causal Forest*)	Extend random forests or Bayesian trees to estimate heterogeneous treatment effects. These models split the data based on covariates that affect treatment effect variation and aggregate effects across similar individuals to maximize heterogeneity in treatment effects. Causal Forests use an "honest" splitting strategy, separating data for tree-building and treatment effect estimation to reduce bias. They also provide valid confidence intervals for CATE estimates through residualization, orthogonalization, and asymptotic theory. Bayesian Causal Forests quantify uncertainty by modeling a posterior distribution over treatment effects
	Neural network models (e.g., *TARNet*)	Learn data representations that balance treated and control groups to address confounding. Then, they use these representations to predict treatment-specific outcomes, enabling flexible CATE estimation in high-dimensional data
Model-agnostic (Meta-learners)	Plug-in learners (*S-*, *T-learner*)	Use standard ML models to predict outcomes with and without treatment. S-learner trains one model including treatment as an additional feature; T-learner trains two separate models for treated and untreated groups. CATE is computed as the difference in predictions ("plug" predictions into the formula for computing treatment effect)
	Two-step learners (*X-*, *R-*, *DR-learner*)	Estimate nuisance functions first, such as outcome models and the propensity score, and then use these to construct pseudo-outcomes that isolate the treatment effect (through residualization and orthogonality). These pseudo-outcomes are used in a second stage to estimate the CATE. This structure allows for better handling of confounding, treatment imbalance, and model mis-specification, and thus, provides certain guarantees for robustness

can be impractical or ethically prohibited in many organizational settings. In these instances, IS researchers may make a contribution in line with Path I, that is, leverage causal ML as a powerful method for theory building. For example, while understanding the causal relationship between social media usage and filter bubbles is of great importance, conducting randomized field experiments in this sensitive domain may be ethically prohibited and, as a consequence, many identified relationships remain purely associative (Kitchens et al. 2020). In such cases, causal ML can facilitate theory building by flexibly accounting for high-dimensional confounders within large-scale social media data, thereby shedding light on the underlying causal mechanisms that were previously inaccessible.

In some settings, the assumptions required for causal inference may be violated due to the nature of the phenomenon or the structure of the data, which requires researchers to adapt the method of causal ML to be able to study a particular setting. For instance, Wang et al. (2022) extend the causal forest to incorporate instrumental variables for estimating heterogeneous treatment effects in observational studies with potential unobserved confounding factors. In another example, Turjeman and Feinberg (2024) extend the causal forest to study the effect of data breaches on behavioral changes of customers. Since data breaches received a lot of publicity, obvious control groups were missing, giving rise to the *temporal causal forest* that constructs meaningful control groups while addressing the assumptions for causal inference. Thus, Path I also encompasses the potential for methodological advancement, where researchers adapt or extend causal ML algorithms to navigate the unique challenges of IS and uncover causal effects that would otherwise remain hidden.

Building on our previous method section, IS researchers following Path I must rigorously evaluate the underlying assumptions of causal ML to avoid biased results. Feuerriegel et al. (2024) offer useful guidance for assessing assumption plausibility and performing sensitivity analyses and robustness checks. While this can be straightforward for experimental data, it is substantially more challenging for observational data. For example, a primary hurdle is the unconfoundedness assumption. Researchers should rely on tools like *Directed Acyclic Graphs (DAGs)* to reason about the underlying causal structure and consult domain knowledge to identify which variables must be included in the analysis to satisfy unconfoundedness (see, e.g., Tafti and Shmueli 2025) or, alternatively, utilize instrumental variables. Another example is the positivity assumption, for which researchers should inspect the distribution of propensity scores for extreme values near 0 or 1. If such values are present, specific subgroups may require exclusion to ensure reliable inference. Ultimately, when applying causal ML to observational data, IS researchers should transparently report the limitations alongside their findings.

Within Path I, IS researchers could leverage causal ML methods for a variety of research endeavors, for example:

- *Estimating causal relationships in complex settings where traditional methods fail:* In many IS contexts, such as the above-mentioned study of filter bubbles on social media, randomized experiments are often precluded by ethical or practical constraints, potential confounders are high-dimensional, and treatment heterogeneities may be complex and non-linear. In these scenarios, causal ML methods enable researchers to estimate nuanced causal effects across varying levels of granularity where traditional methods are limited.
- *Reanalyzing past experiments:* Given the capacity of causal ML to identify heterogeneous treatment effects, researchers can revisit archives of past A/B tests to systematically detect previously overlooked subgroup effects. Such "rescue" analyses can identify interventions that, while ineffective on average, are highly beneficial for specific segments, making them viable for personalized deployment. Additionally, by estimating CATEs and reweighting them to match different distributions, researchers can extrapolate results from historical experiments to new contexts.
- *Adapting causal ML to the unique requirements of IS research:* IS phenomena often present structural challenges that violate standard causal assumptions, such as interference between users on digital platforms (violating SUTVA) or systematic lack of access to IT artifacts for certain subgroups (violating overlap). There is a significant opportunity to develop specialized causal ML architectures tailored to these digital environments, similar to the above-mentioned adaptations of the causal forest, to navigate the specific dependencies and data-generating processes inherent in sociotechnical systems.

3.2 Path II: Studying causal ML as a Phenomenon in Sociotechnical Systems

With the core focus on the interaction between technology, individuals, and organizations, the IS field is uniquely positioned to make a contribution along Path II, that is, study causal ML as an emerging phenomenon. Specifically, IS researchers can provide insights into the development, application, and impact of causal ML "in the wild" using a plethora of methodologies, ranging from analytical modeling and behavioral experiments to case studies and design science. For example, behavioral experiments can reveal how human experts and managers integrate predictions from causal ML into their decision-making, while case studies may shed light on how organizations adapt their infrastructure and workflows to effectively deploy causal ML.

When making contributions along Path II, it is essential to establish the significance of the underlying problem and the novelty of the contribution relative to the existing body of knowledge in IS. Following Padmanabhan et al. (2022), authors should critically evaluate whether replacing the specific term "causal ML" with "ML" or even "Information Technology" alters the paper's core message. If not, it is likely that the findings mirror those of prior research, potentially rendering the contribution incremental. It is thus important for authors to reflect on what is unique and different with causal ML in a context and carefully adhere to the cumulative tradition of the field.

To provide inspiration and a starting point for IS researchers in Path II, we present several exemplary research directions in the following. We do not suggest that these are the only important directions. Instead, we present them as directions we view as potentially meaningful for research on causal ML in sociotechnical systems.

- *Designing highly personalized information systems:* Causal ML may enable systems that present managers with highly individualized causal effects. This capability raises fundamental questions about how such systems should be designed to support managerial decision making effectively. IS researchers have previously studied such questions in the context of traditional ML (e.g., Berger et al. 2021) and can build on a strong foundation documenting substantial heterogeneity in managers' responses to system design choices (e.g., Vessey and Galletta 1991), making IS researchers uniquely positioned to examine how estimated individualized effects should be presented, contextualized, and compared with aggregate views. Furthermore, researchers could investigate how existing infrastructure should be adapted to accommodate the integration of causal ML, for example, by facilitating automated A/B-testing to collect experimental training data. Drawing inspiration from the operational paradigms established for traditional ML (Kreuzberger et al. 2023), there is a clear need to define the principles of "CausalMLOps."
- *Human and organizational requirements for successful deployment:* Beyond technical and system design considerations, substantial uncertainty remains regarding how organizational processes, roles, and employee capabilities must evolve to support the effective deployment of causal ML. Similar to earlier transitions toward predictive ML (Lebovitz et al. 2021), managers may need to adapt their mental models and decision making as the focus shifts from interpreting predictions to reasoning about estimated causal effects at the individual or subgroup level. As demonstrated by prior work on such transitions (e.g., Jöhnk et al. 2021), this

uncertainty creates an opportunity for IS research to examine how organizational training, governance, and accountability structures must evolve to ensure that managers understand the scope, limitations, and appropriate use of causal ML in practice.

- *Unintended side effects on individuals:* With a predominant use of causal ML in practice being personalization (Feuerriegel et al. 2025), its role in shaping economic outcomes for individuals is growing. For instance, when organizations allocate interventions, such as targeted price discounts or feature changes on platforms or in apps, to customers based on the highest estimated causal effect on retention, those individuals may be systematically favored. IS researchers should investigate to what extent causal ML may trigger unintended side effects of moral and ethical nature, such as fairness and discrimination. In this regard, researchers can again build on previous work in the context of traditional ML (e.g., von Zahn et al. 2022) to provide managers with principled approaches to navigate such side effects and ensure that the deployment of causal ML aligns with broader societal and organizational values (Berente et al. 2021).

These exemplary research directions only scratch the surface of how causal ML can shape our discipline. As the methods continue to evolve, we encourage IS researchers to engage with this promising field and make distinct contributions to both theory and practice.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Athey S, Imbens GW (2017) The state of applied econometrics: causality and policy evaluation. J Econ Perspect 31(2):3–32

Berente N, Gu B, Recker J, Santhanam R (2021) Managing artificial intelligence. MIS Q 45(3):1433–1450

Berger B, Adam M, Rühr A, Benlian A (2021) Watch me improve — algorithm aversion and demonstrating the ability to learn. Bus Inf Syst Eng 63(1):55–68

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J (2018) Double/debiased machine learning for treatment and structural parameters. Econometrics J 21(1):C1–C68. https://doi.org/10.1111/ectj.12097

Cinelli C, Feller A, Imbens G, Kennedy E, Magliacane S, Zubizarreta J (2025) Challenges in statistics: a dozen challenges in causality and causal inference. arXiv preprint arXiv:250817099

Curth A, Van der Schaar M (2021) Nonparametric estimation of heterogeneous treatment effects: from theory to learning algorithms. In: International conference on artificial intelligence and statistics, PMLR, pp 1810–1818

Feuerriegel S, Frauen D, Melnychuk V, Schweisthal J, Hess K, Curth A, Bauer S, Kilbertus N, Kohane IS, van der Schaar M (2024) Causal machine learning for predicting treatment outcomes. Nat Med 30(4):958–968

Feuerriegel S, Shrestha YR, Von Krogh G (2025) A new machine learning approach answers what-if questions. MIT Sloan Manag Rev 66(3):65–69

Gordon BR, Moakler R, Zettelmeyer F (2023) Close enough? A large-scale exploration of non-experimental approaches to advertising measurement. Mark Sci 42(4):768–793

Gregor S (2006) The nature of theory in information systems. MIS Q 30(3):611–642

Gupta A, Kannan K, Sanyal P (2018) Economic experiments in information systems. MIS Q 42(2):595–606

Hahn PR, Murray JS, Carvalho CM (2020) Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects. Bayesian Anal 15(3):965–1056. https://doi.org/10.1214/19-BA1172

Holland PW (1986) Statistics and causal inference. J Am Stat Assoc 81(396):945–960. https://doi.org/10.1080/01621459.1986.10478354

Hu J, de Haan E, Skiera B (2024) Uplift modeling with quasi-loss-functions. Data Min Knowl Discov 38(4):2495–2519

Imbens GW, Rubin DB (2015) Causal Inference for statistics, social, and biomedical sciences: an introduction. Cambridge University Press

Jöhnk J, Weißert M, Wyrtki K (2021) Ready or not, AI comes — an interview study of organizational AI readiness factors. Bus Inf Syst Eng 63(1):5–20

Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. Science 349(6245):255–260

Kennedy EH (2023) Towards optimal doubly robust estimation of heterogeneous causal effects. Electron J Stat 17(2):3008–3049

Kitchens B, Johnson SL, Gray P (2020) Understanding echo chambers and filter bubbles: the impact of social media on diversification and partisan shifts in news consumption. MIS Q 44(4):1619–1649

Kreuzberger D, Kühl N, Hirschl S (2023) Machine learning operations (MLOps): overview, definition, and architecture. IEEE Access 11:31,866-31,879

Künzel SR, Sekhon JS, Bickel PJ, Yu B (2019) Metalearners for estimating heterogeneous treatment effects using machine learning. Proc Natl Acad Sci U S A 116(10):4156–4165. https://doi.org/10.1073/pnas.1804597116

Lebovitz S, Levina N, Lifshitz-Assaf H (2021) Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. MIS Q 45(3):1501–1526

Maruping L, Yin D, Chen A, Kankanhalli A, Burton-Jones A, Brown S (2025) Editor's comments: quantitative behavioral is research — a look back and a look forward. MIS Q 49(1):iii–xviii

Nie X, Wager S (2021) Quasi-oracle estimation of heterogeneous treatment effects. Biometrika 108(2):299–319. https://doi.org/10.1093/biomet/asaa076

Okasa G (2022) Meta-learners for estimation of causal effects: Finite sample cross-fit performance. https://arxiv.org/abs/2201.12692, 2201.12692

Padmanabhan B, Fang X, Sahoo N, Burton-Jones A (2022) Editor's comments: machine learning in information systems research. MIS Q 46(1):iii–xix

Pearl J (2009) Causality: models, reasoning, and inference, 2nd edn. Cambridge University Press

Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 66(5):688–701. https://doi.org/10.1037/h0037350

Shalit U, Johansson FD, Sontag D (2017) Estimating individual treatment effect: generalization bounds and algorithms. In: Proceedings of the 34th international conference on machine learning, vol 70, pp 3076–3085, https://proceedings.mlr.press/v70/shalit17a.html

Shi B, Mao X, Yang M, Li B (2025) What, why, and how: An empiricist's guide to double/debiased machine learning. Inf Syst Res, https://doi.org/10.1287/isre.2024.0888

Strittmatter A (2023) What is the value added by using causal machine learning methods in a welfare experiment evaluation? Labour Econ 84:102,412

Tafti A, Shmueli G (2025) Causal inference grounded in causal diagrams: benefits, limitations, and opportunities for the information systems field. MIS Q 49(2):409–428

Turjeman D, Feinberg FM (2024) When the data are out: measuring behavioral changes following a data breach. Mark Sci 43(2):440–461

Van Goffrier G, Maystre L, Gilligan-Lee CM (2023) Estimating long-term causal effects from short-term experiments and long-term observational data with unobserved confounding. In: Conference on causal learning and reasoning, PMLR, pp 791–813

Vessey I, Galletta D (1991) Cognitive fit: an empirical study of information acquisition. Inf Syst Res 2(1):63–84

von Zahn M, Feuerriegel S, Kuehl N (2022) The cost of fairness in AI: evidence from e-commerce. Bus Inf Syst Eng 64(3):335–348

von Zahn M, Bauer K, Mihale-Wilson C, Jagow J, Speicher M, Hinz O (2025) Smart green nudging: reducing product returns through digital footprints and causal machine learning. Mark Sci 44(4):954–969

Wager S, Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. J Am Stat Assoc 113(523):1228–1242

Wang G, Li J, Hopp WJ (2022) An instrumental variable forest approach for detecting heterogeneous treatment effects in observational studies. Manag Sci 68(5):3399–3418