REVIEW Open access

Causality in Materials Informatics — Conceptual Progress, Limitations, and Future Directions

Nguyen Thanh Huy[1*], Le Thi Bich[2], Pham Quang Minh[1]

Abstract

Materials informatics has emerged as a central paradigm in contemporary materials science, leveraging machine learning and data-driven modeling to accelerate materials discovery, optimization, and deployment. Despite substantial advances in predictive accuracy, most existing approaches remain fundamentally correlational, limiting their reliability under distribution shifts, experimental interventions, and real-world deployment scenarios. This reliance on correlation constrains scientific interpretability and undermines the capacity of AI systems to function as genuine instruments of materials reasoning. Causality offers a principled framework for overcoming these limitations by explicitly modeling cause-and-effect relationships among composition, processing, structure, and properties. This narrative review synthesizes conceptual progress in integrating causal inference into materials informatics, examining foundational causal frameworks, advances in causal discovery, and hybrid causal–machine learning approaches, and emerging applications across materials domains such as nanocatalysis, ferroelectrics, and electrochemical energy storage. We critically analyze persistent challenges—including data scarcity, assumption violations, limited external validity, and computational and epistemic constraints—that currently hinder widespread adoption. Drawing exclusively on peer-reviewed literature published, the review emphasizes thematic and epistemic developments rather than algorithmic prescriptions. We argue that causality represents a structural shift in how AI systems contribute to materials science: from correlational predictors to intervention-aware, mechanism-aligned reasoning tools. By articulating future directions centered on hybrid modeling, domain-knowledge integration, and interdisciplinary collaboration, this review positions causality as a necessary foundation for robust, generalizable, and scientifically legitimate materials informatics.

Keywords Materials informatics, Data-driven discovery, Machine learning, Causal inference, Causality, Material design

*Correspondence:
Nguyen Thanh Huy
huy.nguyen@gmail.com
[1] Department of Materials Science and Data Engineering, Faculty of Engineering, Vietnam National University, Hanoi, Vietnam
[2] Department of Artificial Intelligence Systems, Faculty of Engineering, Can Tho University, Can Tho, Vietnam

Introduction

Materials informatics has emerged as a transformative paradigm at the intersection of materials science, data science, and computational modeling, fundamentally reshaping how materials are discovered, characterized, and optimized. By leveraging large, heterogeneous datasets alongside advanced machine learning (ML) and artificial intelligence (AI) techniques, the field seeks to dramatically accelerate the traditionally slow and resource-intensive materials development cycle [1-5]. The formal launch of the materials genome initiative in 2011 marked a decisive inflection point, catalyzing community-wide investments in high-throughput experimentation, large-scale simulations, open materials databases, and data-driven modeling infrastructures [6-8]. Since then, predictive ML models for properties such as band gaps,

mechanical strength, phase stability, and catalytic activity have become routine components of modern materials research workflows.

This data-centric shift represents a profound departure from classical materials science, which historically relied on iterative trial-and-error experimentation guided by physical intuition and computationally expensive first-principles simulations. The growing availability of curated experimental datasets, high-throughput density functional theory (DFT) repositories, and text-mined knowledge from the scientific literature has enabled models that can rapidly interpolate and extrapolate across vast compositional and structural spaces [6, 9, 10]. As a result, materials informatics has delivered impressive gains in predictive accuracy and screening efficiency, often identifying promising candidates orders of magnitude faster than conventional approaches.

Yet, despite these advances, a foundational epistemic limitation persists: the majority of ML models in materials informatics are optimized to learn statistical correlations rather than genuine cause-and-effect relationships [11-14]. While correlational models can achieve high predictive performance within the bounds of their training data, they remain fragile under distribution shifts, extrapolation, or intentional interventions—conditions intrinsic to real-world materials design and processing [10, 15, 16]. Spurious correlations, hidden confounders, and dataset-specific biases can yield predictions that are accurate yet scientifically misleading, offering little guidance for mechanism-driven understanding or rational design.

Causality offers a principled framework to address these shortcomings by explicitly modeling how changes in one variable bring about changes in another [17-20]. In the context of materials science, causal reasoning is central to the discipline's explanatory goals: processing conditions cause microstructural evolution; composition causes phase formation; defects cause changes in electronic, mechanical, or catalytic properties [9, 15, 16, 21]. Unlike purely correlational models, causal models support counterfactual reasoning ("what would happen if…?"), intervention analysis, and robustness under changing conditions—capabilities that align closely with the needs of materials discovery, optimization, and scale-up.

Recent work has begun to explore how causal inference concepts—such as causal graphs, interventions, and structural assumptions—can be integrated into materials informatics pipelines. For example, in nanocatalysis, disentangling the causal effects of synthesis parameters from confounding structural descriptors enables more targeted catalyst design and more reliable performance optimization [15]. More broadly, causal perspectives promise to elevate AI systems from high-accuracy predictors to scientifically meaningful reasoning tools that can support hypothesis generation, experimental planning, and mechanistic insight. The conceptual distinction between correlational and causal paradigms in materials informatics is summarized in **Table 1**.

Table 1. Correlational vs. causal paradigms in materials informatics

Dimension	Correlational ML	Causal- driven materials informatics
Primary objective	Predictive accuracy	Intervention-aware understanding
Core assumption	Statistical association suffices	Explicit cause–effect structure
Robustness to distribution shift	Limited	Improved via invariance
Interpretability	Post-hoc, often fragile	Mechanism-aligned
Supports counterfactuals	No	Yes
Role in materials design	Screening and ranking	Rational intervention and optimization
Scientific legitimacy	Instrumental	Epistemic

Against this backdrop, the present review critically examines the emerging role of causality in materials informatics. The objectives are threefold: (1) to synthesize and clarify the conceptual progress made in applying causal reasoning to data-

driven materials research, (2) to identify the theoretical, methodological, and practical limitations that currently constrain this integration, and (3) to outline future directions capable of bridging the gap between correlational prediction and causal understanding. The review is deliberately scoped to peer-reviewed journal articles published. It is organized thematically rather than chronologically or by algorithmic class, reflecting the view that causality in materials informatics is primarily an epistemic and conceptual challenge rather than a purely technical one.

Fundamentals of causality in materials informatics

Causality occupies a foundational position in scientific reasoning, providing the conceptual machinery required not only to predict outcomes but to understand *why* those outcomes occur and how they would change under deliberate intervention [14, 17, 19]. Unlike purely statistical associations, causal relationships enable counterfactual reasoning—asking what would happen if a system were altered in a specific way—which is essential for mechanism-driven discovery and rational design. In materials informatics, this distinction is particularly consequential. The central scientific questions of the field are inherently causal: how does a change in composition induce a change in phase stability, how do processing conditions shape microstructure, and how do structural motifs give rise to functional properties such as strength, conductivity, or catalytic activity [10, 15, 16].

Most contemporary materials informatics pipelines, however, are built upon machine learning models optimized for predictive accuracy rather than causal validity. Deep neural networks, kernel methods, and ensemble models are highly effective at identifying complex, nonlinear patterns in high-dimensional data, yet they operate largely as correlational engines [11-13]. As a result, they may inadvertently exploit spurious correlations, dataset-specific regularities, or latent confounders that do not reflect true physical dependencies. Such models often perform well in interpolation regimes but degrade sharply when confronted with extrapolation, domain shift, or process interventions—precisely the regimes that define materials innovation and scale-up [10, 15, 16].

Causal inference frameworks offer a principled alternative by explicitly formalizing the relationship between interventions and outcomes. One influential perspective is the potential outcomes framework, originally articulated by Merchant *et al.* [14]. This framework conceptualizes causality in terms of counterfactuals: for each unit—such as a material sample—there exists a set of potential outcomes corresponding to different treatments or interventions. In a materials context, this translates to questions such as: What would the mechanical strength of this alloy have been if a different composition or heat-treatment protocol had been applied [13, 16]? While only one outcome is observed for each sample, causal inference seeks to estimate the unobserved alternatives under well-defined assumptions.

Complementary to this view are structural causal models (SCMs), most prominently associated with Eigenmann *et al.* [17]. SCMs represent causal relationships through directed acyclic graphs (DAGs), where nodes correspond to variables—composition, processing parameters, microstructural descriptors, properties—and directed edges encode assumed causal directions. This graphical formalism enables transparent reasoning about confounding, mediation, and intervention effects, offering a bridge between data-driven learning and physical intuition [11, 21, 22]. In materials science, SCMs have been used to disentangle competing mechanisms in complex systems; for example, causal analysis of microscopy-derived data in ferroelectric materials has revealed distinct atomistic pathways governing functional responses, even when observational correlations alone proved ambiguous [10].

For causal effects to be identifiable from data, key assumptions must be satisfied, notably conditional exchangeability and positivity [14, 17, 18]. Conditional exchangeability requires that, given a suitable set of observed covariates, the treatment assignment (e.g., choice of processing condition) is effectively independent of the potential outcomes. Positivity requires that all relevant treatments occur with nonzero probability across the covariate space. In materials research, where randomized experiments are often infeasible, and datasets are observational by construction, these assumptions are rarely guaranteed. Nevertheless, they provide a conceptual benchmark that motivates the careful integration of domain knowledge, experimental design principles, and statistical adjustment techniques—such as propensity score matching—to approximate causal estimation [22]. Recent syntheses have emphasized that machine learning should be viewed not

as a replacement for causal reasoning, but as an enabling tool that can support causal inference when guided by strong scientific priors [11, 13]. The principal causal frameworks relevant to materials informatics and their conceptual roles are summarized in **Table 2**.

Table 2. Conceptual causal frameworks and their roles in materials informatics

Framework	Core concept	Materials-specific interpretation	Primary contribution
Potential outcomes	Counterfactual reasoning	Property under alternative composition/process	Intervention evaluation
Structural causal models (SCMs)	DAG-based causation	Process → structure → property pathways	Confounding control
Causal discovery	Structure inference from data	Hidden dependencies in synthesis pipelines	Hypothesis generation
Stable/invariant learning	Environment-invariant predictors	Robust descriptors across datasets	Generalization
Bayesian causal inference	Probabilistic causation	Uncertainty-aware causal effects	Risk-sensitive design

Conceptual progress in causal discovery methods

Causal discovery—the task of inferring causal structure directly from data—has emerged as a critical frontier in materials informatics, particularly for systems with complex, multiscale, or only partially understood governing mechanisms [11, 21, 22]. Rather than prespecifying causal graphs, causal discovery methods aim to uncover plausible causal relationships by combining statistical tests, optimization principles, and, increasingly, machine learning techniques [9, 10, 6].

Constraint-based approaches represent one major class of methods. These algorithms exploit patterns of conditional independence in the data to progressively construct a DAG consistent with the observed dependencies [21]. In manufacturing and materials processing contexts, structured reviews have shown that such approaches can reveal process–property linkages that remain opaque to purely correlational analyses, thereby enhancing interpretability and process control [21]. However, their performance is sensitive to sample size, noise, and violations of underlying assumptions—conditions frequently encountered in real-world materials datasets.

Score-based methods offer an alternative formulation, framing causal discovery as an optimization problem over the space of possible DAGs [10, 15]. By assigning scores that balance data fit and model complexity, these methods search for structures that best explain the observed data. Recent developments have improved scalability and differentiability, enabling applications to work with higher-dimensional materials datasets. Such methods have been used to analyze relationships among synthesis parameters, structural descriptors, and functional outcomes, yielding causal hypotheses that can be experimentally tested [10, 15].

Hybrid strategies that integrate domain knowledge with data-driven discovery are increasingly viewed as the most promising pathway forward [10, 15, 16]. In materials informatics, prior physical constraints—such as known thermodynamic dependencies or processing hierarchies—can be encoded to restrict the space of admissible causal graphs. This integration has proven particularly effective in nanocatalysis, where causal inference has been used to extract interpretable cause–effect relationships between synthesis conditions and emergent structural features, guiding more targeted experimental exploration [15]. Similarly, causal machine learning applied to nanoparticle synthesis has

enabled data-efficient discovery by identifying which experimental interventions are most likely to yield desired outcomes [16].

Beyond structure learning, recent conceptual advances have focused on robustness and uncertainty. Stable learning approaches aim to identify causal features that remain invariant across environments, thereby improving generalization under distribution shifts [12]. Bayesian causal methods further extend this framework by providing principled uncertainty quantification over causal estimates, an essential capability for high-stakes materials design decisions [19]. Collectively, these developments signal a shift toward viewing causality not as an auxiliary add-on to materials informatics, but as a core organizing principle capable of transforming predictive models into reliable scientific instruments [9, 10].

Applications of causality in materials design

The incorporation of causal reasoning into materials informatics has begun to move the field beyond high-accuracy prediction toward actionable scientific insight, with tangible implications for materials design and discovery [3, 4, 10, 15, 16, 23]. Across diverse material classes and application domains, causal methods have demonstrated their capacity to disentangle complex dependencies, prioritize meaningful interventions, and support rational design strategies that are difficult to achieve through correlational modeling alone.

One of the most illustrative application domains is nanocatalysis, where performance outcomes often emerge from tightly coupled interactions among synthesis conditions, nanoscale structure, and surface chemistry. In this context, causal inference has been used to explicitly identify how specific processing parameters give rise to distinct structural motifs, which in turn causally determine catalytic activity and selectivity [15]. By separating genuine cause–effect relationships from incidental correlations, such analyses enable more targeted catalyst optimization and reduce reliance on exhaustive trial-and-error experimentation.

Similarly, in ferroelectric materials, causal analysis applied to high-resolution microscopy and functional data has provided a principled means to order and discriminate between competing atomistic mechanisms [10]. Rather than attributing observed behavior to a mixture of correlated descriptors, causal frameworks allow researchers to identify which mechanisms are necessary, sufficient, or merely incidental, thereby sharpening mechanistic interpretation and informing materials engineering decisions.

Electrochemical energy storage offers another compelling case. In battery research, AI and ML methods have been widely adopted for property prediction, lifetime estimation, and materials screening [3]. Recent reviews emphasize that embedding causal elements—such as explicitly modeling degradation pathways or intervention effects of operating conditions—can substantially improve the interpretability and robustness of these models. This is particularly important for batteries, where deployment environments differ markedly from laboratory conditions, and where correlational predictors often fail under real-world stressors.

Causality has also begun to influence the design of generative models for materials discovery. Generative frameworks augmented with causal structure have been shown to improve extrapolation beyond observed data regimes, addressing a key limitation of conventional deep generative models [23]. By constraining generation according to causal relationships —rather than unconstrained statistical patterns—such approaches offer a pathway toward more physically plausible and design-relevant candidate materials. More broadly, retrospective analyses of machine learning developments in materials science over the past decade indicate a gradual but discernible shift toward causal thinking to achieve better generalization and scientific reliability [4].

At the ecosystem level, bibliometric and meta-analytical studies document a steady increase in the integration of AI across materials engineering, with causal inference emerging as a prominent and growing thematic trend [2]. Parallel advances in materials representations—such as graph-based encodings of crystal structures and chemically informed descriptors—have further facilitated causal modeling by aligning data representations with physically meaningful variables

[8]. Notably, conceptual insights from adjacent domains, including healthcare and policy analytics, where causal models underpin actionable predictions, have increasingly been recognized as transferable to materials informatics, offering methodological inspiration for intervention-driven materials design [13].

Challenges and limitations in current approaches

Despite these promising developments, the application of causality in materials informatics remains constrained by several conceptual, methodological, and practical limitations [11–14, 19]. Chief among these is data scarcity and data quality. Many materials datasets are small, heterogeneous, or biased toward successful experiments, limiting the statistical power required for reliable causal discovery [3, 9, 11]. Unlike controlled clinical or industrial trials, materials data are rarely generated under randomized conditions, which exacerbates confounding and complicates causal identification.

The predominance of observational data poses a fundamental challenge to key causal assumptions, particularly conditional exchangeability and positivity [14, 17, 18]. In materials research, choices of composition, processing routes, or characterization techniques are often guided by expert intuition or prior results, leading to systematic selection biases. As a result, causal estimation often relies on strong, sometimes unverifiable assumptions, raising concerns about the stability and credibility of inferred effects.

Model complexity introduces additional tensions. While causal machine learning models can capture rich, nonlinear dependencies, increasing complexity risks overfitting, reduced transparency, and diminished interpretability—ironically undermining one of the primary motivations for adopting causal approaches [12, 13, 19]. Experiences from the social and health sciences underscore that ML-based causal inference demands careful validation, sensitivity analysis, and explicit articulation of assumptions to avoid misleading conclusions [11].

External validity remains another unresolved concern. Even when causal effects are credibly identified within a given dataset, their transferability across material systems, processing regimes, or environmental conditions is not guaranteed [17, 18]. Materials systems are often highly context-dependent, and causal mechanisms operative in one compositional or thermodynamic regime may not persist elsewhere. This limitation reinforces the need for tight integration between causal models and domain knowledge, yet encoding such knowledge into automated or semi-automated systems remains technically and conceptually challenging [11, 13]. Key limitations and conceptual mitigation strategies are synthesized in **Table 3**.

Table 3. Limitations, failure modes, and mitigation strategies in causal materials informatics

Limitation	Manifestation in materials data	Consequence	Conceptual mitigation
Data scarcity	Sparse, biased datasets	Unstable causal estimates	Hybrid modeling + priors
Confounding	Non-random experiment design	Spurious causation	Explicit DAG construction
Model complexity	Opaque causal ML models	Loss of interpretability	Structural constraints
Limited external validity	System-specific mechanisms	Poor transferability	Invariance-based learning
Reproducibility	Pipeline sensitivity	Epistemic fragility	Transparent reporting

Finally, ethical and practical considerations shape adoption. Causal analyses can be computationally demanding, particularly when combined with high-dimensional representations or Bayesian uncertainty quantification [3, 4]. Reproducibility is also a concern, as causal conclusions may depend sensitively on modeling choices, variable definitions, or preprocessing decisions [8]. Together, these challenges suggest that while causality holds substantial promise for advancing materials informatics, its successful integration requires not only algorithmic innovation but also careful epistemic discipline, transparent reporting, and sustained collaboration between data scientists and materials experts.

Results and Discussion

The integration of causality into materials informatics represents a paradigm shift from correlation-based machine learning to a more mechanistic understanding of material behaviors and properties [11, 13-15, 18, 21, 22, 24]. As outlined in the main text, conceptual progress has been marked by the adoption of causal discovery methods, such as constraint-based and score-based algorithms, which have proven effective at uncovering process-structure-property relationships in complex systems, such as nanocatalysts and ferroelectric materials [10, 15, 16]. These advances address the limitations of traditional predictive models, which often fail to generalize due to spurious correlations arising from confounding variables or shifts in data distribution [11-13]. For instance, in high-dimensional materials data, causal inference techniques like propensity score matching and structural causal models (SCMs) enable the identification of true causal effects, thereby facilitating more reliable predictions and interventions [14, 17, 21, 22]. **Figure 1** conceptually summarizes the transition from correlational materials AI to causality-driven reasoning systems.

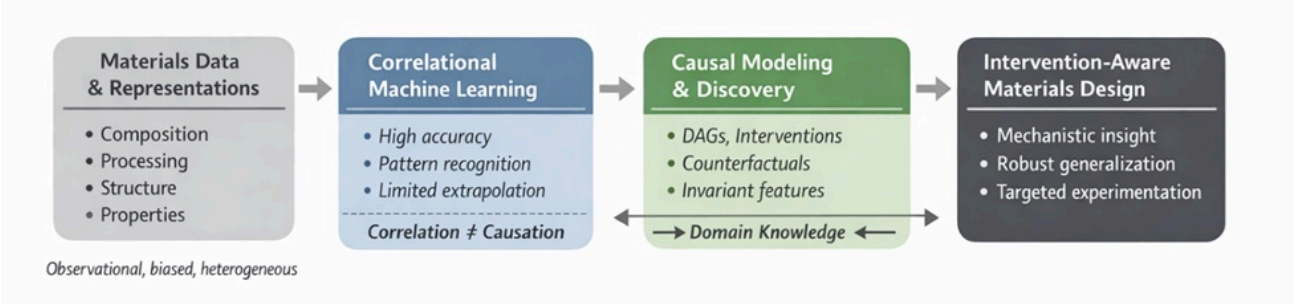

Figure 1. Causal materials informatics pipeline: from correlation to intervention-aware reasoning

However, applying causality in materials informatics is not without challenges. One key issue is reliance on observational data, which predominates in materials science due to the high cost and time required for experimental interventions [3, 9, 11]. Observational datasets often violate key assumptions of causal inference, such as conditional exchangeability and the absence of unmeasured confounding, leading to biased estimates [17-19]. This is particularly evident in manufacturing contexts, where a structured literature review highlighted that causal discovery enhances process optimization but is limited by data quality and the need for domain expertise to validate discovered graphs [21]. Furthermore, the complexity of materials systems—characterized by multiscale interactions from atomic to macroscopic levels—complicates the construction of accurate DAGs [8, 10, 13]. In battery materials research, for example, causal models have improved property predictions, but scalability to large datasets remains a hurdle, as machine learning models can overfit or produce non-interpretable results [3, 4].

Another critical discussion point is the interplay between causality and machine learning in materials design. Hybrid approaches that combine causal inference with deep learning, such as causal ML or Bayesian networks, offer promising avenues for robust generalization [12, 19, 20]. These methods incorporate uncertainty quantification, which is essential for high-stakes applications such as material failure prediction in the aerospace and energy sectors [2, 23]. Yet, ethical considerations arise, particularly regarding the reproducibility of causal findings and the potential biases in training data [3, 11, 13]. For instance, if datasets are skewed toward certain material classes (e.g., metals over polymers), causal discoveries may not transfer across domains, limiting the field's inclusivity [5, 8]. Bibliometric analyses reveal a growing trend toward AI integration in materials engineering, with causality emerging as a niche yet impactful area [2, 7]. This suggests that while progress is evident, broader adoption requires standardized benchmarks and interdisciplinary collaboration between materials scientists, statisticians, and computer scientists [4, 6].

The limitations identified—data scarcity, assumption violations, and computational demands—underscore the need for cautious interpretation of causal results [9, 11, 12]. In practice, causal models should be validated through sensitivity analyses and experimental confirmation, as seen in nanoparticle synthesis, where causal ML-guided experiments were required but iterative refinement was needed [16]. Moreover, the field's emphasis on high-level overviews rather than

actionable details aligns with the goal of conceptual advancement, avoiding the pitfalls of over-prescription in a rapidly evolving domain [10, 15, 17]. Overall, the discussion highlights that causality not only enhances predictive accuracy but also fosters mechanistic insights, paving the way for accelerated discovery of materials for sustainable technologies [3, 23, 25].

Conclusion

This narrative review demonstrates that causality has progressed from a peripheral concept to a foundational epistemic framework within materials informatics. By moving beyond correlational machine learning toward causal reasoning, the field is beginning to address long-standing challenges related to robustness, interpretability, and scientific legitimacy. Conceptual advances in causal inference frameworks, causal discovery, and hybrid causal–machine learning approaches have enabled more reliable interpretation of process–structure–property relationships and have shown tangible value across materials domains such as nanocatalysis, ferroelectrics, and energy storage systems.

At the same time, this review highlights that causality in materials informatics remains constrained by structural limitations, including data scarcity, pervasive confounding, limited external validity, and computational and epistemic complexity. These challenges underscore that causal modeling is not a purely algorithmic upgrade, but a shift in how materials knowledge is represented, validated, and operationalized. Without careful integration of domain knowledge, transparent assumptions, and rigorous validation, causal claims risk reproducing the very fragilities they seek to overcome.

Looking forward, several strategic directions emerge. First, hybrid frameworks that integrate causal inference with advanced AI—such as physics-informed learning, generative modeling, and uncertainty-aware architectures—offer a promising path for handling multiscale materials phenomena. Second, interdisciplinary collaboration among materials scientists, statisticians, and machine learning researchers is essential for developing standardized benchmarks, shared datasets, and evaluation protocols tailored to causal questions. Third, extending causal materials informatics into emerging domains—including sustainable materials, climate-critical technologies, and quantum materials—presents an opportunity for high-impact discovery aligned with global scientific priorities.

Ultimately, embedding causality within materials informatics reframes the role of AI in materials science: not merely as a predictor of outcomes, but as a partner in scientific reasoning. By enabling intervention-aware, mechanism-aligned insights, causality lays the groundwork for AI systems that are not only accurate but trustworthy, generalizable, and scientifically meaningful.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 01 May 2024 Revised: 01 Jun 2024 Accepted: 10 Jul 2024
Published online: 18 January 2025

References

Jha D, Gupta V, Liao W-K, Choudhary A, Agrawal A. Towards physics-informed explainable machine learning and causal models for materials research. Comput Mater Sci. 2023;233:112735.
https://doi.org/10.1016/j.commatsci.2023.112735.

Schölkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, et al. Toward causal representation learning. Proc IEEE. 2021;109(5):612-34.
https://doi.org/10.1109/JPROC.2021.3058954.

Naser MZ. Causality and causal inference for engineers: Beyond correlation, regression, prediction and artificial intelligence. Wiley Interdiscip Rev Data Min Knowl Discov. 2024;14(1):e1533.
https://doi.org/10.1002/widm.1533.

Pei Z, Yin J, Liaw PK, Raabe D. Toward the design of ultrahigh-entropy alloys via mining six million texts. Nat Commun. 2023;14(1):54.
https://doi.org/10.1038/s41467-022-35719-4.

Zheng G, Ivanov D, Brintrup A. An adaptive federated learning system for information sharing in supply chains. Int J Prod Res. 2024;62(15):1-23.
https://doi.org/10.1080/00207543.2024.2432469.

Pyzer-Knapp EO, Pitera JW, Staar PWJ, Takeda S, Laino T, Sanders DP, et al. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. NPJ Comput Mater. 2022;8(1):84.
https://doi.org/10.1038/s41524-022-00765-z.

Vowels MJ, Camgoz NC, Bowden R. D'ya like DAGs? A survey on structure learning and structure validation. ACM Comput Surv. 2022;54(4):82.
https://doi.org/10.1145/3444948.

Bran AM, Cox S, Schilter O, Baldassari C, White AD, Schwaller P. ChemCrow: Augmenting large-language models with chemistry tools. arXiv preprint arXiv:2304.05376. 2023.

Ting JYC, Barnard AS. Data-driven causal inference of process-structure relationships in nanocatalysis. Curr Opin Chem Eng. 2022;36:100802.
https://doi.org/10.1016/j.coche.2022.100802.

Hund L, Schroeder B. A causal perspective on reliability assessment. Reliab Eng Syst Saf. 2020;195:106678. https://doi.org/10.1016/j.ress.2019.106678.

Lopez R, Hütter J-C, Pritchard J, Regev A. Large-scale differentiable causal discovery of factor graphs. Adv Neural Inf Process Syst. 2022;35:19290-303.

Batra R, Song L, Ramprasad R. Emerging materials intelligence ecosystems propelled by machine learning. Nat Rev Mater. 2021;6(8):655-78. https://doi.org/10.1038/s41578-020-00255-y.

Von Lilienfeld OA, Müller K-R, Tkatchenko A. Exploring chemical space with machine learning. Nat Rev Chem. 2020;4(7):347-58. https://doi.org/10.1038/s41570-020-0183-z.

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624(7990):80-5. https://doi.org/10.1038/s41586-023-06735-9.

Vukovic M, Thalmann S, De Wael M, Van Assche F. Causal Discovery in Manufacturing: A Structured Literature Review. J Manuf Mater Process. 2022;6(1):10. https://doi.org/10.3390/jmmp6010010.

Cioffi R, Travaglioni M, Piscitelli G, Petrillo A, De Felice F. Artificial Intelligence and Machine Learning Applications in Smart Production: Progress, Trends, and Directions. Sustainability. 2020;12(2):492. https://doi.org/10.3390/su12020492.

Eigenmann M, Mukherjee S, Maathuis M. Evaluation of causal structure learning algorithms via risk estimation. Proc 36th Conference of Uncertainty in Artificial Intelligence. 2020;124:151-60.

Guo R, Cheng L, Li J, Hahn PR, Liu H. A survey on causal inference. ACM Trans Knowl Discov Data. 2021;15(5):74.

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):1-46.

Hasan SU, Gani MO. A survey on causal discovery methods for temporal and non-temporal data. Trans Mach Learn Res. 2023.

Feng K, Wang W, Hu Y, He J, Zhang L, Zheng W, et al. Deep learning of causal structures in high dimensions under data limitations. Nat Mach Intell. 2023;5(10):1156-67. https://doi.org/10.1038/s42256-023-00744-z.

Brouillard P, Lachapelle S, Lacoste A, Lacoste-Julien S, Drouin A. Differentiable causal discovery from interventional data. Adv Neural Inf Process Syst. 2020;33:21865-77.

Zhang Z, Liu J, Ding T, Tang M, Chen K, Pan R. A Survey on Causal Inference for Recommendation. Innovation. 2023;4(3):100461. https://doi.org/10.1016/j.xinn.2023.100461.

Zhou Y, Chen H. A survey on causal discovery. arXiv preprint arXiv:2211.11164. 2022.

Xu L, Mak S, Minaricova M, Brintrup A. On implementing autonomous supply chains: a multi-agent system approach. Comput Ind. 2024;161:104120. https://doi.org/10.1016/j.compind.2024.104120.