Applied Physics Letters ARTICLE pubs.aip.org/aip/apl

Causal discovery from data assisted by large language models

Cite as: Appl. Phys. Lett. **127**, 121904 (2025); doi: 10.1063/5.0272287
Submitted: 23 March 2025 · Accepted: 9 September 2025 ·
Published Online: 24 September 2025

Kamyar Barakati,[1,a)] Aleksander Molak,[2] Christopher T. Nelson,[3] Xiaohang Zhang,[4] Ichiro Takeuchi,[4] and Sergei V. Kalinin[1,5,a)]

AFFILIATIONS

[1]Department of Materials Science and Engineering, University of Tennessee, Knoxville, Tennessee 37996, USA
[2]CausalPython.io, Warsaw, Poland
[3]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA
[4]Department of Materials Science and Engineering, University of Maryland, College Park, Maryland 20742, USA
[5]Pacific Northwest National Laboratory, Richland, Washington 99354, USA

[a)]**Authors to whom correspondence should be addressed:** kbarakat@vols.utk.edu and sergei2@utk.edu

ABSTRACT

Knowledge-driven discovery of novel materials necessitates the development of causal models for property emergence. While in the classical physical paradigm, the causal relationships are deduced based on physical principles or via experiment, the rapid accumulation of observational data necessitates learning causal relationships between dissimilar aspects of material structure and functionalities based on observations. For this, it is essential to integrate experimental data with prior domain knowledge. Here, we demonstrate this approach by combining high-resolution scanning transmission electron microscopy data with insights derived from large language models (LLMs). By applying ChatGPT to domain-specific literature, such as arXiv papers on ferroelectrics, and combining the obtained information with data-driven causal discovery, we construct adjacency matrices for directed acyclic graphs that map the causal relationships between structural, chemical, and polarization degrees of freedom in Sm-doped $BiFeO_3$. This approach enables us to hypothesize how synthesis conditions influence material properties and guides experimental validation. The ultimate objective of this work is to develop a unified framework that integrates LLM-driven literature analysis with data-driven discovery, facilitating the precise engineering of ferroelectric materials by establishing clear connections between synthesis conditions and their resulting material properties.

The functionalities of solids and their structure–property relationships are governed by a complex interaction between chemical and physical degrees of freedom, which are, in turn, controlled by synthesis conditions. These effects are often interconnected through complex feedback loops that arise during the coupled dynamics of structure evolution, yielding a network of interdependencies that are critical in determining the material's final properties. Very often, processing control variables—such as temperature, pressure, and chemical environment and their process histories—differ from those that directly govern local functionalities, such as atomic-scale and extended defects, phase distributions, strain, concentration, and physical order parameter fields. Consequently, these critical local properties can only be influenced indirectly, complicating efforts to precisely control material behavior through synthesis.[1,2]

A prime example of these complexities is found in ferroelectric materials, where chemical disorder can significantly impact polarization fields. The paradigmatic effect here will be domain nucleation, pinning, field, and strain controlled dynamics of domain walls, all affected by the impurities and structural defects.[3–5] However, discontinuities in the polarization field can simultaneously drive chemical segregation, leading to memory effects that cause the material to "remember" its previous states. This bidirectional interaction between chemical and physical properties underscores the challenge of controlling material behavior, as interventions on one degree of freedom can lead to unintended consequences in another, complicating the control and optimization of ferroelectric properties.[6,7]

Understanding and discovering the causal mechanisms is essential for making accurate interventions and developing counterfactual strategies that can predict and manipulate material behavior. Causal models provide a pathway to go beyond the limitations of purely data-driven methods, offering insights that can guide more effective material design.[8] While experiments that involve direct interventions on

specific degrees of freedom best answer *interventional* causal questions—those examining the immediate impact of one variable on another—the number of observable variables often surpasses the scope of feasible experimental controls.[9,10] This disparity necessitates advanced approaches to causal discovery and—sometimes—addressing counterfactual questions that cannot be resolved through direct experimentation alone.[11–13]

Previously, we have explored the causal relationships in ferroelectric materials using the pairwise causal models[14] and linear non-Gaussian acyclic model (LiNGAM),[15] both integrated with physics-based analyses to incorporate domain-specific insights. However, these methods face limitations when applied to complex systems, as they rely solely on observational data, and implementations used in prior work lacked the ability to incorporate prior domain knowledge effectively.

Here, we address these challenges by integrating large language models (LLMs)[16] into the causal discovery process. By using LLMs to analyze domain-specific literature, such as arXiv papers on ferroelectric materials, we improve the accuracy and interpretability of causal models. This approach systematically links synthesis conditions to material properties by combining literature insights with experimental data, providing a more informed framework for the design and optimization of ferroelectric materials.

As a model system, we selected the previously studied multiferroic material $BiFeO_3$ (BFO), modified through Sm substitution at the Bi sites and grown as an epitaxial composition-spread thin-film. Pure BFO is characterized by its rhombohedral ferroelectric structure with polarization oriented along the ⟨111⟩ direction within the pseudocubic framework. However, substituting around 14% of Bi with Sm alters the energy landscape of the material, inducing a phase transition to an orthorhombic configuration and effectively suppressing its ferroelectric properties.[17,18] Previously, materials within this library have been explored by scanning transmission electron microscopy (STEM) and scanning probe microscopy (SPM). These studies have been published in several previous manuscripts, and the resultant datasets were made open and are used in the present study.[14,15]

The details of sample growth, STEM imaging preparation, and initial data analysis and quantification have been outlined in our previous publications.[14,18,19] The atomic positions were extracted using the DCNN-based atom finder and refined using the Gaussian peak fit.[20,21] These analyses yield the intensities and exact position of all atomic columns corresponding to the A and B site cations in the perovskite lattice of both material and substrate. Based on these, we create a set of local descriptors as defined herein.

The descriptors outlined in Table I and Fig. 1 provide a comprehensive foundation for analyzing the local structural, compositional, and polarization properties of Sm substituted BFO (SmBFO). Understanding the relationships between these descriptors is critical for unraveling the underlying mechanisms governing the material's behavior. For instance, the summed A-site intensity ($\boldsymbol{I}14$), derived from four distinct local A-site columns ($I1 - I4$) per unit cell, encodes variations in local cation composition or occupancy (Bi/Sm distribution), while the B-site intensity ($\boldsymbol{I}5$) corresponds to the single Fe column at the center of each unit cell. These compositional metrics can influence or correlate with structural parameters such as lattice distortions (a, α, Vol) and polar displacements (P_x), thereby serving as useful predictors in causal analyses of structure–property relationships.

TABLE I. Key parameters and their descriptions for the SmBFO analysis.

Parameter	Description
I_{14}	Total intensity of A-site cation columns (Bi, Sm/alkali cations), computed as the sum of four local intensities (I_1–I_4) corresponding to the four A-site positions in each unit cell.
I_5	Intensity of the single B-site cation column (Fe/transition metal) in the unit cell
I_1, I_2, I_3, I_4	Individual intensities of A-site cation columns in a unit cell
a	Lattice_parameter
c	Composition
α	Unit_cell_angle
Vol	Projected unit-cell volume/volume
P_x	Electrical polarization component in the x-direction/In_plane_polarization

Together, this vector ($\boldsymbol{I_{14}}$, $\boldsymbol{I_5}$, $\boldsymbol{a}$, $\boldsymbol{c}$, $\boldsymbol{\alpha}$, $\boldsymbol{Vol}$, $\boldsymbol{P_x}$) is defined for each unit cell within the field of view, enabling a detailed characterization of disorder, polarization, lattice distortions, and chemical composition at the lattice site level. This approach enables the exploration of interdependencies and mechanisms that govern the material's properties.

The PC (Peter–Clark)[22] algorithm was implemented using gradient-based causal structure learning (gCastle)[23] as the first step in our workflow to infer directed acyclic graphs (DAGs)[24] that represent causal relationships among the key descriptors of SmBFO without relying on prior domain knowledge. The process begins with skeleton identification, where every descriptor is initially assumed to be connected [Fig. 2(a)]. From there, tests are applied in multiple rounds to determine which connections are genuine and which can be removed. These tests vary based on the nature of the data: when values are numeric, the algorithm checks whether two descriptors remain correlated after accounting for others; when they represent discrete categories, it compares observed distributions to expected patterns; and when the underlying relationship may not follow a simple trend, a kernel-based independence test is used to capture nonlinear relationships. For example, if the algorithm concludes that $\boldsymbol{I_5}$ and $\boldsymbol{P_x}$ are conditionally independent when controlling for the other descriptors, it removes the edge between them, producing the sparser skeleton seen in Fig. 2(b). In the second step, edge orientation is performed by detecting v-structures[25,26]—for instance, $\boldsymbol{I_{14}} \rightarrow \boldsymbol{a} \leftarrow \boldsymbol{Vol}$ and applying deterministic rules to orient the remaining chains, such as $\boldsymbol{I_{14}} \rightarrow \boldsymbol{a} \rightarrow \boldsymbol{Vol}$. The final output is a DAG [Fig. 2(c)], revealing potential causal pathways among the variables and enabling further analysis of their interdependencies and directional influences.

These results provide key insights into how compositional and structural parameters interact to influence polarization and overall material properties. For example, the causal link between $\boldsymbol{I_{14}}$, $\boldsymbol{a}$, and $\boldsymbol{P_x}$ suggests that modifications in the A-site (through Sm substitution at the Bi sites) composition can propagate through structural changes to alter the ferroelectric properties of SmBFO. Similarly, the relationship between unit-cell volume *Vol*, lattice parameter $\boldsymbol{a}$, and composition $\boldsymbol{c}$ underscores how compositional modifications and lattice distortions

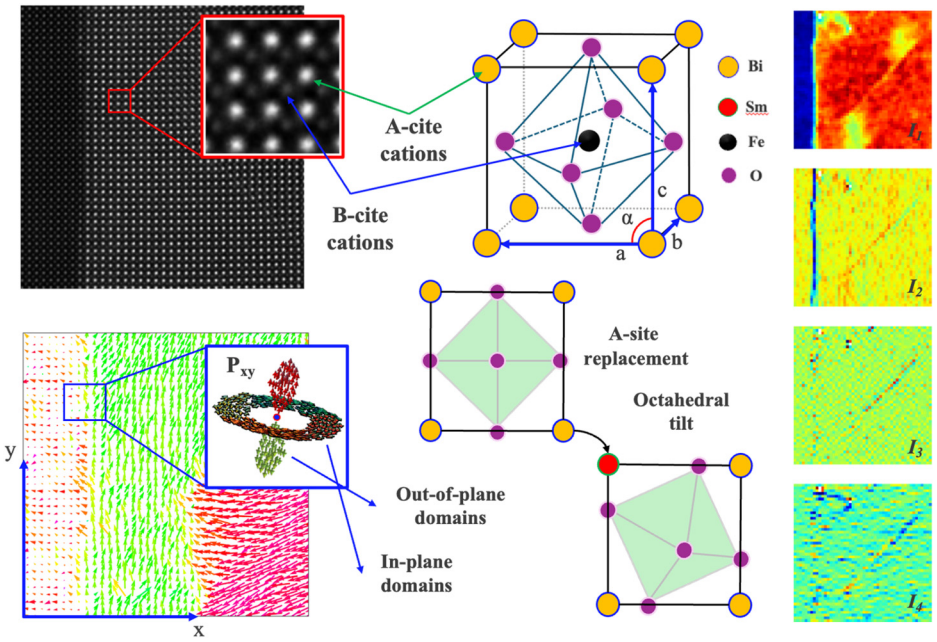

FIG. 1. Key parameters and their visual representation for the SmBFO analysis.

are intrinsically linked in shaping the overall crystal geometry. The bidirectional arrow between a and I_{14} reflects an initially unresolved directionality in the causal structure, arising from Markov equivalence[27] in the observational data; this ambiguity is transparently retained to acknowledge the limits of purely data-driven inference in distinguishing reciprocal causal effects.

Furthermore, this causal analysis can also be performed using a sliding window approach, as demonstrated in a previous study.[15] In this technique, causal relationships are evaluated across subsets of data, providing a localized perspective on causality, which is especially useful for analyzing spatially resolved datasets or temporal sequences.

Pure data-driven methods, such as the PC algorithm, might face a number of limitations stemming from using a finite sample size, the presence of latent or partially latent variables, or violations of the underlying independence test assumptions. Each of these limitations might lead to smaller or larger inaccuracies in the final results, undermining the goal of the analysis. In this sense, the current generation of purely data-driven causal discovery models is not a panacea for all problems and cannot fully replace the need for domain knowledge. At the same time, incorporating domain knowledge into the process can have a protective result against the distortions stemming from some or even all of the above-mentioned limitations, depending on the scenario at hand.

Limitations in problem-solving through mathematical modeling are not specific to causal discovery. Machine learning models cannot find correct solutions to ill-posed questions or those not appropriately expressed, and an in-depth understanding of the underlying chemistry is often needed to define a problem and relate it to a clear goal. Furthermore, relying solely on data-driven approaches can obscure the physical laws that govern materials' properties and behavior, hindering the understanding of fundamental materials phenomena. Data-driven analysis without physical models can be a weak lens for scientific inquiry: models can only be as good as the data they are trained on, and biases in the data can lead to inaccurate predictions. Integrating data science methods with domain knowledge can help alleviate some of these limitations.

To address these limitations, we adopted the approach proposed by Molak,[28] which integrates prior reference data into the causal discovery process. This method allows the incorporation of domain knowledge into the DAG, refining the causal relationships suggested by data alone. Specifically, we leveraged the gCastle toolbox for causal discovery and LangChain[29] for querying large language models (LLMs).[16] Using the ChatOpenAI[30] model (GPT-4-turbo), we queried causal relationships by analyzing domain-specific literature, particularly papers available on arXiv, accessed through the API, which returns structured metadata and abstracts.

The querying process was designed to extract insights from the scientific corpus, evaluating causal relationships between material descriptors (I_{14}, I_5, a, c, α, Vol, P_x) based on their scientifically recognized counterparts in Table I. To ensure consistency with established conventions and prevent misinterpretation, we first queried the Web of Science database[31] to identify widely accepted terminology for these descriptors in published research. To construct a knowledge-informed causal graph, we leveraged arXiv papers on ferroelectric materials, allowing the LLM to infer relationships within a broader context rather than being restricted to a specific material system. The model as represented in Table II was tasked with assessing whether a descriptor, such as I_{14}, influences a structural parameter like a, or vice versa, by interpreting causal relationships across diverse ferroelectric studies. This ensured that the extracted causal dependencies were not biased toward a single composition but rather reflected universally recognized trends in ferroelectric behavior, which were subsequently integrated into the DAG structure for further refinement.

As presented in Fig. 3(a), The LLM-informed DAG has refined the causal relationships identified in the PC algorithm-derived DAG by incorporating information derived from arXiv papers. In the

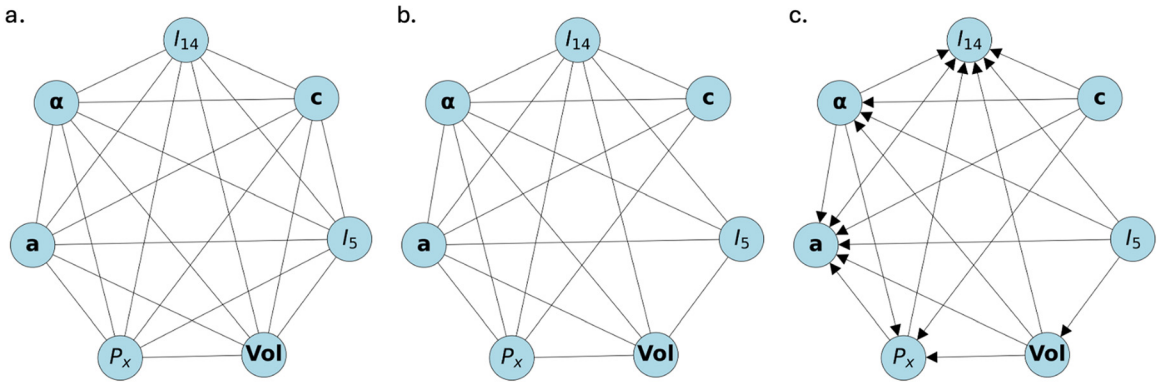

FIG. 2. Stages of the Peter–Clark (PC) algorithm for causal discovery. (a) Fully connected graph where all variables are assumed to have potential dependencies. (b) Skeleton graph after conditional independence tests. (c) Final directed acyclic graphs (DAG) with oriented edges, identifying potential causal relationships among the variables based on v-structures and logical rules.

LLM-informed DAG, $\boldsymbol{I_{14}}$ shows a direct influence on $\boldsymbol{P_x}$ through **a**, providing a clearer link between A-site compositional changes and polarization. Additionally, the dependency of unit-cell volume *Vol* on the lattice parameter ***a*** and composition ***c*** is more explicitly revealed, highlighting the intrinsic coupling among compositional tuning, lattice distortions, and volumetric changes in the crystal structure. These refinements underscore the value of integrating prior knowledge into the causal analysis to enhance the interpretability and reliability of the relationships. Additional details on dataset size and the metadata used in LLM-based reasoning (including publication titles, authors, dates, and abstracts) are provided in the supplementary material.

Finally, the PC algorithm was re-run using the LLM-informed relationships as prior knowledge. This hybrid approach combines the strengths of both data-driven and knowledge-informed methods, ensuring that statistical dependencies align with domain-specific insights. The resulting DAG in Fig. 3(b) incorporates both the rigor of conditional independence tests and the contextual relevance of LLM-derived relationships, leading to the most reliable and interpretable structure. As a result, the bidirectional edge between *a* and I_{14}—originally present due to a Markov equivalence class in the purely data-driven model—was resolved into a directed edge, guided by the incorporation of LLM-informed. However, the integration of LLM-generated insights also introduces potential risks, such as biases in the training data, misinterpretation of complex relationships, or the reinforcement of spurious correlations.[32–34] For example, the appearance of the edge $a \rightarrow c$ (lattice parameter to composition) may seem counterintuitive, since composition is typically considered the causal factor. This likely reflects limitations in the LLM's inference process or biases in the literature it was trained on. Therefore, careful validation remains essential to ensure that the inferred causal pathways are both scientifically sound and meaningful.

TABLE II. Schematic representation of the algorithmic workflow for querying causal relationships between variables using a large language model (LLM).

Algorithm	
Input	Two variables ($\boldsymbol{var_1}$ and $\boldsymbol{var_2}$)
Query to Agent	Asks if $\boldsymbol{var_1}$ causes $\boldsymbol{var_2}$, $\boldsymbol{var_2}$ causes $\boldsymbol{var_1}$, or neither.
Agent OUTPUT	Provides a detailed response explaining the causal relationship (if any).
LLM Interpretation	Converts the agent's response into a structured format **(0,1), (1,0), (0,0), or (-1, -1)**, where **(0,1)** indicates $\boldsymbol{var_1}$ causes $\boldsymbol{var_2}$, **(1,0)** indicates $\boldsymbol{var_2}$ causes $\boldsymbol{var_1}$, **(0,0)** means no causal relationship, and **(-1, -1)** means there is insufficient information to determine causality. The response should strictly follow this format without making assumptions beyond the provided data.
OUTPUT	Updates the causal graph by incorporating required and forbidden edges.

The proposed LLM-assisted causal discovery framework presents a transformative approach for understanding complex interdependencies in materials science, particularly in systems with intricate degrees of freedom. However, it is essential to evaluate its performance and limitations by comparing it to well-established causal discovery methods, such as LiNGAM. In our prior work, LiNGAM was utilized to investigate causal relationships in ferroelectric materials, yielding significant findings that served as a foundation for this study.

From the data shown in Fig. 4, both the LLM-informed causal DAG and the LiNGAM-derived causal structure exhibit consistent causal relationships in the ferroelectric composition, where compositional variation ***c*** is excluded (corresponding to $Sm_xBi_{1-x}FeO_3$ with $x = 0$, i.e., pure $BiFeO_3$). Both the LLM-informed causal graph Fig. 4(a) and the LiNGAM-derived structure Fig. 4(b) consistently identify that the in-plane polarization component $\boldsymbol{P_x}$ is jointly influenced by the A-site cation intensity $\boldsymbol{I_{14}}$, B-site cation intensity $\boldsymbol{I_5}$, and the in-plane lattice parameter **a**. This convergence highlights the multifactorial origin of ferroelectric polarization, arising from both compositional (chemical) and structural degrees of freedom. Furthermore, both models agree on a causal relationship where $\boldsymbol{a} \rightarrow Vol$, indicating that variations in the lattice constant directly modulate the unit-cell

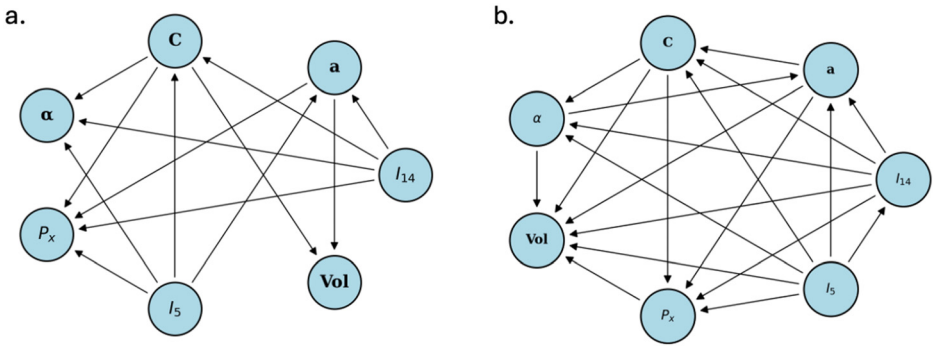

FIG. 3. (a) LLM informed DAG with oriented edges, (b) PC discovered LLM-integrated prior knowledge DAG with oriented edges.

volume. This directionality is physically meaningful, as the unit-cell volume is geometrically derived from the underlying lattice parameters. The PC-discovered DAG integrates prior knowledge, which can enhance interpretability and capture domain-specific insights, but it may introduce bias if the prior knowledge is inaccurate. On the other hand, LiNGAM relies entirely on observed data without incorporating prior domain knowledge, making it a useful tool for uncovering causal relationships. However, its accuracy depends heavily on whether its underlying assumptions—such as linearity and non-Gaussianity—hold in the given dataset (see Table III for a concise comparison of these assumptions). Violations of these assumptions, or the presence of noise or missing data, can lead to spurious causal inferences. Overall, the LLM-integrated approach is most effective when reliable prior information is available, while LiNGAM is suited for purely data-driven discovery—provided its assumptions hold. Despite relying on fundamentally different methodologies, both the LLM-informed and the purely data-driven approaches converge on consistent causal pathways, underscoring the robustness of the inferred relationships. This agreement highlights that key physical dependencies are reproducible across models derived from distinct sources of information. However, important differences remain: the LLM-informed method incorporates external domain knowledge to guide the orientation of edges, enabling it to resolve directionality in cases where statistical evidence alone is insufficient. In contrast, data-driven methods such as LiNGAM rely solely on observed statistical dependencies and are inherently limited by sample size, noise, and identifiability constraints. As a result, they may leave some edges unoriented, miss weak but meaningful connections, or produce alternative structures within a Markov equivalence class.

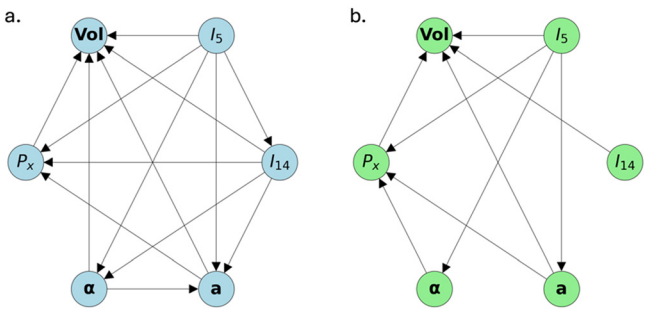

FIG. 4. (a) PC discovered LLM-integrated prior knowledge DAG, (b) LiNGAM-discovered causal matrices for a ferroelectric sample of pure $BiFeO_3$ corresponding to $Sm_xBi_{1-x}FeO_3$ with $x = 0$ (compositional variation **c** is excluded). The invariant composition is excluded as a causal factor, with inferred relationships reflecting dependencies among remaining variables.

By leveraging LLMs, researchers can analyze vast amounts of data from sources like arXiv to uncover causal relationships between synthesis parameters, such as temperature and pressure, and material properties like conductivity and hardness. Initial attempts suggest that without incorporating microscopic degrees of freedom, such as atomic structure and electron behavior, the findings remain too general. Combining data-driven discovery with LLM-based literature analysis allows for a comprehensive understanding of material behavior, facilitating the engineering of materials with specific functionalities. To enhance the effectiveness of this approach, both generalized and specific descriptors of physical quantities are included, acknowledging that similar phenomena are often reported under varying terminologies across the literature. This strategy improves the contextual relevance and retrieval accuracy of LLM-based searches. A causal graph illustrating this process would show growth conditions influencing microscopic variables, which in turn determine material properties, highlighting the importance of detailed microscopic data in the analysis.

As presented in Fig. 5, growth parameters (blue nodes) represent controllable synthesis conditions such as laser energy, deposition rate, and substrate temperature, which directly influence the material fabrication process. Additional parameters like vacuum level, ambient gas pressure, and film thickness shape the physical and chemical environment, affecting critical stages of material formation. The parameter space also includes compositional inputs related to the deposition source, which, while fixed during fabrication, fundamentally govern the resulting material chemistry. Similarly, metrics like film thickness—though often an outcome—are treated as tunable synthesis variables due to their direct control via process parameters and their impact on key structural and functional properties. Material properties (green nodes) depict the resulting characteristics of the synthesized material, including polarization, structural stability, phase purity, and defects, with polarization decomposed into directional components where relevant. These properties collectively determine the material's performance and application potential. The integration of LLMs can significantly enhance the understanding of causal relationships between growth parameters and material properties by providing context-aware priors informed by literature knowledge and domain expertise. While the specific structure shown is not derived from a

TABLE III. Assumptions for LiNGAM and PC (Peter–Clark).

Assumptions	LiNGAM (Original)	PC (Peter–Clark)
Functional form	Linear combination + non-Gaussian noise	Not strictly linear; often uses partial correlation or kernel-based tests.
Noise distribution	Non-Gaussian errors required for identifiability	Usually assumes Gaussian errors; can use kernel-based tests for non-Gaussian data.
Markov and faithfulness	Not explicitly invoked; relies on non-Gaussianity for identifiability	Applies Markov and faithfulness; violations may cause spurious or missing edges.[35,36]
Absence of unobserved confounders	Required (no hidden common causes)	Required (all relevant variables must be included).
Sensitivity to assumption violations	High if linearity or non-Gaussianity is violated	High if assumptions are violated; noise or incomplete data degrade performance.
Use of domain knowledge	Purely data-driven; domain expertise is optional for validation. Other variants, such as DirectLiNGAM might allow for incorporating expert knowledge.[37]	Incorporates prior knowledge (e.g., forbidden or required edges); can enhance interpretability but risk bias if inaccurate.

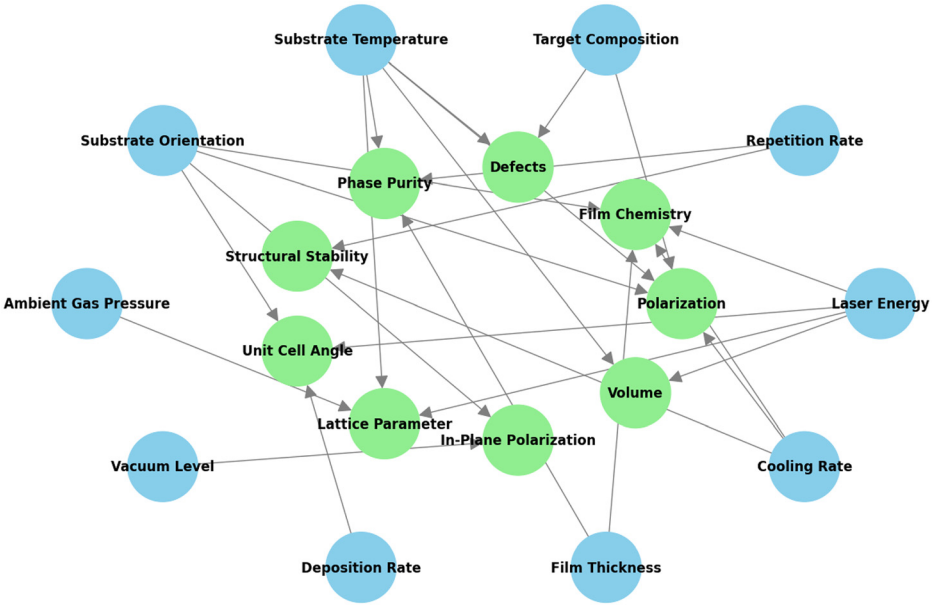

FIG. 5. Causal graph depicting relationships between growth parameters and material properties.

single dataset, it captures recurring variables common to synthesis workflows and serves as a generalizable blueprint for applying causal reasoning across diverse materials systems.

In conclusion, this study demonstrates a progressive refinement in causal discovery by integrating data-driven methods with domain-specific insights. The PC algorithm without prior knowledge establishes a baseline DAG, highlighting statistical dependencies but often overlooking or misrepresenting causal pathways due to the absence of contextual understanding. The LLM-informed DAG leverages insights from published research, refining and clarifying relationships while introducing the need for validation of external knowledge. Finally, the hybrid approach—using the PC algorithm with LLM-informed prior knowledge—combines statistical rigor with domain expertise, yielding the most reliable and interpretable DAG. This integrated workflow emphasizes the importance of combining advanced machine learning techniques with domain-specific knowledge to enhance the reliability and interpretability of causal relationships, providing a powerful framework for understanding complex systems such as SmBFO and guiding future experimental efforts.

See the supplementary material for additional details on the dataset size, metadata used in LLM-based reasoning (including publication titles, authors, dates, and abstracts), experimental setup, and extended results supporting the main findings.

This work (workflow development and concept) was supported (K.B. and S.V.K.) as part of the center for 3D Ferroelectric Microelectronics Manufacturing (3DFeM2), an Energy Frontier Research Center funded by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences under Award Number DE-SC0021118. STEM imaging (C.N.) was performed at the Oak Ridge National Laboratory's Center for Nanophase Materials Sciences (CNMS). The work at the University of Maryland was supported in part by the National Institute of Standards and Technology Cooperative Agreement No. 70NANB17H301 and the Center for Spintronic Materials in Advanced infoRmation Technologies (SMART), one of the centers in nCORE, a Semiconductor Research Corporation (SRC) program sponsored by NSF and NIST.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Kamyar Barakati: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Aleksander Molak:** Conceptualization (equal); Validation (equal); Writing – review & editing (equal). **Christopher T. Nelson:** Supervision (equal); Validation (equal). **Xiaohang Zhang:** Validation (equal); Visualization (equal). **Ichiro Takeuchi:** Validation (equal); Visualization (equal). **Sergei V. Kalinin:** Conceptualization (equal); Formal analysis (equal); Methodology (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in GitHub at https://github.com/KBForgeX/Causal_LLM, Ref. 38.

REFERENCES

[1]A. B. Georgescu, P. Ren, C. Karpovich, E. Olivetti, and J. M. Rondinelli, "Machine-learning based selection and synthesis of candidate metal-insulator transition metal oxides," arXiv:2404.08653 (2024).
[2]V. Venugopal and E. Olivetti, "MatKG: An autonomously generated knowledge graph in material science," Sci. Data **11**(1), 217 (2024).
[3]A. K. Tagantsev, J. Fousek, and L. E. Cross, *Domains in Ferroic Crystals and Thin Films* (Springer, New York, 2010).
[4]P. Muralt, R. G. Polcawich, and S. Trolier-McKinstry, "Piezoelectric thin films for sensors, actuators, and energy harvesting," MRS Bull. **34**(9), 658–664 (2009).
[5]T. M. Shaw, S. Trolier-McKinstry, and P. C. McIntyre, "The properties of ferroelectric films at small dimensions," Annu. Rev. Mater. Sci. **30**, 263–298 (2000).
[6]A. Biswas, A. N. Morozovska, M. Ziatdinov, E. A. Eliseev, and S. V. Kalinin, "Multi-objective Bayesian optimization of ferroelectric materials with interfacial control for memory and energy storage applications," J. Appl. Phys. **130**(20), 204102 (2021).
[7]A. N. Morozovska, E. A. Eliseev, A. Biswas, H. V. Shevliakova, N. V. Morozovsky, and S. V. Kalinin, "Chemical control of polarization in thin strained films of a multiaxial ferroelectric: Phase diagrams and polarization rotation," Phys. Rev. B **105**(9), 094112 (2022).
[8]C. Chen, D. T. Nguyen, S. J. Lee, N. A. Baker, A. S. Karakoti, L. Lauw, C. Owen, K. T. Mueller, B. A. Bilodeau, and V. Murugesan, "Accelerating computational materials discovery with machine learning and cloud high-performance computing: From large-scale screening to experimental validation," J. Am. Chem. Soc. **146**(29), 20009–20018 (2024).
[9]T. D. Cook and D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Houghton Mifflin Company, 2007).
[10]P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search* (MIT Press, 2001).
[11]S. Mysore, Z. Jensen, E. Kim, K. Huang, H.-S. Chang, E. Strubell, J. Flanigan, A. McCallum, and E. Olivetti, "The materials science procedural text corpus: Annotating materials synthesis procedures with shallow semantic structures," arXiv:1905.06939 (2019).
[12]X. Yu, "Machine learning application in the life time of materials," arXiv:1707.04826 (2017).
[13]P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," arXiv:1710.10903 (2017).
[14]M. Ziatdinov, C. T. Nelson, X. H. Zhang, R. K. Vasudevan, E. Eliseev, A. N. Morozovska, I. Takeuchi, and S. V. Kalinin, "Causal analysis of competing atomistic mechanisms in ferroelectric materials from high-resolution scanning transmission electron microscopy data," npj Comput. Mater. **6**(1), 127 (2020).
[15]A. N. M. C. Nelson, M. A. Ziatdinov, E. A. Eliseev, X. Zhang, I. Takeuchi, and S. V. Kalinin, "Mapping causal patterns in crystalline solids," arXiv:2103.01951 (2021).
[16]W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, and Z. Dong, "A survey of large language models," arXiv:2303.18223 (2023).
[17]K. Barakati, Y. Liu, C. Nelson, M. A. Ziatdinov, X. Zhang, I. Takeuchi, and S. V. Kalinin, "Reward driven workflows for unsupervised explainable analysis of phases and ferroic variants from atomically resolved imaging data," arXiv:2411.12612 (2024).
[18]C. T. Nelson, R. K. Vasudevan, X. Zhang, M. Ziatdinov, E. A. Eliseev, I. Takeuchi, A. N. Morozovska, and S. V. Kalinin, "Exploring physics of ferroelectric domain walls via Bayesian analysis of atomically resolved STEM data," Nat. Commun. **11**(1), 6361 (2020).
[19]M. Ziatdinov, N. Creange, X. Zhang, A. Morozovska, E. Eliseev, R. K. Vasudevan, I. Takeuchi, C. Nelson, and S. V. Kalinin, "Predictability as a probe of manifest and latent physics: The case of atomic scale structural, chemical, and polarization behaviors in multiferroic Sm-doped BiFeO3," Appl. Phys. Rev. **8**(1), 011403 (2021).
[20]A. Ghosh, M. Ziatdinov, O. Dyck, B. G. Sumpter, and S. V. Kalinin, "Bridging microscopy with molecular dynamics and quantum simulations: An atomAI based pipeline," npj Comput. Mater. **8**(1), 11 (2022).
[21]M. Ziatdinov, A. Ghosh, C. Y. Wong, and S. V. Kalinin, "AtomAI framework for deep learning analysis of image and spectroscopy data in electron and scanning probe microscopy," Nat. Mach. Intell. **4**(12), 1101–1112 (2022).
[22]M. Kalisch and P. Bühlman, "Estimating high-dimensional directed acyclic graphs with the PC-algorithm," J. Mach. Learn. Res. **8**(3), 613 (2007).
[23]K. Zhang, S. Zhu, M. Kalander, I. Ng, J. Ye, Z. Chen, and L. Pan, "gCastle: A python toolbox for causal discovery," arXiv:2111.15155 (2021).
[24]A. M. Lipsky and S. Greenland, "Causal directed acyclic graphs," JAMA **327**(11), 1083–1084 (2022).
[25]D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques* (MIT Press, 2009).
[26]J. Pearl, *Causality* (Cambridge University Press, 2009).
[27]F. Castelletti, G. Consonni, M. L. Della Vedova, and S. Peluso, "Learning Markov equivalence classes of directed acyclic graphs: An objective Bayes approach," Bayesian Anal. **13**(4), 1235–1260 (2018).
[28]A. Molak, *Causal Inference and Discovery in Python: Unlock the Secrets of Modern Causal Machine Learning with DoWhy, EconML, PyTorch and More* (Packt Publishing Ltd, 2023).
[29]O. Topsakal and T. C. Akinci, presented at the International Conference on Applied Engineering and Natural Sciences (2023).
[30]OpenAI, see https://openai.com/ for "OpenAI."
[31]Clarivate, see https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/ for "Web of Science platform."
[32]Y. Liu, Y. Yao, J.-F. Ton, X. Zhang, R. Guo, H. Cheng, Y. Klochkov, M. F. Taufiq, and H. Li, "Trustworthy LLMs: A survey and guideline for evaluating large language models' alignment," arXiv:2308.05374 (2023).

[33]J. Jiao, S. Afroogh, Y. Xu, and C. Phillips, "Navigating LLM ethics: Advancements, challenges, and future directions," arXiv:2406.18841 (2024).
[34]E. Eigner and T. Händler, "Determinants of LLM-assisted decision-making," arXiv:2402.17385 (2024).
[35]P. Spirtes and J. Zhang, "A uniformly consistent estimator of causal effects under the k-triangle-faithfulness assumption," Statist. Sci. **29**, 662 (2014).
[36]J. Ramsey, J. Zhang, and P. L. Spirtes, "Adjacency-faithfulness and conservative causal inference," arXiv:1206.6843 (2012).
[37]S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvarinen, Y. Kawahara, T. Washio, P. O. Hoyer, K. Bollen, and P. Hoyer, "DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model," J. Mach. Learn. Res. **12**, 1225–1248 (2011).
[38]GitHub (2025). "Causal_LLM," GitHub. https://github.com/KBForgeX/Causal_LLM