ORIGINAL RESEARCH **Open access**

Generalization in Materials AI: A Theoretical Distinction between New Compositions, New Structures, and New Physics

Hiroshi Nakamura[1*], Yuta Kato[1], Aiko Morita[2]

Abstract

The integration of artificial intelligence into materials science has accelerated property prediction and high-throughput screening. Yet, the field's progress hinges on models' ability to generalize beyond their training distributions. Existing literature often addresses generalization in broad terms, focusing on out-of-distribution performance or extrapolation without distinguishing the qualitative nature of material novelty. This conceptual manuscript introduces a novel theoretical framework for categorizing generalization in materials AI into three distinct levels: new compositions (variations within known structural families), new structures (alternative atomic arrangements or topologies), and new physics (emergence of phenomena governed by mechanisms absent from the training data). Drawing on recent advances in graph neural networks, scalable deep learning, and materials representations, we synthesize evidence that current models achieve reasonable interpolation within familiar domains but encounter progressively greater difficulties across these levels. The proposed distinction provides a structured lens for analyzing model limitations, interpreting benchmark results, and guiding the design of future architectures and training strategies. By formalizing these categories, the framework aims to advance theoretical understanding of generalization in materials AI, emphasizing the need for targeted approaches at each level to enable reliable discovery of novel materials.

Keywords Materials AI, Graph neural networks, Generalization, Out-of-distribution prediction, Crystal structures, Extrapolation

*Correspondence:
Hiroshi Nakamura
hiroshi.nakamura@gmail.com
[1] Department of Intelligent Materials Engineering, Faculty of Engineering, Nagoya University, Nagoya, Japan
[2] Department of AI-Driven Materials Systems, Faculty of Engineering, Tohoku University, Sendai, Japan

Introduction

The application of machine learning (ML) to materials science has expanded rapidly in recent years, driven by the convergence of large, open materials databases and advances in representation learning techniques [1, 2]. Models based on deep learning architectures now routinely predict key materials properties—including formation energies, electronic band gaps, elastic constants, and thermodynamic stability—with accuracies that approach or, in some controlled benchmarks, rival those of density functional theory (DFT) calculations [3, 4]. These advances have substantially accelerated high-throughput virtual screening workflows, enabling the evaluation of vast chemical spaces and guiding experimental synthesis and characterization efforts [2, 5]. As a result, ML has become an increasingly integral component of modern materials discovery pipelines.

Despite this progress, the central unresolved challenge in materials AI remains generalization: the ability of models to produce reliable predictions for materials that lie outside the statistical support of the training data [1, 6]. Unlike conventional regression tasks, materials discovery is inherently extrapolative. The space of possible inorganic compounds is astronomically large, sparsely sampled, and structured by complex chemical, structural, and physical constraints [7]. Consequently, predictive success in realistic discovery settings depends not on interpolation within known domains, but on robust performance under varying degrees of novelty.

Existing benchmarks, including widely adopted platforms such as MatBench, have played an important role in standardizing evaluation and demonstrating the potential of ML models for materials prediction [8]. However, these benchmarks predominantly assess interpolation or near-interpolation scenarios, in which test materials share substantial compositional or structural similarity with the training examples. Under such conditions, many models achieve impressive performance metrics. In contrast, systematic evaluations show that predictive accuracy often deteriorates sharply when models are confronted with genuinely out-of-distribution (OOD) samples, particularly those involving unfamiliar chemical environments, rare structural motifs, or atypical bonding patterns [9, 10]. This discrepancy raises concerns about the extent to which current benchmarks reflect the demands of real-world materials discovery.

Research published documents this tension across a range of model architectures. Graph neural networks (GNNs), in particular, have emerged as a dominant paradigm due to their ability to encode crystal structures as graphs and capture local atomic interactions through message passing [11, 12]. Empirical scaling laws further suggest that increasing dataset size and model capacity can enhance predictive accuracy and apparent robustness [2]. Nevertheless, large-scale evaluations consistently reveal that such improvements do not automatically translate into reliable extrapolation. Models frequently struggle when test materials differ qualitatively—not merely quantitatively—from those seen during training, highlighting a persistent gap between interpolation performance and true generalization [6, 13]. Proposed mitigation strategies, including transfer learning, physics-informed constraints, and uncertainty quantification, have shown promise in specific settings. Yet, they tend to yield incremental gains without addressing deeper distinctions, such as material novelty [4, 14].

A critical but underexplored observation is that generalization failure is not monolithic. The difficulty a model faces depends strongly on what kind of novelty is introduced. Some prediction tasks involve relatively modest departures from the training distribution, such as chemical substitutions within a well-characterized structural family (e.g., varying A-site cations in perovskite lattices). Other tasks require extrapolation across fundamentally different crystal structures, coordination environments, or symmetry classes. Still more challenging are scenarios in which models are expected to predict properties governed by physical phenomena absent from the training data, such as topological order or strong electronic correlations [15, 16]. Although these cases are often grouped under the umbrella of OOD generalization, doing so obscures the fact that they pose qualitatively distinct theoretical and representational challenges [1, 17].

In response to this gap, the present manuscript proposes a conceptual framework that distinguishes three hierarchical levels of generalization in materials AI: generalization to new compositions, structures, and physics. This stratification reflects the observation that material novelty is layered rather than uniform, ranging from chemical variation within known topologies to structural innovation and, ultimately, to emergent behavior that may require revised physical descriptions. By explicitly separating these levels, the framework provides a theoretical lens for diagnosing model limitations, comparing architectural inductive biases, and clarifying why certain approaches succeed in some discovery contexts but fail in others. The following sections synthesize relevant literature and elaborate on this layered framework, aiming to advance a more nuanced understanding of generalization in materials machine learning.

Theoretical Background & Literature Synthesis

Machine learning paradigms in materials property prediction

Materials property prediction has undergone a substantial methodological evolution over the past decade, transitioning from traditional descriptor-based approaches to end-to-end deep learning paradigms [1, 3]. Early machine learning models relied heavily on hand-crafted descriptors derived from elemental properties, stoichiometry, and coarse structural parameters, which were then coupled with classical algorithms such as kernel ridge regression or random forests [18]. These approaches offered a degree of interpretability and physical intuition, as individual features could often be traced back to known chemical trends. However, their expressive power was fundamentally constrained, limiting their ability to capture nonlinear, multi-body interactions and complex structure–property relationships in diverse materials systems [19].

Since 2020, graph-based deep learning has emerged as the dominant paradigm for predicting materials properties. In these approaches, materials are represented as graphs in which atoms correspond to nodes and interatomic interactions are encoded as edges, allowing models to learn directly from atomic configurations [11, 20]. Crystal graph convolutional neural networks and their variants have achieved state-of-the-art performance on benchmark datasets derived from the Materials Project and related repositories [12, 21]. More recently, the development of large-scale models trained on millions of DFT calculations has further improved predictive accuracy across a wide range of properties, reinforcing the importance of scale and representation capacity [2]. Collectively, these advances reflect a broader shift toward layered, data-driven representations that reduce reliance on manual feature engineering and instead learn task-relevant abstractions directly from structural information [4, 22].

Materials, representations, and structural encoding

Effective generalization in materials machine learning depends critically on the choice of representation. Representations must encode composition, structure, and relevant physical constraints in a manner that is invariant to symmetry operations while remaining expressive enough to capture subtle differences in atomic environments [23]. Common strategies include incorporating periodic boundary conditions into graph constructions, using distance-based edge features to represent interatomic interactions, and augmenting models with symmetry-aware operations to respect crystallographic invariances [11, 24]. To improve local environmental fidelity, geometric enhancements, such as angular or dihedral information, have been introduced, enabling models to better represent coordination geometry and bonding directionality [7, 25].

Despite these improvements, learned representations often remain implicitly biased toward the dominant motifs present in the training data [8, 26]. Empirical studies indicate that graph neural networks perform well when interpolating within frequently sampled structural families, such as cubic or layered materials, but degrade when confronted with rare space groups, low-symmetry structures, or defect-rich environments [10, 27]. To address these shortcomings, recent work has explored higher-order representations, including line graphs and models that explicitly encode multi-body interactions. While such approaches have demonstrated localized gains, systematic challenges persist in achieving robust generalization across the full diversity of crystal structures encountered in materials discovery [12, 28].

Datasets and benchmarking practices

The rapid progress of materials machine learning has been enabled by the availability of large, curated datasets, including the Materials Project, JARVIS, and Alexandria [29, 30]. These resources have supported the development and comparison of increasingly sophisticated models by providing standardized training and evaluation corpora. Benchmarking practices typically emphasize cross-validation within fixed data distributions, often stratified by chemical composition or structural class to ensure balanced sampling [8, 31]. Under these conditions, models frequently demonstrate strong predictive performance, reinforcing confidence in their interpolation capabilities.

However, such benchmarking protocols rarely probe extrapolative behavior, which is central to realistic materials discovery scenarios [6, 32]. More recent studies published between 2021 and 2023 have begun to explicitly examine out-of-distribution settings, including predictions for hypothetical compounds, structures generated by generative models, or materials occupying sampled regions of chemical space [13, 33]. These investigations consistently reveal sharp performance degradation when test samples fall outside the convex hull of training compositions or involve previously

unseen structural prototypes [9, 34]. These findings underscore a critical mismatch between benchmark success and real-world deployment requirements, motivating the need for evaluation frameworks that more directly assess different forms of generalization.

Challenges in generalization and extrapolation

Generalization in materials artificial intelligence is fundamentally constrained by the intrinsic characteristics of materials data, including sparsity, extreme dimensionality, and adherence to underlying physical laws [1, 5]. Unlike many conventional machine learning domains, materials datasets sample only a minute fraction of the feasible chemical and structural space. At the same time, target properties often depend on complex, nonlinear interactions across multiple length and energy scales. As a result, models trained predominantly on equilibrium crystal structures may exhibit limited reliability when applied to metastable phases, defect-containing materials, or systems synthesized under non-equilibrium conditions [35].

Distinct challenges arise depending on the form of extrapolation required. Extrapolation to new compositions is hindered by the combinatorial diversity of chemical space, where elemental substitutions can induce non-intuitive changes in bonding, electronic structure, and stability. Structural extrapolation, by contrast, demands invariance to crystallographic symmetry, periodicity, and topology, as well as sensitivity to global atomic arrangements rather than purely local environments [10, 24]. Failures in either regime can lead to confident yet physically implausible predictions, undermining the utility of ML models in discovery-oriented workflows.

Recent large-scale analyses demonstrate that even highly expressive models trained on extensive datasets remain vulnerable to out-of-distribution (OOD) shifts when chemical or structural novelty is introduced [2, 6]. Increasing model size and training data volume improves interpolation performance but does not eliminate fundamental extrapolation failures. Interpretability and feature-attribution studies further suggest that learned representations may encode spurious correlations specific to the training distribution rather than robust physical relationships, leading to brittle behavior under distribution shift [3, 14]. These issues are particularly acute in small-data regimes, where limited coverage amplifies overfitting and uncertainty, motivating the use of strategies such as active learning, multi-fidelity modeling, or targeted data acquisition to improve robustness [5, 35].

Strategies to enhance generalization

A growing body of work has proposed methodological strategies to mitigate these limitations and enhance generalization performance. Domain adaptation techniques aim to align feature distributions between training and target domains, while uncertainty estimation frameworks seek to identify predictions made outside the model's effective applicability domain [4, 16]. Incorporating physical priors—such as symmetry constraints, conservation laws, or physically motivated regularization—has also been explored as a means of anchoring learned representations to known scientific principles.

Generative modeling approaches offer an alternative perspective on extrapolation by explicitly sampling beyond the training distribution, generating hypothetical compounds or structures that expand chemical and structural coverage [33]. In parallel, algebraic and geometric enhancements to graph neural networks have been introduced to better encode periodicity, symmetry, and higher-order interactions, addressing some limitations of purely local message-passing schemes [7, 25]. While these approaches have yielded measurable improvements in specific contexts, they often address robustness incrementally rather than confronting the deeper question of which type of novelty a model is expected to handle.

Collectively, the literature indicates substantial progress in predictive accuracy within known domains, yet a persistent gap persists in models' ability to cope with qualitative novelty. Most existing studies implicitly treat generalization as a continuous spectrum of distribution shift, without explicitly distinguishing between fundamentally different modes of extrapolation. This lack of conceptual differentiation obscures the reasons behind model failure and limits the interpretability of benchmark results, motivating the need for a more structured theoretical framework.

Proposed conceptual framework

The proposed framework categorizes generalization in materials artificial intelligence into three conceptually distinct levels, defined by the type of novelty encountered: new compositions, new structures, and new physics. These levels do not simply reflect increasing magnitudes of distribution shift; rather, they correspond to qualitatively different demands on representations, inductive biases, and embedded scientific knowledge.

New compositions refer to generalization within a fixed structural prototype or family, where the elemental makeup varies while the underlying topology remains unchanged. Typical examples include alloying within a rock-salt lattice or A-site substitution in perovskite structures. At this level, models must interpolate across chemical space while preserving structural context. Current graph neural networks often perform adequately under these conditions when training data encompass similar compositions, as local atomic environments remain comparable and message-passing mechanisms can capture relevant interactions [11, 21].

New structures involve extrapolation to materials with fundamentally different atomic arrangements, space groups, or topological features, such as transitions from layered to framework structures or from centrosymmetric to non-centrosymmetric lattices. Successful generalization at this level requires representations and architectures that encode global structural motifs, symmetry operations, and long-range interactions in an invariant manner [7, 12]. Failures frequently arise from inadequate structural encoding or excessive reliance on local patterns prevalent in the training data, leading to degraded performance on unfamiliar prototypes [8, 27].

New physics represents the most demanding level of generalization, encompassing scenarios in which target properties are governed by mechanisms absent from the training distribution. Examples include predicting magnetic ordering, superconductivity, or topological states using models trained primarily on conventional insulating or weakly correlated materials. In such cases, reliable prediction may require implicit modeling of electronic correlations, collective excitations, or emergent behavior that is not captured by standard ground-state descriptors [15, 16]. This level exposes fundamental limitations of purely data-driven approaches and may necessitate theoretical priors or hybrid modeling strategies that extend beyond conventional machine learning paradigms [1, 17].

These three categories form a hierarchy of increasing difficulty. Generalization to new compositions represents the lowest barrier, where compositional interpolation is often sufficient. Structural generalization demands abstraction over crystallographic organization, while generalization to new physics requires reasoning about underlying physical laws and mechanisms. The framework posits that model performance degrades systematically across these levels unless architectures incorporate level-specific inductive biases, such as compositional embeddings, symmetry-aware convolutions, or physics-informed layers, tailored to the nature of the novelty encountered at each level.

Figure 1 presents a schematic representation of the proposed framework. A triangular hierarchy is depicted, with the base labeled “New Compositions.”

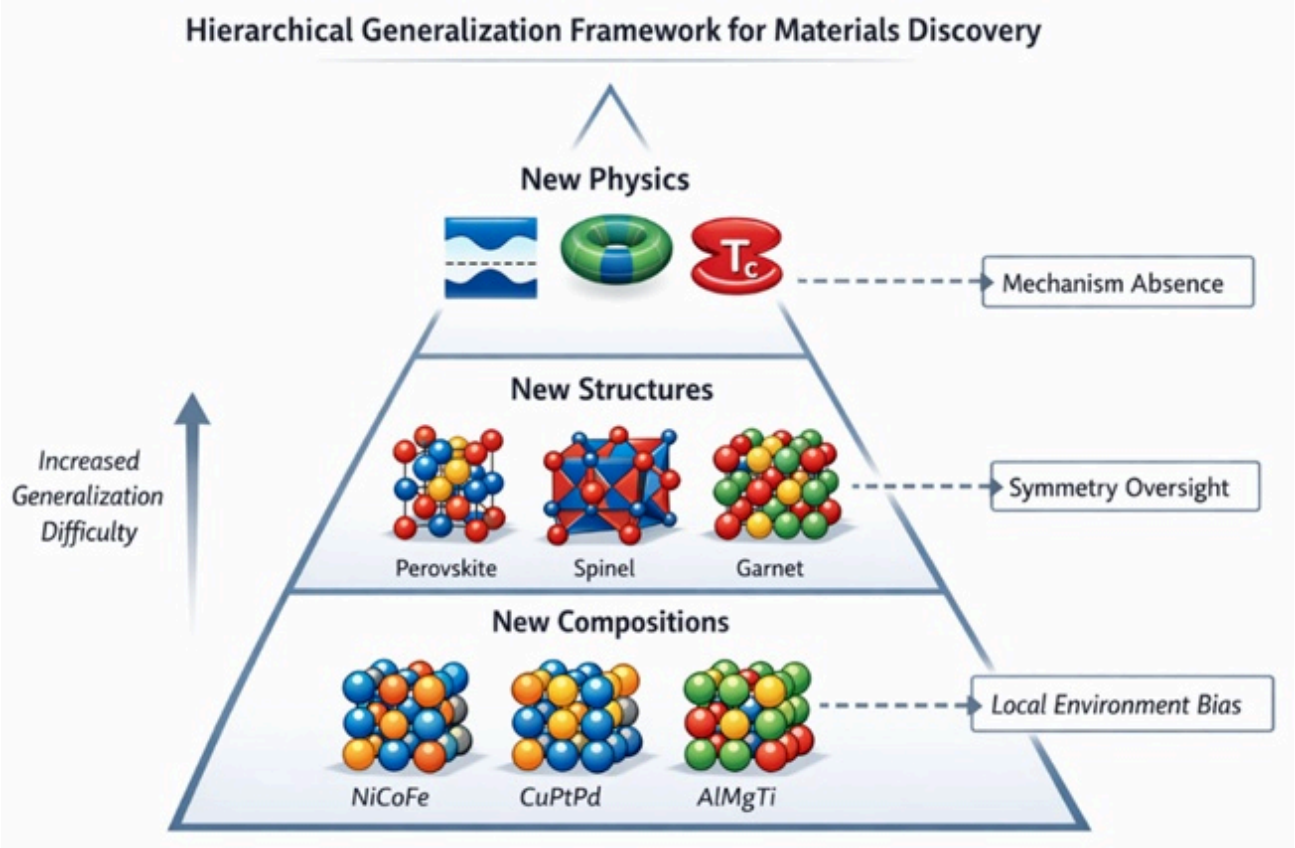

Figure 1. Hierarchical framework from new compositions to new structures to new physics, with increasing generalization difficulty and representative failure modes

Layers of generalization in materials AI

Building on the proposed framework, generalization in materials machine learning can be understood as occurring across four conceptual layers, each corresponding to increasing degrees of novelty and representational demand. These layers synthesize recent literature on materials representations, graph-based architectures, and out-of-distribution performance, and collectively clarify why generalization success varies across tasks.

Layer 1: Compositional generalization within known structural families

Generalization at the first layer involves transferring predictive capability to new chemical compositions within established structural families. This form of generalization relies predominantly on comprehensive coverage of chemical space in the training data and on models' ability to interpolate local atomic environments. When datasets include diverse elemental substitutions within the same topology, graph neural networks demonstrate robust performance because property variations are driven primarily by compositional changes rather than structural rearrangements [2, 11, 21]. Failures at this layer typically arise from sparse sampling of the periodic table or extrapolation beyond the convex hull of training compositions, rather than from inherent architectural limitations [6, 9].

Layer 2: Structural generalization across distinct crystal prototypes

The second layer encompasses generalization to previously unseen structures, where models must extrapolate across different crystal prototypes or symmetry classes. This regime demands inductive biases that capture global topological features, symmetry operations, and long-range interactions beyond local neighborhoods. Standard message-passing graph neural networks often overfit to prevalent motifs in the training data, such as common space groups, resulting in degraded performance on novel structural prototypes. Enhancements, including higher-order geometric encodings, line-

graph constructions, or attention mechanisms operating over extended interaction ranges, are therefore essential to enable structural abstraction and transferability across crystal classes [7, 12, 25, 28].

Layer 3: Physical generalization to novel phenomena

The third layer represents generalization to new physical regimes, where predictive success requires inferring emergent phenomena or collective behaviors that are not explicitly encoded in the ground-state training data. Properties governed by electronic correlations, topological protection, or phase transitions often necessitate multi-fidelity learning across physical regimes or the integration of domain-specific priors, such as symmetry-adapted bases or effective Hamiltonians, to bridge data-driven prediction and theoretical mechanisms [15-17]. In this layer, purely empirical models encounter fundamental limitations, as training distributions rarely sample the relevant physical conditions.

Layer 4: Hierarchical and progressive generalization strategies

The fourth layer concerns strategies for improving generalization across the preceding layers through hierarchical modeling and staged learning. Progressive curricula that align model training with increasing representational complexity—beginning with compositional interpolation, advancing to structural abstraction, and culminating in physics-aware reasoning—offer a systematic pathway toward broader extrapolation capability. Modular architectures and transfer learning paradigms are particularly well suited to this layered progression, as they allow the reuse and refinement of learned representations at successive levels [4, 14, 33]. This hierarchy reflects the increasing complexity of required knowledge representations, ranging from local descriptors to global invariants and, ultimately, theoretical constructs. The relationship between material novelty, dominant failure modes, and required modeling strategies across the proposed layers is summarized schematically in **Figure 2**.

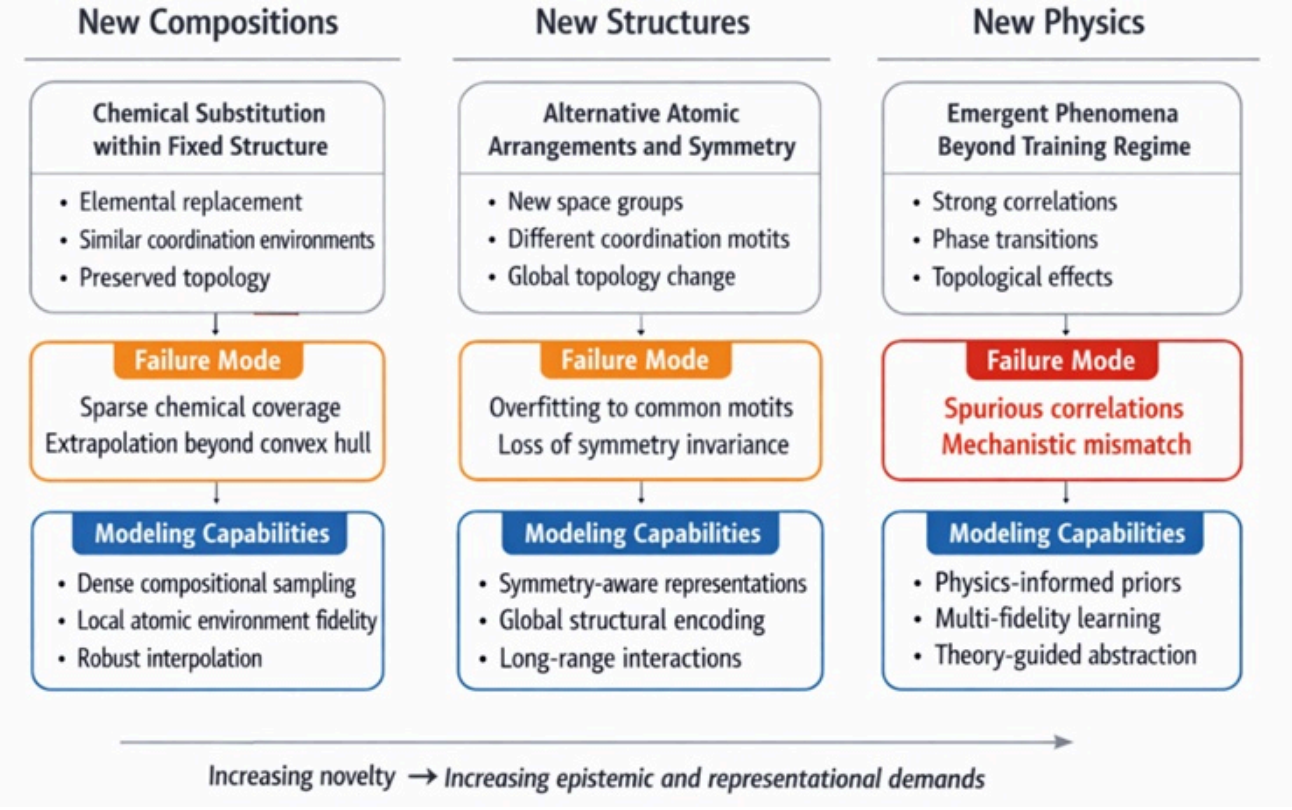

Figure 2. Failure modes and modeling requirements across layers of generalization in materials AI. As novelty increases from new compositions to new structures and new physics, model limitations shift from data sparsity and interpolation failure toward representational bias and fundamental mechanistic mismatch, necessitating progressively stronger inductive biases and physics integration

Results and Discussion

The layered distinction between compositional, structural, and physical generalization provides a structured theoretical lens for interpreting persistent challenges in materials artificial intelligence. Current benchmarks, such as MatBench, predominantly evaluate interpolation within familiar compositional and structural domains, often yielding strong performance metrics that mask vulnerabilities to qualitative novelty [8, 30, 31]. By explicitly separating generalization into layers, the framework clarifies why models excel at compositional screening yet struggle when confronted with unfamiliar structures or emergent physical behavior, consistent with observed distribution shifts in large-scale evaluations [2, 6, 13].

At the lower layers, substantial progress has been achieved through scaled graph neural networks and increasingly diverse datasets, enabling reliable predictions within known structural families [10, 20, 22]. However, limitations in structural abstraction persist, underscoring the need for symmetry-aware and topology-sensitive representations [23, 24, 27]. The physical generalization layer remains the most challenging, as emergent phenomena frequently require reasoning beyond ground-state energetics; physics-informed models and generative extrapolation strategies may help bridge this gap, albeit at the cost of scalability and interpretability [3, 14, 16, 35].

From a practical standpoint, the layered framework informs materials discovery pipelines by aligning modeling objectives with discovery goals. Lower-layer generalization supports high-throughput virtual screening, while advances in higher layers may enable the identification of qualitatively new functional materials, including non-centrosymmetric compounds or strongly correlated systems [5, 15, 33]. The framework also motivates the design of next-generation benchmarks that explicitly probe each layer, moving beyond aggregated out-of-distribution metrics toward targeted evaluation tasks such as cross-prototype transfer or prediction under physical regime shifts. The implications of the proposed framework for model design, training strategy, and evaluation are summarized in **Table 1**.

Table 1. Modeling strategies and inductive biases are required across generalization layers

Layer	Required inductive bias	Suitable modeling strategies	Benchmark implications
New compositions	Local chemical environment sensitivity	Dense datasets, compositional embeddings, message passing	Composition-stratified splits
New structures	Symmetry and topology awareness	Line graphs, equivariant networks, long-range attention	Cross-prototype evaluation
New physics	Physical law consistency	Physics-informed models, multi-fidelity learning	Regime-shift benchmarks

Limitations include partial overlap between layers, the potential for structural changes to induce new physical behavior, and the framework's primary focus on inorganic crystalline materials. Nevertheless, the conceptual structure is extensible to molecular, amorphous, and polymeric systems. Future directions include developing quantitative metrics for layer-specific novelty, such as structural novelty indices or physics-shift measures, and exploring hybrid learning paradigms that combine data-driven models with symbolic or theory-guided reasoning to address the highest-level generalization.

Conclusion

Generalization remains the central determinant of whether artificial intelligence can transition from a powerful interpolation tool to a genuinely transformative engine for materials discovery. While recent advances have demonstrated impressive predictive accuracy in well-sampled domains, models' ability to operate reliably beyond their training distributions remains limited, limiting their scientific and practical impact. Addressing this challenge requires not only improved algorithms and larger datasets, but also a clearer conceptual understanding of the kinds *of novelty* a model is expected to handle.

This work introduces a layered theoretical framework that distinguishes generalization in materials AI into three qualitatively distinct forms of novelty: new compositions, new structures, and new physics. By framing generalization hierarchically, the framework moves beyond undifferentiated notions of out-of-distribution performance and highlights the progressive escalation in representational and epistemic demands. Compositional generalization primarily requires robust interpolation across chemical space, structural generalization necessitates abstraction over global topology and symmetry, and physical generalization demands sensitivity to emergent mechanisms that may not be explicitly encoded in training data.

Articulating these layers clarifies why contemporary models often perform well in benchmark settings yet struggle in discovery-driven scenarios. Success at lower layers can mask vulnerabilities at higher layers, leading to overconfidence in the model's capabilities and misinterpretation of benchmark performance. The framework, therefore, provides a conceptual scaffold for diagnosing model failures, interpreting evaluation results, and aligning architectural design choices with intended deployment contexts. It also underscores that improvements in scale alone are unlikely to overcome fundamental extrapolation barriers without corresponding advances in inductive bias design and knowledge integration.

Beyond retrospective analysis, the framework offers forward-looking guidance for the development of next-generation materials AI systems. By explicitly linking forms of novelty to required inductive biases—such as compositional embeddings, symmetry-aware structural representations, or physics-informed modeling components—it supports more targeted research strategies and principled architecture innovation. Moreover, it motivates the design of evaluation protocols and benchmarks that explicitly probe different layers of generalization, rather than aggregating performance across qualitatively distinct challenges.

In summary, this work advances a structured theoretical perspective on generalization in materials machine learning, reframing it as a layered problem rather than a single continuum. By doing so, it contributes to a more rigorous understanding of the limits and possibilities of data-driven materials modeling. Ultimately, embracing such distinctions is essential for developing AI systems capable of reliably navigating unexplored regions of materials space, thereby accelerating the discovery of materials with genuinely novel structures and functionalities.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 07 Dec 2023 Revised: 15 Jan 2024 Accepted: 03 Mar 2024
Published online: 18 July 2024

References

Morgan D, Jacobs R. Opportunities and challenges for machine learning in materials science. Annu Rev Mater Res. 2020;50:71-103.
https://doi.org/10.1146/annurev-matsci-070218-010015.

Dunn A, Wang Q, Ganose AM, Dopp D, Jain A. Benchmarking materials property prediction methods: the Matbench test set and Automatminer reference algorithm. npj Comput Mater. 2020;6:1-10.
https://doi.org/10.1038/s41524-020-00406-3.

Goodall REA, Lee AA. Predicting materials properties without crystal structure: deep representation learning from stoichiometry. Nat Commun. 2020;11:6280.
https://doi.org/10.1038/s41467-020-19964-7.

Reiser P, Neubert M, Eberhard A, Torresi L, Zhou C, Shao C. et al. Graph neural networks for materials science and chemistry. Commun Mater. 2022;3:93.
https://doi.org/10.1038/s43246-022-00315-6.

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

Batzner S, Musaelian A, Sun L, Geiger M, Mailoa JP, Kornbluth M, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. Nat Commun. 2022;13:2453.
https://doi.org/10.1038/s41467-022-29939-5.

Chen C, Zuo Y, Ye W, Li X, Ong SP. Learning properties of ordered and disordered materials from multi-fidelity data. Nat Comput Sci. 2021;1:46-53.
https://doi.org/10.1038/s43588-020-00002-x.

Karamad M, Sinha S, Shi Y, Siahrostami S, Gates ID, Farimani A, Orbital graph convolutional neural network for material property prediction. Phys Rev Mater. 2020;4:093801.
https://doi.org/10.1103/PhysRevMaterials.4.093801.

Noh J, Gu GH, Kim S, Jung Y. Machine-enabled inverse design of inorganic solid materials: promises and challenges. Chem Sci. 2020;11:4871-81.
https://doi.org/10.1039/D0SC00594K.

Friederich P, Häse F, Proppe J, Aspuru-Guzik A. Machine-learned potentials for next-generation matter simulations. Nat Mater. 2021;20:750-61.
https://doi.org/10.1038/s41563-020-0777-6.

Schmidt J, Pettersson L, Verdozzi C, Botti S, Marques MAL. Crystal graph attention networks for the prediction of stable materials. Sci Adv. 2021;7:eabi7948.
https://doi.org/10.1126/sciadv.abi7948.

von Lilienfeld OA, Burke K. Retrospective on a decade of machine learning for chemical discovery. Nat Commun. 2020;11:1-4. https://doi.org/10.1038/s41467-020-18556-9.

Bartel CJ, Trewartha A, Wang Q, Dunn A, Jain A, Ceder G. A critical examination of compound stability predictions from machine-learned formation energies. npj Comput Mater. 2020;6:97. https://doi.org/10.1038/s41524-020-00393-5.

Wang H-C, Botti S, Marques MAL. Predicting stable crystalline compounds using chemical similarity. npj Comput Mater. 2021;7:12. https://doi.org/10.1038/s41524-020-00484-3.

Furness JW, Kaplan AD, Ning J, Perdew JP, Sun J. Accurate and numerically efficient r²SCAN meta-generalized gradient approximation. J Phys Chem Lett. 2020;11:8208-15. https://doi.org/10.1021/acs.jpclett.0c02405.

Deringer VL, Caro MA, Csányi G. A general-purpose machine-learning force field for bulk and nanostructured phosphorus. Nat Commun. 2020;11:5461. https://doi.org/10.1038/s41467-020-19168-z.

Yao Z, Sánchez-Lengeling B, Jayaraman JD. Inverse design of nanoporous crystalline reticular materials with deep generative models. Nat Mach Intell. 2021;3:76-86. https://doi.org/10.1038/s42256-020-00271-1.

Atz K, Grisoni F, Schneider G. Geometric deep learning on molecular representations. Nat Mach Intell. 2021;3:1023-32. https://doi.org/10.1038/s42256-021-00418-8.

Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. Graph neural networks: a review of methods and applications. AI Open. 2021;2:57-81. https://doi.org/10.1016/j.aiopen.2021.01.001.

Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS, et al. A comprehensive survey on graph neural networks. IEEE Trans Neural Netw Learn Syst. 2021;32:4-24. https://doi.org/10.1109/TNNLS.2020.2978386.

Cheng J, Zhang C, Dong L. A geometric-information-enhanced crystal graph network for predicting properties of materials. Commun Mater. 2021;2:92. https://doi.org/10.1038/s43246-021-00194-3.

Long T, Fortunato NM, Opahle I, Zhang Y, Samathrakis I, Shen C, et al. Constrained crystals deep convolutional generative adversarial network for the inverse design of crystal structures. npj Comput Mater. 2021;7:66. https://doi.org/10.1038/s41524-021-00526-4.

Wang Q, Zhang L. Inverse design of glass structure with deep graph neural networks. Nat Commun. 2021;12:4899. https://doi.org/10.1038/s41467-021-25490-x.

Hsu T, Pham TA, Keilbart N, Weitzner S, Chapman J, Xiao P, et al. Efficient and interpretable graph network representation for angle-dependent properties applied to optical spectroscopy. npj Comput Mater. 2022;8:151. https://doi.org/10.1038/s41524-022-00830-9.

Chen C, Ong SP. A universal graph deep learning interatomic potential for the periodic table. Nat Comput Sci. 2022;2:718-28. https://doi.org/10.1038/s43588-022-00349-3.

Zhong X, Gallagher B, Liu S, Kailkhura B, Hiszpanski A, Han TJJ, et al. Explainable machine learning in materials science. npj Comput Mater. 2022;8:204. https://doi.org/10.1038/s41524-022-00884-7.

Park J, Shim Y, Lee F, Rammohan A, Goyal S, Shim M, et al. Prediction and Interpretation of Polymer Properties Using the Graph Convolutional Network. ACS Polym Au. 2022;2(4):213-22.
https://doi.org/10.1021/acspolymersau.1c00050.

Chen Y, Tang X, Qi X, Li C-G, Xiao R. Learning graph normalization for graph neural networks. Neurocomputing. 2022;493:613-25.
https://doi.org/10.1016/j.neucom.2022.01.003.

Gastegger M, Schütt KT, Müller K-R. Machine learning of solvent effects on molecular spectra and reactions. Chem Sci. 2021;12:11473-83.
https://doi.org/10.1039/D1SC02742E.

Gong S, Yan K, Xie T, Shao-Horn Y, Gomez-Bombarelli R, Ji S, et al. Examining graph neural networks for crystal structures: Limitations and opportunities for capturing periodicity. Sci Adv. 2023;9(45):eadi3245.
https://doi.org/10.1126/sciadv.adi3245.

Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. Chem Mater. 2019;31:3564-72.
https://doi.org/10.1021/acs.chemmater.9b01294.

von Lilienfeld OA, Müller K-R, Tkatchenko A. Exploring chemical compound space with quantum-based machine learning. Nat Rev Chem. 2020;4:347-58.
https://doi.org/10.1038/s41570-020-0189-9.

Huang B, von Lilienfeld OA. Quantum machine learning using atom-in-molecule-based fragments selected on the fly. Nat Chem. 2020;12:945-51.
https://doi.org/10.1038/s41557-020-0527-z.

Christensen AS, Bratholm LA, Faber FA, von Lilienfeld OA. FCHL revisited: faster and more accurate quantum machine learning. J Chem Phys. 2020;152:044107.
https://doi.org/10.1063/1.5126701.

Kalinin SV, Ziatdinov M, Vasudevan S, Xie S. Machine learning in scanning transmission electron microscopy. Nat Rev Methods Primers. 2022;2:1-28.
https://doi.org/10.1038/s43586-022-00095-w.