ORIGINAL RESEARCH **Open access**

Scaling Laws without Physics: A Conceptual Analysis of Model Expansion in Computational Materials Engineering

Hiroshi Tanaka[1], Yuki Sato[1*], Kenji Mori[2], Rina Okabe[1]

Abstract

The rapid evolution of computational materials engineering has ushered in an era where data-driven approaches increasingly dominate discovery pipelines, leveraging vast datasets and expansive model architectures to uncover material properties and behaviors. This conceptual analysis examines the phenomenon of model expansion in materials informatics, focusing on scaling laws that emerge independently of traditional physics-based derivations. By dissecting the interplay between dataset scaling, parameter proliferation, and computational resource demands, we highlight how such expansions influence epistemic gains in materials discovery. A core gap in current paradigms lies in the overreliance on empirical scaling metrics, which often overlook the nuanced trade-offs between model complexity and interpretive insight. To address this, we introduce the "Insight Amplification Cascade" framework, a layered conceptual structure that maps data infrastructures to inference dynamics, emphasizing feedback mechanisms that balance energy costs against discovery yields. This framework integrates representation learning with uncertainty quantification to steer computational workflows toward sustainable scaling. Implications extend to autonomous discovery systems, where model expansion fosters robust inverse design without necessitating physics-grounded priors. Ultimately, this analysis underscores the need for infrastructure-level reforms in materials AI, promoting scalable yet interpretable ecosystems that enhance long-term innovation in computational materials engineering. Through this lens, we advocate for a reevaluation of scaling strategies to prioritize epistemic efficiency over mere parametric growth.

Keywords Materials informatics, Representation learning, AI in materials science, Computational discovery, Data infrastructures, Model scaling

*Correspondence:
Yuki Sato
yuki.sato@gmail.com
[1] Department of Computational Materials Engineering, Faculty of Engineering, University of Tokyo, Tokyo, Japan
[2] Department of Data-Driven Materials Discovery, Faculty of Engineering, Kyoto University, Kyoto, Japan

Introduction

The advent of computational materials engineering has transformed the landscape of materials science, shifting from traditional experimental trial-and-error methods to sophisticated data-driven paradigms that harness computational power for accelerated discovery. This shift is underpinned by the integration of machine learning techniques, which enable the processing of vast multimodal datasets to predict material properties, design novel compounds, and optimize performance characteristics.

Central to this evolution is the role of AI in bridging high-throughput computations with experimental validations, creating closed-loop systems that iteratively refine material candidates [1, 2]. For instance, materials informatics has emerged as a key discipline, employing statistical and algorithmic tools to extract patterns from complex material datasets, thereby facilitating the exploration of expansive chemical spaces [3, 4].

At the heart of these advancements lies the expansion of model architectures, where increases in parameters and

dataset sizes drive purported scaling laws that promise enhanced predictive capabilities. Unlike physics-derived scaling laws, such as those in thermodynamics or quantum mechanics, these computational scaling phenomena arise from empirical observations in training dynamics, often manifesting as power-law relationships between model size and performance metrics [5, 6]. However, in computational materials engineering, this expansion occurs without explicit physics constraints, relying instead on data representations and learning algorithms to infer underlying material behaviors. This "scaling without physics" paradigm raises critical questions about the sustainability and epistemic value of such approaches, particularly as parameter growth escalates computational energy costs while potentially diminishing marginal insights [7, 8].

High-throughput computation has been instrumental in generating the data foundations for these models, enabling the simulation of thousands of material configurations through density functional theory and molecular dynamics [9, 10]. Coupled with machine learning, these infrastructures support inverse materials design, where desired properties guide the search for optimal compositions rather than vice versa [11, 12]. Yet, challenges persist in dataset scaling: the quality, diversity, and multimodal nature of materials data directly impact model robustness, often leading to biases or overfitting in underrepresented domains [13, 14]. Parameter growth, while amplifying model capacity, introduces complexities in training and inference, where larger architectures demand proportional increases in computational resources, exacerbating energy consumption without guaranteed proportional gains in discovery [15, 16].

Epistemic constraints further complicate this landscape. In traditional materials science, knowledge accrual is grounded in physical principles, providing interpretable pathways from data to insight. In contrast, data-driven models in materials engineering often operate as black boxes, where scaling enhances accuracy but obscures causal relationships [17, 18]. This opacity hinders the integration of uncertainty quantification, essential for reliable predictions in high-stakes applications like alloy design or perovskite optimization [19, 20]. Moreover, the coupling of simulations with experiments in autonomous discovery systems amplifies these issues, as real-time feedback loops require models that scale efficiently while maintaining fidelity to material realities [21, 22].

Computational design paradigms have attempted to mitigate these limitations through advanced architectures, such as graph neural networks that encode material structures as relational graphs, enabling scalable representations of atomic interactions [23, 24]. Nonetheless, the literature reveals a persistent gap: while model expansion drives short-term gains in predictive power, it often overlooks long-term infrastructure trade-offs, including data curation costs and environmental impacts [25, 26]. This underscores the need for a conceptual reevaluation, one that prioritizes interpretive scaling over brute-force parametrization.

Positioning this analysis within the broader context, we propose a novel framework that conceptualizes model expansion as an epistemic cascade, disentangling data scaling from insight generation. By focusing on the dynamics of representation-inference interactions, this framework offers a systems-level perspective on how computational materials engineering can achieve sustainable growth, steering discovery pipelines toward balanced trade-offs between energy expenditure and knowledge yield. To synthesize these interdependent scaling dimensions and their infrastructural and epistemic consequences, **Table 1** maps the primary axes of physics-agnostic model expansion in computational materials engineering.

Table 1. Scaling Dimensions in Physics-Agnostic Model Expansion for Computational Materials Engineering

Scaling Dimension	Primary Driver	Computational Effect	Epistem Effect
Dataset Scaling	High-throughput simulations, databases	Increased training time, storage demand	Expand chemica search space
Parameter Scaling	Deeper/wider architectures	Higher FLOPs, memory load	Capture comple correlatic
Representation Scaling	Latent embeddings, graph encodings	Feature dimensionality growth	Enhanc structur abstracti
Feedback Scaling	Closed-loop discovery systems	Iterative retraining overhead	Accelera hypothes refineme

Multimodal Scaling	Cross-source data fusion	Alignment + fusion complexity	Broade inferenc contex

Theoretical Background & Literature Synthesis

Materials data infrastructures

The foundation of computational materials engineering rests on robust data infrastructures that aggregate and curate vast repositories of material properties, structures, and behaviors. These infrastructures have evolved to support multimodal datasets, incorporating experimental measurements, simulation outputs, and metadata from diverse sources [1, 27]. High-throughput computation plays a pivotal role, generating standardized data through automated workflows that span crystal structures, electronic properties, and thermodynamic stabilities [4, 9]. However, scaling these datasets introduces challenges in quality control and interoperability, as inconsistencies in representation formats can impede effective machine learning integration [13, 28].

In this context, dataset scaling is not merely quantitative but qualitative, influencing the breadth of chemical spaces explored. For example, initiatives in materials informatics emphasize the creation of comprehensive databases that enable data-driven discovery, yet they highlight the energy costs associated with data generation and storage [7, 29]. The literature underscores the need for efficient data pipelines that minimize redundancy while maximizing coverage, particularly in underrepresented material classes like high-entropy alloys or perovskites [3, 11].

Representation learning architectures

Representation learning has become central to handling the complexity of materials data, transforming raw atomic configurations into latent spaces amenable to machine learning [2, 10]. Graph neural networks exemplify this, modeling materials as graphs where nodes represent atoms and edges denote bonds, facilitating scalable feature extraction [6, 23]. These architectures allow for parameter-efficient scaling, where growth in model depth enhances the capture of hierarchical features without proportional increases in computational overhead [30, 31].

Yet, the expansion of parameters in these models often leads to diminishing returns in insight, as larger representations may entangle relevant signals with noise [16, 18]. Studies in deep learning for materials emphasize the balance between architectural complexity and interpretability, advocating for embeddings that preserve physical symmetries while enabling inverse design [14, 20]. This synthesis reveals a tension: while representation learning drives model expansion, it must align with epistemic goals to yield meaningful discovery gains [12, 24].

AI-Guided discovery

Systems AI integration in materials discovery has led to autonomous systems that orchestrate closed-loop experimentation, iteratively refining hypotheses through data-model feedback [25, 26]. These systems leverage machine learning to prioritize candidates in vast search spaces, accelerating the identification of functional materials [8, 21]. However, without physics-based anchors, scaling in these systems relies on probabilistic inference, where parameter growth enhances exploration but risks epistemic overfitting [5, 22].

The literature highlights how uncertainty quantification integrates into these pipelines, providing confidence estimates that guide resource allocation [17, 19]. In high-throughput settings, AI-guided discovery balances computational costs against potential gains, emphasizing adaptive strategies that scale models dynamically based on interim insights [15, 32].

Computational design paradigms

Inverse design paradigms invert traditional forward modeling, using target properties to navigate material spaces via optimization algorithms [11, 27]. Machine learning accelerates this by scaling search efficiencies, yet it introduces trade-offs in energy consumption as larger models demand intensive training [7, 28]. Conceptual analyses in the field stress the importance of hybrid approaches that couple simulations with data-driven inference, fostering scalable design without exhaustive enumeration [4, 9].

Emerging paradigms in computational materials engineering focus on sustainable scaling, where model expansion is tempered by infrastructure constraints,

ensuring that discovery pipelines remain viable in resource-limited environments [29, 31].

Uncertainty & interpretability

Uncertainty quantification emerges as a critical component in scaling data-driven models, addressing the epistemic risks of overconfident predictions in materials AI [13, 17]. Interpretability frameworks aim to dissect model decisions, revealing how parameter growth influences inference pathways [18, 24]. The synthesis of literature indicates that while scaling enhances predictive scope, it often amplifies interpretability challenges, necessitating conceptual tools that map uncertainty to discovery steering [19, 22].

Proposed conceptual framework

To address the conceptual gaps in model expansion within computational materials engineering, we introduce the Insight Amplification Cascade (IAC) framework. This original structure conceptualizes scaling laws as emergent dynamics in a layered system, disentangling data infrastructures from inference mechanisms to optimize epistemic yields. The IAC comprises three interconnected layers: the Data Aggregation Layer, which scales multimodal inputs; the Representation Expansion Layer, which grows parameters to encode complex interactions; and the Discovery Steering Layer, which balances energy costs against insight generation through feedback loops.

At its core, the IAC maps computational workflows as cascading amplifications, where dataset scaling initiates a chain reaction propagating through model parameters to discovery outcomes. Feedback loops within the framework enable iterative refinement, allowing the system to adaptively modulate expansion based on intermediate epistemic assessments. For instance, in materials informatics, this manifests as dynamic resource allocation, where high-energy computations are prioritized for high-yield inferences.

A key dynamic in the IAC can be conceptualized as the trade-off between parameter proliferation and insight density, expressed as

$$I \approx \alpha * \left(\frac{P^{\beta}}{E^{\gamma}}\right) \quad (1)$$

where I represents epistemic insight, P denotes model parameters, E signifies energy cost, and α, β, γ are scaling coefficients capturing system efficiencies. This formula highlights how insight amplification diminishes marginally with unchecked parameter growth unless mitigated by energy-aware steering.

Another interaction within the IAC captures the feedback between representation learning and uncertainty propagation, which may be expressed as

$$U = f(R) * \int Ddt \quad (2)$$

where U is uncertainty, R is representation complexity, D is dataset flux, and f denotes a functional mapping that amplifies or dampens based on layer interactions. This underscores the temporal dynamics of scaling, emphasizing sustainable cascades over static expansions. The layered structure, amplification mechanisms, and steering trade-offs constituting the Insight Amplification Cascade are structurally consolidated in **Table 2**.

Table 2. Insight Amplification Cascade (IAC): Layered Dynamics and Trade-Off Mechanisms

IAC Layer	Scaling Mechanism	Amplification Function	Trade-Off Risk
Data Aggregation Layer	Dataset volume + diversity growth	Expands searchable materials space	Data noise bias propagatio
Representation Expansion Layer	Parameter + embedding scaling	Encodes multiscale interactions	Interpretabi collapse
Discovery Steering Layer	Feedback + optimization loops	Converts inference → discovery	Energy overuse, search lock
Energy–Insight Interface	Compute expenditure	Enables large-scale inference	Sustainabil burden
Uncertainty Modulation Layer	UQ + confidence modeling	Regulates epistemic reliability	Over/unde confidenc drift

As conceptualized in **Figure 1**, the Insight Amplification Cascade (IAC) framework provides a blueprint for analyzing discovery ecosystems. The figure's layered cascade—from the Data Aggregation Layer through the Representation Expansion Layer to the Discovery Steering Layer—dissects how data steers design. Its annotated feedback arrows (e.g., "Active data selection," "Uncertainty propagation") and visual metaphors, such as narrowing funnels representing bias accumulation and a balance scale illustrating the energy-cost versus epistemic-gain trade-off, encapsulate the framework's utility in engineering more adaptive systems.

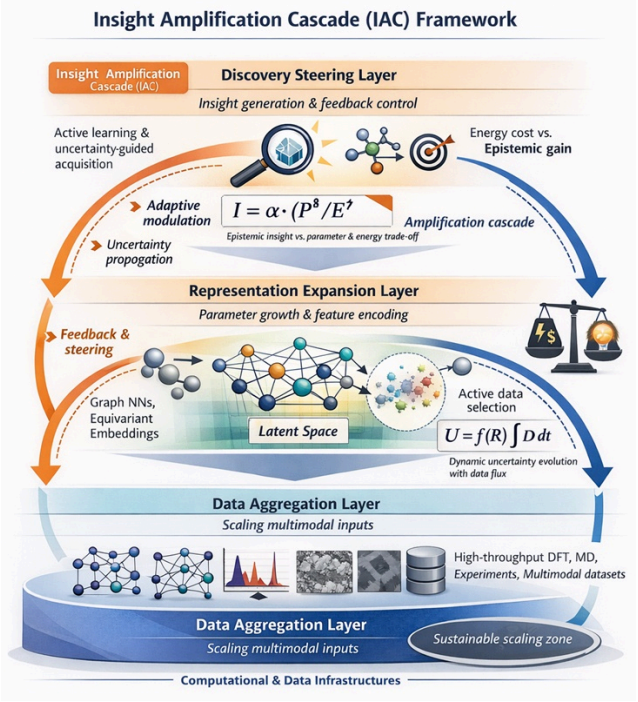

Figure 1. Insight Amplification Cascade (IAC) Framework: A layered architecture showing the progression from data aggregation through representation expansion to steered discovery, mediated by feedback loops and epistemic trade-offs.

Analytical implications

The Insight Amplification Cascade (IAC) framework carries several analytical implications for how model expansion is understood and managed within computational materials engineering. Rather than treating scaling as a monotonic path toward superior performance, the framework reveals it as a multi-stage process governed by cascading dependencies and feedback-mediated trade-offs [25, 26].

One central implication concerns the non-linear relationship between parameter count and epistemic yield. While larger models can encode increasingly intricate correlations within materials data [2, 10], the marginal gain in interpretable insight frequently decays faster than the marginal gain in predictive accuracy [7, 16]. This divergence becomes particularly pronounced when parameter growth outpaces dataset diversity or when representation spaces become overly entangled [13, 18]. The IAC suggests that meaningful scaling requires active modulation at the Representation Expansion Layer—through mechanisms such as sparsity induction, modular decomposition, or adaptive depth—to prevent insight density from collapsing under complexity overhead [30, 31].

A second implication arises from the energy–insight trade-off formalized earlier as $I \approx \alpha * \left(\frac{P^{\beta}}{E^{\gamma}}\right)$. This relationship implies the existence of an optimal operating regime for a given discovery task. Operating far below this regime (small models, limited data) results in under-exploitation of available chemical space [8, 14]; operating far above it produces diminishing epistemic returns at rapidly escalating environmental and economic cost [7, 29]. The framework therefore encourages the development of steering heuristics that continuously estimate the instantaneous values of β and γ during training and inference. Such heuristics could, in principle, be informed by monitoring gradient flow patterns, representation collapse indicators, or uncertainty drift across successive training epochs [17, 19].

A third analytical consequence pertains to feedback loop design. The IAC positions feedback not as an auxiliary mechanism but as the primary steering force that determines whether model expansion becomes amplifying or saturating. In closed-loop discovery pipelines, feedback can operate at multiple timescales: short-term (within a single optimization cycle), medium-term (across batches of candidate evaluation), and long-term (across generations of model retraining) [21, 25]. Each timescale introduces distinct control parameters—learning rate schedules, acquisition function shapes, active learning thresholds—that influence the overall cascade trajectory [26, 32]. Poorly tuned feedback can lock the system into low-yield attractors, where additional parameters merely reinforce existing biases rather than open new discovery channels [5, 22].

A further implication emerges when considering multimodal and multiscale materials data. The Data Aggregation Layer of the IAC must contend with heterogeneous information fluxes: atomic-scale simulation trajectories, mesoscale microstructure statistics, macroscale mechanical testing results, and spectroscopic fingerprints [1, 27]. Scaling laws that appear robust in unimodal settings frequently break down under multimodal expansion unless the Representation Expansion Layer explicitly learns cross-modal alignment and hierarchical abstraction [23, 24]. The framework therefore implies that future infrastructure investments should prioritize cross-modal adapters and multiresolution encoders over sheer parameter count [28, 31].

Finally, the uncertainty propagation expression $\overset{U}{\cdot} = \overset{f(R)}{\int Ddt}$ highlights a temporal dimension often overlooked in scaling discussions [17, 19]. Uncertainty is not a static property of a model snapshot but a dynamic quantity that evolves with data flux and representation growth. Rapid parameter expansion in the presence of slow data accumulation tends to inflate epistemic uncertainty even when aleatoric uncertainty decreases, creating a paradoxical situation where “better” models appear less confident [13, 20]. The IAC implies that sustainable scaling strategies must incorporate explicit temporal regularizers—such as memory replay buffers, continual learning constraints, or uncertainty-preserving distillation—to maintain healthy uncertainty scaling behavior [19, 22].

Results and Discussion

The conceptual shift proposed by the IAC—from viewing model expansion as a primarily quantitative phenomenon to treating it as an epistemic cascade—has broad ramifications for the organization of computational materials research ecosystems.

First, it challenges the prevailing narrative that bigger is almost always better. Although large-scale foundation models have delivered impressive results in language, vision, and even scientific domains, their direct transplantation into materials engineering encounters structural obstacles [2, 23]. Materials data remain fundamentally sparse, noisy, and multiscale compared with text or image corpora [13, 28]. Moreover, the epistemic objective in materials discovery is rarely pure prediction; it is the generation of actionable, interpretable hypotheses that can be tested experimentally or simulated at higher fidelity [11, 27]. The IAC framework suggests that unconstrained scaling often moves the field away from this objective rather than toward it [7, 16].

Second, the framework calls attention to infrastructure-level decisions that are currently under-discussed in the literature. Choices about data curation protocols, compute allocation policies, model checkpointing strategies, and uncertainty reporting standards all shape the effective β, γ, and f(R) coefficients of the cascade [1, 29]. At present, many materials informatics efforts optimize for short-term benchmark scores rather than long-term epistemic efficiency [5, 25]. Shifting incentives toward cascade-aware metrics—such as insight-per-kilowatt-hour, new-discovery-per-parameter, or uncertainty-normalized generalization—could realign community priorities with sustainable progress [26, 32].

Third, the IAC highlights the strategic importance of feedback architecture design. In autonomous discovery systems, the quality of the feedback loop is frequently more consequential than model size [21, 25]. A modestly sized model embedded in a tightly integrated, high-bandwidth experimental–computational feedback loop can outperform a significantly larger model operating in an open-loop or loosely coupled setting [8, 26]. This observation suggests that future resource allocation should favor investments in real-time interfacing, rapid characterization pipelines, and adaptive experimental design over raw FLOPs [15, 32].

Fourth, the framework exposes a subtle risk in the current enthusiasm for foundation models in science. While large pretrained backbones offer transfer learning advantages [2, 30], they also risk imposing generic inductive biases that are poorly matched to materials-specific symmetries and invariances [10, 23]. The IAC implies that foundation-style scaling should be pursued only when accompanied by strong domain-specific alignment mechanisms—physics-informed tokenization, symmetry-equivariant pretraining objectives, or multiscale contrastive losses—otherwise the cascade may amplify irrelevant correlations instead of chemically meaningful ones [14, 24].

Fifth, environmental and societal considerations cannot be decoupled from technical scaling discussions. The exponential growth in training energy documented across machine learning subfields already raises serious sustainability questions [7, 29]; in materials engineering, where discovery timescales span years and experimental

validation remains costly, inefficient scaling imposes compounded burdens. By making energy–insight trade-offs explicit, the IAC provides a conceptual scaffold for discussing responsible scaling practices, including carbon-aware training schedules, model pruning at inference time, and the prioritization of smaller, task-specialized models when appropriate [26, 31].

Taken together, these points suggest that the next phase of progress in computational materials engineering will depend less on achieving ever-larger parameter counts and more on engineering smarter cascades—systems that amplify insight through deliberate orchestration of data, representation, inference, and feedback rather than through brute-force expansion [25, 27].

Conclusion

The accelerating adoption of large-scale data-driven models has brought computational materials engineering to an inflection point. Scaling laws that operate independently of explicit physics offer extraordinary potential to accelerate discovery, yet they simultaneously introduce epistemic, computational, and infrastructural vulnerabilities that are only partially addressed by current paradigms.

The Insight Amplification Cascade framework provides a conceptual lens for navigating this transition. By recasting model expansion as a layered, feedback-driven process rather than a simple quantitative scaling relationship, the IAC exposes critical trade-offs and steering opportunities that are frequently obscured in performance-centric discussions. Its analytical implications—ranging from the non-monotonicity of insight density to the temporal dynamics of uncertainty—offer guidance for designing more sustainable and epistemically productive computational workflows.

Ultimately, the long-term success of data-driven materials discovery will hinge on the field's ability to move beyond size-dominated scaling narratives toward cascade-aware architectures that balance predictive power, interpretive clarity, energy efficiency, and discovery velocity. The IAC is intended as a modest contribution to that reorientation: a systems-level map that invites researchers to ask not merely "how large can we make the model?" but "how intelligently can we amplify insight?"

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 09 Jun 2021 Revised: 29 Sep 2021 Accepted: 01 Nov 2021
Published online: 18 March 2022

References

Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C. Machine learning in materials informatics: Recent applications and prospects. npj Comput Mater. 2017;3(1):54. https://doi.org/10.1038/s41524-017-0056-5.

Schmidt J, Marques MRG, Botti S, Marques MAL. Recent advances and applications of machine learning in solid-state

materials science. npj Comput Mater. 2019;5(1):83. https://doi.org/10.1038/s41524-019-0221-0.

Tao Q, Xu P, Li M, Lu W. Machine learning for perovskite materials design and discovery. npj Comput Mater. 2021;7(1):23. https://doi.org/10.1038/s41524-021-00493-1.

Hara P, Puggioni D, Rondinelli JM. High-throughput density functional perturbation theory and machine learning predictions of infrared, piezoelectric, and dielectric responses. npj Comput Mater. 2020;6(1):65. https://doi.org/10.1038/s41524-020-0337-2.

Chan H, Cherukara MJ, Loeffler TD, Narayanan B, Sankaranarayanan SKRS. Machine learning enabled autonomous microstructural characterization in 3D samples. npj Comput Mater. 2020;6(1):25. https://doi.org/10.1038/s41524-020-0287-6.

Du P, Zeng H, Zhao C, Lu L, Guo L. Microstructure design using graphs. npj Comput Mater. 2021;7(1):22. https://doi.org/10.1038/s41524-021-00490-4.

Zhang Y, Ling C. A strategy to apply machine learning to small datasets in materials science. npj Comput Mater. 2018;4(1):25. https://doi.org/10.1038/s41524-018-0081-2.

Kumar JN, Li Q, Tang KYT, Buonassisi T, Gonzalez MA, Allen JM. Machine learning enables polymer cloud-point engineering via inverse design. npj Comput Mater. 2019;5(1):52. https://doi.org/10.1038/s41524-019-0191-2.

Jacobs R, Mayeshiba T, Afflerbach B, Miles L, Williams M, Turner A, et al. The materials simulation toolkit for machine learning (MAST-ML): An automated open source toolkit to accelerate data-driven materials research. Comput Mater Sci. 2020;175:109599. https://doi.org/10.1016/j.commatsci.2020.109599.

Lu S, Zhou Q, Guo Y, Zhang Y, Wu Y, Wang J. Graph convolutional neural networks improve accuracy of defect classification in crystal structures. npj Comput Mater. 2020;6(1):160. https://doi.org/10.1038/s41524-020-00427-z.

Kaufmann K, Maryanovsky D, Mellor WM, Zhu C, Rosengarten AS, Harrington TJ, Oses C, et al. Discovery of high-entropy ceramics via machine learning. npj Comput Mater. 2020;6(1):42. https://doi.org/10.1038/s41524-020-0317-6.

Saidi WA, Shadid W, Veser G. Deep learning in electron microscopy. npj Comput Mater. 2021;7(1):56. https://doi.org/10.1038/s41524-021-00523-y.

Lu H-J, Zou N, Jacobs R, Afflerbach B, Lu X-G, Morgan D. Error assessment and optimal cross-validation approaches in machine learning applied to impurity diffusion. Comput Mater Sci. 2019;169:109075. https://doi.org/10.1016/j.commatsci.2019.109075.

Xue D, Balachandran PV, Yuan R, Hu T, Qian X, Dougherty ER, et al. Accelerated search for biomimetic materials with machine learning. npj Comput Mater. 2021;7(1):18. https://doi.org/10.1038/s41524-021-00491-3.

Kaufmann K, Vecchio KS. Searching for high entropy alloys: A machine learning approach. Acta Mater. 2020;198:231-40. https://doi.org/10.1016/j.actamat.2020.07.054.

Huang W, Martin P, Zhuang HL. Machine-learning phase prediction of high-entropy alloys. Acta Mater. 2019;169:225-36. https://doi.org/10.1016/j.actamat.2019.03.012.

Möller JJ, Körner W, Krugel G, Urban DF, Elsässer C. Compositional optimization of hard-magnetic phases with machine-learning models. Acta Mater. 2018;153:53-61. https://doi.org/10.1016/j.actamat.2018.04.028.

Alcobaça E, Mastelini SM, Botari T, Pimentel BA, Cassar DR, de Carvalho et al. Explainable machine learning algorithms for predicting glass transition temperatures. Acta Mater. 2020;188:92-102. https://doi.org/10.1016/j.actamat.2019.12.037.

Zhang H, Fu H, Zhu S, Yong W, Sun J, He Y, et al. Machine learning assisted composition effective design for precipitation strengthened copper alloys. Acta Mater. 2021;215:117118. https://doi.org/10.1016/j.actamat.2021.117118.

He J, Li J, Liu C, Wang C, Wang X, Liu B, et al. Machine learning identified materials descriptors for ferroelectricity. Acta Mater. 2021;209:116815. https://doi.org/10.1016/j.actamat.2021.116815.

Zou C, Li J, Wang WY, Zhang D, Froese R, Tang D, et al. Integrating data mining and machine learning to discover high-strength ductile titanium alloys. Acta Mater. 2021;202:211-21. https://doi.org/10.1016/j.actamat.2020.10.056.

Curnan MT, Saidi WA, Yang JC, Han JW. Grain boundary mobilities in polycrystals: A data-driven approach. Acta Mater. 2021;209:116804. https://doi.org/10.1016/j.actamat.2021.116804.

Keith JA, Vassilev-Galindo V, Cheng B, Chmiela S, Gastegger M, Müller KR, et al. Combining machine learning and computational chemistry for predictive insights into chemical systems. Chem Rev. 2021;121(15):9816-72. https://doi.org/10.1021/acs.chemrev.1c00107.

Boyd PG, Lee Y, Smit B. Computational development of the nanoporous materials genome in search of novel hyper-confined fluids. Chem Rev. 2017;117(13):9145-65. https://doi.org/10.1021/acs.chemrev.7b00098.

Flores-Leonar MM, Mejía-Mendoza LM, Domratcheva T, Farias-Rodríguez V, Vargas R, Aspuru-Guzik A, et al. High-throughput experimentation meets artificial intelligence: A new pathway to the discovery of materials. Nat Mach Intell. 2020;2(7):367-76. https://doi.org/10.1038/s42256-020-0203-8.

Lu S, Zhou Q, Ouyang Y, Guo Y, Li Q, Wang J. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. Nat Commun. 2018;9(1):3405. https://doi.org/10.1038/s41467-018-05761-6.

Ren Z, Tian SIP, Noh J, Oviedo F, Xing G, Li J, et al. Inverse design of solid-state materials via a continuous representation. Matter. 2019;1(6):1528-42. https://doi.org/10.1016/j.matt.2019.08.017.

Tian SIP, Kim SY, McDermott MJ, Bartel CJ, Ceder G. Data-driven materials discovery from large chemistry spaces. Matter. 2020;3(4):961-3. https://doi.org/10.1016/j.matt.2020.08.019.

Wu CT, Chang HT, Wu CY, Chen SW, Huang SY, Liu M, et al. Machine learning recommends affordable new Ti alloy with bone-like modulus. Mater Today. 2020;34:41-50. https://doi.org/10.1016/j.mattod.2019.12.003.

Juneja R, Singh AK. Machine learning for advanced functional materials. Adv Intell Syst. 2020;2(6):2000023. https://doi.org/10.1002/aisy.202000023.

Xiao D, Hu C, Du L, Chen C, Pan J, Chen G. Machine learning for materials research and development. Adv Intell Syst. 2021;3(6):2000235. https://doi.org/10.1002/aisy.202000235.

Dan Y, Zhao T, Zhang X, Zhao H, Li X, Li D. Data-driven construction of a design space for two-stage bulk gas separation processes. npj Comput Mater. 2018;4(1):64. https://doi.org/10.1038/s41524-018-0114-z.