ORIGINAL RESEARCH **Open access**

Pretraining on Matter: Conceptual Limits of Foundation Models for Materials Science

Claire Dupont[1*], Julien Martin[1]

Abstract

The advent of foundation models, large-scale pre-trained architectures adapted from natural language processing paradigms, has permeated computational materials science, promising accelerated discovery through data-driven inference. In materials engineering, these models leverage multimodal datasets encompassing atomic structures, properties, and simulations to enable representation learning across scales. However, inherent conceptual limits arise from the interplay between materials' physical hierarchies—spanning quantum to macroscopic levels—and the inductive biases embedded in pretraining strategies. This manuscript synthesizes recent advancements in machine learning architectures, such as graph neural networks and multimodal integration, within materials informatics ecosystems. It identifies epistemic boundaries where foundation models falter in capturing causality, uncertainty, and domain-specific invariances, potentially leading to misaligned discovery pipelines. To address these, we introduce the Matter Pretraining Boundary Framework (MPBF), a conceptual architecture that delineates layers of data assimilation, representational abstraction, and inference steering to mitigate limits in autonomous materials design. Implications extend to high-throughput computation, inverse design, and simulation-experiment coupling, fostering more robust computational workflows in materials engineering. By interpreting these limits through systems-level dynamics, the framework guides infrastructure trade-offs, enhancing the reliability of data-driven paradigms without empirical validation.

Keywords Autonomous discovery, Materials informatics, Uncertainty quantification, Graph neural networks, Representation learning, Foundation models

*Correspondence:
Claire Dupont
claire.dupont@gmail.com
[1] Department of Materials Data Analytics, Faculty of Engineering, University of Bordeaux, Bordeaux, France

Introduction

The evolution of data-driven paradigms in materials science

Over the past decade, computational materials engineering has undergone a structural transformation from simulation-centric inquiry toward fully data-driven discovery ecosystems. Historically, materials innovation relied on physics-based modeling frameworks—most prominently density functional theory (DFT), molecular dynamics, and thermodynamic phase calculations—to predict stability landscapes and functional properties. While these methods remain foundational, their computational intensity and scaling limitations have increasingly necessitated complementary algorithmic strategies capable of navigating exponentially expanding chemical and configurational design spaces [1-3].

The emergence of materials informatics marks a critical inflection point in this transition. By integrating machine learning with high-throughput simulation infrastructures and experimental repositories, researchers have constructed predictive pipelines capable of screening millions of

candidate materials with unprecedented speed. Data resources derived from combinatorial synthesis, automated characterization, and computational databases now serve as substrates for supervised and self-supervised learning systems. These infrastructures enable rapid prediction of alloy compositions, phase stabilities, bandgaps, ionic conductivities, and mechanical performance metrics, significantly reducing dependence on sequential trial-and-error experimentation [4-7].

Beyond acceleration, the data-driven paradigm has reshaped the epistemic logic of materials discovery itself. Rather than iteratively validating physics-derived hypotheses, contemporary workflows increasingly deploy predictive inference to prioritize experimental directions. In this inversion, algorithmic screening precedes mechanistic interpretation, effectively reordering the traditional discovery sequence.

Central to this evolution is the maturation of deep learning architectures adapted to the structural particularities of materials systems. Graph neural networks (GNNs), which encode atomic connectivity through node–edge representations, have emerged as a cornerstone for modeling crystalline lattices, amorphous networks, and molecular assemblies [8-10]. Through message-passing operations, these networks learn localized chemical environments while preserving relational topology, enabling prediction of formation enthalpies, elastic tensors, electronic densities, and defect energetics without direct recourse to quantum mechanical solvers [11-13].

Simultaneously, advances in representation learning have expanded the scope of materials embeddings beyond purely structural encodings. Multimodal learning frameworks now integrate crystallographic descriptors, spectroscopy signatures, microscopy imagery, and even textual synthesis records into unified latent spaces [14-16]. Such embeddings facilitate cross-modal inference—linking processing conditions to microstructural evolution or defect distributions to functional performance.

These developments signal a broader architectural ambition: the construction of generalizable materials models capable of transferring learned knowledge across compositional families and structural classes—from metallic alloys to polymeric networks and ceramic oxides. In this sense, the field is progressing toward foundation-like infrastructures wherein pretrained models function as universal predictors within materials design pipelines.

Challenges in scaling computational discovery

Despite these advances, scaling data-driven methods to foundation model regimes—characterized by massive pretraining on diverse corpora—introduces conceptual hurdles specific to materials science. Unlike domains such as computer vision or language, materials data exhibit sparsity, heterogeneity, and adherence to physical laws that constrain extrapolation [17-19]. Pretraining objectives, often borrowed from self-supervised learning in other fields, may inadequately capture the multi-scale nature of materials phenomena, where atomic-level interactions propagate to macroscopic properties [20-22]. This mismatch can manifest in poor generalization, particularly in inverse design tasks where models must navigate from desired properties back to viable structures [23-25].

Furthermore, uncertainty quantification remains a critical yet underexplored aspect in these paradigms. In computational workflows, epistemic uncertainties arising from incomplete data coverage or model approximations can propagate through discovery pipelines, potentially steering autonomous systems toward suboptimal explorations [26-29]. High-throughput frameworks, while efficient, often overlook these uncertainties, leading to brittle inferences in closed-loop setups that couple simulations with experimental feedback [29, 30]. The literature highlights instances where machine learning interatomic potentials, though accurate for specific regimes, fail to maintain fidelity across phase transitions or under external perturbations [16, 23].

Bridging representation and inference in materials AI

To address these challenges, recent efforts have focused on hybrid approaches that infuse physical priors into learning architectures. For example, incorporating symmetry-aware features or kinetic theories enhances model robustness, enabling better handling of disordered systems like high-entropy alloys [5, 9, 26]. Yet, as foundation models grow in ambition, aiming for "universal" materials predictors, conceptual limits become evident in their epistemic scope. These limits stem not from computational power but from the fundamental tension between data-driven induction and the deductive nature of physical principles [2, 6, 15].

This manuscript positions a novel conceptual framework to interpret these boundaries, emphasizing the dynamics of pretraining on matter-specific data. By dissecting the layers of data assimilation, model abstraction, and discovery steering, it offers systems-level insights into optimizing foundation models for materials engineering. The Matter Pretraining Boundary Framework (MPBF) delineates these interactions, providing a lens for infrastructure trade-offs in computational ecosystems.

Theoretical Background & Literature Synthesis

Foundational architectures in materials machine learning

The integration of machine learning into materials science has been catalyzed by the emergence of architectures explicitly designed to encode structural, compositional, and relational information intrinsic to materials systems. Among these, graph neural networks (GNNs) have assumed a foundational role due to their capacity to represent crystalline and molecular structures as graphs, where atoms are modeled as nodes and interatomic interactions as edges. Through iterative message-passing operations, GNNs capture local chemical environments while propagating contextual information across coordination shells, enabling the prediction of thermodynamic, electronic, and mechanical properties with high fidelity [8, 10, 24].

This architectural paradigm has demonstrated versatility across diverse materials classes. Applications include formation enthalpy prediction in transition metal nitrides, elastic tensor estimation in complex oxides, and thermal transport modeling in disordered solid solutions [4, 27]. Importantly, the success of GNNs lies not merely in predictive accuracy but in their structural alignment with materials physics: translational invariance, permutation symmetry, and locality constraints are inherently respected within graph formulations. Consequently, these models serve as bridges between atomistic simulation traditions and data-driven inference infrastructures.

Complementary deep learning strategies have further enriched this architectural landscape. Subgroup discovery networks, symbolic regression systems, and genetic programming frameworks have been deployed to enhance interpretability—an increasingly critical concern in materials informatics. By extracting low-dimensional descriptors governing complex phenomena such as glass formability, metamaterial resonance behavior, or catalytic activity landscapes, these methods transform opaque predictive systems into interpretable scientific instruments [2, 6]. In doing so, they operationalize a dual function: prediction and mechanistic hypothesis generation.

The convergence of these approaches reflects a broader epistemic shift. Materials ML architectures are no longer evaluated solely on predictive performance but also on their capacity to encode domain knowledge, expose governing descriptors, and align with theoretical priors. This shift foregrounds interpretability, causality, and physical plausibility as co-equal design objectives alongside accuracy.

Representation learning and latent materials spaces

Representation learning extends architectural advances by focusing on how materials knowledge is encoded rather than solely how predictions are produced. Latent embeddings generated through deep representation frameworks aim to capture intrinsic materials features—composition, symmetry, bonding topology, and microstructural morphology—within continuous vector spaces.

Variational autoencoders (VAEs), generative adversarial networks (GANs), and diffusion-based generative models have demonstrated particular promise in this domain. These systems enable microstructure reconstruction, defect morphology synthesis, and composition generation, thereby addressing chronic data scarcity in specialized materials regimes [18, 19, 29]. By learning probability distributions over materials representations, generative models facilitate inverse exploration—sampling hypothetical materials that satisfy targeted structural or property constraints.

A notable extension of representation learning lies in the development of machine-learned interatomic potentials. Spline-based neural potentials and equivariant neural networks integrate classical force-field formulations with data-driven flexibility, enabling high-fidelity simulations across extended timescales [16, 23]. These hybrid potentials preserve physical continuity and energy conservation while learning complex interaction landscapes from quantum datasets.

This synthesis between classical physics and learned representations underscores a defining systems principle: architectural legitimacy in materials AI depends on symmetry compliance and physical invariance. Models that ignore rotational equivariance, conservation laws, or bonding constraints risk producing representations that are statistically predictive yet physically incoherent [14, 19]. Consequently, contemporary research increasingly prioritizes physics-informed architectures that embed conservation laws, tensor symmetries, and geometric equivariances directly into network design.

Pretraining strategies and their adaptation to materials domains

Pretraining has emerged as a transformative paradigm within artificial intelligence, enabling models to learn generalized representations from large unlabeled corpora prior to task-specific adaptation. In materials science, this paradigm is operationalized through exposure to expansive repositories of crystal structures, computed energetics, spectroscopy data, and phase diagrams [17, 22, 25].

Self-supervised learning objectives have proven especially effective in this context. Tasks such as masked atom prediction, crystal graph completion, and contrastive structure matching compel models to infer hidden structural regularities without explicit property labels [5, 24, 26]. Through these mechanisms, pretrained materials models acquire transferable chemical priors applicable to downstream tasks including bandgap prediction, defect energetics, and catalytic screening.

However, materials pretraining diverges fundamentally from linguistic or visual foundation modeling. Domain constraints impose multimodal complexity: datasets span density functional theory (DFT) outputs, diffraction spectra, microscopy images, and textual synthesis protocols. Integrating these heterogeneous modalities necessitates fusion architectures capable of aligning disparate representational scales—from quantum orbitals to mesoscale textures [15, 18].

High-throughput computational infrastructures play a critical enabling role in this pretraining ecosystem. Automated DFT workflows, cluster expansions, and surrogate modeling pipelines generate millions of computed datapoints across chemical spaces, furnishing the scale required for representation generalization [4, 20, 26]. Machine-learning-accelerated simulations further reduce computational cost, enabling iterative dataset expansion.

Yet the literature consistently identifies scaling asymmetries. Simulated datasets inherit approximations embedded in their generating theories. Exchange–correlation functional biases, pseudopotential assumptions, and convergence heuristics propagate into pretrained models, embedding systematic epistemic distortions [12, 13, 16]. Consequently, models pretrained exclusively on simulated corpora may exhibit degraded transferability to experimental domains.

Uncertainty quantification has emerged as a partial corrective mechanism. Bayesian neural networks, ensemble modeling, evidential deep learning, and feature attribution screening enable confidence estimation across predictions [7, 9, 27]. Rather than treating outputs as deterministic, these systems produce probabilistic property distributions, allowing downstream workflows to incorporate epistemic risk into decision processes.

Autonomous discovery and closed-loop systems

Autonomous discovery platforms represent the operational culmination of materials machine learning infrastructures. These systems integrate predictive models, generative design engines, and experimental or simulation feedback into iterative closed loops capable of self-directed exploration [3, 28, 30].

Within such architectures, machine learning models propose candidate materials, which are subsequently evaluated through robotic synthesis, high-throughput experimentation, or accelerated simulation. Performance outcomes are reintegrated into model training, enabling adaptive search trajectory refinement [1, 7, 21]. This cyclical paradigm compresses traditional discovery timelines, transforming hypothesis testing into continuous optimization.

Inverse design constitutes a central operational logic within autonomous systems. By mapping desired properties—ionic conductivity, catalytic selectivity, fracture toughness—back to candidate structures, generative models navigate design spaces inaccessible to forward screening [25, 29]. Latent space optimization, reinforcement learning, and Bayesian optimization frequently guide this reverse mapping.

Despite these advances, closed-loop infrastructures remain vulnerable to epistemic instabilities. Errors introduced in early predictive stages can propagate through synthesis prioritization, experimental allocation, and retraining cycles, amplifying uncertainty across iterations [11, 14, 28]. Feedback amplification effects may lead to premature convergence around local optima or data-dense regions, constraining exploratory diversity.

Machine-learned interatomic potentials illustrate this tension. While quantum-trained potentials enable rapid molecular dynamics simulations, extrapolation into chemically novel regimes often produces unreliable energetics [16, 23]. Similarly, generative microstructure models can reproduce training morphologies yet struggle to capture long-range evolutionary kinetics or processing histories [18, 22, 29].

These limitations foreground the need for steering logics—governance frameworks that regulate exploration, uncertainty propagation, and design prioritization within autonomous loops.

Epistemic risks in multimodal integration

The expansion of multimodal datasets has introduced both representational richness and epistemic complexity. Integrating crystallographic data, microscopy imagery, spectroscopy signals, and textual synthesis metadata enables holistic materials modeling beyond single-modality constraints [15, 17, 19].

Foundation-like multimodal architectures aim to unify these data streams within shared latent spaces. In principle, such fusion enables cross-scale inference—linking atomic defects to mesoscale textures or processing conditions to functional performance.

However, modality alignment introduces structural risks. Disparities in spatial scale, noise distributions, and sampling densities can produce representational distortions. Atomic-resolution simulation data may dominate latent embeddings, marginalizing lower-resolution experimental signals. Conversely, imaging datasets may encode morphological correlations lacking atomistic causality.

Empirical studies in defect prediction, crystallographic texture design, and anisotropy engineering illustrate these tensions. Models trained on multimodal corpora often perform robustly within controlled distributions yet degrade under extrapolative generalization [1, 13, 30]. This reflects incomplete modality coverage and latent misalignment.

Uncertainty quantification frameworks remain underdeveloped in multimodal contexts. While predictive variance can be estimated within single modalities, cross-modal uncertainty propagation remains poorly formalized [9, 27]. As a result, confidence metrics may underestimate epistemic gaps introduced during data fusion.

Synthesis: Architectural acceleration vs epistemic constraint

Taken collectively, the literature reveals a computational ecosystem characterized by accelerating architectural sophistication coupled with persistent epistemic constraints. Foundational architectures enable structure-aware inference; representation learning encodes latent materials knowledge; pretraining scales generalization; and autonomous loops operationalize discovery acceleration.

Yet these advances introduce new interpretive burdens. Representation spaces may encode simulation biases. Multimodal fusion may distort cross-scale relationships. Closed-loop optimization may amplify uncertainty. Generative exploration may prioritize statistical plausibility over physical realizability.

Thus, pretraining and foundation-like modeling function simultaneously as discovery accelerants and epistemic risk multipliers. The representation–inference interface emerges as a critical governance frontier requiring interpretive frameworks capable of diagnosing bias propagation, uncertainty amplification, and modality misalignment.

In this context, conceptual infrastructures—rather than purely algorithmic innovations—become essential. Interpretive systems that map architectural capabilities to epistemic limits are necessary to guide responsible scaling, infrastructure design, and autonomous discovery governance [2, 6, 10].

Proposed conceptual framework

The Matter Pretraining Boundary Framework (MPBF)

To interpret the conceptual limits of foundation models in materials science, we propose the Matter Pretraining

Boundary Framework (MPBF), an original architecture that structures the interplay between data assimilation, representational abstraction, and inference steering. MPBF conceptualizes pretraining as a bounded process, where materials' physical hierarchies impose epistemic constraints on model generalization. Unlike existing paradigms that focus on empirical scaling, MPBF emphasizes systems-level dynamics, delineating layers that capture trade-offs in computational workflows. The functional roles, boundary conditions, and discovery implications associated with each MPBF layer are summarized in **Table 1.**

Table 1. Matter Pretraining Boundary Framework (MPBF): Layers, functions, boundary conditions, and discovery implications in foundation materials models

MPBF Layer	Core Function in Pretraining Ecosystem	Dominant Data / Model Mechanisms	Ep Bo Co
Data Assimilation Layer (DAL)	Aggregates and harmonizes multimodal matter-specific corpora for foundation model pretraining	Crystal graphs; DFT trajectories; spectroscopy; microscopy; synthesis metadata; multi-fidelity datasets	Data simu tr m iml pro noise gap chem
Representational Abstraction Layer (RAL)	Encodes structural, compositional, and symmetry-aware invariances into latent embeddings	Graph neural encoders; equivariant networks; masked structure prediction; contrastive learning; cross-modal fusion	abs inv defi eros collap doma misa (qu r
Inference Steering Layer (ISL)	Governs how pretrained representations drive discovery decisions and optimization loops	Inverse design engines; generative models; Bayesian optimization; active learning; candidate ranking systems	Infer fe amp un under ob misa
			explo
Cross-Layer Coupling Dynamics	Regulates error propagation and corrective feedback across the MPBF stack	Inner loop: model fine-tuning; outer loop: data reacquisition; uncertainty propagation channels	Casc com unc reinfo den mis re s

At its core, MPBF comprises three interconnected layers: the Data Assimilation Layer (DAL), which ingests multimodal materials data; the Representational Abstraction Layer (RAL), which distills invariances through pretraining objectives; and the Inference Steering Layer (ISL), which guides discovery pipelines via feedback mechanisms. These layers form a pipeline where data flows upward, abstractions refine downward, and steering modulates laterally, as conceptualized in **Figure 1**.

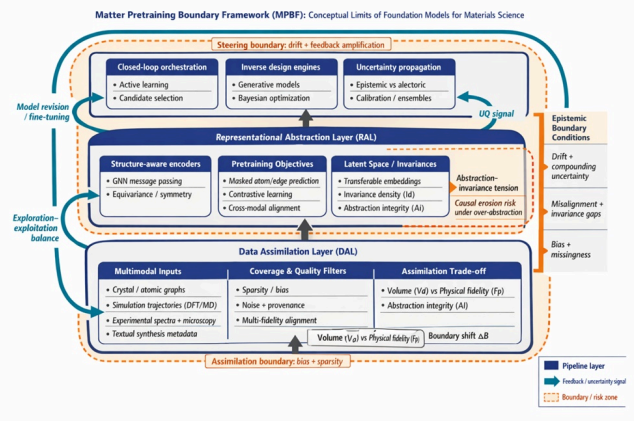

Figure 1. Matter Pretraining Boundary Framework (MPBF) for foundation models in materials science.

MPBF conceptualizes pretraining as a bounded process constrained by matter-specific hierarchies and epistemic risks. The Data Assimilation Layer (DAL) aggregates multimodal corpora and exposes a volume–fidelity trade-off that shifts boundary conditions under sparsity and bias. The Representational Abstraction Layer (RAL) encodes structure-aware invariances through pretraining objectives

but faces abstraction–causality tensions under cross-scale compression. The Inference Steering Layer (ISL) governs closed-loop discovery (inverse design, active learning, uncertainty propagation) while managing drift and feedback amplification. Teal feedback pathways depict inner (model revision) and outer (data re-acquisition) loops, and orange envelopes denote boundary zones where generalization can degrade without uncertainty-aware steering.

In the DAL, heterogeneous inputs—such as atomic graphs, simulation trajectories, and property spectra—are harmonized, but limits arise from sparsity and physical inconsistencies. This layer's dynamics can be conceptualized as a trade-off between data volume V_d and fidelity F_p, expressed as $\frac{\Delta B}{} = \alpha\left(Vd - \beta Fp\right),$ where ΔB represents the boundary shift due to assimilation imbalances, α and β symbolize scaling factors for domain coverage and physical priors, respectively. This captures the interaction between expansive pretraining corpora and the need for matter-specific constraints, preventing overgeneralization.

The RAL builds abstractions via graph-based or multimodal encodings, embedding symmetries and hierarchies. Here, conceptual limits manifest in the abstraction-invariance tension, where over-abstraction erodes causal links. This may be expressed as $\begin{matrix} Ai \\ = \gamma \int R\left(e\right) de \\ - \delta Uq, \end{matrix}$ with A_i denoting abstraction integrity, R(e) the representational embedding function over epistemic elements e, γ an integration coefficient, and δU_q the uncertainty penalty from quantum-to-macro transitions. This formula interprets how pretraining objectives interact with multi-scale uncertainties, informing robust representations.

Finally, the ISL incorporates steering logics for autonomous systems, modulating feedback in closed-loop designs. Limits in this layer stem from inference drift, where pre-trained biases misalign with discovery goals. The steering dynamics capture the feedback between model outputs O_m and pipeline corrections C_d, formalized as $\begin{matrix} T_r Sf \\ = \epsilon\left(Om \circ Cd\right) \\ + \zeta Tr, \end{matrix}$ where S_f is the steering fidelity, ∘ denotes compositional interaction, ϵ and ζ weight feedback and trade-off resolutions T_r respectively. This highlights epistemic risk structures in inverse design and high-throughput loops.

Figure 1 illustrates these layers as a cyclic diagram, with DAL at the base feeding into RAL via upward arrows, ISL encircling for lateral modulation, and dashed boundaries indicating epistemic limits. Feedback loops, shown as bidirectional paths, emphasize iterative refinement without empirical closure. Through MPBF, computational steering logics emerge as interpretive tools, balancing data-driven induction with materials' inherent structures.

Analytical implications

The MPBF offers interpretive lenses for dissecting the dynamics of foundation models in materials science, revealing infrastructure trade-offs that influence discovery workflows. By framing pretraining limits through layered interactions, it highlights how data assimilation imbalances can cascade into representational distortions, affecting high-throughput screenings and autonomous systems [3, 4, 20]. For instance, in representation learning for alloys, the DAL's fidelity constraints underscore the risk of overfitting to simulated datasets, where physical priors are underrepresented, leading to epistemic gaps in phase stability predictions [5, 7, 26].

Systems-level insights from MPBF illuminate feedback loops in closed-loop experimentation, where ISL modulations counteract inference drift. This interpretive approach suggests that incorporating uncertainty-aware steering can enhance robustness in inverse design, mitigating the propagation of biases from pretraining [9, 13, 28]. Consider the interaction between multimodal integration and discovery steering: epistemic risk structures arise when RAL abstractions fail to align with matter's hierarchical scales, potentially steering models toward non-physical extrapolations [15, 18, 19].

A key trade-off captured by MPBF involves the balance between pretraining scale and domain invariance. This can be conceptualized $\begin{matrix} s\ T_s \\ = \eta\ \left(\frac{S_p}{I_d}\right) \\ - \theta\ E_r \end{matrix}$, where T_s denotes the trade-off scalar, S_p the pretraining scale, I_d the invariance density from materials symmetries, η and θ adjustment coefficients, and E_r the epistemic risk from unmodeled hierarchies. This expression interprets how excessive scale without invariance tuning amplifies risks in applications like defect modeling or thermal transport [8, 11, 21].

Further implications extend to computational steering in high-entropy materials, where MPBF's layers guide the

fusion of machine learning potentials with kinetic theories [8, 11, 21]. By interpreting representation-inference interactions, MPBF fosters workflows that prioritize causal alignment over sheer predictive accuracy, informing infrastructure designs that couple simulations with experiments more effectively [16, 23, 30]. Ultimately, these analytical implications steer toward resilient ecosystems, where conceptual limits are not barriers but navigational tools for data-driven materials engineering.

Results and Discussion

The MPBF advances a nuanced understanding of foundation models' conceptual boundaries in materials science, integrating literature on graph neural networks, uncertainty quantification, and autonomous discovery [3, 8, 10, 27]. While pretraining paradigms accelerate inference, MPBF's interpretive framework exposes vulnerabilities in data-model-discovery pipelines, such as those in microstructure reconstruction or alloy design [1, 18, 25]. This systems perspective encourages reevaluation of multimodal datasets, ensuring that representational abstractions respect physical constraints without empirical overreach [15, 17].

In broader computational contexts, MPBF's trade-off formalizations highlight opportunities for hybrid architectures, blending learned potentials with classical simulations to address epistemic risks [9, 16, 23]. However, it also underscores persistent challenges in scaling to disordered systems, where feedback loops must adapt to evolving uncertainties [11, 14, 26]. By focusing on workflow dynamics rather than benchmarks, MPBF contributes to epistemic risk management, potentially influencing fields beyond materials, like chemical informatics [2, 6].

Limitations of this conceptual approach include its reliance on interpretive reasoning, which, while integrative, defers empirical validation to future implementations. Nonetheless, MPBF provides a foundational tool for steering data-driven paradigms toward sustainable discovery infrastructures.

Conclusion

Foundation models hold transformative potential for computational materials engineering, yet their conceptual limits demand structured interpretation to realize robust applications. The MPBF delineates these boundaries through layered dynamics, offering insights into data assimilation, abstraction, and steering that mitigate epistemic risks [1-30]. By formalizing trade-offs and interactions, it guides infrastructure enhancements in high-throughput, inverse design, and closed-loop systems, fostering a balanced integration of machine learning with materials' physical realities. This framework paves the way for more reliable, interpretive workflows in the evolving landscape of data-driven discovery.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 10 Dec 2023 Revised: 21 Mar 2024 Accepted: 05 Apr 2024
Published online: 18 September 2024

References

Ahmadikia B, Paraskevas O, Van Hyning W, Hestroffer JM, Beyerlein IJ, Thrampoulidis C. Data-driven texture design for reducing elastic and plastic anisotropy in titanium alloys. Acta Mater. 2024;265:119585.

Liu G, Sohn S, O'Hern CS, Gilbert AC, Schroers J. Effective subgrouping enhances machine learning prediction in complex materials science phenomena: Inoue's subgrouping in discovering bulk metallic glasses. Acta Mater. 2024;265:119590.

Trehern W, Ortiz-Ayala R, Atli KC, Arroyave R, Karaman I. Data-driven shape memory alloy discovery using artificial intelligence materials selection (AIMS) framework. Acta Mater. 2022;228:117751.

Zhang J, Kong Y, Chen L, Koutná N, Mayrhofer PH. Predicting the formation enthalpy and phase stability of (Ti,Al,TM)N (TM = III-VIB group transition metals) by high-throughput ab initio calculations and machine learning. Acta Mater. 2024;276:120139.

Hou S, Sun M, Bai M, Lin D, Liu W. A hybrid prediction frame for HEAs based on empirical knowledge and machine learning. Acta Mater. 2022;228:117742.

Zhao S, Zhang Y, Zhang Y, Zhang W, Kitipornchai S. Genetic programming-assisted micromechanical models of graphene origami-enabled metal metamaterials. Acta Mater. 2022;228:117791.

Yuan R, Liu Z, Xu Y, Yin R, Lookman T. Optimizing electrocaloric effect in barium titanate-based room temperature ferroelectrics: Combining landau theory, machine learning and synthesis. Acta Mater. 2022;235:118054.

Liu P, Huang H, Jiang X, Zhang Y, Su Y. Evolution analysis of γ' precipitate coarsening in Co-based superalloys using kinetic theory and machine learning. Acta Mater. 2022;235:118101.

Wei X, van der Zwaag S, Jia Z, Wang C, Xu W. On the use of transfer modeling to design new steels with excellent rotating bending fatigue resistance even in the case of very small calibration datasets. Acta Mater. 2022;235:118103.

He J, Li J, Liu C, Wang C, Bai Y. Machine learning identified materials descriptors for ferroelectricity. Acta Mater. 2021;209:116815.

Wang B-Q, Zhao T-Y, Ding H-R, Liu Y-T, Li X-B. Partial melting nature of phase-change memory Ge-Sb-Te superlattice uncovered by large-scale machine learning interatomic potential molecular dynamics. Acta Mater. 2024;276:120123.

He H, Zhao J, Byggmästar J, He R, Djurabekova F. Threshold displacement energy map of Frenkel pair generation in β-Ga2O3 from machine-learning-driven molecular dynamics simulations. Acta Mater. 2024;276:120087.

Kim G, Diao H, Lee C, Samaei AT, Chen W. First-principles and machine learning predictions of elasticity in severely lattice-distorted high-entropy alloys with experimental validation. Acta Mater. 2019;181:124-38.

Sherman S, Simmons J, Przybyla C. Mesoscale characterization of continuous fiber reinforced composites through machine learning: Fiber chirality. Acta Mater. 2019;181:447-59.

Pagan DC, Phan TQ, Weaver JS, Benson AR, Beaudoin AJ. Unsupervised learning of dislocation motion. Acta Mater. 2019;181:510-8.

Mei H, Cheng L, Chen L, Wang F, Kong L. Development of machine learning interatomic potential for zinc. Comput Mater Sci. 2024;233:112723.

Vita JA, Trinkle DR. Spline-based neural network interatomic potentials: Blending classical and machine learning models. Comput Mater Sci. 2024;232:112655.

Zhang Y, Seibert P, Otto A, Raßloff A, Kästner M. DA-VEGAN: Differentiably augmenting vae-gan for microstructure reconstruction from extremely small data sets. Comput Mater Sci. 2024;232:112661.

Borges Y, Huber L, Zapolsky H, Patte R, Demange G. Insights from symmetry: Improving machine-learned models for grain boundary segregation. Comput Mater Sci. 2024;232:112663.

Hu Y, Zhang Y, Wen B, Dai F-Z. Grain boundary engineering in Nickel-rich cathode: A combination of high-throughput first-principles and interpretable machine learning study. Acta Mater. 2024;276:120144.

Thapa R, McKenzie ME, Musterman E, Kaman J, Jain H. Machine learning based insights of seeded congruent crystal growth of LiNbO3 in glass. Acta Mater. 2024;276:120115.

Ju S-P, Huang C-C, Chen H-Y. Illuminating the mechanical responses of amorphous boron nitride through deep learning: A molecular dynamics study. Comput Mater Sci. 2024;232:112664.

Li T, Hou Q, Cui J-C, Yang J-H, Fu B-Q. Deep learning interatomic potential for thermal and defect behaviour of aluminum nitride with quantum accuracy. Comput Mater Sci. 2024;232:112656.

Eremin RA, Humonen IS, Kazakov AA, Lazarev VD, Budennyy SA. Graph neural networks for predicting structural stability of Cd- and Zn-doped γ-CsPbI3. Comput Mater Sci. 2024;232:112672.

Xu K, Zhang L, Bai C-Y, Tu J, Luo J-R. Machine learning aided process design of Fe-Cr-Ni-Al/Ti multi-principal element alloys for excellent mechanical properties. Comput Mater Sci. 2024;232:112660.

Vazquez G, Sauceda D, Arróyave R. Deciphering chemical ordering in High Entropy Materials: A machine learning-accelerated high-throughput cluster expansion approach. Acta Mater. 2024;276:120137.

Zhang X, Wang A, Shao C, Bao H. Understanding thermal transport in magnesium solid solutions through first-principles approaches and machine learning feature screening. Acta Mater. 2024;276:120160.

Guerrero-Rivera R, Godínez-Garcia FJ, Hayashi T, Wang Z, Ortiz-Medina J. Machine-Learning driven STM images prediction of doped/defective graphene: Towards optimized tools for 2D nanomaterials characterization. Comput Mater Sci. 2024;242:113076.

Schenk O, Becker M, Deng Y, Niemietz P, Broeckmann C. Prediction of the microstructure of cold-compacted Astaloy 85Mo with deep generative models. Comput Mater Sci. 2024;242:113064.

Frieden Templeton W, Miner JP, Ngo A, Fitzwater L, Narra SP. Expediting structure–property analyses using variational autoencoders with regression. Comput Mater Sci. 2024;242:113056.