

ORIGINAL RESEARCH

Open access

Algorithmic Screening Frontiers: How Model Priors Reshape Searchable Materials Space

Nguyen Thanh Huy^{1*}, Pham Quang Minh¹, Le Thi Bich²

Abstract

Computational materials engineering has evolved into a data-intensive discipline where high-throughput computation, representation learning, and autonomous discovery systems enable systematic exploration of vast chemical spaces. Central to this evolution is the recognition that model priors—inductive biases, architectural assumptions, and regularization structures embedded in machine learning pipelines—actively reshape the effective searchable materials space rather than merely operating within it. Despite advances in materials informatics, graph neural networks, and closed-loop experimentation, the systemic influence of these priors on screening frontiers remains conceptually underexplored. This article presents the Priors-Adaptive Frontier Reshaping (PAFR) Framework, an original systems-level conceptualization that formalizes how priors modulate data-to-discovery pipelines through layered interactions between representation spaces, inference dynamics, and feedback loops. By integrating insights from multimodal datasets, uncertainty quantification, and simulation–experiment coupling, the framework elucidates computational workflow dynamics and epistemic risk structures that govern algorithmic screening efficiency. The PAFR Framework offers interpretive guidance for designing more robust infrastructures in materials discovery, highlighting trade-offs in prior selection, search space expansion, and steering logics. These insights advance a deeper understanding of representation–inference interactions in data-driven materials engineering.

Keywords Materials informatics, Representation learning, Computational discovery, Data infrastructures, Model priors, Algorithmic screening

*Correspondence:

Nguyen Thanh Huy
huy.nguyen@gmail.com

¹ Department of Materials Data Science, Faculty of Engineering, Vietnam National University, Hanoi, Vietnam

² Department of Computational Engineering Systems, Faculty of Engineering, Can Tho University, Can Tho, Vietnam

Introduction

The field of computational materials engineering has undergone a profound transformation over the past decade, driven by the convergence of high-throughput computational methods, machine learning architectures, and large-scale data infrastructures. Traditional physics-based simulations, while foundational, face scalability limitations when confronting the combinatorial explosion of materials space, estimated to contain 10^{10} to 10^{100} stable compounds. In response, data-driven paradigms have emerged as essential complements, enabling rapid

property prediction and inverse design across diverse material classes [1, 2].

Machine learning applications in materials science have proliferated, particularly through representation learning techniques that encode atomic structures, stoichiometry, and electronic properties into high-dimensional latent spaces [3, 4]. Graph neural networks and deep learning models now routinely process crystal graphs and multimodal datasets, achieving predictive capabilities that accelerate screening campaigns [5, 6]. Concurrently, autonomous discovery systems and closed-loop experimentation platforms integrate simulation with robotic

synthesis, creating self-optimizing pipelines that reduce reliance on exhaustive enumeration [7, 8].

High-throughput computation infrastructures, powered by density functional theory workflows and automated benchmarking, have generated petabyte-scale repositories of materials properties [9, 10]. These ecosystems facilitate the construction of foundation models tailored to scientific domains, where transfer learning across property spaces enhances generalization [11, 12]. However, current discovery models encounter fundamental limits. Many approaches treat the searchable space as static, overlooking how model priors—ranging from inductive biases in neural architectures to regularization priors in Bayesian formulations—dynamically constrain or expand accessible regions of chemical space.

Epistemic constraints arise prominently here. Uncertainty quantification in materials AI reveals that predictive confidence is highly sensitive to training data distribution and architectural choices, leading to domain-of-applicability mismatches [13–15]. Representation learning can introduce unintended biases when priors favor certain symmetry groups or bonding motifs, subtly warping the effective frontier away from experimentally viable candidates [16, 17]. Moreover, inverse materials design frameworks often embed strong assumptions about target property landscapes, inadvertently narrowing the exploration horizon despite claims of broad screening [18, 19].

These limitations highlight a critical conceptual gap: algorithmic screening is not merely a search problem but a dynamic reshaping process governed by model priors. Priors interact with data representations to redefine boundaries, steering computational resources toward emergent subspaces while deprioritizing others. This reshaping manifests in workflow dynamics where feedback from early discoveries refines priors, creating adaptive loops that enhance overall discovery efficiency.

The present work addresses this gap through the Priors-Adaptive Frontier Reshaping (PAFR) Framework. Positioned at the intersection of materials informatics and computational systems design, the PAFR Framework provides an integrative lens for analyzing how priors fundamentally alter searchable materials space. It emphasizes infrastructure-level trade-offs, representation–inference couplings, and steering logics essential for next-generation autonomous pipelines. By focusing on conceptual dynamics rather than empirical benchmarks, the

framework advances theoretical foundations for more resilient data-driven materials engineering.

Theoretical Background & Literature Synthesis

Materials data infrastructures

Modern materials data infrastructures form the epistemic substrate upon which algorithmic screening ecosystems are constructed. Rather than functioning as passive repositories, these infrastructures operate as dynamic knowledge architectures that aggregate, harmonize, and operationalize multimodal datasets spanning high-throughput computational outputs, experimental measurements, synthesis protocols, and historical characterization records. Their integrative capacity enables the co-registration of structural descriptors, compositional gradients, thermodynamic stability metrics, and functional performance indicators within interoperable data environments, thereby supporting scalable training regimes for predictive and generative models [9–11].

At the infrastructural core, automated workflow engines orchestrate event-driven data pipelines that link simulation platforms with downstream learning systems in near-real time. Density functional theory calculations, phase diagram simulations, and molecular dynamics outputs are systematically ingested, standardized, and routed toward surrogate model training or screening modules [8]. This coupling transforms infrastructures into epistemic accelerators—platforms that not only store data but actively condition the pace and direction of discovery.

Yet scale alone does not guarantee epistemic robustness. The distributional geometry of stored data exerts a formative influence on downstream inference. Historical biases toward well-studied alloys, energy materials, or catalysis systems generate dense knowledge clusters, while emergent or computationally expensive chemistries remain sparsely mapped. Such uneven coverage introduces representational asymmetries that propagate into prior embeddings, uncertainty estimations, and screening outputs [12]. Underrepresented chemical domains may therefore be algorithmically marginalized, not through explicit exclusion but through infrastructural silence.

Conceptually, robust materials data infrastructures must be understood as curation-intensive systems rather than accumulation engines. Mechanisms for bias auditing, density mapping, and adaptive data acquisition become essential to preserve representational fidelity. Frontier expansion in computational materials science thus depends not merely on data volume but on the epistemic topology of infrastructural coverage—how evenly, deeply, and reflexively materials spaces are encoded.

Representation learning architectures

Representation learning constitutes the translational layer through which raw materials descriptors are rendered computationally intelligible. Atomic coordinates, crystallographic symmetries, bonding topologies, and microstructural morphologies are transformed into compact latent embeddings that encode property-relevant information across multiple scales. These embeddings function as epistemic proxies—compressed yet information-dense representations that enable predictive inference, similarity search, and inverse design.

Stoichiometry-derived encoders and graph-based architectures have demonstrated strong cross-property transferability, enabling models trained on formation energies, for instance, to generalize toward electronic or mechanical properties [3, 4, 16]. Graph neural networks extend this capability by formalizing materials as relational systems in which nodes represent atoms and edges encode bonding environments. Message-passing operations propagate local chemical information while preserving global structural coherence, thereby supporting high-fidelity property prediction and generative exploration [6, 19].

However, representation is never neutral. Architectural priors embedded within encoding schemes—such as locality assumptions, symmetry constraints, or bonding heuristics—precondition how materials manifolds are geometrically structured within latent space. Message-passing depth determines interaction horizons; equivariance constraints regulate rotational sensitivity; pooling operations compress structural variance. Each design choice introduces an interpretive lens through which materials reality is computationally refracted.

This preconditioning has systemic consequences. Regions of materials space that conform to encoded priors become densely representable and algorithmically accessible, while

structures deviating from architectural assumptions risk latent distortion or compression. The geometry of the latent manifold thus becomes an infrastructural artifact—shaped as much by model design as by underlying materials physics. Understanding screening outcomes therefore requires examining how representational architectures sculpt the navigability of materials landscapes.

AI-guided discovery systems

AI-guided discovery systems extend representation learning into dynamic exploration regimes, embedding predictive models within closed-loop infrastructures that iteratively steer experimentation and computation. Unlike passive high-throughput screening, which enumerates candidate materials exhaustively, AI-guided systems deploy adaptive sampling strategies that prioritize information gain, uncertainty reduction, or performance optimization.

Active learning frameworks rank candidate materials according to epistemic uncertainty or expected model improvement, directing simulation or experimental resources toward high-value regions of search space [7, 20, 21]. Bayesian optimization, reinforcement learning, and surrogate-assisted acquisition functions further refine this prioritization, enabling discovery pipelines to evolve in response to accumulating evidence. The search process becomes path-dependent—shaped by prior predictions, feedback signals, and infrastructural constraints.

Hybridized approaches integrate semi-supervised learning for synthesis classification with evolutionary algorithms that recombine structural motifs across candidate populations [18, 22]. These systems accelerate targeted exploration while embedding procedural knowledge—such as synthesis feasibility—into screening criteria. Discovery thus emerges not solely from predictive accuracy but from the orchestration of learning, experimentation, and design within recursive feedback cycles.

Conceptually, AI-guided discovery reframes materials innovation as an emergent systems phenomenon. Knowledge production arises from the interaction of models, infrastructures, and human oversight rather than from isolated algorithmic performance. Closed-loop coupling introduces reflexivity: models influence the data they later consume, thereby reshaping their own epistemic environment. This recursive dynamic foregrounds the importance of governance mechanisms capable of

monitoring exploration diversity and mitigating self-reinforcing bias trajectories [8].

Computational design paradigms

Computational design paradigms invert the traditional discovery workflow by mapping desired functional properties back to candidate structures. Rather than predicting performance from known materials, inverse design frameworks generate hypothetical compounds optimized for target attributes such as catalytic activity, electronic bandgaps, or mechanical resilience.

Deep generative models—including variational autoencoders, generative adversarial networks, and diffusion architectures—have demonstrated capacity to synthesize novel reticular frameworks, molecular crystals, and porous materials with tunable properties [19, 23, 24]. Evolutionary algorithms complement these models by iteratively mutating and recombining candidate structures, navigating high-dimensional design spaces through fitness-driven selection.

These paradigms rely heavily on prior specification. Physics-informed constraints, symmetry rules, and thermodynamic feasibility filters are embedded to ensure chemically plausible outputs. While such priors enhance search efficiency, they simultaneously bound the horizon of generative diversity. The design space becomes sculpted by what the model is permitted to imagine.

Consequently, computational design must be interpreted as a negotiation between creativity and constraint. Strong priors accelerate convergence but risk excluding unconventional yet viable materials. Weak priors expand exploratory breadth but may yield infeasible candidates. Conceptual analysis of design workflows therefore requires interrogating how prior architectures regulate innovation bandwidth and shape the topology of accessible materials futures.

Uncertainty & interpretability

Uncertainty quantification and interpretability frameworks provide the epistemic safeguards necessary for trustworthy algorithmic screening. Predictive models operating in high-dimensional materials spaces inevitably encounter distributional gaps where training data are sparse or absent. Without uncertainty awareness, such models risk overconfident extrapolation—projecting reliability into epistemic voids.

Domain-of-applicability mapping, Bayesian ensembling, and probabilistic calibration techniques delineate reliability boundaries, enabling screening systems to distinguish between informed inference and speculative prediction [13–15, 17]. When integrated into active learning loops, uncertainty signals guide adaptive sampling toward knowledge gaps, transforming ignorance into an exploration resource.

Interpretability methods further illuminate the internal logic of predictive systems. Feature attribution analyses, attention mapping, and counterfactual perturbation frameworks reveal how structural descriptors influence model outputs. These techniques expose latent steering mechanisms—hidden priors that weight certain compositional or topological features more heavily than others.

From a conceptual standpoint, uncertainty and interpretability function as epistemic auditing layers. They render visible the otherwise opaque influence of prior assumptions embedded across infrastructures, representations, and discovery loops. By surfacing these influences, interpretability frameworks enable prior-aware screening strategies that actively counteract exclusionary bias structures.

Unchecked priors can systematically narrow exploration trajectories, excluding chemically plausible yet underrepresented materials subspaces. Integrating uncertainty diagnostics with interpretability analytics therefore becomes essential for equitable coverage of materials landscapes. Trustworthy screening emerges not from predictive accuracy alone but from reflexive awareness of the epistemic limits shaping algorithmic judgment.

Proposed conceptual framework

The Priors-Adaptive Frontier Reshaping (PAFR) Framework conceptualizes algorithmic screening as a dynamic, multi-layered process wherein model priors actively transform the searchable materials space. Unlike static screening views, the PAFR Framework positions priors as steering operators that modulate data representations, inference trajectories, and discovery selection logics through interconnected pipelines and feedback mechanisms.

The framework comprises four structural layers. The Data Assimilation Layer ingests multimodal materials datasets, standardizing structural, compositional, and simulation-derived features. The Priors Embedding Layer injects inductive biases—ranging from architectural symmetries in graph networks to regularization terms—directly into representation learning modules, preconditioning latent spaces. The Representation and Inference Layer performs forward mapping to property predictions while quantifying uncertainty gradients. Finally, the Discovery Steering Layer applies selection heuristics to reshape the effective frontier, prioritizing candidates based on combined prior confidence and uncertainty signals.

Data → Model → Discovery pipelines operate as follows: raw datasets flow through prior-conditioned encoders to generate a transformed representation manifold; inference propagates forward under these priors, producing a probability-weighted candidate pool; steering logics then extract a refined subspace for targeted validation or synthesis recommendation. Feedback loops close the system by feeding new experimental or simulation outcomes back to the Priors Embedding Layer, enabling adaptive refinement of biases and expansion of previously inaccessible frontiers.

This architecture captures key computational steering logics, including uncertainty-guided exploration, prior regularization for generalization, and loop-mediated subspace growth. Two formal expressions articulate core dynamics within the PAFR Framework. The reshaping of searchable space may be expressed as

$$S_{eff} = P(\theta) \circ R(D) \quad (1)$$

where $P(\theta)$ denotes the operator applying model priors parameterized by θ . \circ represents composition with the representation transformation R , and D is the input data distribution. This equation conceptualizes how priors do not merely filter but geometrically distort the accessible volume of materials space.

A second dynamic captures feedback interaction:

$$\frac{dF}{dt} = \kappa \cdot (U(M) + \lambda \cdot \Delta D_{fb}) \quad (2)$$

Here F represents frontier expansion rate, $U(M)$ denotes model uncertainty contribution, ΔD_{fb} quantifies feedback from new discoveries, and κ, λ are coupling coefficients reflecting infrastructure tuning. This formulation highlights the temporal evolution of screening horizons driven by epistemic signals.

As conceptualized in **Figure 1**, the PAFR Framework is depicted as a cyclic layered diagram. The central searchable space appears as a deformable manifold whose boundaries expand or contract under prior modulation arrows originating from the Priors Embedding Layer. Forward arrows illustrate the primary data-to-discovery pipeline, while dashed backward loops connect discovery outcomes to prior refinement and uncertainty propagation nodes. Layer interfaces are annotated with key operations such as composition, steering, and feedback integration, visually emphasizing the systemic interdependence of components. The systemic interactions between prior embedding, representation transformation, inference propagation, and feedback-driven steering are synthesized within the Priors-Adaptive Frontier Reshaping architecture illustrated in **Figure 1**.

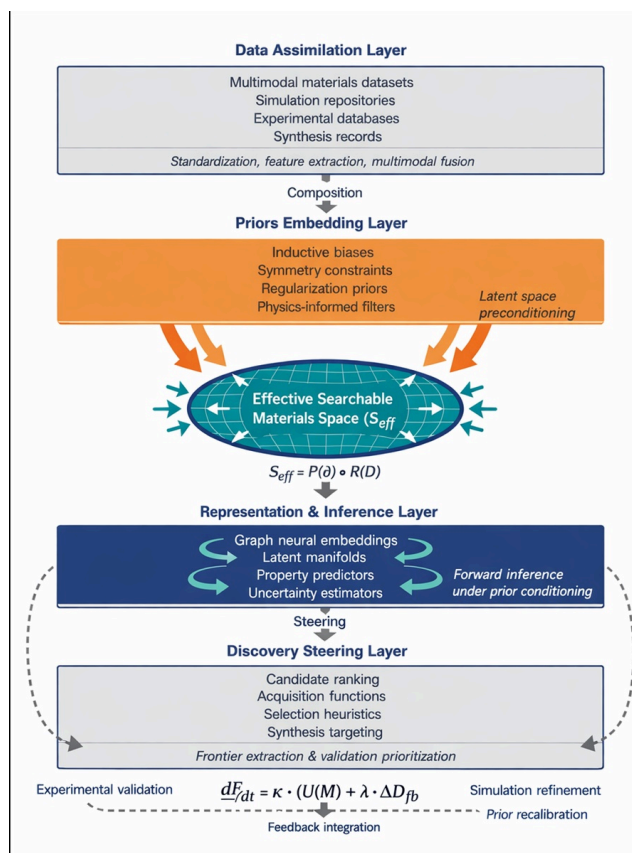


Figure 1. Priors-Adaptive Frontier Reshaping (PAFR) Framework: Layered Dynamics of Prior-Conditioned Algorithmic Screening

The structural logic and operational roles of these interacting layers are systematically synthesized in [Table 1](#).

Table 1. Structural Layers and Operational Functions within the PAFR Framework

Framework Layer	Primary Function	Embedded Priors	Epistemic Role
Data Assimilation Layer	Multimodal data ingestion and standardization	Dataset selection biases, sampling density priors	Define foundational knowledge coverage
Priors Embedding Layer	Injection of inductive and architectural assumptions	Symmetry constraints, regularization terms, physics-informed filters	Pre-condition latent geometric space
Representation & Inference Layer	Encoding and predictive propagation	Message-passing assumptions, feature compression priors	Shape manifold for navigation and uncertainty gradient estimation
Discovery Steering Layer	Candidate prioritization and acquisition logic	Optimization heuristics, uncertainty thresholds	Direct exploration density
Feedback Integration Loops	Adaptive updating of priors and datasets	Reinforcement priors, recalibration weights	Enable reflexive learning

The PAFR Framework provides systems-level insights into infrastructure trade-offs, such as balancing prior expressivity against overfitting risk, and representation fidelity against computational cost. By foregrounding these interactions, the framework advances interpretive understanding of how algorithmic choices at the model level propagate to macroscopic discovery frontiers in computational materials engineering.

Analytical implications

The PAFR Framework yields several analytical implications for computational materials engineering, particularly in how it informs the design and operation of algorithmic screening workflows. By treating model priors as active reshaping agents, the framework highlights opportunities for enhancing pipeline robustness through targeted prior engineering. For instance, in high-throughput computation ecosystems, priors can be tuned to prioritize underrepresented data manifolds, thereby mitigating coverage gaps that limit frontier expansion [9, 11]. This tuning involves balancing inductive biases—such as those in graph neural networks that favor periodic structures—against the need for broader chemical diversity, ensuring that screening does not converge prematurely on familiar subspaces [3, 6].

In terms of epistemic risk mitigation, the PAFR Framework underscores the interplay between uncertainty signals and prior regularization. Uncertainty quantification mechanisms, when integrated into the Priors Embedding Layer, allow for real-time assessment of risk structures, where high-entropy regions signal potential prior mismatches [13, 17]. This integration enables workflows to dynamically allocate computational resources, steering away from overconfident but erroneous predictions and toward exploratory sampling that refines the effective space [15, 20]. Such risk-aware steering logics reduce the likelihood of epistemic blind spots, particularly in multimodal datasets where simulation–experiment discrepancies amplify uncertainties [8, 12].

Infrastructure design trade-offs emerge prominently within the framework’s layered architecture. The trade-off between representational expressivity and computational scalability, for example, dictates how priors compress data flows without losing critical inference pathways [4, 16]. Opting for strong priors, like physics-informed constraints in inverse design, can streamline discovery but at the cost of flexibility in handling novel materials classes [19, 24]. Conversely, weaker priors promote broader exploration yet demand larger infrastructures to manage increased variance, illustrating a systemic balance that PAFR conceptualizes as tunable parameters in pipeline dynamics [7, 10].

Key trade-offs emerging from prior engineering across screening infrastructures are comparatively outlined in [Table 2](#).

Table 2. Prior-Induced Trade-Offs in Algorithmic Screening Frontiers

Prior Design Dimension	Strong Prior Configuration	Weak Prior Configuration	Discovery Advantages
Structural Constraints	Physics-informed symmetry enforcement	Minimal structural assumptions	Efficient convergence in known domains
Representation Compression	High latent regularization	Flexible embedding spaces	Computational efficiency
Acquisition Heuristics	Exploitative optimization	Exploratory sampling	Rapid candidate validation
Generative Design Filters	Strict feasibility priors	Open generative search	Chemical plausibility outputs
Uncertainty Thresholding	Conservative confidence gating	Permissive uncertainty tolerance	Reduced positive bias
Feedback Coupling Strength	Rapid prior updating	Gradual recalibration	Accelerated learning cycles

Representation–discovery synergies are another key implication, where the framework elucidates how prior-conditioned embeddings facilitate bidirectional mappings. In autonomous systems, these synergies manifest as enhanced closed-loop performance, with feedback loops updating representations to align more closely with discovery objectives [21, 22]. For perovskite design or superhard materials prediction, this means priors can be adapted to emphasize property-relevant features, fostering synergies that expand searchable frontiers beyond stoichiometric limits [18, 24].

Finally, steering logic optimizations under PAFR involve formalizing adaptive mechanisms to maximize frontier reshaping efficiency. One such dynamic can be conceptualized as

$$\eta = \underset{\beta}{\operatorname{argmax}} \theta (I(\operatorname{Seff}) - \beta \cdot C(\theta)) \quad (3)$$

where η denotes optimized steering efficiency, $I(\operatorname{Seff})$ measures information gain in the reshaped space, $C(\theta)$ captures prior complexity cost, and β balances the trade-off. This expression captures the interaction between discovery yield and infrastructure overhead, providing an interpretive tool for logic refinement in data-driven ecosystems. Overall, these implications position PAFR as a guide for constructing more adaptive, risk-informed workflows in materials engineering.

Results and Discussion

The Priors-Adaptive Frontier Reshaping (PAFR) Framework contributes to broader discourses in computational materials engineering by reframing algorithmic screening as an intrinsically prior-mediated process rather than a neutral optimization exercise. Within conventional discovery narratives, predictive accuracy and throughput efficiency dominate evaluation criteria. PAFR shifts this evaluative axis by foregrounding how embedded priors—architectural, infrastructural, and procedural—actively sculpt the topology of searchable materials space. Screening outcomes are thus not merely discovered but structurally conditioned by the interpretive assumptions encoded across learning systems.

This reconceptualization bears direct implications for the design of next-generation materials informatics infrastructures. Efforts to embed artificial intelligence more deeply into discovery pipelines—particularly in domains such as energy storage, catalysis, quantum materials, and biomaterials—have emphasized scale, automation, and predictive performance [25, 26]. PAFR complements these trajectories by introducing reshaping awareness as a parallel design principle. Infrastructure architectures may therefore evolve toward reflexive configurations that continuously audit how priors redistribute exploration density, amplify certain compositional trajectories, and suppress others.

A central contribution of the framework lies in formalizing representation–inference couplings as drivers of frontier geometry. Representation learning systems do not simply encode materials—they delimit the latent manifolds within which inference operates. When coupled with optimization engines and generative search, these manifolds become navigational terrains whose curvature, density, and continuity are prior-dependent. Recognizing this dependency supports the development of foundation

models capable of maintaining robustness under domain shift, particularly when transferring knowledge across heterogeneous materials classes or experimental regimes [27, 28]. In this sense, PAFR offers an interpretive bridge between materials informatics and parallel developments in chemistry, condensed matter physics, and molecular engineering, where representation fidelity increasingly governs discovery viability.

Beyond technical infrastructures, the framework also reframes discovery as a socio-technical assemblage. Human oversight, experimental feasibility judgments, and funding priorities implicitly function as external priors that intersect with algorithmic ones. Screening pipelines thus operate within multi-layered prior ecologies where institutional, epistemic, and computational influences converge. Understanding frontier reshaping therefore requires a systems perspective that extends beyond model architectures to encompass governance and decision ecosystems.

Limitations

Notwithstanding its interpretive contributions, several limitations of the PAFR Framework warrant critical reflection. As a purely conceptual construct, the framework does not prescribe implementation protocols or algorithmic specifications. Its analytical utility lies in interpretive structuring rather than procedural guidance. Consequently, translation into bespoke computational workflows may require auxiliary operationalization layers—such as metric design, prior quantification strategies, or simulation coupling schemas.

The framework also assumes relatively idealized feedback conditions. Closed-loop discovery infrastructures are often constrained by asynchronous data acquisition cycles, experimental throughput limitations, and computational latency. These bottlenecks may dampen or distort reshaping dynamics relative to the theoretical fluidity conceptualized within PAFR [8, 29]. For instance, delays in experimental validation could freeze exploration trajectories around provisional priors, introducing temporal inertia into frontier evolution.

Moreover, emphasizing priors as the primary reshaping agents risks underrepresenting other structural determinants. Hardware architectures, energy constraints, and scaling costs influence which models can be deployed and how extensively they can explore materials space.

Similarly, human decision-making—whether in dataset curation, parameter selection, or campaign prioritization—injects discretionary biases that may operate independently of formal priors [7, 10]. While PAFR acknowledges these influences, its analytic center of gravity remains anchored in prior dynamics, potentially simplifying the full heterogeneity of discovery infrastructures.

Future conceptual extensions

Future theoretical expansions of the framework could address these limitations by incorporating multi-agent epistemic dynamics. In ensemble and federated learning ecosystems, multiple models—each governed by distinct priors—interact through consensus mechanisms, competitive optimization, or knowledge distillation. Frontier reshaping in such environments would emerge from negotiated priors rather than singular ones, introducing collective epistemic geometries that evolve through model-to-model dialogue.

Temporal extensions also represent a fertile avenue for conceptual development. In real-time discovery settings—such as autonomous laboratories or adaptive synthesis platforms—priors are not static but continuously updated in response to streaming data. Temporal priors may accelerate convergence in emergent domains while preserving exploratory diversity through dynamic recalibration [30, 31]. Modeling reshaping as a time-dependent process would enable richer analysis of how discovery frontiers expand, contract, or bifurcate across iterative cycles.

Additional extensions could integrate uncertainty-conditioned priors, where confidence estimates regulate the strength with which priors influence search trajectories. Such hybridization would link reshaping intensity to epistemic reliability, preventing premature frontier narrowing in sparsely validated domains.

Collectively, these future directions position PAFR not as a closed theoretical system but as an expandable interpretive scaffold—capable of accommodating distributed intelligence, temporal adaptation, and reflexive governance within computational materials ecosystems.

Conclusion

This manuscript introduced the Priors-Adaptive Frontier Reshaping (PAFR) Framework as a novel conceptual lens

for interpreting how model priors dynamically redefine searchable materials space within computational and data-driven engineering. Moving beyond performance-centric narratives, the framework positions priors as structural forces that contour the epistemic geometry of discovery itself.

By integrating layered infrastructures, representation architectures, inference engines, and feedback steering mechanisms, PAFR elucidates how screening pipelines evolve through recursive prior–data interactions. Analytical implications reveal pathways for mitigating epistemic risk, optimizing infrastructural design, and fostering synergistic human–machine workflows. Formalized reshaping dynamics further provide a theoretical vocabulary for articulating how exploration density, diversity, and directionality are governed within algorithmic ecosystems.

At a broader level, the framework advances theoretical understanding of representation–discovery couplings, emphasizing that innovation trajectories are shaped not only by data availability but by the interpretive priors through which that data is operationalized. Such insight is particularly salient as materials science transitions toward autonomous discovery infrastructures and foundation-scale learning systems.

Ultimately, PAFR contributes an epistemic design perspective to computational materials engineering—one

that encourages reflexive awareness of how discovery systems are architected, steered, and bounded. By illuminating the prior structures that shape exploration horizons, the framework offers interpretive tools for navigating complex chemical landscapes with greater robustness, inclusivity, and strategic foresight in the pursuit of next-generation materials solutions.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 05 Sep 2021 Revised: 21 Jan 2022 Accepted: 06 Apr 2022
Published online: 18 September 2022

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. *Nature*. 2018;559(7715):547–55.
<https://doi.org/10.1038/s41586-018-0337-2>.

Ramprasad R, Batra R, Pilia 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>.

Choudhary K, DeCost B, Chen C, Jain A, Tavazza F, Cohn R, et al. Recent advances and applications of deep learning methods in materials science. *npj Comput Mater*. 2022;8(1):59.
<https://doi.org/10.1038/s41524-022-00734-6>.

Goodall REA, Lee AA. Predicting materials properties without crystal structure: deep representation learning from stoichiometry. *Nat Commun.* 2020;11(1):6280.
<https://doi.org/10.1038/s41467-020-19964-7>.

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>.

Feng S, Fu H, Zhou H, Wu Y, Lu Z, Dong H. A general and transferable deep learning framework for predicting phase formation in materials. *npj Comput Mater.* 2021;7(1):10.
<https://doi.org/10.1038/s41524-020-00488-z>.

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

Aspuru-Guzik A, Persson KA, Flores-Leonar MM, Salinas O, Allioux D, Farfan J, et al. Autonomous experimentation systems for materials development: A community perspective. *Matter.* 2021;4(9):2702-26.
<https://doi.org/10.1016/j.matt.2021.06.036>.

Uhrin M, Huber SP, Yu J, Marzari N, Pizzi G. Workflows in AiiDA: Engineering a high-throughput, event-based engine for robust and modular computational workflows. *Comput Mater Sci.* 2021;187:110086.
<https://doi.org/10.1016/j.commatsci.2020.110086>.

Zakutayev A, Wunderlich S, Schwarting M, Perkins JD, Schafer R, Snyder GJ, et al. The materials research platform: Defining the requirements from user stories. *Matter.* 2019;1(6):1419-30.
<https://doi.org/10.1016/j.matt.2019.10.024>.

Gupta V, Choudhary K, Tavazza F, Campbell C, Liao W-k, Choudhary A, et al. Cross-property deep transfer learning framework for enhanced predictive analytics on small materials data. *Nat Commun.* 2021;12(1):6595.
<https://doi.org/10.1038/s41467-021-26921-5>.

Chang R, Wang YX, Ertekin E. Towards overcoming data scarcity in materials science: unifying models and datasets with a mixture of experts framework. *npj Comput Mater.* 2022;8(1):242.
<https://doi.org/10.1038/s41524-022-00929-x>.

Sutton C, Boley M, Ghiringhelli LM, Rupp M, Vreeken J, Scheffler M. Identifying domains of applicability of machine learning models for materials science. *Nat Commun.*

2020;11(1):4428.
<https://doi.org/10.1038/s41467-020-17112-9>.

Umehara M, Stein HS, Guevarra D, Newhouse PF, Boyd DA, Gregoire JM. Analyzing machine learning models to accelerate generation of fundamental materials insights. *npj Comput Mater.* 2019;5(1):34.
<https://doi.org/10.1038/s41524-019-0172-5>.

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-z>.

Nyshadham C, Rupp M, Bekker B, Shapeev AV, Mueller T, Rosenbrock CW, et al. Machine-learned multi-system surrogate models for materials prediction. *npj Comput Mater.* 2019;5(1):51.
<https://doi.org/10.1038/s41524-019-0189-9>.

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

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/s41524-021-00495-8>.

Fung V, Zhang J, Hu G, Ganesh P, Sumpter BG. Inverse design of two-dimensional materials with invertible neural networks. *npj Comput Mater.* 2021;7(1):200.
<https://doi.org/10.1038/s41524-021-00670-x>.

Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Comput Mater.* 2019;5(1):21.
<https://doi.org/10.1038/s41524-019-0153-8>.

Jennings PC, Lysgaard S, Hummelshøj JS, Vegge T, Bligaard T. Genetic algorithms for computational materials discovery accelerated by machine learning. *npj Comput Mater.* 2019;5(1):46.
<https://doi.org/10.1038/s41524-019-0181-4>.

Huo H, Rong Z, Kononova O, Sun W, Botari T, He T, et al. Semi-supervised machine-learning classification of materials synthesis procedures. *npj Comput Mater.* 2019;5(1):62.
<https://doi.org/10.1038/s41524-019-0204-1>.

Yao Z, Sánchez-Lengeling B, Bobbitt NS, Bucior BJ, Lee SGH, Day GM, et al. 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>.

Avery P, Wang X, Oses C, Gossett E, Proserpio DM, Toher C, et al. Predicting superhard materials via a machine learning informed evolutionary structure search. *npj Comput Mater.* 2019;5(1):89.

<https://doi.org/10.1038/s41524-019-0226-8>.

Liu J, Liu Y, Wang Z, Tang J, Wen Y, Huang J, et al. Artificial intelligence-based material discovery for clean energy future. *Adv Intell Syst.* 2022;4(6):2200073.

<https://doi.org/10.1002/aisy.202200073>.

Suwardi A, Wang F, Xue K, Han MY, Teo EY, Wang P, et al. Machine learning-driven biomaterials evolution. *Adv Mater.* 2022;34(1):2102703.

<https://doi.org/10.1002/adma.202102703>.

Brunton SL, Proctor JL, Kutz FJ. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Sci Adv.* 2017;3(4):e1602614.

<https://doi.org/10.1126/sciadv.1602614>.

Wolloch M, Levchenko SV, Righi MC. Computational synthesis of 2D materials: A high-throughput approach to materials design. *Comput Mater Sci.* 2022;204:111182.

<https://doi.org/10.1016/j.commatsci.2021.111182>.

Wolloch M, Losi G, Chehaimi O, Yalcin F, Ferrario M, Righi MC. High-throughput generation of potential energy surfaces for solid interfaces. *Comput Mater Sci.* 2022;207:111302.

<https://doi.org/10.1016/j.commatsci.2022.111302>.

Lv C, Chen C, Chuang YC, Tseng TL. A real-time surface inspection method for precision steel balls based on machine vision. *Adv Mater.* 2022;34(10):2101474.

<https://doi.org/10.1002/adma.202101474>.

Li Z, Yoon J, Zhang R, Rajabipour F, Srubar WV III, Dabo I, et al. Machine learning in concrete science: Applications, challenges, and best practices. *npj Comput Mater.* 2022;8(1):127.

<https://doi.org/10.1038/s41524-022-00810-x>.