ORIGINAL RESEARCH **Open access**

Discovery Acceleration vs Epistemic Depth: Speed–Understanding Trade-Offs in Computational Design

Bruno Martins[1*], Lucas Pereira[1], Renata Azevedo[2]

Abstract

The field of computational and data-driven materials engineering has witnessed a paradigm shift toward accelerated discovery pipelines, leveraging machine learning and high-throughput computations to navigate vast materials spaces. However, this emphasis on speed often comes at the expense of epistemic depth, where understanding of underlying mechanisms is sidelined by predictive efficiency. This manuscript introduces a conceptual framework that examines the inherent trade-offs between discovery acceleration and epistemic comprehension in computational design ecosystems. By integrating insights from materials informatics, representation learning, and uncertainty quantification, we propose a systems-level architecture that balances rapid iteration with interpretive rigor. The framework delineates how data infrastructures, model architectures, and feedback loops influence the speed–understanding continuum, highlighting computational steering logics that mitigate epistemic risks without compromising efficiency. Implications extend to autonomous discovery systems, inverse design strategies, and multimodal datasets, fostering more resilient AI-guided materials engineering. Ultimately, this approach advocates for hybrid paradigms where acceleration serves as a scaffold for deeper mechanistic insights, potentially transforming how computational tools are deployed in materials research.

Keywords Materials informatics, Uncertainty quantification, Representation learning, AI in materials science, Computational discovery, Data infrastructures

*Correspondence:
Bruno Martins
bruno.martins@outlook.com
[1] Department of Materials Modeling and Data Science, Faculty of Engineering, University of Minho, Braga, Portugal
[2] Department of Computational Materials Systems, Faculty of Engineering, University of Porto, Porto, Portugal

Introduction

The advent of computational and data-driven approaches has fundamentally reshaped materials engineering, enabling the exploration of complex materials landscapes at scales previously unattainable through traditional experimental methods. High-throughput computation, coupled with machine learning algorithms, has facilitated the rapid screening of candidate materials for applications ranging from energy storage to catalysis [1, 2]. This evolution is rooted in the integration of vast datasets with predictive models, where materials informatics serves as the backbone for identifying patterns and correlations that guide design decisions [3, 4]. For instance, graph neural networks and other deep learning architectures have emerged as powerful tools for encoding atomic and molecular structures, allowing for efficient property predictions across diverse chemical spaces [2, 4].

Yet, as these ecosystems mature, a critical tension arises between the pursuit of discovery acceleration and the maintenance of epistemic depth. Acceleration, characterized by swift iterations through computational workflows, prioritizes volume and velocity in generating hypotheses or candidates [5, 6]. In contrast, epistemic depth involves a nuanced understanding of causal relationships, structural dependencies, and emergent behaviors within materials systems [7, 8]. This trade-off is

not merely operational but systemic, embedded in the design of data-driven infrastructures that often favor black-box predictions over interpretable insights [9, 10]. The rise of autonomous discovery systems exemplifies this dynamic, where closed-loop experimentation integrates simulation with real-time feedback to expedite materials optimization [6, 11]. However, such systems risk amplifying uncertainties if epistemic foundations are overlooked, leading to brittle designs vulnerable to extrapolation failures [12, 13].

Central to this discourse is the role of representation learning in bridging computational efficiency with conceptual clarity. Representations, whether graph-based or multimodal, act as intermediaries that compress high-dimensional materials data into actionable forms [4, 14]. Advances in foundation models for science have further amplified this capability, enabling transfer learning across disparate datasets and accelerating inverse design tasks [15, 16]. Nonetheless, the epistemic cost of these accelerations manifests in challenges such as domain shift, where models trained on synthetic data falter in experimental contexts [17, 18]. Uncertainty quantification emerges as a mitigating factor, providing mechanisms to gauge confidence in predictions and steer discovery toward regions of reliable knowledge [7, 9, 12].

The limitations of current discovery models underscore the need for a more balanced paradigm. Traditional high-throughput approaches, while prolific in candidate generation, often lack the interpretive layers necessary for long-term scientific advancement [19, 20]. For example, machine learning frameworks that optimize for speed may inadvertently prioritize superficial correlations over mechanistic understanding, resulting in discoveries that are efficient but epistemically shallow [21, 22]. This is particularly evident in simulation–experiment coupling, where discrepancies between computed and observed properties highlight gaps in epistemic fidelity [16, 18]. Moreover, the proliferation of multimodal materials datasets introduces complexities in data fusion, where acceleration demands streamlined processing at the potential expense of contextual depth [23, 24].

Addressing these constraints requires a conceptual reevaluation of computational design pipelines. Rather than viewing speed and understanding as opposing forces, an integrative framework can conceptualize them as interdependent dimensions within a unified system. This involves dissecting the interactions between data infrastructures, model architectures, and discovery logics to identify leverage points for optimization [25]. By doing so, materials engineering can evolve toward infrastructures that not only accelerate but also enrich epistemic outcomes, ensuring that rapid discoveries are grounded in robust interpretations.

This manuscript positions a novel conceptual framework to navigate these trade-offs, emphasizing systems-level insights into how computational elements can be orchestrated for balanced outcomes. Through this lens, we explore the dynamics of discovery steering, where feedback mechanisms adaptively modulate the speed–understanding equilibrium [26, 27]. The framework aims to provide a blueprint for future computational ecosystems, fostering designs that are both agile and insightful.

Theoretical Background & Literature Synthesis

Materials data infrastructures: Foundations of computational acceleration

Materials data infrastructures constitute the epistemic substrate upon which contemporary computational materials engineering is constructed. These infrastructures function not merely as repositories but as dynamic knowledge ecosystems that aggregate, standardize, and operationalize data across experimental, computational, and theoretical domains. Early materials databases were largely archival in orientation; however, the proliferation of high-throughput density functional theory (DFT) workflows and automated characterization pipelines has transformed them into active engines of discovery. Contemporary infrastructures now support bidirectional data flows—ingesting simulation outputs while simultaneously informing model training, screening protocols, and inverse design systems.

A defining feature of modern infrastructures is their ability to support multimodal integration. Structural crystallography data, spectroscopic signatures, phase diagrams, thermodynamic descriptors, and kinetic parameters are increasingly harmonized within unified schemas. This multimodality expands the dimensionality of searchable materials space, enabling correlations that were previously inaccessible within siloed datasets. Importantly, such

integration also enables cross-scale reasoning, where atomistic simulations inform mesoscale microstructure predictions and, in turn, macroscopic performance modeling.

Open-source ecosystems have been instrumental in enabling this expansion. Toolkits for automated data ingestion, normalization, and feature extraction reduce barriers to interoperability across computational platforms. Standardized ontologies and APIs facilitate seamless connectivity between databases, machine learning pipelines, and visualization systems. Infrastructural interoperability thus becomes a force multiplier: rather than isolated knowledge silos, the field evolves toward federated discovery environments where datasets co-evolve with algorithms.

Yet, infrastructural acceleration introduces epistemic vulnerabilities. The drive toward scale frequently prioritizes volumetric expansion over curation depth. As high-throughput pipelines generate millions of calculated compounds, validation mechanisms—experimental cross-checks, uncertainty annotation, provenance tracking—struggle to keep pace. This produces what may be termed *epistemic dilution*: datasets expand numerically while interpretive confidence per datapoint diminishes. Without robust provenance metadata—calculation parameters, convergence criteria, exchange–correlation functionals—downstream models risk learning artifacts rather than physical regularities.

Network-analytic approaches have emerged as infrastructural sense-making tools. By mapping compositional and structural connectivity across known materials, these analyses reveal synthesizability corridors, metastability clusters, and transformation pathways. Such maps function as navigational overlays atop raw databases, guiding search algorithms toward chemically feasible regions. However, they also expose infrastructural asymmetries: densely studied alloy families appear as hyperconnected hubs, while exploratory chemistries remain peripheral data deserts.

To address these imbalances, next-generation infrastructures are incorporating layered metadata architectures. Provenance graphs, uncertainty tags, synthesis histories, and experimental validation flags enrich raw property datasets with contextual depth. Rather than static repositories, infrastructures evolve into epistemically annotated knowledge fabrics—balancing throughput with traceability, and scale with interpretive richness.

Representation learning architectures: Encoding materials reality

Representation learning architectures translate heterogeneous materials data into structured latent encodings that computational systems can manipulate. This encoding process is epistemically consequential: it determines which aspects of materials reality become computationally legible and which are compressed, abstracted, or lost.

Graph neural networks (GNNs) have become central to this representational transformation. By modeling atoms as nodes and interatomic interactions as edges, GNNs capture relational and topological features intrinsic to crystalline and molecular systems. Message-passing operations propagate information across lattice structures, enabling the emergence of global descriptors from local bonding environments. Such architectures have demonstrated remarkable efficacy in predicting formation energies, elastic tensors, and catalytic activity.

Beyond graph paradigms, representation learning encompasses convolutional encoders for microstructural imaging, transformer architectures for sequence-like materials descriptors, and variational autoencoders for generative exploration. Each architecture introduces distinct representational biases. Convolutional networks privilege spatial locality, transformers emphasize relational attention, and autoencoders compress variance into generative manifolds. Consequently, the choice of architecture shapes the epistemic geometry of latent space—determining which similarities are computationally proximal and which distinctions are blurred.

Multi-fidelity representation models extend this paradigm by integrating datasets of varying accuracy. Low-cost approximate simulations provide broad coverage, while high-accuracy quantum calculations refine local regions of interest. Through hierarchical fusion, models achieve a balance between scalability and precision. This stratified encoding mirrors epistemic layering in scientific reasoning—broad heuristics guiding fine-grained validation.

However, representational power introduces trade-offs. Deep encoders capable of capturing nonlinear interactions often do so at the cost of interpretability. Latent

embeddings become mathematically expressive yet physically opaque. Invariance constraints—rotational, translational, permutational—enhance robustness but may restrict adaptability to unconventional chemistries or defect-rich systems. Over-compression risks latent space degeneracy, where distinct materials collapse into indistinguishable embeddings.

Hybrid neuro-symbolic architectures are emerging as a response to these tensions. By embedding physical equations, symmetry constraints, or causal graphs within neural frameworks, such systems aim to retain interpretive scaffolding alongside statistical expressiveness. Representation learning thus evolves from pure abstraction toward epistemically grounded encoding—balancing acceleration with intelligibility.

AI-Guided discovery systems: Autonomous acceleration

AI-guided discovery systems operationalize materials data and representations into iterative innovation engines. These systems automate the cycle of hypothesis formulation, candidate screening, and performance validation—compressing timelines that historically spanned decades into computationally tractable loops.

Active learning constitutes a foundational paradigm within these systems. Here, models iteratively select data points that maximize informational gain—often guided by predictive uncertainty. Rather than random sampling, computational resources concentrate on epistemically informative regions, accelerating convergence toward optimal materials candidates.

Closed-loop experimentation extends this paradigm into cyber-physical integration. Robotic synthesis platforms, automated characterization instruments, and real-time data assimilation pipelines form adaptive discovery circuits. AI models propose compositions; robotic systems synthesize them; characterization outputs feed back into model retraining. Such loops have demonstrated success in catalyst optimization, battery materials screening, and thin-film synthesis.

Despite their efficiency, these systems raise epistemic questions. Iterative optimization privileges short-horizon performance metrics—activity, stability, conductivity—potentially at the expense of mechanistic understanding. Discovery becomes outcome-oriented rather than explanation-oriented. Post-hoc interpretability analyses attempt to recover causal insights, but these reconstructions may lag behind accelerated discovery trajectories.

Autonomous agents coordinating these pipelines introduce additional complexity. Reinforcement learning policies must balance exploration of unknown chemistries with exploitation of promising regions. Over-exploitation risks premature convergence; over-exploration dilutes efficiency. Decision logics thus encode epistemic value judgments—how much uncertainty is worth pursuing relative to performance gains.

The integration of foundation models further amplifies both capability and risk. Pre-trained on massive materials corpora, these models enable transfer learning across domains. Yet, they may also propagate latent biases embedded within training distributions. Autonomous discovery systems therefore require epistemic governance layers—mechanisms that audit decision pathways, uncertainty propagation, and bias inheritance.

Computational design paradigms: From screening to inversion

Computational design paradigms define how discovery objectives are operationalized within algorithmic search spaces. Traditional forward modeling predicts properties from known structures. Inverse design inverts this logic—inferring structures capable of delivering target functionalities.

Machine learning enables efficient navigation of inverse spaces. Generative models propose candidate materials conditioned on desired bandgaps, catalytic activities, or mechanical strengths. Bayesian optimization frameworks iteratively refine these proposals, converging toward high-performance regions with minimal evaluations.

High-throughput screening remains complementary to inverse design. By exhaustively exploring combinatorial composition spaces, screening workflows establish baseline performance landscapes. These landscapes then inform inverse algorithms, constraining generative exploration within physically plausible bounds.

Yet, acceleration introduces epistemic asymmetries. Screening pipelines often prioritize candidate volume over mechanistic explanation. Materials are flagged as

promising without elucidating the physicochemical rationales underlying their performance. Inverse models mitigate this by embedding physical priors—thermodynamic stability constraints, symmetry rules, or synthesis feasibility metrics—but uncertainties persist.

Extrapolation risk is particularly acute. When inverse models venture beyond training distributions, predictions may appear plausible yet lack physical grounding. Epistemically balanced paradigms therefore interleave generative acceleration with interpretive checkpoints—simulation validation, uncertainty auditing, and mechanistic probing.

Uncertainty & interpretability: Epistemic counterweights to speed

As computational discovery accelerates, uncertainty quantification and interpretability emerge as critical stabilizing forces. They function as epistemic counterweights—ensuring that predictive speed does not eclipse reliability.

Uncertainty quantification frameworks assess confidence in model outputs. Bayesian neural networks, ensemble learning, and Gaussian processes provide probabilistic prediction intervals. These intervals guide adaptive sampling strategies, directing computational or experimental effort toward regions of high epistemic ambiguity.

In multi-fidelity contexts, uncertainty acts as a bridging metric. Low-accuracy predictions are selectively escalated to high-accuracy simulations when uncertainty thresholds are exceeded. This hierarchical validation preserves computational efficiency while safeguarding reliability.

Interpretability frameworks interrogate the internal logic of models. Feature attribution methods, saliency mapping, and latent space probing reveal which structural motifs drive predictions. In materials science, such interpretability reconnects black-box outputs to chemical intuition—linking coordination environments, defect structures, or electronic descriptors to performance metrics.

However, transparency often conflicts with acceleration. Highly interpretable models—symbolic regressions, physics-informed networks—may sacrifice predictive throughput relative to deep architectures. Embedding interpretability directly into model design, rather than as post-hoc analysis, represents an emerging frontier.

Ultimately, uncertainty and interpretability form an epistemic governance layer within AI-accelerated discovery. They ensure that rapid design cycles remain scientifically accountable—anchoring computational innovation within interpretable, trustworthy knowledge systems.

Synthesis perspective

Across infrastructures, representations, discovery systems, design paradigms, and epistemic safeguards, a unifying tension emerges: the balance between acceleration and understanding. Data infrastructures scale discovery; representation architectures encode complexity; AI systems automate iteration; design paradigms target performance; uncertainty frameworks stabilize inference. Together, they form an interconnected epistemic ecosystem in which speed and insight co-evolve—sometimes synergistically, sometimes in tension.

This synthesized theoretical grounding establishes the conceptual substrate for subsequent framework development, situating accelerated materials discovery within a multilayered architecture of data, models, systems, and epistemic governance.

Proposed conceptual framework

To address the speed–understanding trade-offs in computational materials design, we introduce the Epistemic Momentum Framework (EMF). This original architecture conceptualizes discovery pipelines as dynamic systems where acceleration and epistemic depth are modulated through interconnected layers: data assimilation, representational transduction, inference propulsion, and feedback recalibration. At its core, EMF views materials discovery not as a linear process but as a momentum-driven cycle, where computational velocity propels exploration while epistemic mass anchors interpretive stability.

The data assimilation layer ingests multimodal inputs, from high-throughput simulations to experimental validations, forming a foundational reservoir. This layer emphasizes infrastructural agility, allowing rapid ingestion without immediate epistemic filtering, to maintain acceleration. Transitioning to representational transduction, raw data is encoded into hybrid spaces that blend graph-based topologies with uncertainty-aware embeddings. Here, the

framework prioritizes transformations that preserve mechanistic cues, ensuring representations serve dual roles in speed and comprehension.

Inference propulsion constitutes the accelerative engine, where machine learning architectures drive predictive and inverse tasks. Unlike conventional models, EMF incorporates steering logics that dynamically adjust inference based on epistemic gradients—regions where understanding deficits are detected via uncertainty signals. Finally, feedback recalibration closes the loop, recalibrating parameters through interpretive reflections that inform subsequent iterations. This layer introduces adaptive damping to prevent epistemic erosion, such as by injecting human-in-the-loop validations or symbolic constraints during high-speed phases.

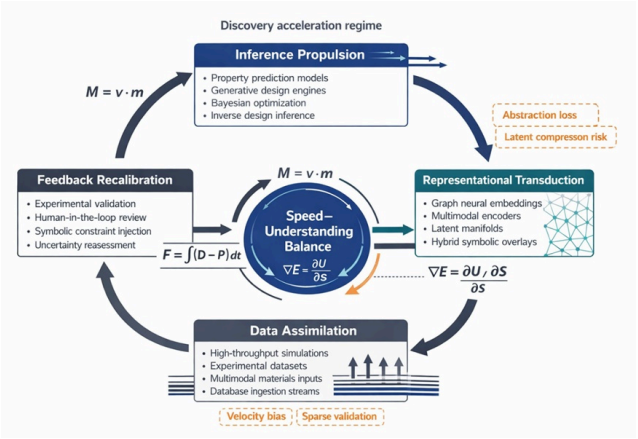

Figure 1. The framework in the study

As conceptualized in **Figure 1**, the EMF is depicted as a cyclic diagram with four quadrants representing the layers, connected by arrows denoting momentum flows. Data assimilation occupies the bottom quadrant, feeding upward into representational transduction on the right, then to inference propulsion at the top, and leftward to feedback recalibration, completing the cycle. Overlaid are vectors illustrating speed (longer arrows in propulsion) and depth (thicker nodes in recalibration), with a central equilibrium point symbolizing trade-off optimization.

A key dynamic within EMF can be conceptualized as the momentum balance, expressed as $M = v * m$, where v represents discovery velocity (computational iterations per unit time), and m denotes epistemic mass (accumulated interpretive knowledge). This captures the interaction between acceleration and depth, implying that high velocity with low mass yields unstable trajectories, while balanced momentum sustains robust discovery.

Another aspect may be expressed as the steering gradient, $\nabla E = \partial U/\partial S,$ where U is uncertainty in representational space, and S is speed metric (e.g., throughput rate). This formalizes how epistemic risks guide adjustments, directing the system toward minima where understanding compensates for rapid pacing.

Finally, feedback efficacy can be captured as $F = \int (D - P)\, dt,$ integrating discrepancies between discovered (D) and predicted (P) outcomes over time, to recalibrate layers and mitigate cumulative epistemic drift.

Through these elements, EMF provides a blueprint for computational ecosystems that harmonize speed with understanding, steering materials design toward sustainable advancements.

Analytical implications

The Epistemic Momentum Framework (EMF) offers a lens through which to examine the systemic interactions in computational materials design, revealing how trade-offs manifest across pipelines. By dissecting these implications, we uncover opportunities for refining discovery workflows, emphasizing the interplay of computational elements in modulating speed and understanding.

The layered momentum interactions and their systemic trade-offs across discovery infrastructures are synthesized in **Table 1**.

Table 1. Speed–Understanding Trade-Off Dynamics Across Computational Materials Discovery Layers

Table Content

Discovery Layer	Acceleration Mechanisms	Epistemic Depth Contributions	Trade-Risl
Data Assimilation	High-throughput ingestion; automated pipelines	Multimodal knowledge aggregation	Provena dilutic validat lag

Representational Transduction	Latent compression; scalable encoders	Mechanistic feature encoding	Abstrac loss degene
Inference Propulsion	Predictive ML; generative design	Rapid hypothesis testing	Correla bias extrapol error
Feedback Recalibration	Closed-loop validation	Interpretive refinement	Delay correct
System Integration	Autonomous orchestration	Cross-layer synthesis	Momen imbala

Discovery Pipeline Dynamics In EMF, pipelines are characterized by momentum flows that dictate the progression from data to outcomes. High-velocity regimes, prevalent in high-throughput systems, accelerate candidate identification but can dilute epistemic mass if feedback is insufficient [5, 6]. This dynamic implies that infrastructures must incorporate adaptive throttling, where data assimilation layers buffer against overload, preserving interpretive capacity during rapid iterations [14, 24]. For instance, in inverse design contexts, propulsion inference benefits from representations that embed epistemic anchors, such as prior physical constraints, to counteract drift [15, 28]. The implication here is a shift toward pipelines that dynamically scale momentum, ensuring acceleration aligns with accumulating knowledge rather than overriding it.

Representation–Inference Interactions Representations in EMF act as transducers that influence inference efficacy, where the choice of encoding impacts the speed–understanding equilibrium. Graph-based architectures, while efficient for acceleration, may introduce epistemic bottlenecks if they overly compress contextual details [2, 4]. Analytical insights suggest enhancing these interactions through hybrid transduction, merging neural efficiencies with symbolic overlays to facilitate deeper inference without deceleration [1, 21]. This can be expressed as the transduction efficiency, $T = R / (U * V)$, where R is representational fidelity (capturing mechanistic details), U is uncertainty propagation, and V is velocity of inference. Such a formulation highlights how minimizing uncertainty amplifies efficiency, steering systems toward balanced interactions that enrich epistemic outcomes [7, 9].

Epistemic Risk Structures Risk structures within EMF encapsulate the vulnerabilities arising from trade-offs, particularly in uncertainty-prone environments. Autonomous systems, for example, amplify risks when feedback recalibration lags behind propulsion, leading to epistemic voids in extrapolated spaces [11, 13]. Implications point to risk mitigation via structured quantification, where steering logics preemptively identify high-risk gradients [12, 25]. This structure may be captured as risk accumulation, $RA = \sum (\Delta E * t)$, summing epistemic deviations (ΔE) over time (t) across layers, underscoring the need for timely recalibration to prevent compounding errors [7, 18]. By formalizing these, EMF advocates for infrastructures that proactively distribute risks, fostering resilient designs.

Infrastructure Trade-Offs At the infrastructure level, EMF implies a reconfiguration of computational ecosystems to optimize trade-offs. Data infrastructures that prioritize multimodal integration accelerate discovery but risk epistemic fragmentation if fusion logics are underdeveloped [23, 29]. Analytical perspectives suggest layered trade-off management, where feedback loops enforce equilibrium, such as by allocating resources to interpretability modules during high-speed phases [8, 22]. A further dynamic can be conceptualized as the trade-off frontier, $F = \max(S)$ subject to $D \geq \theta$, where S is speed, D is depth, and θ is a threshold for minimal understanding. This captures the boundary conditions for sustainable infrastructures, guiding the evolution of materials engineering tools [17, 20].

These implications collectively illuminate how EMF can inform the design of next-generation systems, emphasizing interpretive enhancements amid accelerative demands.

Results and Discussion

The EMF extends beyond theoretical abstraction by integrating with established computational paradigms, offering a cohesive view of speed–understanding tensions. In materials informatics, where machine learning drives property predictions, the framework's momentum balance addresses gaps in current models that favor predictive speed over mechanistic clarity [1-3]. For instance, active learning systems, while effective for acceleration, often overlook the epistemic mass needed for generalizable insights, a shortfall that EMF's steering logics could rectify [5, 7, 25].

Representation learning, a cornerstone of data-driven approaches, benefits from EMF's transduction layer, which promotes interactions that preserve depth [4, 15, 21]. This is particularly relevant in handling uncertainties, where traditional quantification methods may not fully capture systemic risks, suggesting a role for EMF in enhancing interpretability [8, 9, 12]. Autonomous discovery and closed-loop systems further align with EMF's cycles, implying that incorporating recalibration could mitigate epistemic erosion in real-time workflows [6, 11, 18].

Broader field implications include the potential for EMF to influence inverse design and high-throughput paradigms, where trade-offs are acute [17, 19, 28]. By formalizing dynamics like feedback efficacy, the framework provides tools for computational steering that balance efficiency with rigor [22, 26, 27]. Challenges remain, such as scaling EMF to vast datasets without computational overhead, but its systems-level focus positions it as a versatile architecture for evolving materials ecosystems [14, 24, 29].

Ultimately, EMF encourages a paradigm where acceleration amplifies, rather than supplants, understanding, paving the way for more integrated computational strategies.

Conclusion

The exploration of speed–understanding trade-offs in computational materials design underscores the need for frameworks like EMF to harmonize accelerative and epistemic elements. By delineating layers and dynamics, EMF provides interpretive insights into pipeline optimizations, steering toward resilient discovery systems. As materials engineering advances, adopting such architectures could enhance the fidelity of AI-guided processes, ensuring that rapid innovations are underpinned by robust comprehension. This conceptual shift holds promise for transforming computational ecosystems into balanced infrastructures that drive sustainable progress in the field.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 02 Nov 2021 Revised: 16 Apr 2022 Accepted: 03 May 2022
Published online: 18 September 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.

Schmidt J, Marques MRG, Botti S, Marques MAL. Recent advances and applications of machine learning in solid-state materials science. npj Computat Mater. 2019;5(1):83.

Kim S, Yoon H, Lee J, Kim S, Lee S, Cho S, et al. Deep learning accelerates the detection of live bacteria using thin-film transistor arrays. Adv Intell Syst. 2019;1(6):1900048.

Chen C, Ye W, Zuo Y, Zheng C. Graph networks as a universal machine learning framework for molecules and crystals. Chem

Mater. 2019;31(9):3564-72.

Tran K, Ulissi ZW. Active learning across intermetallics to guide discovery of electrocatalysts for CO2 reduction and H2 evolution. Nat Catal. 2018;1(9):696-703.

Ward L, O'Keeffe SC, Stevick J, Wolverton C, Chard K, Gainaru C, et al. A long-lived automated materials discovery infrastructure. npj Comput Mater. 2021;7(1):175.

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.

Ward L, Wolverton C. Atomistic calculations and materials informatics: A review. Curr Opin Solid State Mater Sci. 2017;21(3):167-76.

Pilania G, Gubernatis JE, Lookman T. Multi-fidelity machine learning models for accurate bandgap predictions of solids. Comput Mater Sci. 2017;129:156-63.

Pilania G, Mannodi-Kanakkithodi A, Uberuaga BP, Ramprasad R, Gubernatis JE, Lookman T. Machine learning bandgaps of double perovskites. Sci Rep. 2016;6(1):19375.

Montoya JH, Winther KT, Flores RA, Bligaard T, Hummelshøj JS, Aykol M. Autonomous intelligent agents for accelerated materials discovery. Chem Sci. 2020;11(32):8517-32.

Ling J, Hutchinson M, Antono E, Paradiso S, Meredig B. High-Dimensional materials and process optimization using data-driven experimental design with uncertainty. Integr Mater Manuf Innov. 2017;6(3):207-17.

Meredig B, Antono E, Church C, Hutchinson M, Ling J, Paradiso S, et al. Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance in materials discovery. Mol Syst Des Eng. 2018;3(6):819-25.

Ward L, Dunn A, Faghaninia A, Zimmermann NER, Bajaj S, Wang Q, et al. Matminer: An open source toolkit for materials data mining. Comput Mater Sci. 2018;152:60-9.

Ahmad Z, Xie T, Maheshwari C, Grossman JC, Viswanathan V. Machine learning enabled computational screening of inorganic solid electrolytes for suppression of dendrite formation in lithium metal anodes. ACS Cent Sci. 2018;4(8):996-1006.

Kim E, Huang K, Saunders A, McCallum A, Ceder G, Olivetti E. Materials synthesis insights from scientific literature via text extraction and machine learning. Chem Mater. 2017;29(21):9436-44.

Xue D, Balachandran PV, Hogden J, Theiler J, Xue D, Lookman T. Accelerated search for materials with targeted properties by adaptive design. Nat Commun. 2016;7(1):11241.

Sun S, Hartono NTP, Ren ZD, Oviedo F, Buscemi A, Layurova M, et al. Accelerated optimization of TiO2-mediated UV photocatalysis through automated flow synthesis and machine learning. Joule. 2020;4(1):130-46.

Oliynyk AO, Antono E, Sparks TD, Ghadbeigi L, Gaultois MW, Meredig B, et al. High-Throughput machine-learning-driven synthesis of full-heusler compounds. Chem Mater. 2016;28(20):7324-31.

Ward L, Agrawal A, Choudhary A, Wolverton C. A general-purpose machine learning framework for predicting properties of inorganic materials. npj Comput Mater. 2016;2(1):16028.

Ling J, Jones R, Templeton J. Machine learning strategies for systems with invariance properties. J Comput Phys. 2016;318:22-35.

Meredig B. Five high-impact research areas in machine learning for materials science. Chem Mater. 2019;31(23):9579-81.

Meredig B, Agrawal A, Kirklin S, Saal JE, Doak JW, Thompson A, et al. Combinatorial screening for new materials in unconstrained composition space with machine learning. Phys Rev B. 2014;89(9):094104.

Meredig B, Antono E, Church C, Hutchinson M, Ling J, Paradiso S, et al. High-Throughput computational screening of 12,000 experimentally stable compounds. Mol Syst Des Eng. 2020;5(2):485-94.

Balachandran PV, Xue D, Theiler J, Hogden J, Lookman T. Adaptive strategies for materials design using uncertainties. Sci Rep. 2016;6(1):19660.

Oliynyk AO, Adutwum LA, Rudyk BW, Pisavadia H, Lawler S, Mar A. Exploring the cationic states of I–III–IV2 semiconductors: An example of computational materials discovery for photovoltaic applications. Cryst Growth Des. 2018;18(6):3661-9.

Saal JE, Oliynyk AO, Meredig B. Machine learning in materials discovery: Confirmed predictions and lessons learned. Annu Rev Mater Res. 2020;50(1):49-69.

Kim E, Huang K, Jegelka S, Olivetti E. Virtual screening of inorganic materials synthesis parameters with deep learning.

npj Comput Mater. 2017;3(1):53.

Aykol M, Hegde VI, Suram L, Hung L, Malik K, Zhao L, et al. Network analysis of synthesizable materials discovery. Nat Commun. 2019;10(1):2018.