ORIGINAL RESEARCH **Open access**

Closed-World Training in an Open Materials Universe

Natalia Petrova[1*], Elena Stoyanova[1], Ivan Dimitrov[2]

Abstract

In the rapidly evolving field of computational and data-driven materials engineering, machine learning models are increasingly trained on curated datasets that represent a closed-world approximation of material properties and behaviors. However, the broader materials universe encompasses vast, unexplored compositional spaces, dynamic environmental interactions, and emergent phenomena that defy static boundaries. This conceptual manuscript addresses the inherent tension between closed-world training paradigms—characterized by finite, labeled data regimes—and the open, infinite nature of materials discovery. We introduce a novel conceptual framework, termed the Adaptive Boundary Inference Architecture (ABIA), which integrates representation learning, uncertainty-aware feedback mechanisms, and multi-scale inference logics to navigate this disparity. ABIA conceptualizes training as a dynamic process where model boundaries adapt through iterative interactions between data representations and discovery pipelines, fostering resilience to out-of-distribution materials. By synthesizing recent advances in graph neural networks, foundation models, and autonomous systems, the framework highlights computational steering strategies that balance exploitation of known data with exploration of open spaces. Implications extend to enhanced inverse design, multimodal integration, and epistemic risk management in materials informatics, ultimately advancing sustainable and efficient materials engineering workflows. This work underscores the need for interpretive systems that transcend traditional closed-loop constraints, promoting a more holistic approach to data-driven discovery in an unbounded materials landscape.

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

*Correspondence:
Natalia Petrova
natalia.petrova@gmail.com
[1] Department of Computational Materials Engineering, Faculty of Engineering, University of Sofia, Sofia, Bulgaria
[2] Department of Data-Driven Materials Systems, Faculty of Engineering, Technical University of Sofia, Sofia, Bulgaria

Introduction

The advent of computational and data-driven paradigms has fundamentally reconfigured the epistemology and practice of materials engineering. Once dominated by empiricism, serendipitous discovery, and iterative laboratory experimentation, the field now operates within an increasingly predictive and automation-enabled paradigm [1-3]. Machine learning algorithms, high-throughput simulations, and integrated informatics infrastructures collectively enable the exploration of chemical and structural design spaces at scales previously unattainable through experimental means alone. In this transformed landscape, materials discovery is no longer constrained by laboratory throughput but by the representational and inferential capacities of computational systems.

At the core of this transformation lies the coupling of machine learning architectures with large-scale simulation ecosystems. Density functional theory (DFT), molecular dynamics, and combinatorial screening pipelines generate expansive property datasets, which are subsequently mined by predictive models to forecast stability, performance, and synthesis feasibility. Such infrastructures facilitate accelerated identification of candidate materials

for applications spanning energy storage, catalysis, quantum devices, and structural engineering.

Yet, despite these advancements, a foundational epistemic tension persists. Most computational training paradigms operate under a closed-world assumption—an operational regime in which models are optimized on finite, bounded datasets presumed to adequately represent the phenomena of interest [4, 5]. These datasets define the representational universe through which machine learning systems interpret materials reality. Feature spaces, compositional ranges, structural motifs, and environmental conditions are pre-encoded, forming an enclosed epistemic domain within which inference occurs.

This closed-world training paradigm stands in stark contrast to the open, unbounded nature of the materials universe. Materials reality encompasses undiscovered compositions, metastable polymorphs, defect-rich microstructures, and context-dependent behaviors shaped by synthesis pathways and environmental exposures. The combinatorial vastness of chemical possibility—often estimated to exceed 10^100 plausible stable compounds—renders any training dataset inherently incomplete [6]. Consequently, predictive systems operate within epistemic islands embedded in an ocean of unknown materiality.

Evolution of data-driven paradigms in materials science

The emergence of materials informatics has been instrumental in operationalizing computational discovery. Early infrastructural efforts focused on aggregating simulation outputs and experimental measurements into centralized databases, enabling supervised learning for property prediction [7, 8]. Density functional theory repositories and curated experimental archives provided the statistical substrate necessary for regression and classification models, catalyzing the first wave of data-driven materials screening [9, 10].

Subsequent architectural innovations introduced deep learning frameworks capable of modeling non-linear, high-dimensional structure–property relationships. Graph-based neural representations proved especially transformative, encoding atomic connectivity, coordination environments, and electronic interactions within relational learning architectures [11-14]. These approaches enabled predictive modeling across crystalline families and compositional gradients, accelerating discovery in domains such as battery electrodes, heterogeneous catalysis, thermoelectrics, and high-entropy alloys [15, 16].

However, the rapid scaling of predictive accuracy masked an implicit assumption: that training datasets sufficiently spanned relevant materials variability. In practice, datasets remained skewed toward experimentally tractable systems, energetically favorable phases, or historically studied chemistries. This representational bias introduced latent vulnerabilities. When confronted with anomalous compositions, metastable states, or underexplored bonding regimes, model performance often degraded, revealing the limits of closed-environment learning [17, 18].

The magnitude of the materials universe amplifies this limitation. Even the most comprehensive databases capture only infinitesimal fractions of chemically plausible space, reinforcing the structural incompleteness of training infrastructures [6].

The closed–open dichotomy in computational workflows

Closed-world training environments excel in interpolation. Dense sampling within bounded compositional and structural regimes enables high-fidelity predictions, robust benchmarking, and reproducible performance metrics [19, 20]. Within these epistemic enclosures, machine learning systems function as precision instruments, refining predictions across known domains.

In contrast, the open materials universe introduces profound epistemic uncertainty. Undiscovered phases, non-equilibrium synthesis outcomes, defect-mediated properties, and environment-dependent transformations lie beyond encoded knowledge boundaries [21, 22]. When predictive systems encounter such regimes, inference shifts from interpolation to extrapolation—a transition marked by elevated epistemic risk.

This dichotomy manifests operationally across computational discovery pipelines. High-throughput screening infrastructures, while expansive, remain bounded by predefined chemical libraries and simulation heuristics. Rare events, emergent phenomena, and unconventional bonding configurations may remain systematically excluded from exploration [23, 24].

Inverse design strategies further illuminate this tension. Generative frameworks effectively propose candidate

materials within constrained latent manifolds [25]. Yet, as design objectives expand into open chemical space, combinatorial explosion undermines generative tractability and physical plausibility [26]. The challenge thus extends beyond computational capacity—it is infrastructural and epistemological.

Addressing this divide requires reconceptualizing inference systems not as static predictors but as adaptive explorers capable of renegotiating their epistemic boundaries. This entails rethinking how representations, uncertainty signals, and discovery workflows co-evolve within computational ecosystems [27, 28].

Gaps in current discovery infrastructures

Despite rapid progress in multimodal data integration and large-scale foundation models [5, 15], contemporary computational infrastructures remain disproportionately optimized for predictive fidelity within closed regimes. Performance metrics emphasize accuracy, mean absolute error reduction, and benchmark ranking—criteria that privilege interpolation while obscuring extrapolative fragility [3, 7].

Uncertainty quantification methods offer partial corrective mechanisms. Bayesian inference, ensemble modeling, and probabilistic embeddings estimate predictive confidence and identify knowledge gaps [17, 19]. However, these mechanisms are frequently appended as post-hoc evaluative layers rather than embedded within core training logics. As a result, uncertainty signals rarely exert structural influence over discovery trajectories [18, 29].

Moreover, the integration of simulation pipelines with experimental validation loops remains infrastructurally fragmented. Autonomous laboratories and closed-loop experimentation systems often inherit the same closed-world biases embedded within their training data and screening heuristics [8, 22, 23]. Rather than expanding epistemic horizons, such systems risk reinforcing dominant paradigms, allocating computational and experimental resources toward well-charted materials families [4, 6]. Key structural differences between closed-world learning and open-universe discovery—and the corresponding ABIA responses—are summarized in **Table 1**.

Table 1. Closed-World Training vs Open-Universe Discovery: Failure Modes and ABIA Design Responses

Dimension	Closed-World Training Regime (Typical)	Open Materials Universe Reality	Common Mod Prac
Data coverage	Finite, curated, biased toward tractable systems	Vast, sparse, long-tail novelty	Overcon in underrepr chemi
Representation space	Fixed embeddings / descriptor assumptions	Evolving structural motifs and contexts	Mani disconti represe colla
Inference objective	In-distribution accuracy; benchmark performance	Robustness under shift; anomaly tolerance	OOD brit extrapo instal
Uncertainty role	Often post-hoc (appended UQ layer)	Primary signal of “unknown unknowns”	Uncer ignore explorati
Screening workflow	Predefined libraries; static search heuristics	Emergent phenomena, rare events, metastability	System missed d regi
Simulation–experiment coupling	Fragmented or sequential	Coupled, context-dependent validation	Bias prop through loop aut

Inverse design	Constrained latent spaces	Combinatorial explosion; feasibility limits	Physi implau candic narrow
Multimodal integration	Partial fusion; modality mismatch	Heterogeneous, noisy, misaligned evidence	Amplifie through misalig

This infrastructural inertia constrains exploratory breadth. Discovery pipelines may optimize efficiency while inadvertently suppressing novelty—a paradox at the heart of data-driven materials engineering.

Positioning of the present work

In response to these systemic tensions, this work introduces the Adaptive Boundary Inference Architecture (ABIA) as a conceptual framework for reconciling closed-world computational training with open-universe materials exploration.

ABIA advances three integrative premises:

1. **Boundary Awareness** – Computational systems must explicitly model the epistemic limits of their training domains.
2. **Adaptive Inference** – Learning architectures should dynamically recalibrate representations and predictive logics when encountering open-space anomalies.
3. **Discovery Steering** – Uncertainty, representational sparsity, and simulation–experiment discrepancies should function as navigational signals guiding exploratory expansion.

Structurally, ABIA conceptualizes materials discovery as a layered interaction between data ingestion infrastructures, adaptive model architectures, and boundary-steering discovery workflows. Rather than treating openness as noise or error, the framework interprets it as a generative dimension of scientific inquiry—an informational frontier where discovery potential is maximized.

By situating computational materials engineering within this boundary-adaptive paradigm, the present manuscript seeks to provide an interpretive lens through which predictive systems can evolve from closed optimization engines into open discovery orchestrators.

Theoretical Background & Literature Synthesis

The theoretical foundations of computational materials engineering emerge from the convergence of machine learning, materials informatics, statistical physics, and systems science, forming an interdisciplinary scaffold for understanding discovery infrastructures in data-intensive environments [1, 2]. Within this synthesis, materials design is no longer interpreted as a linear experimental endeavor but rather as a multiscale inference ecosystem, where computational representations, algorithmic reasoning, and physical validation interact across hierarchical strata. These strata span atomistic simulations, mesoscale structural modeling, and macroscopic performance prediction, collectively constituting nested epistemic layers within materials data ecosystems [3, 4].

Such ecosystems are structurally bounded. Training datasets, simulation regimes, and experimental archives form enclosed knowledge domains—operationally efficient but epistemically delimited. The closed–open divide thus arises: computational pipelines are optimized within finite representational universes, yet they are deployed to interrogate an effectively infinite materials possibility space. Understanding this divide requires examining the infrastructures through which knowledge is encoded, inferred, and operationalized.

Representation learning in materials contexts

Representation learning constitutes the epistemic entry point through which materials reality becomes computationally tractable. By encoding crystallographic structures, chemical compositions, and microstructural morphologies into machine-readable embeddings, representation learning transforms physical matter into navigable mathematical manifolds [11-13].

Graph neural networks (GNNs) have emerged as particularly influential in this regard. Their relational architectures enable the encoding of bond topologies, coordination environments, and lattice symmetries, supporting predictive modeling across crystalline and polycrystalline systems [14, 27, 29]. Unlike handcrafted descriptors, these learned representations embed invariance to translational, rotational, and permutational symmetries, enhancing cross-material transferability and reducing descriptor bias [10, 20].

However, representational robustness is contingent on exposure. In open materials universes characterized by compositional novelty and structural sparsity, embeddings trained on finite datasets encounter generalization limits. Representation collapse, manifold discontinuities, and extrapolative instability emerge when models encounter chemistries absent from training regimes [5, 18]. Literature identifies data augmentation as a partial mitigation strategy—introducing structural perturbations, simulated defects, or compositional interpolations to approximate openness [19, 27, 30]. Yet such approaches remain epistemically tethered to closed training distributions, extending but not transcending their boundaries.

Thus, representation learning simultaneously enables discovery and constrains it, functioning as both epistemic bridge and bottleneck.

Machine learning architectures and their limitations

Advancements in machine learning architectures have expanded the predictive bandwidth of computational materials science. Deep neural networks, attention mechanisms, and graph-based encoders facilitate high-dimensional property inference, enabling rapid screening of candidate materials across vast design spaces [7, 15, 17].

Foundation models represent a significant architectural evolution. Pre-trained on expansive multimodal corpora—including crystallographic databases, simulation outputs, and literature text—they provide transferable embeddings that accelerate downstream tasks such as bandgap prediction, phase stability estimation, and catalytic activity screening [5, 15]. Their scale introduces emergent representational coherence, enabling cross-task generalization previously unattainable in narrow models.

Parallel progress in explainable AI (XAI) has enhanced interpretability within these architectures. Attribution mapping, saliency analysis, and feature importance decomposition illuminate the internal reasoning pathways of predictive systems, exposing biases toward overrepresented chemistries or structural motifs [17, 21]. Such transparency is critical for scientific legitimacy, aligning algorithmic inference with mechanistic plausibility.

Despite these advancements, architectural optimization remains anchored in closed-world assumptions. Training objectives privilege in-distribution accuracy, reinforcing performance within known domains while attenuating resilience under distributional shift [6, 18]. Benchmarking initiatives, though methodologically rigorous, reveal performance degradation when models are extrapolated to novel compositions, metastable phases, or underexplored synthesis conditions [6, 28, 31].

This fragility underscores a structural asymmetry: architectures scale computationally faster than their epistemic coverage expands.

High-throughput and autonomous discovery systems

High-throughput computational infrastructures operationalize machine learning predictions within automated discovery pipelines. Density functional theory (DFT) screening, combinatorial simulation, and rapid property estimation enable large-scale exploration of compositional spaces, compressing discovery timelines from decades to years [8, 23, 24].

Autonomous discovery systems extend this paradigm by embedding AI within closed-loop experimental platforms. Here, machine learning models dynamically guide synthesis, characterization, and validation cycles, forming self-optimizing experimentation ecosystems [22, 23]. Robotic laboratories, adaptive synthesis planning, and real-time characterization feedback exemplify this convergence of computation and physical experimentation.

Active learning plays a pivotal role in these systems. By prioritizing high-uncertainty or high-information samples, active learning algorithms optimize experimental allocation, reducing redundant trials while maximizing knowledge gain [9, 22]. This strategic sampling partially mitigates data sparsity, enabling efficient expansion of training datasets.

Yet autonomy does not equate to openness. Discovery systems inherit epistemic constraints from their foundational datasets and simulation priors. Screening pipelines may systematically overlook emergent phenomena—metastable polymorphs, unconventional bonding motifs, or non-equilibrium synthesis pathways—lying beyond encoded knowledge regimes [4, 26, 32].

Consequently, literature increasingly emphasizes adaptive boundary expansion: infrastructures capable not only of optimizing within known domains but of structurally redefining their search universes during operation [16, 25].

Inverse design and multimodal integration

Inverse design reconfigures the discovery logic of materials engineering by reversing the traditional structure-to-property mapping. Rather than predicting performance from known materials, generative frameworks propose candidate structures optimized for target functionalities—superconductivity, catalytic efficiency, mechanical resilience, or optical response [25, 33].

Generative adversarial networks (GANs), variational autoencoders (VAEs), and diffusion-based models populate latent design manifolds, sampling candidate materials from learned distributions. These models expand exploration capacity, identifying compositions and structures absent from empirical databases.

However, generative efficacy is contingent on prior completeness. In open universes lacking comprehensive training coverage, generative outputs risk plausibility gaps—structures that are mathematically coherent yet physically unrealizable [25, 26]. This exposes a generative paradox: design space expansion may outpace physical validity.

Multimodal integration offers a partial corrective. By combining structural data, experimental measurements, simulation outputs, and scientific text, multimodal models enrich contextual learning and improve generative grounding [5, 20, 21]. Text-mined synthesis pathways, for example, can constrain generative proposals toward experimentally feasible regimes.

Yet modality fusion introduces alignment challenges. Divergent noise structures, resolution mismatches, and representational incongruence across modalities propagate uncertainty through inference pipelines [18, 19, 34]. Hybrid orchestration frameworks—where computational steering dynamically reconciles modality discrepancies—are thus emerging as critical infrastructural innovations [3, 7, 8].

Uncertainty quantification and epistemic risks

Uncertainty quantification (UQ) underpins reliability in AI-driven materials inference. Bayesian neural networks, Monte Carlo dropout, Gaussian processes, and ensemble learning approaches quantify predictive confidence, distinguishing aleatoric variability from epistemic ignorance [17-19].

Aleatoric uncertainty reflects intrinsic noise—measurement error, thermal fluctuations, or synthesis variability—while epistemic uncertainty arises from knowledge insufficiency. In open discovery universes, epistemic uncertainty dominates, manifesting as unknown unknowns embedded within unexplored compositional and structural regimes.

Interpreting epistemic uncertainty as merely predictive variance underutilizes its epistemological value. Literature increasingly frames it as a navigational signal—an indicator of boundary regions where knowledge infrastructures should expand [17, 28]. When coupled with simulation–experiment interfaces, uncertainty gradients reveal mismatches between modeled predictions and empirical realities, exposing zones of epistemic fragility [10, 22, 23].

Such coupling transforms uncertainty from a passive confidence metric into an active discovery driver, steering exploration toward high-impact knowledge frontiers.

Synthesis: Closed precision vs open exploration

Across representation learning, architectural scaling, autonomous experimentation, inverse design, and uncertainty modeling, a structural pattern emerges. Computational materials systems achieve extraordinary precision within delimited epistemic spaces yet exhibit fragility when confronting the open-endedness of real materials universes [1, 2, 6].

Closed training infrastructures enable optimization, benchmarking rigor, and reproducibility. However, they simultaneously constrain exploratory breadth, embedding systemic blind spots within discovery pipelines. High-throughput automation accelerates search but does not

inherently expand its epistemic horizon. Generative models proliferate candidates but depend on prior completeness. Uncertainty quantification signals risk yet requires infrastructural mechanisms to operationalize response.

This synthesized landscape reveals a foundational infrastructural trade-off:

Precision scales within closure; Discovery scales through openness

Bridging this divide requires conceptual architectures capable of dynamically negotiating system boundaries—frameworks that treat openness not as noise to suppress but as a structural dimension of scientific exploration [3, 4, 15].

Proposed conceptual framework

To address the conceptual gap between closed-world training and the open materials universe, we introduce the Adaptive Boundary Inference Architecture (ABIA). ABIA is structured as a multi-layered system that conceptualizes training not as a static optimization but as a dynamic interplay between bounded data regimes and expansive discovery horizons. At its core, ABIA comprises three structural layers: the Representation Ingestion Layer, the Inference Adaptation Layer, and the Discovery Steering Layer. These layers form interconnected pipelines where data flows from ingestion to model refinement and onward to exploratory outputs, modulated by feedback loops that incorporate uncertainty signals.

The Representation Ingestion Layer processes multimodal inputs—such as structural graphs, property vectors, and simulation-derived features—into adaptive embeddings. Unlike fixed representations, this layer employs dynamic encoding schemes that allow for incremental updates, capturing interactions between known and potential unknowns. The pipeline then transitions to the Inference Adaptation Layer, where models adjust boundaries through iterative refinement, balancing exploitation of closed data with probes into open spaces. Finally, the Discovery Steering Layer directs outputs toward inverse design or autonomous workflows, using epistemic cues to prioritize uncharted territories.

Feedback loops are integral, enabling bidirectional information flow: uncertainties from inference feedback to refine representations, while discovery outcomes inform adaptation strategies. This creates computational steering logics that interpret system states, such as representation completeness or inference divergence, to guide resource allocation. ABIA operationalizes the closed–open tension as a layered architecture with uncertainty-driven feedback loops that expand inference boundaries during discovery **(Figure 1)**.

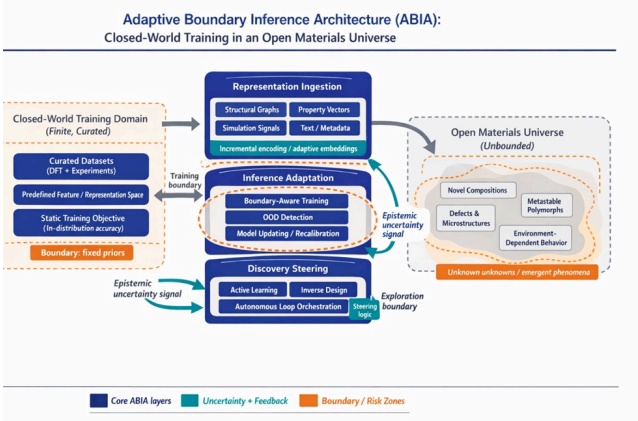

Figure 1. Adaptive Boundary Inference Architecture (ABIA) for closed-world training under open-universe materials discovery.

ABIA formalizes the closed–open tension as a three-layer architecture—representation ingestion, inference adaptation, and discovery steering—linked by uncertainty-aware feedback loops. The left enclosure depicts finite curated training regimes (bounded datasets, predefined representations, static objectives), while the right region represents the unbounded materials universe (novel compositions, metastable structures, defect-rich microstructures, and context-dependent behaviors). Teal feedback pathways operationalize epistemic uncertainty and validation mismatch as navigational signals that trigger boundary adaptation, enabling models to preserve closed-domain precision while expanding robustness to out-of-distribution materials.

A key dynamic within ABIA can be conceptualized as the boundary adaptation function, which captures the interaction between closed training entropy and open exploration potential. This may be expressed as $B(t) = \alpha \cdot Hc(D) + \beta \cdot \int Uo(x)\,dx$, where $H_c(D)$ denotes the entropy of the closed dataset D, $U_o(x)$ represents uncertainty density over open variables x, and α, β are weighting factors for balance. This formula interpretive illustrates how training

boundaries evolve over time t, emphasizing trade-offs in allocating computational focus.

Another aspect formalizes feedback loop efficiency as $F = \partial I \partial R \cdot (1 - \gamma \cdot \Delta E)$, where R is representation fidelity, I is inference output, ΔE captures epistemic divergence, and γ modulates risk sensitivity. This expression highlights how adaptations mitigate risks in open contexts without empirical tuning.

Finally, discovery steering logic may be captured as $S = \sum kwk \cdot P_k(U)$, with $P_k(U)$ as priority functions over uncertainty-driven paths k, weighted by w_k for pipeline alignment. These formulas underscore ABIA's interpretive power in navigating closed-open tensions.

Analytical implications

The Adaptive Boundary Inference Architecture (ABIA) offers a range of analytical implications for computational and data-driven materials engineering, particularly in how it reinterprets workflow dynamics across representation, inference, and discovery stages [1-3]. By framing closed-world training as an adaptive process, ABIA provides insights into optimizing resource allocation in high-throughput systems, where computational costs often scale with data volume [4, 8, 23].

Subheading: Implications for Representation-Inference Interactions In ABIA, representations are not static but evolve through interactions with inference layers, allowing for interpretive analysis of how embedding spaces handle open universe intrusions [11-13, 27]. This implies a shift toward resilient encodings that prioritize relational flexibility over rigid feature sets [14, 20, 29]. For instance, in multimodal contexts, ABIA suggests steering logics that harmonize disparate data streams, reducing fragmentation in property predictions [5, 21]. Analytically, this can be expressed as the representation resilience metric, conceptualized as $Rr = \eta \cdot (1 - \sigma c \sigma o)$, where σo and σc denote variance in open and closed embeddings, respectively, and η \eta η scales interpretive robustness. This formula captures the trade-off between maintaining closed-world accuracy and accommodating open variations, guiding infrastructure designs that minimize epistemic drift [17-19].

Subheading: Systems-Level Insights into Discovery Pipelines ABIA's feedback loops imply enhanced steering in discovery pipelines, where uncertainty signals direct exploration beyond closed boundaries [22, 23, 25]. This fosters analytical views on inverse design, interpreting generative processes as boundary-expanding operations rather than mere optimizations [25, 26, 33]. In autonomous systems, implications extend to coupling simulations with experimental feedbacks, where ABIA logics interpret discrepancies as opportunities for layer realignment [8, 23, 24]. A further implication formalizes this as the steering efficiency interaction, $= \kappa \cdot \log(P_k P_e) - \lambda \cdot Cu$, with P_e and P_k as exploration and known probabilities, C_u as uncertainty cost, and κ, λ as balancing coefficients. This expression interpretive highlights how pipelines can dynamically weigh exploitation against exploration, informing trade-offs in computational budgets [6, 16, 28].

Subheading: Epistemic Risk Structures and Infrastructure Trade-Offs Analytically, ABIA elucidates epistemic risk structures by integrating uncertainty quantification into core layers, implying proactive management of open-world vulnerabilities [17-19, 21]. This leads to insights on infrastructure trade-offs, such as the balance between model complexity and adaptability in graph-based architectures [7, 10, 15, 27]. For example, in materials informatics ecosystems, ABIA implies that closed-loop constraints can be mitigated through adaptive inference, enhancing overall system resilience [3, 4, 22]. Another dynamic can be captured as the risk mitigation function, $Mr = \mu \cdot Uq - \nu \cdot (Db - Da)$, where U_q is quantified uncertainty, D_b and D_a represent boundary and actual data divergences, and μ,ν \mu, \nu μ,ν modulate risk sensitivity. This formula underscores interpretive strategies for minimizing losses in open scenarios, applicable to uncertainty-aware workflows [9, 18, 19]. Overall, these implications promote a holistic view of materials engineering infrastructures, emphasizing interpretive integrations over isolated optimizations [1, 2, 6].

Results and Discussion

The conceptual framework of ABIA advances the discourse in computational materials engineering by providing a

systems-level lens for navigating the closed-open dichotomy [1, 3, 4]. It integrates disparate elements from representation learning to discovery steering, offering interpretive coherence amid the field's rapid evolution [2, 5, 7, 15]. Key strengths lie in its emphasis on feedback dynamics, which align with emerging trends in autonomous and active learning systems [8, 9, 22, 23]. However, interpretive challenges arise in scaling ABIA logics to real-world infrastructures, where computational constraints may limit adaptive iterations [6, 24, 28].

In synthesizing literature, ABIA complements graph neural network advancements by interpreting their relational strengths as foundations for boundary adaptation [11-14, 27, 29]. This extends to inverse design, where generative mechanisms gain from uncertainty-driven steering [25, 26, 33]. Yet, multimodal integration remains a focal point, as ABIA's layers imply nuanced handling of data heterogeneities that current architectures often overlook [20, 21]. Epistemic risks, central to open universe navigation, benefit from ABIA's risk structures, aligning with quantification methods but advocating for deeper embedding [17-19].

Broader field implications include fostering collaborative ecosystems, where ABIA-inspired workflows could standardize discovery pipelines across informatics platforms [8, 16]. Trade-offs, such as between precision and exploration, highlight the need for balanced computational strategies [3, 6]. Ultimately, ABIA encourages a shift toward resilient, interpretive systems that embrace the materials universe's openness, potentially accelerating innovations in sustainable engineering [4, 15, 23].

Conclusion

In summary, this conceptual manuscript has explored the tension inherent in closed-world training within the expansive open materials universe, proposing the Adaptive Boundary Inference Architecture (ABIA) as a unifying framework. Through layered structures, feedback mechanisms, and steering logics, ABIA provides interpretive insights into enhancing computational workflows, from representation adaptation to discovery optimization. Analytical implications underscore trade-offs and dynamics that foster robustness, while integrating uncertainty and multi-scale interactions.

Looking ahead, ABIA's conceptual contributions pave the way for more adaptive materials engineering paradigms, emphasizing systemic resilience over static models. By transcending closed boundaries, it holds promise for advancing data-driven discoveries in an unbounded landscape.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 06 Jul 2023 Revised: 03 Aug 2023 Accepted: 20 Oct 2023
Published online: 18 March 2024

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.

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559(7715):547-55.

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.

de Pablo JJ, Jackson NE, Webb MA, Chen LQ, Moore JE, Morgan D, et al. New frontiers for the materials genome initiative. npj Comput Mater. 2019;5(1):41.

Tshitoyan V, Dagdelen J, Weston L, Dunn A, Rong Z, Kononova O, et al. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature. 2019;571(7763):95-8.

Dunn A, Wang Q, Ganpule S, Sapkota D, Ceder G. Benchmarking materials property prediction methods: The Matbench test set and Automatminer reference algorithm. npj Comput Mater. 2020;6(1):138.

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.

Hu J, Stefanov S, Song Y, Sadhukhan K, Okunishi E, Apley DW, et al. MaterialsAtlas.org: Amaterials informatics web app platform for materials discovery and survey of state-of-the-art. npj Comput Mater. 2022;8(1):65.

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

Jørgensen PB, Jacobsen KW, Schmidt MN. Neural message passing with edge updates for predicting properties of molecules and materials. npj Comput Mater. 2018;4:1-8.

Fung V, Hu G, Ganesh P, Sumpter BG. Machine learned features from density functional theory simulations for accurate adsorption energy prediction. Nat Commun. 2021;12(1):88.

Choudhary K, DeCost B. Atomistic line graph neural network for improved materials property predictions. npj Comput Mater. 2021;7(1):185.

Dai M, Demirel MF, Liang Y, Hu JM. Graph neural networks for an accurate and interpretable prediction of the properties of polycrystalline materials. npj Comput Mater. 2021;7(1):103.

Goodall RE, Lee AA. Predicting materials properties without crystal structure: deep representation learning from stoichiometry. Nat Commun. 2020;11(1):6280.

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624(7990):80-5.

Kaufmann K, Vecchio KS. Searching for high entropy alloys: A machine learning approach. Acta Mater. 2020;198:231-40.

Zhong X, Gallagher B, Liu S, Hiszpanski A, Kailkhura B, Han TY-J. Explainable machine learning in materials science. npj Comput Mater. 2022;8(1):204.

Bartel CJ, Trewartha A, Wang Q, Dunn A, Jain A, Ceder G. A critical examination of compound stability predictions from machine-learned formation energies. npj Comput Mater. 2020;6(1):97.

Xu P, Ji X, Li M, Lu W. Small data machine learning in materials science. npj Comput Mater. 2023;9(1):42.

Chen C, Zuo Y, Ye W, Li X, Ong SP. Learning properties of ordered and disordered materials from multi-fidelity data. Nat Comput Sci. 2021;1(1):46-53.

Zhang L, Wang YC, Wan B, Wang F, Zhou J, Dai X, et al. Accelerating discovery of hidden material properties from large datasets using machine learning. Adv Intell Syst. 2022;4(3):2100153.

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.

Szymanski NJ, Rendy B, Fei Y, Kumar RE, He T, Milikisiyants D, et al. An autonomous laboratory for the accelerated synthesis of novel materials. Nature. 2023;624(7990):86-91.

Meredig B, Antono A, McGrath R, Christensen M, Gent W, Trickey W. From materials discovery to production: informatics enables the product pipeline. Digit Discov. 2022;1(4):356-64.

Dan Y, Zhao Y, Li X, Li S, Hu M, Hu J. Generative adversarial networks (GAN) based efficient sampling of chemical space for inverse design of inorganic materials. npj Comput Mater. 2020;6(1):84.

Fung V, Ganesh P, Sumpter BG. Graph neural network predictions of metal-organic framework CO2 adsorption properties. Digit Discov. 2022;1(6):733-42.

Gibson J, Hire A, Hennig RG. Data-augmentation for graph neural network learning of the relaxed energies of unrelaxed structures. npj Comput Mater. 2022;8(1):211.

Rosen AS, Iyer A, Chen Y, Curtarolo S, Aykol M. Machine learning for high-throughput experimental exploration of metal halide perovskites. Digit Discov. 2023;2(6):1851-66.

Ong PV, Johnson LE, Hosono H, Sushko PV. Structure and stability of CaH 2 surfaces: On the possibility of electron-rich surfaces in metal hydrides for catalysis. J Mater Chem A. 2017;5(11):5550-8.

Li Q, Fu N, Omee SS, Hu J. MD-HIT: Machine learning for material property prediction with dataset redundancy control. npj Comput Mater. 2024;10:245.

Frey NC, Wang J, Rackers JA, Weiser J, Kingsbury RW, Ren H, et al. SchNetPack 2.0: A neural network toolbox for atomistic machine learning. Digit Discov. 2023;2:993-1003.

Zheng P, Roungpaisarnkit P, Zubatiuk T, Wei Y, Cencer MM, Sigalov S, et al. A graph neural network with negative message passing for graph color prediction. Digit Discov. 2024;3:128-37.

Chen L, Tran H, Batra R, Kim C, Ramprasad R. Machine learning models for the prediction of energy, forces, and stresses for heterogeneous materials. npj Comput Mater. 2021;7:13.

Wang AY-T, Murdock RJ, Kauwe SK, Oliynyk AO, Gurlo A, Brgoch J,et al. Machine learning for materials scientists: An introductory guide toward best practices. Chem Mater. 2020;32:4954-65.