ORIGINAL RESEARCH **Open access**

Latent Anisotropy: Directional Bias in Materials Embedding Spaces

Daniel Brooks[1*], Ethan Moore[1], Amelia Carter[2]

Abstract

In the evolving landscape of computational and data-driven materials engineering, embedding spaces serve as foundational representations that encode material properties, structures, and behaviors into vectorial forms amenable to machine learning workflows. These spaces facilitate high-throughput screening, inverse design, and autonomous discovery by bridging atomic-scale simulations with macroscopic predictions. However, inherent directional biases—termed latent anisotropy—emerge from the interplay of data modalities, architectural choices in neural networks, and inference dynamics, potentially skewing discovery pathways toward certain material classes or property regimes. This conceptual manuscript identifies a critical gap in understanding how such biases propagate through materials informatics pipelines, influencing the epistemic reliability of AI-assisted materials research. We introduce the Anisotropic Representation Cascade (ARC) framework, which conceptualizes embedding spaces as multi-layered systems where directional preferences arise from representation encoding, propagation through graph-based architectures, and feedback in closed-loop systems. By integrating insights from uncertainty quantification and multimodal data fusion, ARC elucidates trade-offs in computational steering logics that balance exploration breadth with directional fidelity. Implications extend to enhancing robustness in foundation models for materials science, fostering more equitable navigation of chemical spaces, and informing infrastructure designs that mitigate bias amplification in simulation-experiment couplings. This work underscores the need for interpretive tools in data-driven paradigms to ensure unbiased acceleration of materials innovation.

Keywords Materials informatics, Graph neural networks, Representation learning, Materials embedding spaces, Latent anisotropy, Directional bias

*Correspondence:
Daniel Brooks
daniel.brooks@gmail.com
[1] Department of Computational Materials Engineering, Faculty of Engineering, University of Manchester, Manchester, United Kingdom
[2] Department of Data-Driven Materials Science, Faculty of Engineering, University of Birmingham, Birmingham, United Kingdom

Introduction

The advent of data-driven paradigms in materials engineering

The integration of computational methods with data-centric approaches has transformed materials engineering from a predominantly empirical discipline into a predictive science capable of accelerating discovery cycles [1-3]. Traditionally, materials development relied on trial-and-error experimentation guided by domain expertise and thermodynamic principles. However, the advent of high-throughput computation has enabled the generation of vast datasets encompassing electronic structures, phase stabilities, and functional properties across diverse chemical compositions [4-6]. This shift is exemplified by frameworks like the Automatic FLOW for Materials Discovery (AFLOW), which leverage machine learning to predict thermodynamic stability without exhaustive simulations [4]. Concurrently, machine learning techniques have permeated materials informatics, allowing for the extraction of patterns from multimodal datasets that include crystallographic information, spectroscopic data, and simulation outputs [7-9].

Such paradigms are underpinned by the ability to represent materials in computationally tractable forms. Embedding spaces, where materials are mapped to high-dimensional vectors, emerge as pivotal constructs that encapsulate intrinsic attributes like atomic coordination and electronic configurations [10-12]. These representations enable efficient querying of property landscapes, facilitating tasks such as alloy design and catalyst optimization [5, 13, 14]. For instance, generative models like adversarial networks have been employed to sample chemical spaces inversely, identifying compositions with targeted properties [12]. Yet, as these systems scale, the complexity of data-driven workflows introduces subtle distortions that can influence the directionality of exploratory searches.

Challenges in representation and inference dynamics

A core challenge in computational materials engineering lies in the fidelity of representations derived from heterogeneous data sources. Materials datasets often exhibit multimodality, combining ordered crystalline structures with disordered amorphous phases, necessitating learning frameworks that handle varying fidelities [8, 15, 16]. Graph neural networks (GNNs) have proven instrumental in this regard, modeling atomic interactions as relational graphs to predict properties like bandgaps or mechanical responses [9-11]. However, these architectures inherently impose structural assumptions, such as rotational invariance or locality biases, which may not uniformly capture anisotropic phenomena inherent to materials like layered compounds or ferroelectrics [17, 18].

Moreover, inference in embedding spaces is not isotropic; directional preferences can arise from training objectives that prioritize certain subspaces, leading to uneven exploration of the materials universe [19-21]. This is compounded in closed-loop systems where autonomous agents iteratively refine models based on experimental feedback, potentially reinforcing initial biases [14, 17, 22]. Uncertainty quantification plays a mitigating role, providing measures of confidence that guide decision-making in high-stakes applications like energy materials design [16]. Despite these advances, a systematic understanding of how latent directional biases—stemming from embedding construction—cascade through discovery pipelines remains underexplored, risking skewed outcomes in inverse design tasks [12, 23, 24].

Integration of computational ecosystems

The broader ecosystem of materials AI encompasses foundation models that generalize across scientific domains, drawing from large-scale pretraining on diverse corpora [20, 22]. These models, akin to language models in natural language processing, adapt to materials-specific tasks through fine-tuning, enabling predictions in underrepresented regimes [7, 19]. High-throughput infrastructures further amplify this by coupling simulations with machine learning, automating workflows from data acquisition to validation [1, 2, 6]. Yet, the interplay between simulation fidelity and experimental coupling introduces epistemic risks, where biases in embedding spaces could propagate, favoring certain directional trajectories in property optimization [25, 26].

Autonomous discovery systems exemplify this integration, employing reinforcement learning to navigate design spaces [14, 17]. In such setups, embedding biases might steer agents toward local optima aligned with dominant data modes, overlooking novel anisotropies in less-sampled regions [9, 11]. Addressing these requires a conceptual lens that dissects the directional flow within embedding architectures, ensuring balanced representation across material hierarchies [8, 9].

Positioning the current framework

This manuscript positions latent anisotropy as a foundational concept in materials embedding spaces, advocating for a systems-level analysis of bias propagation. By synthesizing computational insights, we introduce the Anisotropic Representation Cascade (ARC) framework, which interprets directional biases through layered interactions in data-model-discovery pipelines, offering interpretive guidance for robust AI infrastructures in materials engineering.

Theoretical Background & Literature Synthesis

Foundational concepts in materials representation learning

Representation learning forms the bedrock of data-driven materials engineering, transforming raw structural and property data into latent spaces that support predictive

modeling [1, 2, 21]. Early efforts focused on feature engineering, where descriptors like atomic fingerprints or symmetry functions encoded local environments for kernel-based predictions [6, 18, 26]. These evolved into end-to-end learning paradigms, where neural architectures directly infer embeddings from input representations such as graphs or voxels [9-11]. Graph neural networks, in particular, leverage message-passing schemes to aggregate neighborhood information, capturing relational dependencies critical for polycrystalline or alloy systems [5, 11, 13].

The transition to deep learning has enabled handling of complex, non-Euclidean data, with models learning hierarchical features that reflect material anisotropies at multiple scales [19, 20]. For example, multi-fidelity approaches integrate low- and high-accuracy data to broaden applicability across ordered and disordered materials [8]. This synthesis highlights how embedding spaces inherently encode directional preferences, influenced by the choice of invariance constraints in network design [17, 18].

Bias and anisotropy in embedding architectures

Directional bias, or anisotropy, in embedding spaces arises from architectural and data-induced asymmetries [7, 10, 15]. In GNNs, convolutional operations may favor certain graph topologies, leading to uneven sensitivity to directional features like bond angles or lattice orientations [9, 11]. Literature on explainable machine learning underscores this, revealing how feature attributions expose latent preferences in predictions for magnetic or thermal properties [7, 15]. Similarly, generative models exhibit biases in sampling, often clustering outputs around high-density regions of trained data, which can skew inverse design toward familiar compositional motifs [12, 24].

Uncertainty quantification frameworks address these by modeling epistemic uncertainties, which signal directional unreliability in unexplored spaces [16, 25]. In high-entropy alloys or metallic glasses, such biases manifest as overconfidence in isotropic assumptions, potentially missing anisotropic behaviors under stress or temperature gradients [5, 13, 15]. Synthesizing these, anisotropy is not merely a artifact but a systemic property emerging from representation-model interactions [3, 19, 21].

Data-driven pipelines and feedback mechanisms

Computational pipelines in materials discovery integrate representation learning with downstream tasks, forming closed loops that couple simulations and experiments [1, 14, 17]. High-throughput systems automate data generation and model refinement, but feedback loops can amplify initial embedding biases, directing discovery toward biased subspaces [4, 6, 23]. Autonomous agents, employing reinforcement strategies, navigate these pipelines by optimizing actions based on embedding-derived rewards, yet risk entrenching directional preferences if uncertainties are inadequately propagated [14, 17, 22].

Multimodal datasets exacerbate this, as fusion of disparate sources—like spectroscopic and crystallographic data—introduces alignment biases that favor certain modalities [8, 16]. Foundation models mitigate through broad pretraining, but still inherit directional artifacts from corpus imbalances [20]. This synthesis reveals pipelines as dynamic systems where anisotropy cascades, influencing the steering of discovery logics [2, 3].

Epistemic implications for infrastructure design

At the infrastructure level, anisotropy in embeddings poses epistemic risks to the reliability of AI-driven materials ecosystems [7, 17, 21]. Inverse design workflows, for instance, rely on invertible mappings in embedding spaces, but directional biases can constrain the invertibility, limiting access to anisotropic material classes [12, 24]. Simulation-experiment couplings further highlight trade-offs, where biased embeddings may misalign computational predictions with empirical realities, necessitating adaptive logics [4, 14, 23].

Literature on interpretable models emphasizes dissecting these risks through attribution and visualization, fostering designs that incorporate bias-aware steering [7, 9, 15]. Overall, synthesizing these strands, theoretical backgrounds point to a need for frameworks that interpret anisotropy as an integral dynamic, informing balanced infrastructures in computational materials science [3, 19, 22].

Proposed conceptual framework

The Anisotropic Representation Cascade (ARC) framework

To address latent anisotropy in materials embedding spaces, we propose the Anisotropic Representation Cascade (ARC) framework. ARC conceptualizes embedding spaces as multi-layered cascades where directional biases emerge and propagate through interconnected stages of data ingestion, model processing, and discovery steering. At its core, ARC delineates three structural layers: the Encoding Layer, where raw material data is projected into vector spaces; the Propagation Layer, involving neural architectures that diffuse information with inherent directional preferences; and the Steering Layer, which governs feedback loops in discovery pipelines. These layers interact via cascade dynamics, where biases from upstream layers influence downstream inferences, creating feedback that refines or amplifies anisotropy.

In the Encoding Layer, multimodal data—such as atomic coordinates and property tensors—are mapped to embeddings, introducing initial directional biases based on data fidelity and modality alignment. This can be conceptualized as a bias injection function, where the embedding vector $\overrightarrow{e}$ for a material is shaped by a directional operator D, such that $\overrightarrow{e} = D(X)$, with X representing input features. Here, D captures the interaction between data modalities, potentially skewing $\overrightarrow{e}$ toward dominant axes in the feature space.

The Propagation Layer extends this through graph-based or deep architectures, where message-passing or convolutional operations enforce locality and symmetry, but at the cost of anisotropic distortions. ARC interprets this as a flow network, with biases accumulating along propagation paths. For instance, the directional bias amplification may be expressed as $\Delta b = \sum_{i,} w_i \cdot \partial_i$ where w_i are layer weights and ∂_i denotes partial derivatives along embedding dimensions, illustrating trade-offs between architectural depth and bias diffusion.

Finally, the Steering Layer integrates these into data-model-discovery pipelines, employing closed-loop feedbacks to adjust exploration trajectories. Uncertainty signals serve as modulators, countering bias by redirecting inference toward underrepresented directions. A key dynamic in ARC is the feedback equilibrium, captured as $F = \alpha (U - B)$, where F is the feedback adjustment, U is uncertainty, B is accumulated bias, and α a scaling factor representing system adaptability. This formula highlights computational steering logics that balance exploitation of biased paths with exploration of anisotropic frontiers.

The structural components and bias injection dynamics across ARC layers are systematized in **Table 1**.

Table 1. Structural Layers of the ARC Framework and Associated Bias Formation Mechanisms

ARC Layer	Core Computational Function	Bias Formation Mechanism	Directio Manifesta
Encoding Layer	Multimodal data → vector embeddings	Modality imbalance, feature weighting, fidelity disparity	Axis-skev latent representa
Propagation Layer	Neural diffusion & relational learning	Architectural inductive biases, locality constraints	Directio featur amplifica
Steering Layer	Exploration optimization & design navigation	Reward shaping, confidence prioritization	Explorat funneli
Feedback Loops	Model retraining & adaptive updates	Uncertainty underutilization	Reinforc anisotro pathwa

ARC's pipelines emphasize iterative cascades: data flows into encodings, propagates through models, and steers discoveries, with loops recycling outputs to refine upstream layers. This fosters resilient infrastructures, mitigating epistemic risks in inverse design by promoting directional diversity. As conceptualized in **Figure 1**, the framework is depicted as a cascading diagram with layered nodes connected by directed edges, feedback arrows looping between layers, and shaded regions indicating bias accumulation zones, alongside symbolic notations for the introduced formulas.

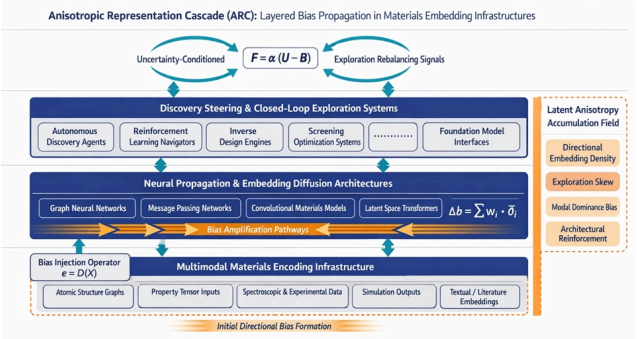

Figure 1. Conceptual Architecture of the Anisotropic Representation Cascade (ARC) Framework in AI-Driven Materials Discovery

Conceptual architecture of the *Anisotropic Representation Cascade (ARC)* framework illustrating how directional biases emerge and propagate across layered materials embedding infrastructures. The Encoding Layer introduces modality-dependent anisotropies during vector construction, which diffuse through neural propagation architectures where structural inductive biases amplify directional distortions. The Steering Layer integrates these embeddings into closed-loop discovery systems, where reinforcement and optimization logics steer exploration trajectories. Feedback loops modulated by uncertainty signals counterbalance accumulated bias, forming a dynamic equilibrium governing anisotropic navigation of materials design spaces.

Analytical implications

Interpretive dynamics of bias propagation

The ARC framework illuminates interpretive dynamics within materials embedding spaces, where latent anisotropy manifests as directional flows that shape inference pathways. In data ingestion stages, biases introduced during encoding can cascade, favoring representations aligned with prevalent data modalities—such as crystalline over amorphous structures—thereby influencing the breadth of explorable chemical spaces [8, 9, 16]. This propagation is particularly evident in graph-based architectures, where relational aggregations amplify directional preferences, potentially constraining model adaptability to anisotropic material behaviors like directional conductivity or mechanical asymmetry [9-11]. ARC's layered structure allows for analyzing these as systemic interactions, revealing how upstream encoding distortions modulate downstream predictions without invoking empirical validations.

Computationally, this implies trade-offs in pipeline efficiency: deeper propagation layers enhance feature abstraction but risk bias entrenchment, necessitating steering mechanisms that dynamically recalibrate directional vectors. Such implications extend to uncertainty integration, where epistemic measures act as counterbalances, fostering interpretive resilience in multimodal fusions [7, 16, 25]. By viewing anisotropy as an emergent property of cascade interactions, ARC guides infrastructure designs toward modular architectures that permit bias auditing at layer interfaces.

Trade-offs in discovery steering logics

A key analytical lens provided by ARC concerns trade-offs inherent to discovery steering logics in closed-loop systems. Feedback loops, as modeled in the framework, interact with accumulated biases to steer exploration, often prioritizing high-confidence directions at the expense of novel anisotropic regimes [14, 17, 22]. This can be conceptualized as a steering trade-off function, where the directional preference P is balanced against exploration entropy E, expressed as $T = \beta (P \cdot E - 1)$, with β denoting a logic-specific coefficient that reflects system priorities. Here, T captures the interaction between bias-driven focus and uncertainty-driven diversification, highlighting how over-reliance on biased embeddings may narrow discovery funnels in inverse design tasks [12, 24].

In high-throughput contexts, these trade-offs manifest in workflow dynamics, where simulation-experiment couplings must navigate biased spaces to align computational directives with empirical constraints [4, 6, 23]. ARC interprets this as an optimization landscape where steering logics mitigate cascade effects, promoting equitable navigation across material hierarchies. For foundation models, implications involve pretraining strategies that incorporate directional regularizations, ensuring broader applicability without amplifying inherited anisotropies **[19, 20]**.

Epistemic risk structures in computational ecosystems

Analytically, ARC elucidates epistemic risk structures arising from latent anisotropy, framing them as interconnected vulnerabilities in materials AI ecosystems [3, 21]. Directional biases in embeddings can introduce risks of misrepresentation, where inference paths overlook subtle anisotropies critical for applications like energy storage or

catalysis [5, 13, 14]. This risk amplification may be expressed as $R = \gamma \int B(d)\, dd$, integrating bias density $B(d)$over embedding dimensions d d d, with ɣ scaling for ecosystem complexity. Such a formulation interprets risks as accumulative, underscoring the need for interpretive safeguards in autonomous systems [1, 2, 17].

Infrastructure-level implications emphasize hybrid designs that couple bias-aware embeddings with adaptive feedbacks, reducing epistemic gaps in data-driven paradigms [7, 23, 25]. By dissecting these structures, ARC offers insights into fostering robust discovery, where risk mitigation enhances the interpretive fidelity of computational steering across diverse materials informatics workflows.

Results and Discussion

Linking the ARC framework to broader computational paradigms

The Anisotropic Representation Cascade (ARC) framework can be situated within the broader epistemic evolution of computational materials engineering as an interpretive superstructure layered atop existing representation and inference paradigms. While contemporary scholarship has extensively examined architectural innovation—particularly the performance scaling of graph neural networks, message-passing systems, and transformer-derived materials encoders—the interpretive consequences of directional bias formation within embedding infrastructures have remained comparatively under-theorized. ARC addresses this omission by reframing representation learning not merely as a feature extraction exercise but as a directional system whose latent geometries actively shape discovery trajectories.

In traditional embedding paradigms, latent spaces are often treated as isotropic continua where distance metrics faithfully encode materials similarity. ARC challenges this assumption by asserting that embedding manifolds are structurally deformed through cascading interactions between data modalities, inductive priors, and optimization logics. Within graph neural operations, for instance, message propagation schemes privilege relational proximities defined by bonding topologies or spatial adjacency. While such operations enhance predictive fidelity, they also impose anisotropic weighting across embedding dimensions, amplifying sensitivities to certain structural motifs while attenuating others [9-11]. ARC reframes these architectural biases as directional propagation currents within latent space, transforming static representations into dynamic anisotropic fields.

This interpretive repositioning aligns naturally with the rise of explainable artificial intelligence in materials science. Attribution techniques—such as saliency mapping, feature occlusion, and graph attention visualization—have been deployed to interrogate prediction rationales [7, 15]. However, ARC extends explainability from local feature importance to systemic bias tracing. Rather than asking which features influence a prediction, the framework asks how directional embedding currents guide entire discovery pathways. This shift from interpretability to directional epistemology broadens transparency discourse, enabling infrastructural auditing of embedding geometries themselves.

Generative modeling paradigms further illustrate ARC's integrative reach. Adversarial networks, variational autoencoders, and diffusion architectures sample latent manifolds to propose candidate materials [12, 24]. These sampling processes implicitly assume that learned distributions proportionally reflect the breadth of chemical space. ARC complicates this assumption by demonstrating that anisotropic density clusters—formed through cascade dynamics—act as gravitational attractors in generative sampling. Consequently, inverse design systems may repeatedly explore composition regimes aligned with embedding density gradients rather than true novelty frontiers. Interpreting bias as a directional constraint enables recalibration of sampling logics, introducing entropy-weighted or anisotropy-penalized exploration schemes.

Connections to multi-fidelity learning reinforce the framework's systemic relevance. Disordered systems, amorphous phases, and metastable compounds challenge representation continuity due to sparse or noisy data [8, 16]. ARC posits that anisotropy intensifies under such fidelity discontinuities, as embedding encoders disproportionately weight high-confidence crystalline datasets. Propagation layers then diffuse this imbalance, embedding structural order as a directional prior across downstream tasks. By foregrounding cascade amplification, ARC provides a conceptual apparatus for diagnosing fidelity-induced representation distortions and designing propagation filters that preserve anisotropic diversity.

Through these linkages, ARC advances a holistic synthesis that bridges representation learning, generative design, explainable AI, and multi-fidelity modeling into a unified directional systems perspective. Rather than treating these paradigms as modular innovations, the framework reveals their interdependence within cascading embedding ecologies that collectively steer computational discovery [3, 19, 22].

Addressing infrastructure challenges in data-driven ecosystems

Beyond theoretical alignment, ARC offers a critical infrastructural lens through which to interrogate the operational architectures of data-driven materials ecosystems. Contemporary discovery infrastructures increasingly operate as closed-loop systems integrating high-throughput simulation, machine learning inference, and experimental validation [4, 14, 17]. While such coupling accelerates iteration cycles, it also creates recursive reinforcement environments where embedding biases are repeatedly re-ingested, retrained, and amplified.

ARC's feedback equilibrium formulation provides a conceptual diagnostic for this phenomenon. In autonomous discovery settings, reinforcement learning agents optimize navigation strategies using reward functions derived from embedding predictions. If those embeddings encode anisotropic distortions, steering policies inherit directional entrenchments, privileging already-dense exploration corridors. Over successive loops, this produces epistemic path dependence—an infrastructural inertia that constrains exploratory divergence despite algorithmic adaptability.

Uncertainty-modulated steering emerges within ARC as a countervailing infrastructural logic. By embedding epistemic uncertainty into feedback channels, systems can dynamically redirect exploration toward underrepresented embedding vectors. This resonates with reinforcement learning strategies incorporating curiosity-driven rewards or entropy maximization, yet ARC reframes such mechanisms as anisotropy mitigation tools rather than purely efficiency enhancers [14, 17].

High-throughput simulation–experiment couplings further illuminate infrastructural vulnerabilities. Validation pipelines often prioritize candidates predicted with high model confidence, inadvertently reinforcing embedding density biases [6, 23]. ARC interprets this as validation anisotropy, wherein empirical confirmation disproportionately accumulates along dominant directional axes. Over time, this skews dataset composition itself, recursively shaping future encodings. Addressing this requires infrastructural diversification protocols that allocate experimental bandwidth to anisotropically sparse regions.

Foundation models introduce an additional infrastructural scale. Pretraining on multimodal scientific corpora enables cross-domain generalization but simultaneously embeds corpus-derived directional imbalances [20]. ARC's cascade logic maps these imbalances onto encoding-layer anisotropies that propagate through fine-tuning and task adaptation. Directional diversification in corpus design—through balanced materials classes, synthesis conditions, and property regimes—therefore becomes not merely a data curation concern but an infrastructural necessity for equitable embedding geometries.

Collectively, these analyses position ARC as an infrastructural critique framework capable of diagnosing bias entrenchment across autonomous labs, high-throughput platforms, and foundation model ecosystems. By exposing cascade reinforcement loops, it informs governance strategies balancing scalability, reliability, and epistemic inclusivity [5, 13, 26].

Interpretive extensions and conceptual limitations

While ARC establishes a robust interpretive scaffold, its conceptual architecture invites extension into emergent computational frontiers. Quantum-informed embeddings, for instance, represent a nascent representational regime where wavefunction descriptors, electron density fields, and entanglement metrics inform latent vectorization [18]. Directional anisotropy in such spaces may arise not from graph topology but from probabilistic amplitude distributions, introducing fundamentally different cascade dynamics. Extending ARC to quantum embeddings would require reconceptualizing propagation operators beyond message passing toward coherence-sensitive diffusion models.

Similarly, the rise of multimodal foundation models integrating text, imagery, spectroscopy, and simulation outputs introduces cross-modal anisotropy vectors whose cascade behaviors remain unexplored. ARC provides a preliminary grammar for such analyses but would benefit from formalization into tensorial anisotropy metrics capable

of quantifying directional bias across heterogeneous latent substrates.

The framework's non-empirical stance constitutes both a strength and a limitation. By abstaining from quantitative benchmarking, ARC preserves theoretical generality, enabling applicability across architectures and discovery infrastructures. However, operationalizing its constructs—such as bias density fields or cascade amplification coefficients—will require methodological translation into measurable observables. Future work may develop anisotropy indices, embedding curvature diagnostics, or directional entropy metrics to empirically instantiate ARC's interpretive propositions.

Despite this, the conceptual orientation aligns with foundational theoretical syntheses in computational science that precede measurement formalization [3, 21, 25].

Conclusion

The Anisotropic Representation Cascade (ARC) framework advances a systems-level conceptualization of latent anisotropy within materials embedding spaces, positioning directional bias not as an incidental modeling imperfection but as a cascading infrastructural phenomenon spanning encoding, propagation, and discovery steering layers. Through this layered architecture, ARC elucidates how multimodal data asymmetries, neural inductive priors, and closed-loop optimization logics interact to generate directional embedding currents that shape exploration trajectories across computational materials ecosystems.

By synthesizing insights from representation learning, graph architectures, generative modeling, autonomous discovery, and uncertainty quantification, the framework establishes an interpretive bridge linking micro-level embedding distortions to macro-level discovery outcomes. Its feedback equilibrium formulation highlights the role of uncertainty as a regulatory force capable of counterbalancing anisotropy amplification, offering conceptual guidance for designing resilient, bias-aware infrastructures.

Importantly, ARC reframes computational steering as a directional governance problem. Discovery systems do not merely search chemical space—they are guided through it by anisotropic embedding geometries that privilege certain materials classes, property regimes, and structural motifs. Recognizing and auditing these geometries becomes essential for ensuring equitable exploration, particularly in inverse design contexts where underrepresented regions may harbor transformative materials innovations.

Looking forward, the framework opens avenues for interpretive tool development, including anisotropy visualization maps, cascade auditing dashboards, and bias-regularized embedding protocols. Extensions into quantum embeddings, multimodal foundation models, and autonomous laboratory ecosystems further underscore its scalability as a theoretical lens.

Ultimately, ARC underscores that accelerating materials innovation requires more than computational power and data abundance—it demands directional awareness. By illuminating how biases form, propagate, and steer discovery infrastructures, the framework contributes to building epistemically robust, exploration-balanced AI ecosystems capable of navigating the full diversity of chemical and structural possibility spaces in the era of data-centric materials engineering.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 29 Sep 2023 Revised: 21 Oct 2023 Accepted: 03 Jan 2024
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.

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.

Morgan D, Jacobs R. Opportunities and challenges for machine learning in materials science. Ann Rev Mater Res. 2020;50:71-103.

Schmidt J, Shi R, Berri S, Chen L, Botti S, Marques MAL. Predicting the thermodynamic stability of solids with machine learning based on the API automatic FLOW for materials discovery (AFLOW). Chem Mater. 2017;29(17):7354-64.

Wen C, Zhang Y, Wang C, Xue D, Bai Y, Antonov S, et al. Machine learning assisted design of high entropy alloys with desired property. Acta Mater. 2019;170:109-17.

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

Luo B, Li Y, Zhang S, Feng Y, Zhang D, Shi Z, et al. Explainable machine learning in materials science. npj Comput Mater. 2022;8(1):204.

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.

Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. Chem Mater. 2019;31(9):3564-72.

Fung V, Zhang J, Juarez E, Sumpter BG. Benchmarking graph neural networks for materials chemistry. npj Comput Mater. 2021 Jun 3;7(1):84.

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.

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

Yang C, Zhang Y, Wen C, Yin K, Zhao Y, Su Y. Machine learning-based alloy design system for facilitating the rational design of high entropy alloys with enhanced hardness. Acta Mater. 2022;222:117431.

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.

Zhang H, Zhao Y, Wang C, Su Y. Interpretable machine-learning strategy for soft-magnetic property and thermal stability in Fe-based metallic glasses. Acta Mater. 2020;200:803-10.

Chen C, Zuo Y, Ye W, Li X, Deng Z, Ong SP. A critical review of machine learning of energy materials. Adv Energy Mater. 2020;10(8):1903242.

de Witt CS, Peng B, Kamienny PA, Torr P, Böhmer W, Whiteson S. Deep multi-agent reinforcement learning for decentralized continuous cooperative control. arXiv preprint arXiv:2003.06709. 2020;19.

Glielmo A, Zeni C, De Vita A. Efficient nonparametric n-body force fields from machine learning. Phys RevB. 2018;97(18):184307.

Court CJ, Yildirim B, Jain SS, Cole JM. Paradigm shift without the shift: Interpreting and advancing deep learning for materials. Matter. 2022;5(5):1340-2.

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

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

Kitchin JR. Machine learning in molecular and materials science and chemistry. Nat Comput Sci. 2021;1(5):333-4.

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.

Bruix A, Margraf JT, Andersen M, Reuter K. First-principles-based multiscale modelling of heterogeneous catalysis. Nat Catal. 2019;2(8):659-70.

Lu Z, Lu S, Zhao Y, Wang C, Su Y. Interpretable machine-learning strategy for failure prediction in ceramic matrix composites. Acta Mater. 2021;217:117177.

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.