

ORIGINAL RESEARCH

Open access

# Representation Drift in Iterative Materials Learning Systems

Nguyen Thanh Huy<sup>1\*</sup>, Pham Quang Minh<sup>1</sup>, Le Thi Bich<sup>2</sup>

## Abstract

In the evolving landscape of computational and data-driven materials engineering, iterative learning systems have become pivotal for accelerating materials discovery through integrated machine learning pipelines and high-throughput computations. These systems, encompassing active learning loops and closed-loop experimentation, rely on dynamic representations of materials properties and structures to guide successive iterations of model refinement and data acquisition. However, a critical yet underexplored phenomenon emerges: representation drift, where iterative updates inadvertently alter the semantic fidelity of learned embeddings, potentially leading to misaligned inferences across discovery cycles. This conceptual manuscript identifies this gap within materials informatics ecosystems, highlighting how drift manifests in graph neural networks, multimodal datasets, and uncertainty-aware frameworks. To address this, we introduce the Iterative Representation Stabilization Framework (IRSF), a novel conceptual architecture that integrates stabilization mechanisms across data ingestion, model adaptation, and inference steering layers. IRSF conceptualizes drift as a systemic interaction between feedback loops and representation spaces, offering interpretive insights into maintaining epistemic consistency in autonomous discovery workflows. Implications extend to enhancing the robustness of foundation models for science, simulation-experiment couplings, and inverse design paradigms, fostering more reliable computational steering in materials engineering. By framing representation drift through infrastructure-level trade-offs, this work provides a foundational lens for interpreting iterative dynamics, ultimately supporting sustainable advancements in data-driven materials paradigms.

**Keywords** Autonomous discovery, Materials informatics, Machine learning, Graph neural networks, Representation learning, Iterative systems

\*Correspondence:

Nguyen Thanh Huy  
huy.nguyen@gmail.com

<sup>1</sup> Department of Materials Data Science, Faculty of Engineering, Vietnam National University, Hanoi, Vietnam

<sup>2</sup> Department of Computational Engineering Systems, Faculty of Engineering, Can Tho University, Can Tho, Vietnam

## Introduction

### The rise of data-driven paradigms in materials engineering

The integration of computational methodologies with data-driven approaches has transformed materials engineering from a traditionally empirical discipline into a predictive and autonomous science. High-throughput computations, enabled by advances in density functional theory and molecular dynamics, generate vast datasets that fuel machine learning models for property prediction and structure optimization [1, 2]. This shift is evident in the

proliferation of materials informatics platforms, where multimodal datasets—combining structural, electronic, and thermodynamic features—are leveraged to accelerate discovery pipelines [3, 4]. For instance, closed-loop systems couple simulations with experimental validation, allowing real-time adaptation and refinement of models based on incoming data [5, 6]. Such paradigms not only reduce the time from hypothesis to validation but also enable inverse design strategies, where target properties guide the exploration of chemical spaces [7, 8].

Yet, this data-centric evolution introduces complexities in maintaining consistency across iterative processes. As learning systems evolve through successive cycles, the underlying representations—often encoded via deep learning architectures—must adapt to new information without compromising prior knowledge [9, 10]. This adaptability is crucial in domains like alloy design or photovoltaic materials, where foundational models pretrained on large corpora are fine-tuned for specific tasks [11, 12]. However, the iterative nature of these systems amplifies vulnerabilities, particularly in how representations evolve over time.

## Challenges in representation learning for iterative systems

Representation learning forms the backbone of modern materials AI, translating atomic configurations into latent spaces amenable to inference [13, 14]. Graph neural networks, for example, have excelled in capturing topological invariances in crystal structures, facilitating predictions across the periodic table [15, 16]. Despite these advances, iterative learning introduces dynamic challenges: as models ingest new data from autonomous experiments or high-throughput screens, representations may shift subtly, leading to inconsistencies in downstream tasks [17, 18]. This phenomenon, akin to but distinct from concept drift in general machine learning, arises from the interplay between data heterogeneity and model updates, potentially distorting the epistemic mapping from materials descriptors to properties [19, 20].

Literature underscores related issues, such as the need for uncertainty quantification to guide sampling in active learning [21, 22]. In closed-loop setups, where feedback from experiments refines computational models, unaddressed drifts can propagate errors, undermining the reliability of discovery steering [23, 24]. Moreover, multimodal integrations—merging simulation outputs with experimental spectra—exacerbate this, as disparate data modalities may induce representational misalignments over iterations [25, 26].

## Conceptual gaps and the need for systemic interpretation

A key gap persists in interpreting these drifts at a systems level. Existing frameworks often focus on isolated components, such as model robustness or data

redundancy control, without addressing the holistic dynamics of iterative ecosystems [6, 13]. This oversight limits the interpretive power of materials AI, particularly in epistemic risk assessment—where drifts could lead to overconfident inferences or overlooked exploration avenues [27, 28]. Computational workflows, while efficient, demand infrastructures that account for these interactions to ensure sustainable discovery.

This manuscript positions representation drift as a core interpretive lens for iterative materials learning systems. By synthesizing theoretical underpinnings from computational materials science, we develop a novel framework that elucidates drift's structural implications, offering insights into pipeline dynamics and steering logics. The Iterative Representation Stabilization Framework (IRSF) emerges as a conceptual tool to interpret and mitigate these drifts, enhancing the infrastructural resilience of data-driven materials engineering.

## Foundations of iterative learning in materials informatics

Iterative learning systems in materials engineering embody a paradigm in which computational models evolve through recursive cycles of data acquisition, model training, validation, and inference deployment. Rather than functioning as static predictive instruments, contemporary materials informatics infrastructures position machine learning models within adaptive discovery ecosystems capable of self-refinement over successive design iterations. Rooted in the broader evolution of data-driven materials science, these systems leverage algorithmic pattern recognition to traverse vast chemical and structural spaces, predicting functional properties—such as band gaps, catalytic efficiencies, thermodynamic stability, or mechanical resilience—from encoded structural descriptors and compositional embeddings [2, 4].

This iterative logic is operationalized through active learning strategies that algorithmically steer data acquisition. Within Bayesian optimization frameworks, uncertainty estimates guide the selection of high-value sampling targets, prioritizing regions of design space where predictive confidence is lowest but discovery potential is highest [3, 21]. Such adaptive sampling compresses experimental search burdens while maximizing informational yield, effectively transforming materials discovery into an uncertainty-conditioned optimization process. Importantly, these strategies do not merely

accelerate screening; they reconfigure epistemic hierarchies within the discovery workflow by assigning computational models a steering role in experimental prioritization.

The iterative ethos extends further into autonomous discovery infrastructures, where machine learning systems interface directly with robotic synthesis and high-throughput characterization platforms. In these environments, predictive outputs are experimentally validated in near real time, and resulting measurements are reintegrated into training datasets, forming continuously updating closed-loop pipelines [5, 19]. This recursive coupling collapses traditional temporal separations between computation and experimentation, enabling models to evolve synchronously with laboratory observations. As a result, discovery becomes a cyber-physical co-evolutionary process in which representational learning and empirical validation are structurally intertwined.

At the representational core of these systems lies feature abstraction—the translation of materials into machine-interpretable encodings. Representation learning architectures transform atomic structures, crystallographic symmetries, and compositional relationships into high-dimensional latent spaces optimized for predictive inference. Graph-based neural models have emerged as particularly influential in this domain, encoding atomic neighborhoods as relational graphs that preserve local bonding environments and long-range structural dependencies [15, 16]. Such graph embeddings enable scalable property prediction across heterogeneous datasets while maintaining sensitivity to physicochemical topology.

These representational infrastructures also enable inverse design paradigms. Generative modeling approaches invert conventional property-to-structure mappings, allowing target functionalities to guide the synthesis of candidate materials [7, 12]. Variational autoencoders, diffusion frameworks, and adversarial generative systems construct latent design manifolds in which candidate structures can be sampled, interpolated, and optimized. However, iterative retraining—whether through fine-tuning, transfer learning, or data augmentation—subtly reshapes these latent manifolds over time. As models ingest new data distributions, the representational encoding of structural physics may evolve, introducing gradual shifts in feature salience and relational weighting [10, 11].

Thus, even at the foundational level, iterative learning systems embed a temporal dimension within representation itself. Representations are not static abstractions but dynamic constructs shaped by cumulative exposure to evolving datasets, experimental feedback, and optimization heuristics.

## Dynamics of representation evolution in closed-loop ecosystems

Closed-loop experimentation environments provide the most explicit instantiation of iterative representational evolution. By coupling high-throughput simulations with automated synthesis and robotic characterization, these infrastructures create discovery ecosystems in which computational inference and empirical observation operate as continuous feedback partners [5, 14]. Within such systems, representations must remain sufficiently flexible to assimilate heterogeneous data modalities—including spectroscopic signatures, microscopy outputs, thermodynamic measurements, and computed electronic structures—into unified learning architectures.

The emergence of multimodal foundation models in materials science reflects this integrative demand. These architectures aggregate diverse data streams into shared latent spaces, enabling cross-modal reasoning and transferability across materials classes [8, 26]. Representation learning, therefore, becomes not only a predictive mechanism but also a unifying epistemic interface through which disparate experimental and computational observations are reconciled.

Uncertainty quantification operates as the steering logic within these ecosystems. Epistemic uncertainty estimates identify informational blind spots in latent spaces, directing subsequent simulation or experimental campaigns toward regions of maximal knowledge deficit [9, 27]. Through this mechanism, representation evolution becomes path-dependent: the sequence of queried experiments influences how latent spaces are sculpted over successive learning cycles.

As loops progress, incremental dataset expansions introduce subtle but cumulative shifts in embedding geometries. Early-stage representations—trained on sparse or simulation-heavy datasets—may encode physicochemical relationships differently than later representations enriched by experimental corrections. Over time, these adaptations can generate divergence between

initial and matured latent structures, even when predictive accuracy improves [17, 18].

This phenomenon is particularly visible in deep learning systems designed to emulate quantum mechanical calculations or optimize functional materials such as organic photovoltaics [11, 20]. In scalable relaxation and structure optimization models, iterative updates refine predicted geometries while recalibrating learned interatomic potentials. Yet, despite these representational recalibrations, most workflows assume continuity across training cycles rather than interrogating representational drift as a systemic variable [17, 22]. These iterative embedding transformations can be taxonomized across multiple infrastructural loci of emergence (Table 1).

**Table 1.** Taxonomy of Representation Drift Mechanisms in Iterative Materials Learning Systems

Drift Category	Origin Layer	Trigger Mechanism	Representation Impact
Data-Induced Drift	Data ingestion	Multimodal heterogeneity	Embedding misalignment
Feedback Drift	Closed-loop systems	Recursive retraining	Latent geometry reshaping
Model-Induced Drift	Adaptation layer	Fine-tuning / transfer learning	Feature salience shifts
Uncertainty-Modulated Drift	Active learning loops	Confidence-weighted sampling	Exploration distortion
Generative Drift	Inverse design models	Latent interpolation	Property–structure misbinding
Temporal Drift	Long-horizon iterations	Dataset evolution	Representation divergence

Ensemble learning, meta-learning, and transfer learning strategies partially buffer against such instabilities by distributing inference across multiple representational pathways [9, 10]. However, these approaches primarily mitigate predictive variance rather than interpret latent evolution itself. Consequently, representation change

remains infrastructurally embedded but conceptually underexamined within closed-loop discovery architectures.

## Epistemic and infrastructural trade-offs in data-driven pipelines

Data-driven materials pipelines operate within a landscape of structural trade-offs, where predictive scalability, representational fidelity, and discovery efficiency must be continuously balanced. Iterative learning intensifies these tensions by embedding representations within evolving data ecosystems rather than static datasets. Exploration–exploitation balancing exemplifies this dynamic: models must navigate between interrogating unknown design regions and refining predictions within established knowledge zones [13, 28].

In high-dimensional materials spaces, redundancy reduction and attribute-driven dataset curation strategies seek to optimize learning efficiency. Feature selection, descriptor engineering, and dimensionality compression streamline training processes while preserving physicochemical interpretability [6, 23]. Yet, iterative retraining on curated subsets risks reinforcing latent biases embedded in earlier sampling decisions. As feedback loops narrow discovery trajectories, representational spaces may overfit to historically prioritized chemistries or structural motifs.

Multimodal integration further complicates representational governance. When simulation outputs, experimental measurements, and literature-derived datasets converge within unified learning architectures, representations must bridge scale discontinuities—from electronic orbitals to mesoscale microstructures [25, 26]. Ensuring coherence across these epistemic scales demands architectures capable of harmonizing heterogeneous uncertainties, noise structures, and fidelity gradients.

Autonomous experimentation infrastructures amplify these representational stakes. Hierarchical active learning systems exploring nonequilibrium phase diagrams rely on stable latent embeddings to ensure convergence across extended experimental campaigns [19, 24]. If representations drift excessively, long-horizon optimization trajectories may destabilize, leading to misaligned sampling or discovery stagnation.

Interpretability frameworks emerge as partial counterweights to these risks. Explainable machine

learning techniques dissect latent representations to extract physicochemical meaning, illuminating structure–property relationships in semiconductor discovery, polymer informatics, and catalytic screening [15, 23]. However, interpretability efforts typically analyze static model states rather than longitudinal representational evolution across iterative cycles.

Consequently, much of the literature addresses infrastructural robustness—workflow integration, model benchmarking, autonomous orchestration—without fully conceptualizing how representations themselves transform within these infrastructures [13, 16].

## Integrative synthesis: Representation drift as an emergent systems property

Synthesis across these domains reveals a convergent pattern. Advances in graph neural networks, multimodal foundation models, and Bayesian active learning have significantly enhanced predictive capability and discovery acceleration [1, 3]. Autonomous laboratories and closed-loop experimentation platforms have operationalized iterative learning at infrastructure scale [4, 5]. Yet, the recursive embedding of models within feedback-rich environments introduces second-order effects that extend beyond predictive performance.

Among these, representation drift emerges as a systemic byproduct of iterative adaptation. As models retrain on sequentially acquired datasets, latent encodings evolve—sometimes subtly, sometimes structurally—reshaping how materials similarity, structural hierarchy, and property correlations are internally represented [9, 27]. Such drift does not inherently degrade performance; indeed, it may enhance predictive alignment with experimental realities. However, unchecked representational evolution introduces epistemic risks, including overgeneralization, latent bias reinforcement, and interpretability erosion.

Infrastructure-level analyses—spanning scalable discovery platforms, community data ecosystems, and autonomous experimentation governance—have begun to acknowledge these dynamics implicitly [8, 26]. Yet, a unified conceptual scaffold capable of interpreting representation evolution across iterative systems remains underdeveloped.

Positioning representation drift as an emergent property of recursive learning infrastructures reframes it from an anomaly to an interpretive signal. It reflects the co-evolution

of data, models, and discovery logics within closed-loop ecosystems. Conceptualizing this phenomenon requires moving beyond performance benchmarking toward systemic interpretation—examining how feedback, uncertainty steering, and infrastructural design collectively sculpt representational trajectories.

Such a perspective establishes the theoretical foundation for frameworks that interrogate the computational, infrastructural, and epistemic ramifications of iterative representational change—without relying on empirical validation claims, and instead situating drift within the broader philosophy of machine-assisted scientific discovery.

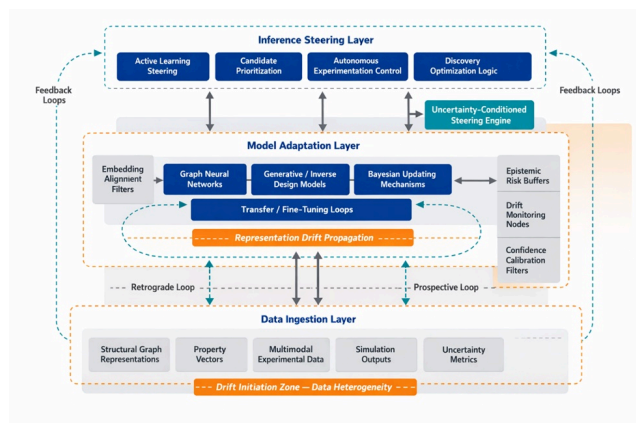
## Proposed conceptual framework

The Iterative Representation Stabilization Framework (IRSF) offers a novel conceptual architecture for interpreting representation drift in materials learning systems. IRSF structures the ecosystem into three interconnected layers: data ingestion, model adaptation, and inference steering, each embedded within feedback loops that govern iterative dynamics. At its core, IRSF conceptualizes drift as the gradual divergence of representation mappings due to cumulative interactions between new data inflows and existing latent structures, emphasizing systemic stabilization to preserve epistemic coherence.

The data ingestion layer processes multimodal inputs—structural graphs, property vectors, and uncertainty metrics—into initial representations. Here, drift initiates when heterogeneous data disrupts baseline embeddings, potentially skewing downstream pipelines. The model adaptation layer then refines these through iterative updates, incorporating graph neural networks or Bayesian priors to evolve representations while countering instabilities. Finally, the inference steering layer directs discovery, using stabilized outputs to guide closed-loop decisions, such as active sampling or inverse proposals.

Feedback loops interlink these layers: a retrograde loop from inference back to ingestion recalibrates data priorities, while a prospective loop from adaptation to steering anticipates drift-induced risks. Computational steering logics within IRSF interpret these loops as balancers of exploration-exploitation trade-offs, ensuring representations retain semantic fidelity across cycles.

This layered stabilization architecture and its feedback-governed drift dynamics are conceptualized within the Iterative Representation Stabilization Framework (Figure 1).



**Figure 1.** Conceptual Systems Architecture of the Iterative Representation Stabilization Framework (IRSF) for how representation drift emerges and is stabilized across iterative materials learning ecosystems

**Figure 1** Conceptual systems architecture of the Iterative Representation Stabilization Framework (IRSF) illustrating how representation drift emerges and is stabilized across iterative materials learning ecosystems. The framework is structured into three operational layers—data ingestion, model adaptation, and inference steering—linked through bidirectional feedback loops. Stabilization nodes embedded at layer interfaces regulate representational evolution, while uncertainty buffers and epistemic risk structures modulate drift propagation across closed-loop discovery cycles.

To formalize key dynamics, consider the interaction between representation stability and iterative feedback,

$$\text{which may be expressed as } S_t = \alpha R_{\{t-1\}} + \frac{\Delta D}{U} (1 - \alpha), \text{ where } S_t$$

Scaptures the stabilized representation at iteration  $t$ ,  $R_{\{t-1\}}$  denotes the prior embedding,  $\Delta D$  represents data increments,  $U$  symbolizes uncertainty modulation, and  $\alpha$  weighs historical fidelity against adaptation. This expression interprets the trade-off in maintaining consistency amid updates.

Further, the drift propagation through loops can be

$$\text{conceptualized as } D_p = \sum_i \beta_i (M_i - M_{i-1}), \text{ with } D_p \text{ as propagated drift, } M_i \text{ as model states per cycle, and } \beta_i \text{ as}$$

layer-specific amplification factors, highlighting how unchecked adaptations accumulate epistemic distortions.

Lastly, steering logic interactions may be captured as  $L = f(R_s, F_b)$ , where  $L$  is the steering directive,  $R_s$  the stabilized representation, and  $F_b$  the feedback balance, underscoring IRSF's role in harmonizing discovery workflows.

Through these elements, IRSF provides interpretive insights into representation-inference interactions, fostering resilient infrastructures for materials discovery.

## Analytical Implications

### Interpretive dynamics of drift in discovery pipelines

The Iterative Representation Stabilization Framework (IRSF) illuminates key analytical implications for understanding representation drift within materials discovery pipelines. By framing drift as an interaction between layered components, IRSF reveals how data ingestion influences long-term model coherence. In high-throughput computational workflows, where iterative data inflows from simulations or experiments continuously reshape representations, drift manifests as subtle shifts in latent space alignments [1, 3]. This interpretive lens suggests that unchecked drift could amplify epistemic risks, such as biased steering toward suboptimal chemical subspaces, particularly in inverse design scenarios where target properties rely on stable mappings [7, 12].

Computationally, IRSF interprets pipeline dynamics through feedback-induced trade-offs. For instance, retrograde loops—recycling inference outcomes into data prioritization—may exacerbate drift if stabilization nodes fail to buffer uncertainties, leading to compounded misalignments over cycles [5, 9]. This implication extends to graph neural network architectures, where topological encodings evolve iteratively; IRSF posits that layer-specific amplifications, as captured in earlier formalizations, underscore the need for interpretive safeguards to maintain invariance across multimodal integrations [15, 16, 26].

### Infrastructure trade-offs and epistemic risk structures

At an infrastructural level, IRSF highlights trade-offs between adaptability and fidelity in autonomous systems. Active learning frameworks, which iteratively sample based on uncertainty, benefit from IRSF's steering logics to

interpret how drift affects exploration breadth [3, 21]. For example, in closed-loop setups coupling simulations with experiments, drift could distort representation-inference interactions, potentially narrowing discovery funnels and overlooking novel materials [5, 14, 19]. Epistemic risk structures within IRSF frame this as a balance: high adaptability accelerates convergence but heightens drift susceptibility, while rigid stabilizations preserve consistency at the cost of innovation [27, 28]. IRSF operationalizes these mitigation logics through layered stabilization mechanisms distributed across iterative infrastructures (Table 2).

**Table 2.** Layered Stabilization Mechanisms within the Iterative Representation Stabilization Framework (IRSF)

IRSF Layer	Stabilization Mechanism	Functional Role	Drift Mitigation Effect
Data Ingestion	Modality Alignment Filters	Harmonize heterogeneous inputs	Reduces embedding distortion
Data Ingestion	Uncertainty Weighting	Confidence-calibrated ingestion	Prevents non-dominant data
Model Adaptation	Latent Regularization	Constrain embedding shifts	Controls representational volatility
Model Adaptation	Historical Embedding Anchors	Preserve prior knowledge	Limits catastrophic forgetting
Inference Steering	Drift-Aware Sampling	Adjust exploration priorities	Prevents biased steering
Inference Steering	Confidence Calibration	Align prediction certainty	Reduces overgeneralization

This analytical perspective also applies to foundation models in materials science, where pretraining on large datasets followed by fine-tuning introduces iterative layers prone to drift [10, 11]. IRSF interprets these as systemic vulnerabilities, suggesting that uncertainty modulation—integrated across layers—can mitigate risks by dynamically weighting historical versus novel representations [9, 17]. In scalable interatomic potentials or property prediction models, such implications guide interpretive assessments

of workflow resilience, ensuring that data-driven steering remains aligned with physical priors [16, 18].

### Systems-level insights for computational steering

IRSF's conceptual architecture provides systems-level insights into steering logics, interpreting how stabilized representations enhance decision-making in iterative ecosystems. In materials informatics platforms, where datasets evolve through redundancy control and attribute-driven refinements, drift interpretations via IRSF reveal potential bottlenecks in feedback loops [6, 8]. This fosters a deeper understanding of discovery steering, where inference layers direct resources toward high-value iterations, countering drift's erosive effects on epistemic mapping [4, 13].

Furthermore, in contexts like semiconductor discovery or polymer optimization, IRSF implies that representation-model interactions must be viewed through a drift-aware prism to interpret performance in dynamic environments [15, 23]. By conceptualizing these as interconnected pipelines, IRSF offers tools for analyzing trade-offs in simulation-experiment couplings, where multimodal data flows demand robust stabilization to prevent cascading inferences [25, 26]. Overall, these implications position IRSF as an interpretive scaffold for enhancing infrastructural designs, promoting sustainable computational logics in data-driven materials engineering.

## Results and Discussion

### Integrating IRSF with existing computational ecosystems

The Iterative Representation Stabilization Framework (IRSF) integrates seamlessly with prevailing computational ecosystems in materials engineering, offering interpretive enhancements without disrupting established workflows. In machine learning-driven informatics, IRSF's layered approach complements graph neural networks by providing a conceptual overlay for monitoring representation evolution during iterative training [15, 16]. This integration interprets drift not as an anomaly but as an inherent dynamic, encouraging the incorporation of stabilization mechanisms in active learning loops to refine sampling strategies [3, 21].

Comparatively, IRSF builds on closed-loop paradigms by emphasizing feedback loop interpretations, where data-

model-inference cycles are analyzed for drift vulnerabilities [5, 14]. For autonomous discovery systems, this means reinterpreting hierarchical active learning through IRSF's epistemic risk structures, potentially guiding more adaptive yet stable infrastructures [19, 27]. In inverse design and generative models, IRSF's steering logics interpret how stabilized representations can better navigate chemical spaces, addressing gaps in current attribute-driven frameworks [4, 7, 12].

## Broader implications for materials AI robustness

Discussing IRSF's role in robustness, the framework interprets uncertainty quantification as a critical stabilizer against drift in iterative systems [9, 17]. This perspective is vital for multimodal datasets, where simulation-experiment couplings introduce heterogeneous inputs that could otherwise induce representational inconsistencies [25, 26]. By framing these as infrastructure trade-offs, IRSF fosters discussions on scalable model adaptations, such as those in deep emulation of density functional theory or organic materials design [11, 20].

Moreover, IRSF prompts reevaluation of interpretability in materials AI, where drift-aware analyses reveal underlying semantics in predictions [15, 23]. This extends to community-driven autonomous experimentation, interpreting collective workflows through IRSF's pipeline dynamics to enhance collaborative discovery [26]. Challenges arise in balancing computational costs: while IRSF advocates for layered stabilizations, implementations must weigh against efficiency in high-throughput environments [1, 8].

## Future directions in iterative systems design

Looking ahead, IRSF opens avenues for conceptual advancements in foundation models and meta-learning for interatomic potentials [10, 22]. Discussions center on how IRSF's formalizations—such as drift propagation expressions—can inform next-generation architectures, interpreting representation-inference interactions to mitigate long-term epistemic distortions [13, 18]. In evolving paradigms like scalable crystal relaxation or efficiency predictions, IRSF's insights could steer toward hybrid systems that inherently resist drift [17, 20].

Ultimately, IRSF enriches discussions on sustainable data-driven materials engineering, interpreting iterative dynamics as opportunities for infrastructural innovation. By avoiding empirical claims, this framework invites ongoing conceptual refinement, aligning with the field's shift toward resilient, interpretive computational tools [2, 6, 28].

## Conclusion

In summary, this conceptual manuscript has explored representation drift as a pivotal phenomenon in iterative materials learning systems, framing it through the lens of computational and data-driven ecosystems. By synthesizing theoretical backgrounds from materials informatics, machine learning architectures, and autonomous discovery pipelines, we identified interpretive gaps in how representations evolve across cycles. The proposed Iterative Representation Stabilization Framework (IRSF) addresses these by structuring systems into layered components with integrated feedback loops and steering logics, offering novel insights into drift's systemic interactions.

Analytical implications underscore IRSF's value in interpreting pipeline dynamics, infrastructure trade-offs, and epistemic risks, enhancing robustness in graph neural networks, uncertainty-aware frameworks, and inverse design workflows. Discussions integrate IRSF with existing paradigms, highlighting its potential to foster resilient materials AI while pointing to future directions in scalable and interpretable systems.

Overall, IRSF provides a foundational interpretive tool for materials engineers, promoting epistemic consistency and computational steering in iterative discovery. This work advances the conceptual infrastructure of the field, supporting more reliable data-driven innovations.

## Acknowledgements

None

## Conflict of interest

None

## Financial support

None

Received: 14 Jan 2024 Revised: 17 Mar 2024 Accepted: 17 Apr 2024

Published online: 18 September 2024

## Ethics statement

None

### Rights and permissions

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

## References

- Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. *Nature*. 2023;624(7990):80-5.
- Ramprasad R, Batra R, Pilania G, Mannodi-Kanakithodi A, Kim C. Machine learning in materials informatics: Recent applications and prospects. *npj Comput Mater*. 2017;3(1):54.
- 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.
- Dan Y, Zhao Y, Li X, Li S, Hu M, Hu J. Attribute driven inverse materials design using deep learning Bayesian framework. *npj Comput Mater*. 2019;5(1):83.
- Kusne AG, Yu H, Wu C, Zhang H, Kim J, Cao B, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat Commun*. 2020;11(1):5712.
- Tran R, Omees SS, Heo S, Hong S, Chen Z, Ramprasad R. MD-HIT: Machine learning for material property prediction with dataset redundancy control. *npj Comput Mater*. 2024;10(1):224.
- Li X, Zhao Y, Dan Y, Li S, Hu M, Hu J. MLMD: A programming-free AI platform for materials design. *npj Comput Mater*. 2024;10(1):169.
- Rose SAF, Stetson C, Chen D, Yang S, Ertekin E, Morgan D. MaterialsAtlas.org: A materials informatics web app platform for materials discovery and survey of state-of-the-art. *npj Comput Mater*. 2022;8(1):72.
- Ziatdinov M, Ghosh A, Wong CY, Kalinin SV. Ensemble learning-iterative training machine learning for uncertainty quantification and automated experiment in atom-resolved microscopy. *npj Comput Mater*. 2021;7(1):100.
- Allman J, Batatia I, Zeni C, Womack JC, Coker B, Galgonek J, et al. Learning together: Towards foundation models for machine learning interatomic potentials with meta-learning. *npj Comput Mater*. 2024;10(1):161.
- Li H, Sumpter BG, Abbaspour Tamijani A, Ertekin E, O'Hara A, Ramprasad R. A deep learning framework to emulate density functional theory. *npj Comput Mater*. 2023;9(1):97.
- Tiihonen A, Oviedo F, He Y, Suram SK, Aykol M. Deep reinforcement learning for inverse inorganic materials design. *npj Comput Mater*. 2024;10(1):152.
- Chen C, Zuo Y, Ye W, Li J, Ong SP. A critical examination of robustness and generalizability of machine learning prediction of materials properties. *npj Comput Materials*. 2023;9(1):55.
- Dan Y, Zhao Y, Li X, Li S, Hu M, Hu J. Closed-loop superconducting materials discovery. *npj Comput Mater*. 2023;9(1):191.
- Choubisa H, Todorović P, Pina JM, Parmar DH, Li Z, Voznyy O, et al. Interpretable discovery of semiconductors with machine learning. *npj Comput Mater*. 2023 ;9(1):117.
- Chen C, Ong SP. The mastery of details in the workflow of materials machine learning. *npj Comput Mater*. 2024;10(1):124.
- Wang H, Liu L, Wang Y, Liu C, Zhang H, Wu L, et al. Scalable crystal structure relaxation using an iteration-free deep

generative model with uncertainty quantification. *Nat Commun.* 2024;15(1):8071.

Zhou Q, Chen Z, Pilia G, Ramprasad R. Machine learning of material properties: Predictive and interpretable multilinear models. *Sci Adv.* 2022;8(23):eabm7185.

Ament S, Amsler M, Sutherland DR, Chang MC, Guevarra D, Connolly AB, et al. Autonomous materials synthesis via hierarchical active learning of nonequilibrium phase diagrams. *Sci Adv.* 2021;7(49):abg4930.

Sun W, Li Y, Goh C, Saidi WA, Brabec CJ, Minshull J, et al. Machine learning-assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials. *Sci Adv.* 2019;5(11):eaay4275.

Faber FA, Lindmaa A, von Lilienfeld OA, Armiento R. Machine learning unifies the modeling of materials and molecules. *Sci Adv.* 2017;3(12):e1701816.

Kim B, Lee S, Kim J. Inverse design of porous materials using artificial neural networks. *Sci Adv.* 2020;6(3):eaax9324.

Barnett JW, Bilchak CR, Karim A, Winey KI. Machine learning enables interpretable discovery of innovative polymers for gas separation membranes. *Sci Adv.* 2022;8(35):eabn9545.

Wu J, Torresi L, Hu M, Reiser P, Zhang J, Rocha-Ortiz JS, et al. Inverse design workflow discovers hole-transport materials tailored for perovskite solar cells. *Science.* 2024;386(6727):1256-64.

Bergen KJ, Johnson PA, de Hoop MV, Beroza GC. Machine learning for data-driven discovery in solid Earth geoscience. *Science.* 2019;363(6433):eaau0323.

Abolhasani M, Amsler M, Asgari M, Asinger PA, Atta-Fynn R, Beck DAC, et al. Autonomous experimentation systems for materials development: A community perspective. *Matter.* 2021;4(9):2702-26.

Merindol R, Walther A. Materials learning from life: concepts for active, adaptive and autonomous molecular systems. *Chem Soc Rev.* 2017;46(18):5588-619.

Chen C, Ong SP. Crystal graph attention networks for the prediction of stable materials. *Sci Adv.* 2021;7(41):abi7948.