ORIGINAL RESEARCH **Open access**

Explainability Drift in Iterative Materials AI Workflows: A Conceptual Failure Analysis

Lucas Andrade[1*], Mariana Lopes[1]

Abstract

The integration of artificial intelligence into materials science has accelerated discovery through iterative workflows that cycle through data acquisition, model refinement, prediction, explanation, and hypothesis-driven experimentation. While explainable artificial intelligence (XAI) methods enhance trust and scientific insight by elucidating model decisions, these explanations are not static. This manuscript introduces the novel concept of explainability drift: the systematic degradation, inconsistency, or divergence in the fidelity, stability, and relevance of XAI-generated explanations across successive iterations of materials AI workflows. Distinct from prediction-focused concept drift, explainability drift arises from evolving data distributions, model updates, feature space expansions, and domain shifts inherent to materials exploration. Through a purely conceptual failure analysis, we delineate the mechanisms underlying explainability drift, including temporal instability in feature attributions, erosion of surrogate model alignment, and semantic misalignment between explanations and emerging material knowledge. Drawing on recent peer-reviewed advances in XAI applications to property prediction, microstructure analysis, and generative design, we synthesize theoretical foundations to highlight why drift undermines iterative efficacy. The proposed conceptual framework organizes explainability drift into multidimensional layers—attributional, structural, and epistemic—offering a structured lens for analyzing failure modes without empirical validation. This framework emphasizes risks such as misguided hypothesis generation, diminished trust in AI-assisted insights, and inefficient navigation of vast materials design spaces. By conceptualizing explainability drift as an intrinsic challenge, the work advocates for theoretical advancements in sustained explainability to support robust, interpretable AI-driven materials innovation.

Keywords Materials discovery, Concept drift, Explainability drift, Iterative AI workflows, Explainable artificial intelligence, Failure analysis

*Correspondence:
Lucas Andrade
lucas.andrade@gmail.com
[1] Department of Materials Data Science and Artificial Intelligence, Faculty of Engineering, Federal University of Minas Gerais, Belo Horizonte, Brazil

Introduction

Artificial intelligence (AI) has fundamentally reshaped materials science by enabling rapid exploration of complex and high-dimensional design spaces, spanning alloy optimization, functional material discovery, and structure–property inference through high-throughput computation and generative modeling. Contemporary materials AI increasingly operates through iterative workflows—including active learning, Bayesian optimization, autonomous experimentation, and closed-loop discovery platforms—in which predictive models are repeatedly retrained or fine-tuned as new data are acquired [1–4]. By tightly coupling simulation, experimentation, and machine learning in adaptive cycles, these paradigms promise to accelerate materials discovery beyond the limits of human-driven trial-and-error.

Despite these advances, the growing reliance on complex model architectures—such as deep neural networks, graph neural networks, and transformer-based representations—introduces substantial challenges for scientific interpretability.

In materials science, where models are expected not only to predict but also to explain underlying physical mechanisms, opacity poses a barrier to scientific trust, hypothesis formation, and experimental decision-making. Without reliable interpretative insight, iterative optimization risks devolving into computational search divorced from domain understanding.

Explainable artificial intelligence (XAI) has emerged as a response to this challenge by providing mechanisms that relate model predictions to physically meaningful features. Post-hoc attribution techniques (e.g., feature importance measures and local approximations), attention mechanisms embedded in modern architectures, and intrinsically interpretable surrogate or physics-constrained models have all been applied to materials problems [5, 6]. These approaches have supported interpretation in domains ranging from ionic conductivity in ceramics to phase stability in perovskites and mechanical behavior in alloys, facilitating human–AI collaboration by highlighting correlative—and occasionally causal—drivers of materials performance [7–9].

However, most applications of XAI in materials science implicitly assume static conditions: explanations are generated for a fixed model trained on a fixed dataset. In contrast, iterative materials AI workflows repeatedly update both models and data, producing sequences of explanations rather than single explanatory snapshots. As new compositions, processing regimes, or characterization modalities are introduced, explanations are recalculated, reshaped, or replaced—raising a fundamental but largely unexamined question: are explanations themselves stable and reliable across iterations?

This manuscript introduces explainability drift as a novel theoretical construct to address this gap. Explainability drift is the progressive change, degradation, or divergence in the quality, consistency, and interpretability of XAI outputs across iterative cycles in materials AI workflows. Unlike classical concept drift—which concerns changes in the predictive relationship between inputs and targets, typically formalized as shifts in P(y∣X)P(y\mid X)P(y∣X)—explainability drift pertains specifically to the interpretive layer of AI systems: how feature attributions evolve, how surrogate explanations lose fidelity, or how explanatory narratives become misaligned with expanding domain knowledge [10–12].

In iterative materials contexts, data updates often involve subtle yet consequential distributional shifts, such as the inclusion of rare phases, extreme operating conditions, or previously unexplored regions of chemical space. These shifts can destabilize explanation mechanisms even when predictive performance appears robust. For example, feature importance rankings derived from attribution methods may fluctuate non-monotonically across iterations, producing contradictory physical interpretations and undermining confidence in AI-guided hypothesis generation. Such instability is particularly problematic in materials science, where explanations are frequently used to justify experimental prioritization and resource allocation.

The conceptual failure analysis developed here identifies three interrelated modes of explainability drift: (i) attributional drift, where dominant explanatory features oscillate across iterations, generating inconsistent design hypotheses; (ii) fidelity drift, wherein surrogate or simplified explanation models progressively diverge from increasingly complex base learners; and (iii) epistemic drift, where explanations fail to capture emergent or latent physical mechanisms as workflows expand into underexplored materials regimes. These failure modes do not operate in isolation; rather, they can interact and cascade, leading to inefficient experimentation, misdirected optimization efforts, and stalled discovery trajectories in autonomous or semi-autonomous systems [13–15].

Although recent literature demonstrates the utility of XAI for single-shot interpretability in materials applications, systematic investigations into the longitudinal stability of explanations remain scarce. Existing studies largely focus on static benchmarks or isolated use cases, leaving unresolved how interpretability behaves under repeated retraining, data augmentation, and architectural evolution [16–18]. Iterative workflows amplify these vulnerabilities, as explanation inconsistencies compound over time and silently propagate through decision loops.

In response, this paper advances a purely theoretical synthesis that positions explainability drift as a critical, yet underrecognized, limitation of iterative materials AI. By integrating insights from XAI theory, materials informatics, and drift

phenomena, the manuscript establishes a conceptual foundation for analyzing, characterizing, and mitigating drift in explanations—without recourse to empirical datasets or simulation studies.

Theoretical Background and Literature Synthesis

Iterative AI workflows in materials science

Iterative artificial intelligence workflows mark a fundamental transition in materials science from static, model-centric prediction toward adaptive, closed-loop discovery systems. In these workflows, initial models trained on curated experimental or computational datasets are continuously updated through active learning, experimental feedback, or generative proposal mechanisms, enabling progressive refinement of hypotheses and design strategies [19–21]. Rather than treating learning as a one-time optimization problem, iterative pipelines conceptualize materials discovery as a dynamic process in which knowledge evolves alongside data acquisition.

Bayesian optimization coupled with surrogate modeling provides a canonical illustration of this paradigm. By iteratively updating acquisition functions that trade off exploration and exploitation, these systems navigate high-dimensional compositional and processing spaces while managing uncertainty under sparse sampling. In materials applications—such as alloy design or catalyst optimization—successive experimental observations reshape posterior beliefs, effectively redefining the statistical landscape in which optimization occurs. Similarly, generative frameworks, including variational autoencoders, autoregressive models, and diffusion-based architectures, extend iterative logic by proposing candidate structures, screening them through proxy simulations or learned property predictors, and reincorporating outcomes into subsequent training cycles [22, 23].

A defining characteristic of these workflows is distributional evolution. Newly synthesized materials, unexpected phase behaviors, or multimodal data integration (e.g., coupling microscopy, spectroscopy, and thermodynamic descriptors) introduce samples that systematically deviate from the assumptions embedded in the original training distribution. Consequently, iterative materials AI pipelines are not merely refining models within a fixed probabilistic regime; they are continuously redefining the feature space, the target landscape, and the epistemic boundaries of the problem itself. This dynamic setting creates unique interpretability challenges that are largely absent in static machine-learning deployments.

Explainable AI techniques and their materials applications

Explainable artificial intelligence (XAI) encompasses a diverse set of methods designed to render model behavior interpretable to human users, typically categorized into ante-hoc (intrinsically interpretable) and post-hoc explanatory approaches [5, 24]. Ante-hoc strategies embed interpretability directly into model architecture, while post-hoc methods derive explanations after model training without altering the underlying predictive mechanism.

In materials science, post-hoc attribution techniques—such as feature importance measures, gradient-based saliency, and Shapley-value approximations—have been widely adopted to identify influential descriptors governing properties such as perovskite stability, dielectric response, and catalytic activity [6]. Example-based explanations and surrogate models further translate complex predictors into simplified, locally faithful representations intended to support scientific reasoning. Concurrently, attention mechanisms within transformer and graph-based architectures offer internal signals that are often interpreted as proxies for elemental interactions, bonding environments, or structural motifs.

Ante-hoc approaches, including symbolic regression, sparse linear models, and physics-informed neural networks, aim to ensure interpretability by constraining learned representations to align with known physical laws, thermodynamic principles, or microstructural hierarchies [7, 25]. These models are often promoted as more trustworthy alternatives to opaque deep learning, particularly in scientific domains where interpretability underpins hypothesis formation and experimental decision-making.

Despite these advances, the dominant evaluation paradigm for XAI in materials science remains static. Explanations are typically assessed at a single training snapshot, assuming a fixed data distribution and stable model architecture. Little theoretical attention has been paid to how explanations behave when models are repeatedly retrained, augmented with new modalities, or embedded within autonomous experimental loops—conditions that increasingly define state-of-the-art materials AI systems.

Concept drift phenomena and analogies in scientific machine learning

Concept drift traditionally refers to temporal or contextual changes in the conditional distribution P(y|X)P(y\mid X)P(y|X), leading to degradation in predictive performance if models are not updated accordingly. Extensive methodological literature addresses this challenge through drift detection mechanisms—often based on statistical tests over error distributions—and adaptation strategies such as ensemble updating, sliding windows, or incremental retraining [10, 25].

In scientific machine learning and materials modeling, analogous forms of drift arise from physical and experimental realities rather than solely from time. Phase transitions, processing variability, measurement noise, and the fusion of heterogeneous data modalities can all induce shifts that invalidate earlier learned mappings between descriptors and properties [11]. These phenomena are increasingly recognized as central obstacles to robust prediction in materials informatics.

However, extending the drift concept beyond prediction reveals a more subtle but equally consequential issue: explainability drift. While predictive accuracy may remain stable or even improve across iterations, the explanatory layer can undergo substantial transformation. Feature attributions may reorder unpredictably, surrogate explanations may lose fidelity as base models grow in complexity, and explanatory narratives may drift away from established domain priors as exploration moves into less familiar regions of chemical space [12]. Unlike conventional concept drift, these effects are not reliably detected by performance metrics, yet they directly affect the scientific validity of AI-mediated reasoning.

Gaps in sustaining explainability across iterative workflows

A synthesis of current literature reveals a pronounced emphasis on one-off explainability, with explanations treated as static artifacts rather than dynamic components of evolving workflows [16, 18]. Theoretical treatments rarely address how repeated model updates compound interpretability inconsistencies, nor how explanation instability interacts with sparse data regimes characteristic of materials discovery. This gap is particularly problematic in materials science, where minor perturbations in data or representation can trigger disproportionate shifts in inferred feature relevance, given the vastness and heterogeneity of chemical space [16].

The absence of longitudinal perspectives on explainability represents a critical theoretical vulnerability. Iterative workflows are designed to accelerate discovery by closing the loop between prediction, explanation, and experimentation. Yet if explanations drift silently across iterations, they risk propagating misleading scientific intuitions, distorting experimental prioritization, and eroding trust in autonomous or semi-autonomous systems. Interpretability, rather than serving as a stabilizing epistemic anchor, may itself become a source of hidden instability.

These unresolved tensions motivate the framework developed in the following section, which reconceptualizes explainability not as a static validation tool but as a dynamic, failure-prone layer within iterative materials AI systems. By unifying insights from iterative learning, XAI methodologies, and drift theory, the proposed framework provides a coherent conceptual basis for analyzing and mitigating explainability drift in autonomous scientific discovery pipelines.

Proposed conceptual framework

The proposed framework conceptualizes explainability drift as a multidimensional phenomenon within iterative materials AI workflows, structured across attributional, structural, and epistemic dimensions to enable systematic failure analysis. In the attributional dimension, drift manifests as instability in local or global feature importance scores (e.g., SHAP values or

gradient-based attributions) across iterations, quantified conceptually by divergence metrics such as Jensen-Shannon distance between successive explanation vectors. Structural drift concerns the degradation of surrogate or proxy model fidelity, in which simpler, interpretable approximations increasingly deviate from the evolving base model as complexity accrues with new data modalities. Epistemic drift captures broader semantic misalignment, wherein explanations fail to incorporate or reflect newly emergent physical mechanisms or domain knowledge uncovered during workflow progression.

Failure modes are analyzed as cascading effects: attributional instability leads to oscillating design hypotheses; structural erosion reduces the trustworthiness of explanations for experimental prioritization; epistemic drift fosters overgeneralization or overlooked causal factors in material property landscapes. Mitigation pathways are theorized through conceptual interventions like periodic explanation recalibration (re-anchoring to domain priors), ensemble explanation aggregation for stability, or hybrid physics-informed explanation modules that constrain drift via thermodynamic consistency checks—all without prescribing implementation details. The framework posits drift as inevitable yet monitorable via theoretical comparison protocols that track explanation coherence relative to workflow state transitions.

This novel structure distinguishes explainability drift from related concepts by centering on the explanatory apparatus's evolution, providing a failure-analysis lens tailored to the iterative, data-evolving nature of materials AI. The iterative materials AI workflow and associated explainability drift are illustrated in **Figure 1**.

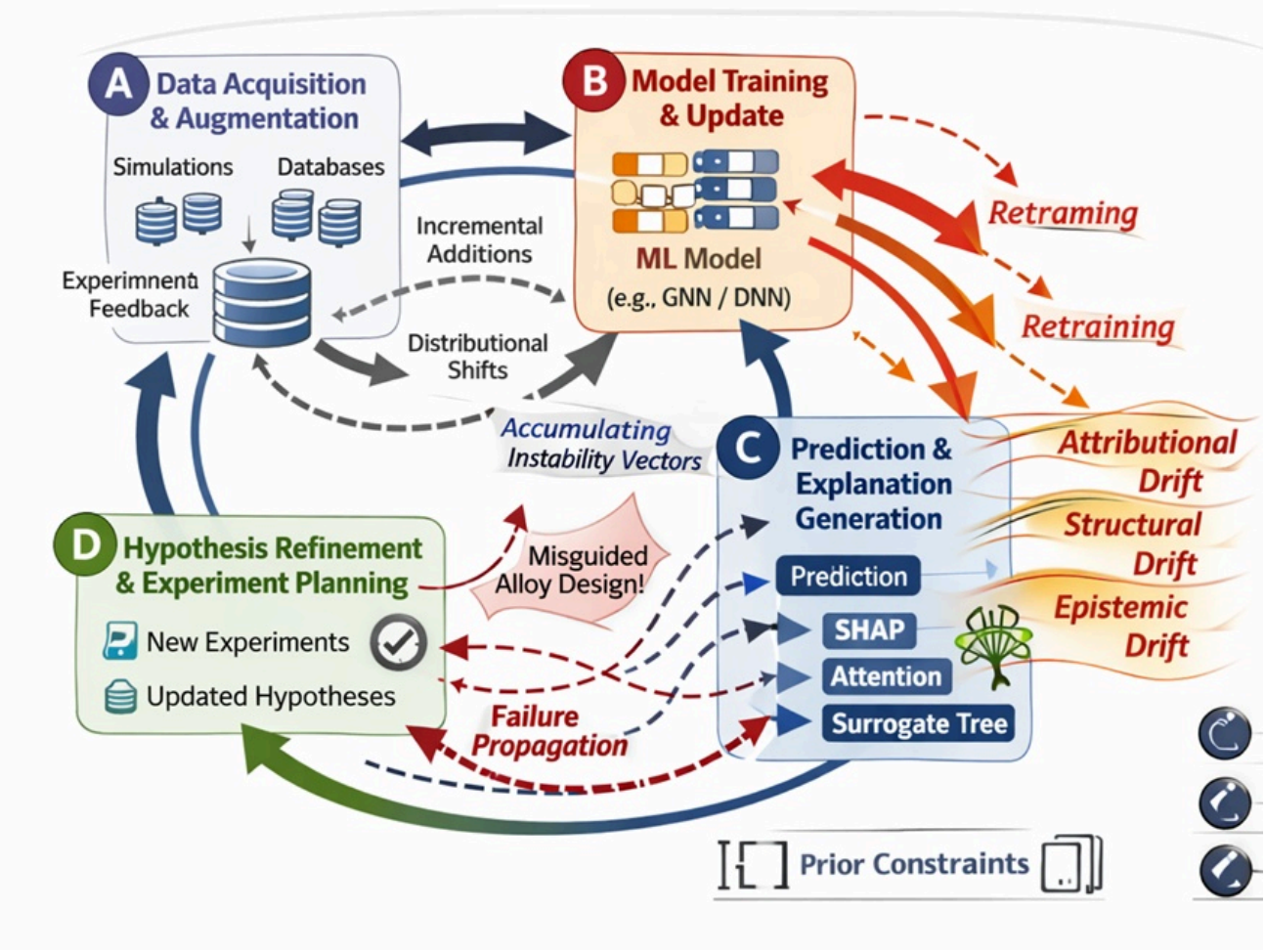

Figure 1. An iterative materials AI workflow with explainability drift

The multidimensional nature of explainability drift, summarized in **Table 1**, highlights how distinct but interacting mechanisms undermine interpretability across iterative workflows.

Table 1. Dimensions of explainability drift in iterative materials AI workflows

Drift dimension	Conceptual definition	Primary mechanism in iterative workflows	Manifestation in materials AI	Scientific consequence
Attributional drift	Instability or reordering of explanatory feature importance across iterations	Incremental distributional shifts in feature space due to active sampling, exploration of rare regimes, or data augmentation	Fluctuating importance of compositional vs. processing descriptors across optimization cycles	Contradictory design heuristics; reduced reproducibility of hypothesis generation
Structural drift	Progressive loss of fidelity between interpretable surrogates and evolving base models	Increasing architectural complexity from multimodal data integration and iterative retraining	Divergence between surrogate explanations and graph-based or deep neural predictors	Erosion of trust in the validity of experimental prioritization
Epistemic drift	Semantic misalignment between explanations and evolving domain knowledge	Expansion into sparsely sampled or emergent regions of materials space	Failure to surface latent physical mechanisms (e.g., phase stability, nonlocal couplings)	Misleading scientific narratives; overlooked causal drivers
Cascading interaction	Nonlinear interaction among drift dimensions	Propagation of early attributional instability into structural and epistemic failures	Amplified explanation incoherence across autonomous loops	Inefficient exploration, stalled discovery trajectories

Propositions: Explainability drift in iterative materials AI workflows

Proposition 1 (Attributional explainability drift)

In iterative materials AI workflows, explainability drift arises primarily from attributional instability induced by incremental distributional shifts within evolving feature spaces. These shifts generate oscillations in feature-importance rankings across iterations, undermining the coherence and reproducibility of hypothesis generation for novel compositions or microstructures. This form of drift is theoretically distinct from predictive concept drift, as it perturbs the interpretability layer independently of conventional accuracy or error metrics.

Proposition 2 (Structural drift of surrogate explanations)

Structural explainability drift intensifies nonlinearly as workflows iterate and integrate increasingly multimodal descriptors (e.g., atomic-scale, microstructural, and thermodynamic representations). As base learners evolve toward higher-order architectures—such as graph-based or hierarchical neural models—the fidelity gap between interpretable surrogate explanations and underlying model behavior progressively widens, leading to systematic degradation of explanatory validity.

Proposition 3 (Epistemic explainability drift)

Epistemic drift emerges when the semantic content of explainability outputs becomes misaligned with evolving domain knowledge, particularly as workflows explore sparsely sampled or weakly constrained regions of chemical space. Under

these conditions, explanations increasingly fail to surface latent physical mechanisms—such as emergent phase-stability relationships or nonlocal structure–property couplings—that are only revealed through iterative scientific feedback.

Proposition 4 (Cascading explainability failure)

Attributional, structural, and epistemic explainability drifts interact multiplicatively rather than additively, producing cascading failure modes within autonomous or semi-autonomous materials discovery loops. Early attributional instability propagates into surrogate misrepresentation, which in turn amplifies epistemic gaps, ultimately degrading experimental prioritization, inflating exploration costs, and constraining discovery trajectories.

Proposition 5 (Theoretical diagnostics of drift)

Explainability drift can be conceptually monitored through divergence-based diagnostics—such as distributional distances computed over successive attribution or explanation manifolds—providing a theoretically grounded indicator of workflow robustness. Such monitoring enables proactive interpretive recalibration anchored in domain priors and physical reasoning, without reliance on empirical retraining thresholds or performance-based triggers.

Proposition 6 (Hybrid mitigation of explainability drift)

Because explainability drift is inherently multidimensional, its mitigation requires hybrid theoretical constructs—such as physics-constrained or domain-regularized explanation ensembles—that enforce cross-dimensional consistency. These constructs preserve interpretability as an epistemic control mechanism rather than a post-hoc artifact, thereby sustaining trustworthiness and scientific validity across iterative materials AI innovation cycles.

Results and Discussion

The proposed framework for explainability drift advances current discourse by reframing interpretability not as a static validation artifact, but as a dynamic, failure-prone layer within iterative materials AI workflows. By decomposing explainability drift into attributional, structural, and epistemic dimensions, the framework exposes mechanisms through which explanation fidelity degrades even when predictive performance remains stable or improves. This distinction is critical, as most existing mitigation strategies for instability in machine learning prioritize recalibrating predictions rather than preserving interpretive coherence. In contrast, the present framework foregrounds explanation dynamics as a first-order epistemic concern in autonomous and semi-autonomous discovery systems.

A central insight of the framework is that incremental data augmentation, often treated as inherently beneficial in active learning or Bayesian optimization, can have nontrivial and compounding effects on explanatory stability. Small shifts in training distributions—introduced through targeted sampling, exploration of extreme compositions, or integration of new characterization modalities—may disproportionately perturb attribution mechanisms, leading to oscillations in feature relevance or narrative explanations. In materials discovery contexts, such oscillations are not merely cosmetic; they can lead to contradictory design heuristics, such as alternating emphasis on compositional tuning and processing parameters across successive iterations. Traditional concept drift paradigms do not capture this failure mode, as they implicitly assume that stable accuracy implies stable understanding. Key failure pathways and corresponding theoretical mitigation strategies are synthesized in **Table 2**, illustrating how drift propagates across interpretability layers.

Table 2. Conceptual failure modes and mitigation pathways for explainability drift

Failure Mode	Underlying drift interaction	Triggering conditions in iterative materials AI	Impact on discovery workflow	Conceptual mitigation strategy
Oscillating design hypotheses	Attributional → Epistemic	Aggressive exploration in Bayesian optimization; inclusion of rare compositions	Conflicting guidance on material optimization targets	Explanation ensemble aggregation to stabilize attribution signals

Surrogate explanation breakdown	Attributional → Structural	Rapid architectural evolution with multimodal descriptors	Loss of local faithfulness in interpretable models	Periodic surrogate recalibration anchored to domain priors
Epistemic blind spots	Structural → Epistemic	Exploration of under-characterized chemical space	Overlooking emergent physical mechanisms	Physics-constrained or domain-regularized explanation frameworks
Cascading interpretability failure	Attributional + Structural + Epistemic	Long-horizon autonomous experimentation loops	Inefficient experimental allocation; declining trust	Conceptual monitoring of explanation divergence across iterations
Over-stabilization risk	Excessive drift suppression	Rigid explanation anchoring	Reduced adaptability to genuinely novel phenomena	Balanced tolerance for explanation evolution

The framework further elucidates how explainability drift becomes particularly consequential in high-impact materials applications characterized by sparse data regimes and emergent phenomena. In domains such as perovskite stability optimization, alloy phase design, or functional ceramics, explanatory signals are often used to justify prioritizing experiments under constrained resources. Attributional drift in these settings can misdirect attention toward features that appear salient only transiently, exacerbating the exploration–exploitation dilemma inherent in vast chemical and structural design spaces. Epistemic drift poses an even bigger risk: as workflows venture into poorly characterized regions of materials space, explanations may lag behind evolving scientific understanding, systematically overlooking latent physical mechanisms that only become apparent through cumulative feedback.

Importantly, the framework highlights that explainability drift is not additive but interactive. Attributional instability can undermine the fidelity of surrogate explanations, thereby amplifying epistemic misalignment by distorting the conceptual narratives through which scientists interpret model behavior. These cascading effects help explain why some autonomous discovery pipelines experience diminishing returns despite increasing data volume and model sophistication. The framework thus provides a unifying conceptual account of failures that are often observed empirically but rarely explicitly theorized.

From a methodological perspective, the framework's purely theoretical nature is both a limitation and a deliberate choice. Rather than prescribing quantitative thresholds or algorithm-specific diagnostics, the framework advocates conceptual monitoring of explanatory divergence as an epistemic early-warning signal. This stance reflects the current immaturity of standardized metrics for explanation quality and stability in scientific machine learning. By decoupling the framework from specific XAI techniques—whether attribution-based, surrogate-driven, or intrinsically interpretable—it retains generality across evolving methodological ecosystems.

Comparisons with existing literature underscore the originality of this contribution. Prior studies on XAI in materials science predominantly focus on interpretability in isolated prediction tasks, whereas research on drift largely centers on temporal performance degradation across general machine learning systems. Few, if any, works integrate these strands into a coherent failure-analysis structure that explicitly accounts for the iterative evolution of explanations themselves. The multidimensional layering proposed here enables systematic reasoning about explainability failures without privileging particular models, materials classes, or experimental platforms [15-17].

The framework also carries broader theoretical implications for the design of iterative AI systems beyond materials science. Sustaining explainability may require rethinking workflow architectures to incorporate periodic explanatory anchoring—for example, re-aligning explanations with known physical invariants, conservation laws, or domain priors. Such anchoring could mitigate cascading drift while preserving adaptive learning. At the same time, the framework cautions against excessive rigidity: complete suppression of explanatory change may stifle discovery in dynamic regimes

where new mechanisms genuinely emerge. This tension highlights the need for a nuanced conceptual understanding of what constitutes acceptable versus pathological drift—an open theoretical question warranting further investigation [19-22].

Overall, the discussion positions explainability drift not as an anomaly to be eliminated, but as an intrinsic property of iterative, data-driven scientific exploration. Recognizing and theorizing this property is a necessary step toward developing resilient, interpretable, and trustworthy AI ecosystems capable of sustaining long-term scientific innovation.

Conclusion

This manuscript introduces explainability drift as a novel theoretical construct that characterizes the progressive degradation and transformation of interpretive fidelity in iterative materials AI workflows. Through a structured conceptual failure analysis, the proposed multidimensional framework—encompassing attributional, structural, and epistemic layers—elucidates how evolving data distributions, model architectures, and domain knowledge can silently compromise explanatory coherence, even in the absence of predictive performance loss.

The propositions derived from this framework articulate causal pathways through which explainability drift emerges, propagates, and cascades within closed-loop discovery systems. Collectively, they emphasize that interpretability in materials AI cannot be treated as a static validation step, but must instead be understood as a dynamic epistemic process requiring dedicated theoretical safeguards. In particular, the analysis underscores the necessity of explanation-centric monitoring and physics-informed stabilization principles to prevent misguided hypothesis generation and inefficient experimental prioritization.

By synthesizing insights from autonomous materials platforms, explainable AI research, and drift theory, this work highlights a critical tension at the heart of iterative AI: while adaptive learning accelerates exploration, unchecked drift in explanations threatens scientific understanding, trust, and long-term discovery trajectories. The purely conceptual foundation developed here provides a vocabulary and analytical structure for addressing this tension, without imposing premature algorithmic prescriptions.

Future theoretical efforts should build on this foundation by formalizing interaction effects between workflow design choices and explainability stability, and by developing principled notions of acceptable drift grounded in scientific epistemology. Ultimately, advancing robust, human-aligned AI in materials science will require systems that not only predict efficiently but also sustain interpretable insight as knowledge and data co-evolve.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 19 Mar 2025 Revised: 16 Apr 2025 Accepted: 20 May 2025
Published online: 18 July 2025

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.
https://doi.org/10.1038/s41586-023-06735-9.

Szymanski NJ, Rendy B, Fei Y, Kumar RE, He T, Milsted D, et al. An autonomous laboratory for the accelerated synthesis of novel materials. Nature. 2023;624(7990):86-91.
https://doi.org/10.1038/s41586-023-06734-w.

Jablonka KM, Ongari D, Moosavi SM, et al. Explainable machine learning in materials science. npj Comput Mater. 2022;8:204.
https://doi.org/10.1038/s41524-022-00884-7.

Kusne AG, Gao H, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. Nat Commun. 2020;11:5966.
https://doi.org/10.1038/s41467-020-19597-w.

Oviedo F, Ferres JL, Buonassisi T, et al. Interpretable and explainable machine learning for materials science and chemistry. Acc Mater Res. 2023;4(6):961-73.
https://doi.org/10.1021/accountsmr.1c00244.

Chen C, Zuo Y, Ye W, et al. A critical review of machine learning in materials science. Adv Mater. 2022;34(12):2106820.
https://doi.org/10.1002/adma.202106820.

Ziatdinov M, Ghosh A, Wong CY, Kalinin SV, Atomai framework for deep learning analysis of image and spectroscopy data in electron and scanning probe microscopy. Nat Mach Intell. 2022;4(12):1101-12.

van Houtum GJ, Vlasea ML. Active learning via adaptive weighted uncertainty sampling applied to additive manufacturing. Addit Manuf. 2021;48:102411.

Ramprasad R, Batra R, Pilania G, et al. Emerging materials intelligence ecosystems propelled by machine learning. Nat Rev Mater. 2021;6(8):655-78.

Butler KT, Davies DW, Cartwright H, et al. Machine learning for molecular and materials science. Nature. 2018.

Ziatdinov M, Liu Y, Morozovska AN, et al. Hypothesis learning in automated experiment: application to combinatorial materials libraries. Nat Commun. 2021;12:3394.

Zhang Z. Computation and data-driven methods toward superhard materials design. University of Houston; 2022.

Hinder F, et al. Concept drift detection and adaptation in machine learning: a systematic review. Front Artif Intell. 2024;7:1330258.

Lu J, Liu A, Dong F, et al. A systematic review on detection and adaptation of concept drift in streaming data using machine learning techniques. WIREs Data Min Knowl Discov. 2024;14(3):e1536.

Webb GI, Lee WS, Goethals B, et al. Understanding concept drift. Mach Learn. 2020;109:1-25.

Tina GM, Ventura C, Ferlito S, De Vito S. A state-of-the-art review on machine-learning based methods for PV. Appl Sci. 2021;11(16):7550.

Ditzler G, Roveri M, Alippi C, et al. Learning in nonstationary environments: a survey. IEEE Comput Intell Mag. 2020;15(2):12-25.

Batra R, Pilania G, Uberuaga BP, et al. Multifidelity machine learning models for structure-property mapping in materials discovery. npj Comput Mater. 2021;7:1-12.

Ishizuki N, Shimizu R, Hitosugi T. Autonomous experimental systems in materials science. Sci Technol Adv Mater Methods. 2023;3(1):2197519.

Park CW, Wolverton C. Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. Phys Rev Mater. 2020;4(6):063801.

Wang H, Xie Y, Li Q, et al. Materials property prediction with uncertainty quantification: A benchmark study. Appl Phys Rev. 2023;10(2):021409.
https://doi.org/10.1063/5.0141401.

Jablonka KM, Ongari D, Moosavi SM, Smit B. Big-data science in porous materials: materials genomics and machine learning. Chem Rev. 2020;120(16):8066-8127.
https://doi.org/10.1021/acs.chemrev.0c00004.

Liu Y, Zhao T, Ju W, Shi S. Small data machine learning in materials science. npj Comput Mater. 2023;9:42.
https://doi.org/10.1038/s41524-023-01000-z.

Hinder F, Vaquet V, Hammer B. One or two things we know about concept drift—a survey on monitoring in evolving environments. Part A: detecting concept drift. Front Artif Intell. 2024;7:1330257.
https://doi.org/10.3389/frai.2024.1330257.

Wang Z, et al. Targeted materials discovery using Bayesian algorithm execution. npj Comput Mater. 2024;10:156.
https://doi.org/10.1038/s41524-024-01326-2.