

ORIGINAL RESEARCH

Open access

# The Problem of Epistemic Debt in Iterative Materials AI Pipelines

Bruno Martins<sup>1\*</sup>, Lucas Pereira<sup>1</sup>, Renata Azevedo<sup>2</sup>, Pedro Costa<sup>1</sup>

## Abstract

Iterative materials AI pipelines, encompassing active learning frameworks and closed-loop discovery systems, have transformed the pace of materials innovation by enabling sequential decision-making under uncertainty. Yet these very systems are susceptible to an underrecognized failure mode: epistemic debt, the gradual accumulation of unexamined assumptions, unresolved uncertainties, and path-dependent constraints that silently erode the future potential of knowledge generation. This failure mode remains largely unacknowledged despite the growing reliance on such pipelines in materials science. This paper articulates epistemic debt as an intrinsic structural risk of iterative materials AI, one that demands explicit recognition if the promise of autonomous discovery is to be realized sustainably. By tracing the mechanisms through which debt accumulates, identifying materials-specific vulnerabilities that exacerbate it, and proposing both a typology and practical management principles, the analysis seeks to shift the conversation from short-term performance metrics to long-term epistemological integrity. Epistemic debt is formally defined as the accumulation of unexamined assumptions, unresolved uncertainties, and path-dependent constraints in an iterative knowledge-generating system that increase the cost of future learning or limit the space of future discoveries. It is conceptually distinct from technical debt, which concerns code maintainability and infrastructure, and from statistical compounding errors, which arise from sampling variance or measurement noise. The mechanisms driving epistemic debt—assumption cascades, path-dependent constraints, unrecognized uncertainty, and feedback loop amplification—interact in ways that are especially pernicious in materials contexts, where small initial datasets, high-dimensional composition spaces, and costly experimental iterations amplify the long-term consequences of early choices. Materials-specific vulnerabilities render these pipelines particularly fragile, as early decisions about representation or sampling can foreclose vast regions of chemical space without immediate visibility. A typology of epistemic debt types is articulated, distinguishing representational, sampling, modeling, and decision debt, each carrying unique signatures and risks within active-learning loops. Detection principles centered on assumption auditing and counterfactual tracing, together with mitigation strategies such as ensemble diversity and deliberate debt refinancing, provide a structured framework for managing this failure mode before it compounds irreversibly. By foregrounding epistemic debt as a distinct category of risk, this analysis offers the materials AI community a new lens through which to evaluate the sustainability of iterative discovery pipelines and to safeguard the integrity of long-term scientific progress.

**Keywords** Epistemic debt, Iterative materials AI, Active learning pipelines, Path dependence, Assumption cascades, Closed-loop discovery

\*Correspondence:

Bruno Martins  
bruno.martins@gmail.com

<sup>1</sup> Department of Materials Modeling and AI, University of Minho, Braga, Portugal

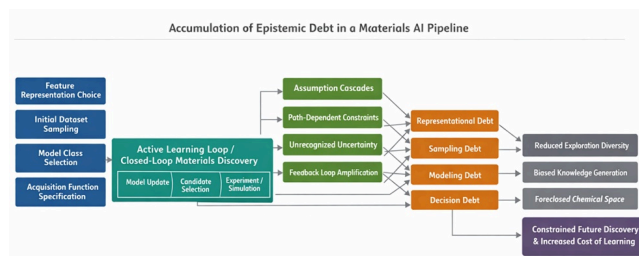
<sup>2</sup> Department of Intelligent Materials Systems, University of Porto, Porto, Portugal

## Introduction

Iterative materials AI pipelines have become indispensable for navigating the immense complexity of materials

discovery. Systems built around active learning, Bayesian optimization, and closed-loop experimental feedback now routinely guide the selection of next experiments or computations, updating predictive models in real time and steering exploration toward promising regions of chemical space. These workflows promise to compress decades of traditional trial-and-error research into months or even weeks. Yet, they introduce a subtle and largely invisible risk: the accumulation of epistemic debt. Early decisions about feature representations, initial data sampling strategies, model architectures, and acquisition functions embed assumptions that propagate forward, constraining the very possibilities available to later stages of the pipeline. Over successive iterations, these constraints compound, raising the marginal cost of new knowledge and, in extreme cases, locking the system into narrow, suboptimal discovery trajectories [1-8].

**Figure 1** provides a hierarchical representation of how early design decisions propagate through iterative materials AI pipelines to generate epistemic debt and constrain downstream discovery.



**Figure 1.** A hierarchical representation of how early design decisions propagate through iterative materials AI pipelines to generate epistemic debt and constrain downstream discovery.

The problem is not merely one of statistical inefficiency or computational overhead. It is fundamentally epistemological. Each iteration in an active-learning loop does not simply add data; it also inherits and reinforces a growing lattice of prior commitments whose validity is rarely revisited [6, 9-13]. When an early choice of featurization is treated as fixed, or when an initial screening set is accepted without scrutiny, subsequent model updates become conditionally dependent on those choices in ways that are difficult to unwind. This path dependence is especially acute in materials science because the search space is combinatorially vast, experimental validation is expensive, and negative results are seldom reported. Consequently, the community observes only the successful

branches of the discovery tree while the foreclosed alternatives remain invisible.

Technical debt in machine learning systems, as classically described, manifests in the form of brittle code, outdated dependencies, or deferred refactoring that eventually slows development velocity [1]. Epistemic debt, by contrast, resides at the level of scientific knowledge itself. It is not about whether the software runs but about whether the knowledge produced remains robust and extensible. The distinction matters because materials AI pipelines are not merely engineering tools; they are knowledge-generating engines whose outputs shape future research agendas. When epistemic debt accrues unnoticed, entire families of materials or design principles may be overlooked not because they are intrinsically unpromising but because the pipeline's early architecture rendered them invisible.

Recent surveys of autonomous materials research highlight the growing sophistication of closed-loop systems, yet rarely address the long-term epistemological costs of the sequential decisions these systems embody [6]. Similarly, foundational works on machine learning for molecular and materials science emphasize performance gains while leaving the accumulating constraints of iterative workflows underexamined [3, 4]. The present analysis, therefore, identifies epistemic debt as a distinct failure mode that operates orthogonally to conventional concerns such as model accuracy or data quality. It arises precisely because iterative pipelines are designed to be adaptive; each adaptation builds upon prior adaptations, creating a chain of conditional dependencies that is seldom audited.

The consequences extend beyond any single project. In a field where publication bias already favors positive outcomes, epistemic debt can become institutionalized. Successful discoveries are celebrated, while the hidden opportunity costs—the alternative compositions or property spaces that were never explored because of an early modeling choice—remain undocumented. Over time, the community's collective knowledge graph develops blind spots that are difficult to detect because the pipelines themselves have shaped what counts as “interesting” data. This paper, therefore, articulates epistemic debt as an intrinsic structural risk of iterative materials AI, one that demands explicit recognition if the promise of autonomous discovery is to be realized sustainably. By tracing the mechanisms through which debt accumulates, identifying materials-specific vulnerabilities that exacerbate it, and proposing both a typology and practical management

principles, the analysis seeks to shift the conversation from short-term performance metrics to long-term epistemological integrity.

## Defining Epistemic Debt

Epistemic debt must be understood as a distinct category of risk that operates at the level of scientific reasoning rather than at the level of software implementation or statistical estimation. It is not a synonym for technical debt, nor is it reducible to the propagation of measurement error. Instead, it captures a specific form of accumulated epistemological liability that arises uniquely within iterative, sequential knowledge-generating systems.

To clarify its conceptual distinctiveness, **Table 1** systematically differentiates epistemic debt from adjacent failure modes such as technical debt, statistical error propagation, and path dependence.

**Table 1.** Structural differentiation of epistemic debt from related failure modes in iterative materials AI

Dimension	Epistemic debt	Technical debt	
Core level of operation	Scientific knowledge structure	Software/infrastructure	
Primary cause	Unexamined assumptions, unresolved uncertainty, and early constraints	Poor code design and deferred maintenance	
Visibility	Latent and often invisible during execution	Observable in system performance	
Reversibility	Low; often requires pipeline restart or branching	High; can be refactored	
Impact on discovery space	Narrows and forecloses	Slows development efficiency	

	future possibilities		
Detectability via standard metrics	Poor (accuracy may improve)	Moderate (system failures, delays)	(c
Typical mitigation strategy	Assumption auditing, counterfactuals, and exploration reserves	Code refactoring and modularization	I sa e

**Definition 1:** Epistemic debt is the accumulation of unexamined assumptions, unresolved uncertainties, and path-dependent constraints in an iterative knowledge-generating system that increase the cost of future learning or limit the space of future discoveries.

This definition emphasizes three interlocking components. First, unexamined assumptions refer to modeling choices—such as the selection of a particular descriptor set or the assumption that a given surrogate model class adequately captures underlying physics—that are made early and then treated as given. Second, unresolved uncertainties are those initial ambiguities about data quality, relevance, or completeness that are not propagated forward in a principled manner. Third, path-dependent constraints are the irreversible narrowing of the exploration space that occurs when early decisions eliminate branches of the decision tree without later re-evaluation. Together, these elements create a growing “debt” because correcting them later requires more effort or, in some cases, becomes practically impossible within the original pipeline architecture.

The distinction from technical debt is fundamental. Technical debt, as analyzed in foundational machine-learning system studies, concerns implementation artifacts: poorly modularized code, inadequate testing, or deferred infrastructure investments that inflate future maintenance costs [1]. Epistemic debt, by contrast, concerns the validity and extensibility of the knowledge claims themselves. A pipeline may have pristine code and flawless continuous integration yet still suffer from severe epistemic debt if its representational choices systematically exclude physically meaningful degrees of freedom.

Epistemic debt is also distinct from statistical compounding error. The latter arises when variance or bias in individual measurements propagates through a calculation, producing

widening confidence intervals. While statistical error is quantifiable and can often be mitigated by additional samples, epistemic debt is structural: it concerns the very framing of the problem space. Even with perfect measurements, an early commitment to a low-dimensional descriptor set can render certain physical mechanisms invisible, irrespective of how many additional data points are collected.

A third contrast is with simple compounding uncertainty in sequential decision-making. While uncertainty propagation is a recognized challenge in Bayesian active learning [7], epistemic debt goes further by incorporating the feedback between uncertainty estimates and the very data that are chosen to reduce them. When early uncertainty is treated as resolved, subsequent acquisition functions become optimistically biased, accelerating the accumulation of debt rather than merely reflecting it.

In materials AI, the consequences are concrete. A pipeline that begins with a small set of DFT calculations on a restricted compositional subspace may implicitly assume that the physics governing that subspace generalizes to neighboring regions [4, 14–20]. Later iterations then optimize within that subspace, never testing whether the initial generalization holds. The resulting knowledge is not wrong per se, but is incomplete in ways that become harder to detect the longer the pipeline runs. Over time, the community inherits a body of literature whose foundations rest on unexamined commitments, creating downstream costs when new research programs attempt to build upon these results.

Epistemic debt is therefore not an incidental byproduct of iteration; it is an intrinsic feature of any system that learns sequentially without periodic epistemic housekeeping. Recognizing it as such reframes the design of materials AI pipelines from purely performance-oriented engineering to a balanced practice that explicitly accounts for long-term epistemological sustainability.

## Mechanisms of Epistemic Debt Accumulation

Epistemic debt in materials AI does not accumulate as a diffuse byproduct of complexity; it emerges through a set of interacting mechanisms that progressively harden early assumptions into structural constraints on downstream inference. What begins as a provisional modeling choice

can, through iteration and reuse, acquire the status of an unquestioned foundation, shaping both the trajectory of discovery and the limits of what the system can subsequently perceive.

One of the most consequential pathways through which this occurs is the gradual entrenchment of early representational decisions. Initial choices about featurization or relevance—often justified by acceptable performance on a limited validation set—are carried forward into later stages without systematic re-examination. Over time, these representations cease to be treated as hypotheses and instead function as de facto ground truth. In materials informatics, this dynamic becomes particularly visible when descriptor sets tuned for a narrow class of compounds are extended to broader chemical spaces without modification, thereby excluding variables that become salient under new conditions [15, 20]. The resulting models inherit a constrained view of the system, one that reflects historical convenience rather than current evidence.

This representational inertia is compounded by the path-dependent nature of exploration in resource-limited environments. Early sampling decisions, especially within active-learning frameworks, shape the regions of composition space that are subsequently accessible. Once an acquisition function has concentrated effort within a particular domain, the accumulation of data in that region reinforces its perceived importance. In contrast, adjacent or unexplored regions become progressively less likely to be sampled. Reversing this trajectory is rarely feasible, as it would require reallocating already expended resources and reinitializing the search process. The consequence is a form of hysteresis in which the direction of inquiry becomes locked by initial conditions, and alternative pathways are effectively removed from consideration [7, 12].

A related source of epistemic debt arises from the treatment of uncertainty. Early-stage uncertainty, particularly of the epistemic kind associated with limited or biased data, is often compressed into point estimates once a model is trained. Subsequent iterations then proceed under the implicit assumption that these estimates are reliable, even when their underlying uncertainty has not been rigorously propagated. This leads to acquisition strategies that operate with unwarranted confidence, narrowing the search prematurely and reducing the system's capacity to recover from early misjudgments. In materials contexts, where data acquisition is costly and iterative refinement is constrained, the incentive to simplify

uncertainty representations is strong. Yet, the long-term effect is an erosion of epistemic flexibility [8, 13].

These dynamics are further intensified by the recursive structure of AI-guided discovery itself. Predictions generated by the model inform the selection of new experiments, and the resulting data are used to update the same model, creating a feedback loop in which early biases can be amplified rather than corrected. If the initial training set underrepresents certain regions or encodes systematic error, the acquisition function will tend to favor regions consistent with those biases, reinforcing them over successive iterations. What begins as a minor distortion can, through repeated reinforcement, evolve into a significant blind spot, narrowing the effective search space and limiting the system's capacity for genuine discovery [10, 20]. Although this risk is acknowledged in principle within the active-learning literature, practical safeguards remain limited, and the amplification mechanism often operates unchecked.

Because materials discovery pipelines operate under strict resource and time constraints, the opportunity to intervene diminishes rapidly as these mechanisms interact. Once epistemic debt has accumulated beyond a critical threshold, corrective action becomes disproportionately costly. In such cases, meaningful recovery may require abandoning the existing trajectory and initiating a new pipeline with altered initial assumptions—a step that is rarely feasible within typical research programs. The result is a system that continues to operate with inherited limitations, not because they are optimal, but because they have become structurally embedded in the process of discovery itself.

## Materials-Specific Vulnerabilities

Materials AI pipelines cannot be understood as straightforward extensions of generic machine-learning systems, because the structure of the materials domain amplifies the accumulation of epistemic debt in ways that are both systematic and difficult to reverse. A defining feature of this amplification lies in the scarcity of initial data. When models are trained on only dozens or hundreds of labeled examples, early representational and modeling choices exert disproportionate influence, effectively shaping the hypothesis space before meaningful statistical correction becomes possible [3, 4]. Under these conditions,

what might otherwise function as provisional assumptions quickly solidify into enduring constraints.

This fragility is intensified by the combinatorial scale and sparsity of chemical space. Even relatively simple compositional systems expand into vast, largely uncharted regions, and the absence of dense sampling means that early exploration decisions implicitly determine which portions of this space are ever considered. Once a pipeline commits to a particular region, there is little intrinsic mechanism to detect the opportunity cost of what remains unexplored [6, 20]. The resulting asymmetry between explored and unobserved regions introduces a structural blind spot that persists across iterations.

The economic reality of materials discovery further compounds this effect. Each experimental validation or high-fidelity simulation carries a substantial cost, discouraging exploratory sampling in favor of exploiting already-characterized regions. This pragmatic bias accelerates the entrenchment of early decisions, as the system becomes increasingly optimized around a narrow subset of possibilities [7, 13]. What emerges is not merely a technical limitation but a feedback between resource constraints and epistemic narrowing.

A related vulnerability arises from the relative weakness of theoretical constraints in many areas of materials science. While first-principles methods provide guidance, they often lack the precision required to invalidate incorrect modeling assumptions in complex, multi-component systems. As a result, early errors can persist undetected across multiple iterations, accumulating influence before sufficient evidence emerges to challenge them [4]. This delayed correction contrasts sharply with domains where strong theoretical priors enforce rapid falsification.

These dynamics are further obscured by publication practices that privilege successful outcomes. The absence of documented failures conceals the frequency and structure of epistemic lock-in, preventing the field from developing a shared understanding of how such trajectories emerge and persist [6]. When combined, these conditions—data scarcity, high-dimensional sparsity, costly iteration, limited theoretical constraint, and selective reporting—create an environment in which early assumptions are both highly influential and rarely revisited. The domain is therefore predisposed not only to the accumulation of epistemic debt but to its persistence.

## A Typology of Epistemic Debt

Translating epistemic debt into a usable analytical construct requires distinguishing its principal forms, each corresponding to a different locus within the materials AI pipeline. One prominent form arises at the level of representation, where the choice of descriptors constrains what relationships the model can, in principle, learn. When featurization schemes are fixed early and propagated without revision, they can exclude critical structural or chemical information. In practice, this becomes evident when simplified elemental descriptors fail to capture coordination-dependent phenomena in complex oxides, leading to systematic limits on generalization even as additional data are introduced [15, 20]. The persistence of such limitations signals that the constraint lies not in data quantity but in representational scope.

A closely related form of debt emerges through the structure of the data itself. Initial sampling decisions, particularly in active-learning contexts, shape the distribution on which all subsequent learning is based. When early exploration is confined to a narrow compositional subset, the model's perception of uncertainty becomes localized, giving the appearance of convergence while vast regions remain unexamined. This misalignment between local confidence and global ignorance reflects a coupling between path-dependent sampling and unpropagated uncertainty, producing a trajectory that appears efficient but is fundamentally incomplete [7, 8].

The model architecture introduces another layer of constraint, one that operates through the expressivity of the surrogate itself. When the chosen model lacks the capacity to represent the true complexity of the property landscape, it imposes a structural ceiling on what can be discovered. This is particularly apparent in cases where relatively simple models are applied to systems characterized by multimodal or highly nonlinear behavior, resulting in systematic residual patterns that correlate with regions the acquisition strategy has deprioritized [10, 12]. In such cases, the limitation is not merely predictive error but a deeper mismatch between model form and problem structure.

Finally, the decision-making layer contributes its own form of debt through the criteria used to guide exploration. Acquisition functions that emphasize short-term gains can bias the search toward regions that are already well understood, reducing diversity and limiting the discovery of

genuinely novel configurations. Over successive iterations, this preference manifests as a contraction of explored space, even when the system is nominally designed for global optimization [13, 20]. The resulting trajectory reflects not only the data and the model but the implicit priorities encoded in the decision rule itself.

Taken together, these forms of epistemic debt do not operate in isolation. They interact across representation, sampling, modeling, and decision-making, creating a layered structure in which early assumptions propagate and reinforce one another. Understanding this interplay is essential if materials AI is to move beyond locally optimized pipelines toward systems capable of sustained, reliable discovery.

**Table 2** consolidates the typology of epistemic debt into an operational framework linking each debt type to its underlying mechanisms, observable signals, and targeted mitigation strategies.

**Table 2.** Integrated mapping of epistemic debt types, generative mechanisms, detection signals, and mitigation strategies in materials AI pipelines

Debt type	Primary generative mechanism(s)	Materials-specific manifestation	Early detection signals
Representational debt	Assumption cascades	Descriptor sets fail to encode local structure or physics	Persistent generalization failure on the subset
Sampling debt	Path dependence + Unrecognized uncertainty	Initial dataset restricted to narrow composition classes	Artificially global uncertainty with convergence
Modeling debt	Feedback loop amplification	Surrogate models are unable to capture multimodal landscapes	Structural residual patterns unexplained by regression

Decision debt	All four mechanisms combined	Acquisition functions over-prioritize exploitation	Decisional myopia
Systemic debt interaction	Cross-mechanism reinforcement	Co-occurrence of multiple debt types in long pipelines	Simultaneous diversification

The typology is not exhaustive but provides a diagnostic scaffold. Different debt types can co-occur, and their relative dominance shifts across pipeline stages. Recognizing the specific type at play enables more targeted detection and mitigation.

## Detection Principles

Detection of epistemic debt cannot rely on conventional performance metrics such as prediction accuracy or convergence speed, because these indicators often improve even as hidden constraints tighten. Instead, detection demands deliberate, structured interrogation of the pipeline’s foundational architecture at regular intervals. Five principles offer a coherent framework for surfacing debt before it becomes irreversible.

### Assumption auditing

At fixed iteration milestones, the pipeline must explicitly revisit every modeling choice made in the first 10%–20% of its run and test whether those choices remain justified by the expanded dataset. For instance, if an initial featurization scheme was selected because it correlated modestly with a handful of formation energies, later iterations should recompute correlation statistics across the entire accumulated corpus and flag any degradation. In active-learning pipelines for materials, where early descriptor selection is common, this audit prevents assumption cascades from going unnoticed [15, 20].

### Counterfactual analysis

The system should periodically simulate what would have happened had a different early decision been taken—different initial sampling distribution, alternative model class, or modified acquisition function—and compare the hypothetical trajectory with the actual one. Even

approximate counterfactuals, generated by replaying the pipeline from a branching point, reveal whether the current path has narrowed the discovery space more aggressively than necessary. Closed-loop Bayesian frameworks already contain the computational machinery for such replay; extending them to epistemic bookkeeping is a natural evolution [7, 8].

### Uncertainty tracking

Rather than simply reporting scalar uncertainty values, pipelines must maintain a longitudinal map of how uncertainty evolves across both sampled and unsampled regions of composition space. A healthy pipeline shows uncertainty decreasing in explored areas while remaining appropriately high elsewhere; a debt-laden one shows artificially suppressed global uncertainty because early choices have rendered large volumes of space invisible to the surrogate model. Persistent mismatch between local and global uncertainty trends signals unrecognized uncertainty at work [13].

### Diversity monitoring

Diversity metrics—such as the average pairwise distance in feature space or the entropy of the sampled composition distribution—must be tracked across iterations. A monotonic decline in diversity, even while model performance improves, indicates that path-dependent constraints have begun to dominate. In high-dimensional materials spaces, this signal is especially diagnostic because the combinatorial explosion makes premature narrowing detectable long before any single property prediction fails [10, 12].

### Retrospective analysis

After a discovery campaign concludes or reaches a natural checkpoint, a dedicated post-mortem traces every major outcome back to its earliest enabling decisions. This exercise identifies which early commitments were necessary and which were merely habitual, thereby quantifying the opportunity cost of the chosen path. Retrospective analyses of past active-learning studies in solid-state materials have already begun to reveal how seemingly innocuous initial screening sets shaped later discoveries; formalizing this practice as standard operating procedure would make epistemic debt visible across the community [6, 20].

Collectively, these principles transform detection from an ad-hoc intuition into a repeatable protocol. They do not require new hardware or vastly more computation; they require only that the pipeline treat its own history as an object of scientific scrutiny rather than an immutable given. When applied consistently, they convert epistemic debt from an invisible tax into a manageable design parameter.

## Mitigation Principles

Once epistemic debt is detected, mitigation must address both its immediate symptoms and its structural drivers. Six interlocking principles provide a comprehensive strategy that can be embedded directly into the design of iterative materials AI pipelines.

### Debt awareness

From the outset, every pipeline specification document and grant proposal should explicitly list epistemic debt as a primary risk category alongside accuracy, scalability, and cost. Treating debt as a design variable rather than an afterthought forces research teams to allocate resources for its management. In practice, this means including debt metrics in project dashboards and requiring quarterly debt-status reports [6].

### Assumption documentation

All early assumptions—about descriptor completeness, model expressivity, acquisition-function hyperparameters, and stopping criteria—must be logged in a machine-readable “epistemic ledger” that travels with the pipeline. Each assumption is tagged with its rationale, the evidence available at the time of adoption, and a scheduled review date. When new data arrive, the ledger automatically flags assumptions whose supporting evidence has weakened. This documentation turns tacit commitments into explicit, auditable objects [1, 2].

### Exploration reserves

A fixed fraction of the iteration budget—typically 15%–25%—should be reserved for purely exploratory moves that deliberately ignore the current acquisition function. These “debt-refinancing iterations” sample regions the model currently deems unpromising, thereby stress-testing path-dependent constraints. In materials discovery, where expensive experiments make random exploration seem

wasteful, the reserve principle restores balance between exploitation and long-term option value [7, 13].

### Ensemble diversity

Rather than committing to a single surrogate model or descriptor set, the pipeline should maintain a living ensemble of at least three parallel representations or model families. Discrepancies among ensemble members serve as early-warning signals of representational or modeling debt. Periodic voting or Bayesian model averaging then reduces the risk that any single path-dependent choice dominates downstream decisions. Ensemble methods are already common in uncertainty quantification; extending them to epistemic bookkeeping is straightforward [8, 10].

### Retrospective learning

Upon completion of a campaign, the team must conduct a formal retrospective that reconstructs the decision tree, quantifies foreclosed alternatives, and publishes the analysis alongside the primary scientific results. These retrospectives become community assets that allow subsequent pipelines to avoid repeating the same debt patterns. Over time, a public repository of epistemic post-mortems would accelerate collective learning about which early choices tend to generate the highest downstream costs [6, 20-25].

### Debt refinancing

When audits reveal excessive debt, the pipeline should support deliberate “refinancing” operations: restarting a subset of the workflow from an earlier checkpoint with revised assumptions while preserving valuable data. Although computationally expensive, refinancing is far cheaper than discarding an entire campaign that has locked itself into a suboptimal basin. In closed-loop systems, this might involve spawning a parallel branch that explores a previously pruned region using the latest accumulated knowledge [12, 20, 26-29].

These mitigation principles do not eliminate epistemic debt—iteration by its nature generates some—but they keep it within tolerable bounds and make its accumulation transparent. Implemented together, they convert iterative materials AI from a process that unconsciously mortgages future discovery potential into one that actively stewards epistemological capital. The result is not only more robust individual pipelines but a healthier collective research

ecosystem in which knowledge remains extensible across projects and generations.

## Relation to Other Failure Modes

Epistemic debt does not exist in isolation; it interacts with, amplifies, and is amplified by several other well-documented failure modes in machine learning and materials informatics. Understanding these relationships clarifies why epistemic debt deserves recognition as a distinct category rather than a mere symptom.

Epistemic debt is the primary mechanism that produces path dependence in iterative pipelines. While path dependence is often invoked loosely to describe any historical contingency, epistemic debt supplies the concrete micro-processes—assumption cascades, unrecognized uncertainty, and feedback amplification—that make early choices irreversible. A pipeline may appear path-dependent simply because later stages inherit the constraints created by earlier epistemic commitments [6, 7].

A particularly insidious interaction occurs with model collapse. In iterative self-training loops, model collapse arises when generated data increasingly resemble the training distribution, eroding diversity. Epistemic debt hastens this collapse by ensuring that the initial distribution is already artificially narrow; each feedback cycle then compounds the narrowing. The result is a pipeline that collapses not merely because of statistical homogenization but because the epistemic foundation was never broad enough to support sustained exploration [10].

Finally, epistemic debt conceals reward hacking. Acquisition functions are essentially reward models that guide the next experiment. When early modeling debt distorts the surrogate's understanding of the true objective landscape, the acquisition function begins to optimize a proxy reward that diverges from the actual material's goal. The pipeline appears to make steady progress toward higher predicted performance, while in reality, it is hacking its own misspecified objective. The debt hides the misspecification until the final validation stage, at which point the accumulated constraints make correction prohibitively expensive [13].

By distinguishing epistemic debt from these related failure modes while acknowledging their tight coupling, the

analysis provides a more granular vocabulary for diagnosing pipeline pathologies. A system suffering from overfitting may improve with regularization, but if the root cause is representational debt, regularization alone will merely mask the deeper constraint. Recognizing the epistemic layer, therefore, enables more precise and ultimately more effective interventions.

## Implications for Materials AI Practice

The recognition of epistemic debt as a distinct failure mode carries concrete implications for how materials AI research is conducted, reviewed, and disseminated.

For authors, three changes are immediately actionable. First, every study reporting an iterative discovery campaign must include an explicit section documenting the major early assumptions and the rationale for each. Second, authors should report at least one meaningful alternative path that was deliberately not taken, together with a brief counterfactual analysis of its potential consequences. Third, a post-hoc epistemic debt assessment—quantifying diversity trends, uncertainty evolution, and assumption stability—should become standard supplementary material. These practices increase transparency without imposing unrealistic burdens [6, 20].

For reviewers, epistemic debt introduces new evaluation criteria. Reviewers should ask whether the pipeline incorporated assumption auditing or exploration reserves, and whether diversity metrics were reported across iterations. When a manuscript claims “autonomous discovery” or “global optimization,” reviewers must probe whether the claim accounts for the foreclosed regions of design space. Questions about early decision points and their long-term consequences should become as routine as questions about model accuracy or data leakage [7, 8].

For the broader community, three systemic shifts are required. First, funding agencies and journals should develop standardized debt-reporting templates analogous to the existing data-availability statements. Second, community benchmarks for iterative materials AI should include retrospective epistemic analyses alongside conventional performance metrics. Third, dedicated workshops and special issues on debt mitigation strategies would accelerate the accumulation of shared best practices. Over time, these changes would embed

epistemic stewardship into the culture of materials informatics, ensuring that the accelerating pace of discovery does not come at the expense of its long-term robustness [2, 6].

## Conclusion

Epistemic debt represents a previously under-theorized yet pervasive failure mode in iterative materials AI pipelines. It arises from the accumulation of unexamined assumptions, unresolved uncertainties, and path-dependent constraints that compound across active-learning cycles and closed-loop discovery campaigns. Unlike technical debt or statistical error, epistemic debt operates at the level of scientific knowledge itself, silently limiting the future space of possible discoveries even as short-term performance metrics appear favorable. The mechanisms of assumption cascades, path-dependent constraints, unrecognized uncertainty, and feedback loop amplification interact with materials-specific vulnerabilities—small datasets, high-dimensional spaces, expensive iterations, weak theoretical guardrails, and publication bias—to make this failure mode especially acute in the materials domain.

The field of materials AI stands at a pivotal juncture. The same sequential, adaptive architectures that promise to revolutionize discovery also create the conditions for irreversible epistemological lock-in. Only by treating epistemic debt with the same seriousness now reserved for model accuracy and computational efficiency can the

community ensure that today's accelerated pipelines do not mortgage the discoveries of tomorrow. The call is therefore clear: every iterative materials AI effort must incorporate explicit epistemic housekeeping. Assumption cascades must be interrupted, path-dependent constraints must be stress-tested, and the hidden opportunity costs of early decisions must be made visible. In this way, the promise of autonomous materials research can be realized not merely faster, but more sustainably and more completely.

## Acknowledgements

None

## Conflict of interest

None

## Financial support

None

## Ethics statement

None

Received: 28 Dec 2021 Revised: 20 Mar 2022 Accepted: 20 Jun 2022  
Published online: 18 January 2023

## 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

- Huyen C. Designing machine learning systems. Sebastopol: O'Reilly Media; 2022.
- Ionescu TB, Schlund S, Schmidbauer C. Epistemic debt: A concept and measure of technical ignorance in smart manufacturing. In: International Conference on Applied Human Factors and Ergonomics; AHFE 2019. Cham: Springer; 2019. p. 81-93.
- Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. *Nature*. 2018;559(7715):547-55.
- Schmidt J, Marques MR, Botti S, Marques MA. Recent advances and applications of machine learning in solid-state

materials science. *npj Comput Mater*. 2019;5(1):83.

Zunger A. Inverse design in search of materials with target functionalities. *Nat Rev Chem*. 2018;2(4):0121.

Montoya JH, Aykol M, Anapolsky A, Gopal CB, Herring PK, Hummelshøj JS, et al. Toward autonomous materials research: Recent progress and future challenges. *Appl Phys Rev*. 2022;9(1).

Kusne AG, Yu H, Wu C, Zhang H, Hattrick-Simpers J, DeCost B, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat Commun*. 2020;11(1):5966.

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.

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.

Bassman Oftelie L, Rajak P, Kalia RK, Nakano A, Sha F, Sun J, et al. Active learning for accelerated design of layered materials. *npj Comput Mater*. 2018;4(1):74.

Balachandran PV, Young J, Lookman T, Rondinelli JM. Learning from data to design functional materials without inversion symmetry. *Nat Commun*. 2017;8(1):14282.

Yuan R, Liu Z, Balachandran PV, Xue D, Zhou Y, Ding X, et al. Accelerated discovery of large electrostrains in BaTiO<sub>3</sub>-based piezoelectrics using active learning. *Adv Mater*. 2018;30(7):1702884.

Tran K, Ulissi ZW. Active learning across intermetallics to guide discovery of electrocatalysts for CO<sub>2</sub> reduction and H<sub>2</sub> evolution. *Nat Catal*. 2018;1(9):696-703.

Pandey D, Niwaria K, Chourasia B. Machine learning algorithms: A review. *Mach Learn*. 2019;6(2):916-22.

Seko A, Togo A, Tanaka I. Descriptors for machine learning of materials data. In: *Nanoinformatics*. Singapore: Springer Singapore; 2018. p. 3-23.

Tabor DP, Roch LM, Saikin SK, Kreisbeck C, Sheberla D, Montoya JH, et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat Rev Mater*. 2018;3(5):5-20.

Murugan NA, Podobas A, Gadioli D, Vitali E, Palermo G, Markidis S. A review on parallel virtual screening softwares for high-performance computers. *Pharmaceuticals*. 2022;15(1):63.

Gay D. *Composite materials: Design and applications*. Boca Raton: CRC Press; 2022.

Lai M, Shin D, Jibril L, Mirkin CA. Combinatorial synthesis and screening of mixed halide perovskite megalibraries. *J Am Chem Soc*. 2022;144(30):13823-30.

Duong DL, Yun SJ, Lee YH. van der Waals layered materials: Opportunities and challenges. *ACS Nano*. 2017;11(12):11803-30.

Bengio Y, Lodi A, Prouvost A. Machine learning for combinatorial optimization: A methodological tour d'horizon. *Eur J Oper Res*. 2021;290(2):405-21.

Liao X, Lu R, Xia L, Liu Q, Wang H, Zhao K, et al. Density functional theory for electrocatalysis. *Energy Environ Mater*. 2022;5(1):157-85.

Stanev V, Oses C, Kusne AG, Rodriguez E, Paglione J, Curtarolo S, et al. Machine learning modeling of superconducting critical temperature. *npj Comput Mater*. 2018;4(1):29.

Raschka S, Mirjalili V. *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. Birmingham: Packt Publishing; 2019.

Vasudevan R, Pilania G, Balachandran PV. Machine learning for materials design and discovery. *J Appl Phys*. 2021;129(7).

Sun W, Bartel CJ, Arca E, Bauers SR, Matthews B, Orvañanos B, et al. A map of the inorganic ternary metal nitrides. *Nat Mater*. 2019;18(7):732-9.

Ramakrishna S, Zhang TY, Lu WC, Qian Q, Low JS, Yune JH, et al. *J Intell Manuf*. 2019;30(6):2307-26.

Wang WY, Li J, Liu W, Liu ZK. Integrated computational materials engineering for advanced materials: A brief review. *Comput Mater Sci*. 2019;158:42-8.

Liu Y, Zhao T, Ju W, Shi S. Materials discovery and design using machine learning. *J Mater*. 2017;3(3):159-77.