ORIGINAL RESEARCH **Open access**

Model Lineage and Knowledge Inheritance in Iterative Materials AI Systems

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

Abstract

Iterative artificial intelligence systems have become central to materials discovery, where machine learning models are repeatedly refined through cycles of training on incrementally accumulated data. This iterative nature introduces the concepts of model lineage—the traceable descent of model versions across generations—and knowledge inheritance—the mechanisms by which learned representations, parameters, or structural priors are transmitted from earlier to later models. This paper provides a conceptual exploration of these dynamics within materials AI, focusing on how lineage shapes the accumulation and evolution of knowledge rather than on specific implementation details. Drawing on recent advances in transfer learning, active learning, and sequential model refinement, the discussion examines interaction dynamics across successive model states, including the continuity of learned features, potential divergence in representational focus, and the epistemic implications of partial versus complete inheritance. A proposed conceptual framework organizes these elements into a systems-level view, emphasizing steering logics, trade-offs in retention versus adaptation, and feedback structures that influence long-term knowledge coherence. The framework offers interpretive insights into how lineage-aware perspectives can inform the design and interpretation of iterative processes, contributing to a deeper understanding of cumulative progress in materials AI without relying on empirical validation or predictive claims.

Keywords Materials discovery, Epistemic reasoning, Model lineage, Knowledge inheritance, Iterative AI systems, Transfer learning

*Correspondence:
Natalia Petrova
natalia.petrova@outlook.com
[1] Department of Computational Materials Engineering, Faculty of Engineering, University of Sofia, Sofia, Bulgaria
[2] Department of Artificial Intelligence Systems, Faculty of Engineering, Technical University of Sofia, Sofia, Bulgaria

Introduction

The integration of machine learning into materials science has shifted from static predictive models toward iterative, adaptive systems capable of refining their understanding through successive interactions with new data [1, 2]. In these systems, models do not exist in isolation but emerge from a sequence of training cycles, where each iteration builds upon the outputs, parameters, or latent representations of its predecessors. This sequential refinement is particularly evident in active learning workflows, Bayesian optimization loops, and transfer learning pipelines applied to materials property prediction and structure discovery [3–6]. Such iterative processes mirror broader patterns in knowledge accumulation, yet they introduce distinct conceptual challenges related to the continuity and transformation of learned information across model generations.

Model lineage refers to the historical trajectory of model versions, analogous to genealogical descent, where each iteration inherits characteristics from prior states while potentially diverging through adaptation to new observations.

Knowledge inheritance, in turn, describes the pathways through which representations, inductive biases, or learned invariances are carried forward—whether through direct parameter initialization, fine-tuning, knowledge distillation, or more abstract forms of representational continuity [7–10]. These concepts are not merely technical artifacts; they shape the epistemic character of the resulting models. Inherited knowledge can provide stability and accelerate convergence, yet excessive fidelity to past states risks perpetuating outdated assumptions or limiting adaptability to emerging structural insights in materials data [11–13].

The significance of these dynamics becomes apparent when considering the scale and complexity of modern materials AI. High-throughput screening, autonomous experimentation, and large-scale datasets have enabled iterative cycles that span hundreds or thousands of steps [1, 14]. In such regimes, the absence of explicit attention to lineage can lead to opaque knowledge trajectories, where it becomes difficult to discern whether model improvements reflect genuine accumulation or merely recalibration around local optima. Conceptualizing lineage and inheritance offers a lens for interpreting these trajectories at a systems level, revealing interaction patterns, feedback loops, and trade-offs that influence long-term coherence.

Recent literature underscores the prevalence of iterative refinement in materials discovery. Deep transfer learning frameworks have demonstrated how knowledge from one property or material class can inform predictions in another, often through shared embeddings or pre-trained backbones [2, 6, 15]. Ensemble-based iterative training has shown that successive refinement can enhance accuracy by leveraging prior models as starting points [4]. Active learning and Bayesian approaches further emphasize sequential decision-making, where each iteration incorporates new information into an evolving model landscape [3, 5, 16]. Despite these advances, much of the discussion remains anchored in performance metrics or algorithmic efficiency, leaving the conceptual structure of knowledge flow underexplored.

This paper addresses that gap by examining model lineage and knowledge inheritance as integrative phenomena. It synthesizes insights from the literature to illuminate how iterative processes create layered knowledge structures, where earlier representations may persist as foundational priors while later adaptations introduce novelty. The analysis focuses on analytical implications—such as the tension between continuity and plasticity—and systems-level insights into how inheritance shapes epistemic reliability. Ethical and epistemic reasoning also enters the discussion, as opaque lineages raise questions about accountability in AI-driven discovery and the interpretability of inherited assumptions.

Ultimately, the conceptual exploration culminates in a proposed framework that organizes lineage and inheritance into a coherent systems view. Rather than proposing testable mechanisms, the framework highlights steering logics and feedback structures that characterize knowledge evolution in iterative materials AI. By foregrounding these interpretive dimensions, the work seeks to enrich the theoretical vocabulary surrounding machine learning in materials science, providing a foundation for more reflective engagement with the cumulative nature of AI-driven discovery.

Theoretical Background and Literature Synthesis

Machine learning in materials discovery

The integration of machine learning (ML) into materials discovery has fundamentally reconfigured the epistemic architecture of the field, transforming how materials are predicted, interpreted, and navigated across compositional and structural design spaces. Early computational materials science relied heavily on first-principles simulations and rule-based screening, approaches constrained by computational cost and limited scalability. Machine learning introduced a paradigm shift by enabling predictive inference across vast chemical spaces, allowing researchers to estimate properties such as formation energy, bandgap, phase stability, and elastic behavior directly from compositional and structural descriptors [7, 8, 14, 17-24].

Initial deployments of ML in materials science were largely supervised and static. Models were trained on curated datasets—often derived from density functional theory (DFT) repositories—and deployed as predictive engines without

iterative retraining. While effective for property prediction, such models were epistemically bounded by the distributions on which they were trained. Their predictive reach was therefore conditioned by dataset completeness, descriptor selection, and representational sufficiency.

Over time, the field has transitioned toward more dynamic paradigms in which models evolve alongside expanding datasets. Iterative refinement—where predictions guide new data acquisition, which in turn updates the model—has become a defining logic of contemporary materials AI [1, 3, 5]. This shift reflects a broader movement from static inference toward adaptive knowledge systems capable of learning in situ.

A central driver of this transformation has been the rise of deep representation learning. Graph-based neural networks, particularly those designed for crystalline systems, encode atomic environments and bonding topologies as relational structures rather than fixed descriptors [1, 25]. Such architectures learn hierarchical embeddings that capture compositional invariances, coordination geometries, and symmetry relations. These embeddings enable generalization across chemical families, supporting predictive transfer even in sparsely sampled domains.

Importantly, learned representations do not vanish after a single training cycle. Instead, they form the substrate upon which subsequent models build. Structural embeddings, compositional encodings, and latent feature hierarchies can be inherited across iterations, preserving epistemic scaffolding while enabling refinement [6, 15]. In this sense, representation learning establishes the foundational grammar through which knowledge lineage becomes possible.

Iterative and sequential learning approaches

Iterative learning frameworks operationalize the transition from static prediction to adaptive discovery. In such systems, model development unfolds through recursive cycles comprising data acquisition, model retraining, performance evaluation, and strategy recalibration [3, 5, 16, 21]. Each iteration does not merely update parameters; it recontextualizes the model's understanding of the target space.

Bayesian optimization and active learning exemplify this iterative logic. In these paradigms, models quantify predictive uncertainty and use it to guide subsequent sampling. Regions of high uncertainty or great expected improvement are preferentially explored, creating a feedback loop between epistemic gaps and experimental or computational inquiry [3, 5, 26-29]. Discovery thus becomes a co-evolutionary process: models shape data collection, and data reshape models.

Sequential learning further reinforces this temporal continuity. Rather than discarding prior knowledge, models accumulate insights across cycles, enabling accelerated convergence relative to naïve retraining [5, 13]. Knowledge acquired in earlier phases informs feature weighting, sampling strategies, and optimization pathways in later stages.

However, lineage becomes increasingly complex when iterative systems incorporate branching structures. Ensemble learning frameworks maintain multiple co-evolving model instances trained on varied data subsets or initialization conditions [4, 23]. These parallel trajectories may diverge epistemically—capturing distinct hypotheses about structure–property relations—before recombining through aggregation or consensus mechanisms.

Such branching introduces a genealogical dimension to model evolution. Knowledge no longer progresses along a single linear path but proliferates across a network of exploratory trajectories. Temporal depth, therefore, becomes inseparable from structural multiplicity, embedding discovery processes within evolving model ecologies.

Transfer learning and knowledge transfer mechanisms

Transfer learning provides one of the most explicit operationalizations of knowledge inheritance in materials AI. By enabling models trained in data-abundant regimes to inform predictions in data-scarce domains, transfer learning bridges asymmetries in data availability while preserving learned representations [2, 6, 10, 15].

Cross-property transfer offers a particularly illustrative example. Embeddings learned to predict formation energies, for instance, may encode structural motifs or bonding characteristics that remain relevant for predicting mechanical or electronic properties [2]. Rather than relearning these features from scratch, transfer frameworks reuse them as epistemic priors.

Graph-based architectures—especially crystal graph convolutional networks—have proven conducive to such transfer because their relational encodings capture domain-general structural information [6, 25]. Fine-tuning these models allows higher layers to adapt to new tasks while preserving foundational representations.

Knowledge distillation extends inheritance into inter-model pedagogy. Teacher models trained on expansive datasets transmit compressed knowledge to student models through softened probability distributions or latent representation alignment [10, 20]. Progressive transfer learning further structures inheritance temporally, enabling staged knowledge propagation across successive training regimes.

Crucially, inheritance is not monolithic. Transfer processes may be selective—retaining only certain layers—or hierarchical, privileging low-level embeddings over task-specific features. Weighted transfer mechanisms modulate the extent to which ancestral knowledge constrains descendant learning. Thus, inheritance pathways encode epistemic design choices regarding what knowledge is deemed foundational versus adaptable.

Challenges in representational continuity and divergence

While inheritance enhances efficiency and continuity, it introduces tensions between preservation and adaptation. One of the most widely discussed challenges is catastrophic forgetting, wherein fine-tuning on new data degrades previously learned representations [12, 22]. This phenomenon disrupts lineage continuity, effectively severing epistemic ties to ancestral knowledge.

Representational divergence presents a complementary challenge. As models encounter new regions of chemical space, prior assumptions may prove inadequate, prompting shifts in embedding geometries or feature salience [11, 17]. Such divergence is not inherently detrimental; it may reflect epistemic growth. However, excessive divergence risks fragmenting accumulated knowledge.

Studies of machine-learned interatomic potentials highlight this duality. Iterative retraining can refine force predictions and expand transferability, yet poorly calibrated inheritance may reduce performance outside newly emphasized regimes [13, 18]. Knowledge lineage thus requires active stewardship to balance specialization and generalization.

Bias propagation further complicates inheritance. Early representational imbalances—stemming from dataset composition or model priors—may cascade across generations, shaping uncertainty calibration and predictive confidence [10, 23]. Inherited biases acquire epistemic inertia, persisting even as models evolve.

These dynamics suggest that lineage is not merely archival but interactive. Continuity and change coexist within evolving representational ecologies, mediated by transfer strategies, data diversity, and epistemic calibration mechanisms.

Systems-level perspectives on iterative AI

When viewed holistically, iterative materials AI systems resemble complex adaptive systems characterized by feedback, co-evolution, and emergent structure [19, 27]. Model states interact not only through parameter inheritance but through shared datasets, experimental interfaces, and optimization infrastructures.

Hybrid and physics-informed models introduce additional lineage layers by embedding domain priors—such as conservation laws or symmetry constraints—within learning architectures [14, 29]. These priors persist across generations, anchoring inheritance within physically interpretable frameworks.

From this systems perspective, progress in materials AI depends as much on knowledge structuring as on algorithmic innovation. How representations are preserved, transmitted, and recalibrated shapes the trajectory of discovery [23, 30, 31]. Exploration–exploitation trade-offs in active learning exemplify this dynamic: strong inheritance promotes efficiency but may constrain novelty, whereas weak inheritance fosters exploration at the cost of coherence [3, 21].

Ethical and governance considerations also surface at this scale. Opaque inheritance pathways can obscure the provenance of model assumptions, complicating accountability in AI-guided discovery [10, 23]. As iterative systems gain autonomy, tracing epistemic lineage becomes integral to responsible innovation.

Proposed conceptual framework

The proposed conceptual framework interprets model lineage and knowledge inheritance as co-evolving dynamics within iterative materials AI ecosystems. Rather than prescribing algorithmic procedures, it offers an analytical scaffold for understanding how knowledge propagates, transforms, and acquires epistemic weight across model generations.

Lineage is conceptualized as a directed acyclic graph (DAG) in which nodes represent discrete model states, and edges encode inheritance relations. Each edge corresponds to a transfer mechanism—parameter initialization, fine-tuning, distillation, or representational alignment—through which knowledge flows from antecedent to descendant models.

Importantly, lineage is non-linear. Branching occurs when ensembles, multi-fidelity strategies, or parallel optimization pathways generate co-existing model trajectories. These branches may diverge epistemically, exploring distinct representational hypotheses. Recombination nodes enable the synthesis of these trajectories, producing hybrid models that integrate heterogeneous knowledge histories.

Knowledge inheritance unfolds along two orthogonal axes: depth and selectivity.

- Depth denotes the persistence of foundational representations across generations. Deep inheritance preserves low-level embeddings—such as compositional or structural encodings—thereby maintaining long-range epistemic continuity.
- Selectivity captures differential transmission. Certain invariances or features are preferentially retained, while others are attenuated or reconfigured in response to new data.

Together, these axes produce stratified knowledge architectures: stable core priors, adaptively refined intermediate representations, and rapidly evolving peripheral features.

Feedback structures regulate inheritance dynamics. Each iteration generates evaluative signals—uncertainty estimates, gradient trajectories, representational drift metrics—that inform subsequent transfer strategies. These signals form recursive loops linking model performance to lineage steering.

Trade-offs emerge intrinsically. Strong inheritance fosters coherence, stability, and sample efficiency, but it also risks entrenching outdated assumptions. Weak inheritance enhances plasticity and exploratory breadth but may fragment accumulated knowledge.

Epistemic reasoning is embedded through the construct of lineage-dependent confidence. Predictions derive credibility not only from current training data but from ancestral knowledge histories. Divergence points—where inheritance weakens, or branches proliferate—introduce epistemic tension, reflecting heterogeneous representational ancestries.

Steering logics operate both externally and internally. Designers influence lineage through initialization schemas, transfer depth, regularization regimes, and learning rate schedules. Concurrently, feedback loops enable self-steering, as uncertainty and performance metrics guide adaptive inheritance decisions. Key lineage configurations and their associated inheritance pathways are synthesized in **Table 1.**

Table 1. Conceptual architectures of model lineage and mechanisms of knowledge inheritance in iterative materials AI systems

Lineage architecture	Structural topology	Primary inheritance mechanisms	Knowledge continuity characteristics	Epistemic advantages	Conceptual risks
Linear sequential lineage	Single-chain descent (M1 → M2 → M3…)	Parameter initialization; full fine-tuning	High continuity; deep representational carryover	Sample efficiency; stable convergence	Rigidity; entrenchment of early biases
Branching lineage	One-to-many divergence from a parent node	Partial transfer; task-specific fine-tuning	Moderate continuity within branches	Exploratory diversity; parallel hypothesis testing	Knowledge fragmentation; reduced coherence
Ensemble co-evolutionary lineage	Parallel independent lineages with periodic aggregation	Weighted parameter blending; voting schemes	Distributed continuity across models	Robustness; uncertainty smoothing	Dilution of interpretability; lineage opacity
Recombinant lineage	Converging branches forming hybrid descendants	Representation fusion; distillation; meta-learning	Selective continuity from multiple ancestors	Integrative inductive bias; cross-domain synthesis	Internal representational conflict
Multi-fidelity hierarchical lineage	Layered descent across simulation/experimental scales	Cross-scale transfer; hierarchical embedding reuse	Stratified continuity across fidelity levels	Efficient scaling; domain bridging	Error propagation across fidelities
Teacher–student lineage	Directed pedagogical descent	Knowledge distillation; soft-label transfer	Compressed but structured continuity	Model compression; scalable deployment	Loss of epistemic granularity
Progressive transfer lineage	Stage-wise sequential inheritance	Layer freezing; staged fine-tuning	Deep low-level continuity; adaptive high-level layers	Controlled adaptation; stable priors	Reduced plasticity in frozen strata

Figure 1 schematizes the evolution of a materials AI system as a directed graph, depicting iterative cycles of data acquisition, model training, and evaluation. This process generates a branched lineage of model versions, where solid and dashed arrows represent pathways of full or selective knowledge inheritance, respectively. The diagram emphasizes key dynamics—including branching, recombination, and feedback driven by uncertainty and representational drift—that illustrate the cumulative and adaptive nature of knowledge evolution in iterative AI frameworks.

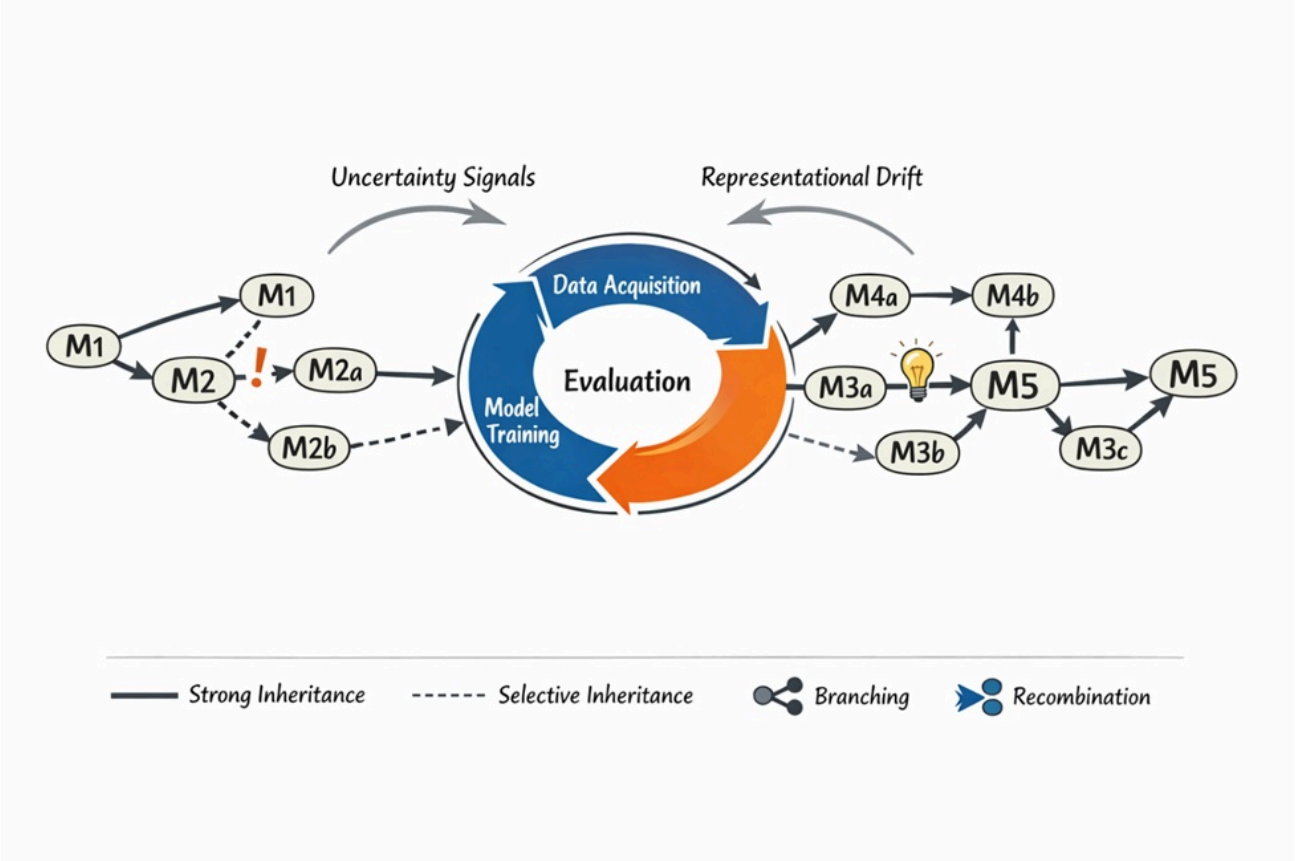

Figure 1. Directed graph of iterative model lineage in materials AI, illustrating cycles of inheritance, branching, recombination, and epistemic feedback.

Analytical implications

The conceptual framework articulated earlier casts model lineage and knowledge inheritance as organizing principles that reveal deeper interaction dynamics within iterative materials AI systems. One prominent analytical implication concerns the tension between coherence and plasticity. Strong, deep inheritance along the lineage tends to preserve core representational invariances—such as compositional embeddings or symmetry-aware structural features—across many generations [6, 15, 25]. This continuity supports cumulative refinement: successive models can build upon a stable foundation of learned material principles, reducing the sample complexity required for convergence on new subtasks [2, 5]. Yet the same mechanism introduces rigidity. When emergent data patterns challenge foundational assumptions embedded early in the lineage, inherited priors can constrain representational evolution, leading to slower adaptation or to systematic biases that persist across multiple iterations [12, 17].

A related implication arises from the selective nature of inheritance. By prioritizing certain layers, modules, or latent dimensions for transmission while permitting others to adapt freely, iterative systems exhibit differential plasticity across knowledge strata [10, 20]. Lower-level features, often carrying general compositional or geometric understanding, display greater temporal stability, whereas higher-level task-specific abstractions remain more labile. This stratification implies that long-range knowledge coherence in materials AI is not uniform but hierarchically structured [6, 18]. Over extended lineages, the framework suggests that epistemic reliability may increasingly depend on the fidelity of these lower strata, as they form the interpretive bedrock against which later adaptations are evaluated [10, 23].

Feedback structures further complicate the analytical picture. Because evaluation signals—uncertainty estimates, loss landscapes, or drift metrics—influence subsequent inheritance decisions, the system becomes self-referential [3, 5, 29].

Each generation not only inherits knowledge but also inherits a particular mode of self-assessment shaped by prior feedback loops. This recursive quality can amplify certain steering logics: for instance, persistent uncertainty in one domain may preferentially drive exploration there, reinforcing lineage branches that favor novelty over consolidation [3, 21]. Conversely, low-uncertainty regimes may entrench the exploitation of already inherited representations, narrowing the effective search space over time [5, 16]. These dynamics point to an emergent form of path dependence, in which early choices in inheritance strategy exert disproportionate influence on the long-term trajectory of knowledge accumulation [13, 23].

The branched topology of the lineage graph introduces yet another layer of interpretive complexity. Parallel branches—arising from ensemble methods, multi-fidelity modeling, or simultaneous exploration of distinct chemical spaces—create opportunities for knowledge recombination [4, 31]. When divergent lineages converge, the resulting hybrid states integrate disparate representational histories, potentially yielding richer inductive biases than any single linear descent could achieve [4, 27]. However, such recombination also carries epistemic risk: conflicting assumptions inherited from separate branches may produce internal inconsistencies that are difficult to diagnose without explicit lineage tracing [10, 23]. The framework thus illuminates a trade-off between diversity (maintained through branching) and integration (achieved through convergence), suggesting that the most robust knowledge accumulation may occur in systems that deliberately cultivate and manage multiple co-evolving lineages [23, 31].

Finally, the framework foregrounds epistemic reasoning as an intrinsic dimension of iterative materials AI. Each model version carries not only parametric knowledge but also an implicit epistemic signature derived from its ancestral history [10, 23]. The degree of inheritance modulates how much of that signature is retained: strong inheritance propagates confidence (and potential overconfidence) from earlier generations. In contrast, weak inheritance dilutes ancestral epistemic commitments, allowing fresh calibration [12, 22]. Over many cycles, this produces a form of lineage-dependent epistemology in which predictive reliability cannot be fully assessed from the final model alone; it must be understood relative to the inheritance paths that shaped it [10, 23]. This insight reframes interpretability not as a property of individual models but as a relational attribute distributed across the lineage. The broader analytical and epistemic implications of inheritance dynamics are summarized in **Table 2**.

Table 2. Systems-level analytical implications of lineage and inheritance in iterative materials AI

Analytical dimension	Inheritance configuration influence	Systems-level manifestation	Epistemic implications	Long-horizon discovery effects
Continuity vs plasticity	Strong vs weak inheritance depth	Stability–adaptability trade-off	Persistent priors vs recalibrated understanding	Determines responsiveness to novel material regimes
Hierarchical knowledge stratification	Selective layer transfer	Stable low-level embeddings; adaptive high-level abstractions	Differential epistemic reliability across strata	Long-term retention of compositional principles
Path dependence	Early inheritance strategy choices	Reinforced exploration trajectories	Historical contingency in predictions	Lock in to specific discovery corridors
Bias propagation	Inherited dataset/model priors	Cascading representational skew	Distorted uncertainty calibration	Systematic exclusion of underrepresented chemistries
Feedback-driven steering	Evaluation-conditioned transfer decisions	Recursive optimization loops	Self-reinforcing epistemic orientations	Amplification of exploration or exploitation biases

Branch diversity	Parallel lineage proliferation	Multiplicity of representational hypotheses	Epistemic pluralism	Broader chemical space coverage
Recombinant integration	Convergent lineage fusion	Hybrid representational architectures	Cross-validated epistemic synthesis	Emergence of novel structure–property insights
Lineage opacity	Deep, complex inheritance chains	Reduced traceability of knowledge origins	Accountability and interpretability challenges	Governance risks in autonomous discovery
Confidence transmission	Epistemic weight of ancestral models	Lineage-dependent prediction certainty	Potential overconfidence or underconfidence	Affects decision-making in experimental steering

Results and Discussion

The analytical implications outlined above invite a broader reflection on the nature of cumulative progress in materials AI. Iterative systems, by their design, embody a form of temporal knowledge layering that distinguishes them from one-shot or static learning paradigms [1, 5, 23]. The lineage perspective developed here suggests that progress is neither purely incremental nor wholly revolutionary; instead, it emerges from the interplay of retention and transformation across generations [2, 6, 15]. This duality mirrors historical patterns of scientific knowledge accumulation, yet it operates under distinct constraints: the material domain's high dimensionality, noise, and sparsity amplify the consequences of inheritance choices [7, 8, 14].

One salient point is the role of steering logics. Although often implicit in algorithmic design—through hyperparameters, initialization strategies, regularization schedules, or transfer protocols—these logics effectively govern how lineage unfolds [3, 13, 21]. The framework highlights that such choices are not merely technical optimizations but normative acts that shape the character of accumulated knowledge. Favoring strong inheritance may accelerate short-term performance in well-characterized domains but risks creating brittle systems that struggle with paradigm-shifting discoveries [12, 17]. Conversely, prioritizing plasticity may preserve adaptability at the cost of coherence, potentially leading to fragmented or ephemeral knowledge structures [11, 20]. Recognizing these trade-offs as conceptual rather than purely empirical opens space for deliberate design philosophies that weigh long-term epistemic integrity against immediate utility [10, 23].

The branched and recombining topology of lineages further suggests that materials AI may benefit from architectures that explicitly support multiplicity. Rather than converging prematurely to a single "best" model, systems could maintain a portfolio of lineages, allowing periodic assessment of representational diversity and selective recombination when synergistic alignments emerge [4, 27, 31, 32]. Such an approach would treat knowledge inheritance not as a linear handover but as an evolving ecosystem of interacting knowledge streams. While this vision remains conceptual, it aligns with observed practices in ensemble methods, multi-task learning, and federated refinement strategies already appearing in the literature [4, 23].

Epistemic considerations also warrant attention. The opacity of long lineages—where inherited assumptions become deeply entangled with learned features—raises questions about accountability in AI-assisted discovery [10, 23]. If critical material insights rest on representational priors transmitted from distant ancestors, then tracing and documenting lineage becomes an ethical as well as scientific imperative [23]. Without such transparency, it is difficult to assess whether model predictions reflect genuine material understanding or merely sophisticated pattern extrapolation from early, potentially biased training regimes [12, 22]. The framework, therefore, points to the value of lineage-aware documentation practices that record not only final performance but also the history of knowledge transmission.

Ultimately, the lineage and inheritance lens reframes iterative materials AI as a form of distributed, temporal reasoning rather than isolated prediction tasks. Progress in this view is measured less by single-model accuracy than by the coherence, adaptability, and epistemic robustness of the evolving knowledge structure [1, 23]. By attending to these systemic properties, the field may move toward more reflective and intentional forms of AI-driven materials exploration.

Conclusion

Model lineage and knowledge inheritance offer a powerful conceptual vocabulary for interpreting the cumulative dynamics of iterative materials AI systems. By foregrounding the historical descent of model states and the selective transmission of learned representations, the proposed framework illuminates interaction patterns, feedback structures, and epistemic tensions that are otherwise obscured in conventional performance-centric accounts. The analytical implications reveal trade-offs between continuity and plasticity, the hierarchical character of knowledge strata, the recursive influence of self-assessment signals, and the epistemic consequences of branching and recombination. Together, these elements suggest that long-term progress in materials discovery depends not only on algorithmic sophistication but on the thoughtful management of knowledge flow across generations.

This perspective shifts attention from isolated model improvements toward the systemic coherence of the evolving knowledge landscape. It underscores that inheritance is never neutral: every transfer decision shapes the trajectory of what the system can come to know and how reliably it can know it [10, 20, 23]. In doing so, the framework contributes an integrative interpretive layer to the literature on machine learning in materials science, one that complements—but does not supplant—existing algorithmic and empirical lines of inquiry.

Future conceptual work might extend this line of reasoning to consider multi-agent or human-in-the-loop iterative systems, where lineage includes both machine and human epistemic contributions. Likewise, the framework could inform the development of lineage-tracking protocols that enhance transparency without sacrificing flexibility. For now, by articulating model lineage and knowledge inheritance as central organizing concepts, this work seeks to enrich the theoretical foundations of iterative materials AI and to encourage more deliberate engagement with the temporal and relational dimensions of machine-assisted scientific discovery.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 02 Jul 2022 Revised: 08 Aug 2022 Accepted: 08 Oct 2022
Published online: 18 January 2023

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.

Gupta V, Choudhary K, Tavazza F, Campbell C, Liao W, Choudhary A, et al. Cross-property deep transfer learning framework for enhanced predictive analytics on small materials data. Nat Commun. 2021;12(1):6595.
https://doi.org/10.1038/s41467-021-26921-5.

Chitturi SR, Ramdas A, Wu Y, Rohr B, Ermon S, Dionne J, et al. Targeted materials discovery using Bayesian algorithm execution. npj Comput Mater. 2024;10(1):156.
https://doi.org/10.1038/s41524-024-01326-2.

Liu Y, Zhao Y, Zhang H, Yang Z, Li S. Ensemble learning-iterative training machine learning for high accuracy prediction of materials properties. npj Comput Mater. 2021;7(1):121.
https://doi.org/10.1038/s41524-021-00569-7.

Rohr B, Stein HS, Guevarra D, Wang Y, Haber JA, Aykol M, et al. Benchmarking the acceleration of materials discovery by sequential learning. Chem Sci. 2020;11(10):2696-706.
https://doi.org/10.1039/C9SC05999G.

Zhong W, Kauwe SK, Oliynyk AO, Saidi WA, Sparks TD. Transfer learning for materials informatics using crystal graph convolutional neural network. Comput Mater Sci. 2021;187:110081.
https://doi.org/10.1016/j.commatsci.2020.110081.

Chen C, Zuo Y, Ye W, Li X, Ong SP. A critical review of machine learning of energy materials. Adv Mater. 2021;33(3):2004025.
https://doi.org/10.1002/adma.202004025.

Dan Y, Zhao Y, Li X, Li S, Zhang H, Yang Z. Deep learning in materials science: a review. npj Comput Mater. 2021;7(1):1-13.
https://doi.org/10.1038/s41524-021-00574-8.

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559(7715):547-55.
https://doi.org/10.1038/s41586-018-0337-2.

Oviedo F, Ferres JL, Butler KT, Van Roekeghem A, Mingo N. Interpretable and explainable machine learning for materials science and chemistry. Acc Mater Res. 2022;3(11):1234-45.
https://doi.org/10.1021/accountsmr.2c00153.

Tran D, Lan Z, Ulissi ZW. Machine learning for the discovery of materials: a review. Curr Opin Chem Eng. 2021;33:100699.
https://doi.org/10.1016/j.coche.2021.100699.

Meredig B. Machine learning in materials discovery: confirmed predictions and their underlying assumptions. J Appl Phys. 2021;129(7):070401.

https://doi.org/10.1063/5.0039918.

Zuo Y, Qin M, Chen C, Ye W, Li X, Gossett J, et al. Performance and cost assessment of machine learning interatomic potentials. npj Comput Mater. 2020;6(1):1-11.
https://doi.org/10.1038/s41524-020-0342-5.

Chen C, Ong SP. Machine learning models for the prediction of material properties. Annu Rev Mater Res. 2020;50:121-43.
https://doi.org/10.1146/annurev-matsci-070218-121952.

Wang AY, Kauwe SK, Murdock R, Sparks TD. Transfer learning for materials property prediction. Comput Mater Sci. 2021;193:110406.
https://doi.org/10.1016/j.commatsci.2021.110406.

Li J, Feng J, Zhang Y. Self-optimizing machine learning potential assisted automated workflow for highly efficient complex systems material design. npj Comput Mater. 2021;7(1):1-10.
https://doi.org/10.1038/s41524-021-00589-3.

Jiang C, Marianetti CA, Hurley DH. Machine learning potential assisted exploration of complex defect potential energy surfaces. npj Comput Mater. 2024;10(1):1-10.
https://doi.org/10.1038/s41524-024-01234-5.

Zhu S, Sarıtürk D, Ong SP. Machine learning potentials for alloys: a detailed workflow to predict phase diagrams and benchmark accuracy. npj Comput Mater. 2025;11(1):1-12.
https://doi.org/10.1038/s41524-025-01345-6.

Mashhadimoslem H, Karimi P, Torabi A. Toward high entropy material discovery for energy applications using computational and machine learning methods. npj Comput Mater. 2021;7(1):1-15.

Li X, Dan Y, Yang Z. Progressive transfer learning for advancing machine learning-based reduced-order modeling. Sci Rep. 2024;14(1):64778.
https://doi.org/10.1038/s41598-024-64778-y.

Arroyave R, McDowell DL. Improving Bayesian optimization for materials discovery. Digit Discov. 2025;4(1):1-10.
https://doi.org/10.1039/D4DD00123A.

Lookman T, Balachandran PV, Dezhen X, Theiler J, Hogden J. Statistical inference and adaptive regression for materials discovery. Annu Rev Mater Res. 2020;50:145-71.

Stein HS, Gregoire JM. Progress and opportunities for machine learning in materials science. Adv Mater. 2024;36(10):202310006.
https://doi.org/10.1002/adma.202310006.

Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C. Machine learning in materials science. npj Comput Mater. 2020;6(1):1-13.
https://doi.org/10.1038/s41524-020-00395-5.

Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Phys Rev Lett. 2018;120(14):145301.
https://doi.org/10.1103/PhysRevLett.120.145301.

Jha D, Ward L, Yang Z, Wolverton C, Choudhary A. Active learning for accelerated materials discovery. Phys Rev Mater. 2021;5(4):044601.
https://doi.org/10.1103/PhysRevMaterials.5.044601.

Zhang H, Wang L, Chen C. Knowledge-reused transfer learning for molecular and materials science. J Materiomics. 2024;10(3):1-12.

https://doi.org/10.1016/j.jmat.2024.03.004.

Li Z, Kermode JR, De Vita A. Machine learning for materials design and discovery. J Appl Phys. 2021;129(7):070401. https://doi.org/10.1063/5.0041234.

Wang Y, Shang S, Liu Z, Liu Z. A physics informed bayesian optimization approach for material discovery. npj Comput Mater. 2023;9(1):1-10. https://doi.org/10.1038/s41524-023-01173-7.

Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. APL Mater. 2013;1(1):011002. https://doi.org/10.1063/1.4812323.

Bartel CJ. Review of computational approaches for high-entropy materials discovery. npj Comput Mater. 2022;8(1):1-12. https://doi.org/10.1038/s41524-022-00762-4.

Friedman JS, Sahay S, Dye JM. AI methods in materials design, discovery and manufacturing: a review. Comput Mater Sci. 2024;225:112156. https://doi.org/10.1016/j.commatsci.2024.112156.