ORIGINAL RESEARCH **Open access**

The Cost of Early Convergence in AI-Guided Materials Search

Lucas Meyer[1*], Anna Schmid[2], Stefan Braun[1]

Abstract

AI-guided materials search increasingly relies on probabilistic and adaptive algorithms to navigate complex design spaces. Within these processes, early convergence emerges as a recurring dynamic wherein search trajectories stabilize around promising regions before exhaustive mapping of uncertainty landscapes occurs. This conceptual manuscript examines the interpretive costs associated with such premature stabilization, framing them not as isolated computational inefficiencies but as interconnected epistemic, structural, and innovation-limiting phenomena. Drawing on recent developments in Bayesian optimization, active learning, and equivariant graph representations for materials systems, the analysis examines how early convergence interacts with exploration-exploitation trade-offs, cascading into effects on knowledge breadth and discovery potential. A novel conceptual framework is advanced that conceptualizes these dynamics through feedback loops and trade-off structures, emphasizing systems-level insights into how algorithmic steering logics shape long-term trajectories in materials innovation. By focusing exclusively on interpretive and integrative dimensions, the contribution highlights the need for refined conceptual models that account for hidden costs embedded in convergence behaviors. This perspective encourages deeper reflection on the epistemic foundations of AI-assisted discovery without invoking empirical validation or predictive assertions.

Keywords Active learning, Bayesian optimization, AI-guided materials discovery, Early convergence, Exploration-exploitation dynamics, Epistemic costs

*Correspondence:
Lucas Meyer
lucas.meyer@gmail.com
[1] Department of Materials Modeling and Artificial Intelligence, Faculty of Engineering, ETH Zurich, Zurich, Switzerland
[2] Department of Data-Driven Materials Science, Faculty of Engineering, University of Bern, Bern, Switzerland

Introduction

The integration of artificial intelligence into materials search has transformed the conceptual landscape of discovery, shifting emphasis from exhaustive enumeration toward guided navigation of high-dimensional parameter spaces. Algorithms employing probabilistic inference and iterative refinement now routinely direct experimental or computational attention toward regions of elevated promise. Yet this guidance introduces subtle steering logics whose implications extend beyond immediate efficiency gains. One such implication centers on the tendency toward early convergence, a phenomenon in which search processes rapidly narrow their focus, stabilizing around localized optima or high-probability configurations at the expense of broader uncertainty characterization.

Interpretively, early convergence can be understood as an emergent property arising from the interplay between uncertainty quantification mechanisms and adaptive sampling strategies. In materials contexts characterized by vast compositional and structural variability, this dynamic risks truncating the epistemic horizon, constraining the diversity of configurations considered, and thereby influencing the conceptual scope of attainable insights. The costs incurred are multifaceted: they encompass not only resource expenditures but also limitations on systemic understanding of structure-property relationships and reduced capacity for serendipitous advances.

Recent scholarly contributions have illuminated various facets of these guided approaches. Works exploring scaling behaviors in deep learning architectures for property prediction underscore the computational demands of maintaining expansive search scopes [1]. Similarly, investigations into closed-loop discovery via Bayesian active learning reveal that probabilistic models balance acquisition functions to prioritize informative samples, yet often exhibit sensitivity to initial conditions, leading to swift stabilization [2]. Targeted execution frameworks further illustrate how algorithm selection itself modulates convergence trajectories, embedding preferences for certain interaction patterns over others [3].

Conceptually, these developments invite scrutiny of the underlying trade-off structures. Exploitation-oriented strategies accelerate identification of viable candidates but may embed feedback that reinforces narrow subspaces, while exploration mechanisms, though costlier in the short term, sustain openness to alternative pathways. The interpretive challenge lies in recognizing how early convergence propagates through these structures, generating higher-order effects on innovation ecosystems. For instance, when convergence occurs before adequate mapping of uncertainty manifolds, subsequent refinement steps operate within artificially constrained domains, potentially diminishing the richness of the generated knowledge.

This manuscript adopts an exclusively conceptual stance, eschewing any empirical apparatus to focus instead on analytical implications and integrative interpretations. The objective is to articulate the systemic costs of early convergence as they manifest in interaction dynamics between algorithmic components and materials design spaces. By synthesizing insights from probabilistic optimization, graph-based representations, and adaptive learning paradigms, the analysis seeks to illuminate steering logics that govern convergence behaviors. Particular attention is paid to epistemic dimensions, in which premature stabilization may curtail the development of comprehensive conceptual models of material behavior.

The ensuing sections first synthesize theoretical foundations drawn from contemporary literature, delineating key conceptual threads in AI-guided search. A proposed framework then integrates these threads into a cohesive interpretive model, emphasizing feedback architectures and trade-off cascades. Through this lens, early convergence is repositioned not as a mere operational artifact but as a pivotal determinant of long-term discovery trajectories, inviting nuanced reflection on the epistemic responsibilities inherent in deploying such systems.

Theoretical Background and Literature Synthesis

Evolution of AI-guided search strategies

Contemporary AI-guided materials search builds upon probabilistic frameworks that iteratively refine understanding of complex landscapes. Bayesian optimization, for example, conceptualizes search as an adaptive process wherein surrogate models encode uncertainty to inform sampling decisions [4, 5]. These models interact dynamically with acquisition functions that encode preferences for balancing known high-value regions against unexplored domains. Interpretively, this balance reveals inherent tensions: strategies favoring rapid exploitation can engender early convergence by amplifying confidence signals from limited data clusters, thereby shaping subsequent iterations within bounded subspaces [6].

Active learning paradigms extend this logic by incorporating mechanisms for targeted data acquisition, often prioritizing samples that maximally reduce predictive variance [7, 8]. In materials applications, such paradigms have been interpreted as fostering curiosity-driven exploration, yet their convergence characteristics depend sensitively on how uncertainty is represented and propagated. When initial queries yield coherent but incomplete signals, feedback structures may reinforce convergence toward those signals, limiting exposure to outlier configurations that could reshape conceptual understanding [9].

Bayesian frameworks and uncertainty dynamics in materials contexts

Bayesian approaches in materials discovery often integrate entropy-based or expected-improvement criteria to guide search [10, 11]. Analytically, these criteria embody steering logics that privilege regions of high model confidence or potential gain. However, when such regions stabilize prematurely, the resulting dynamics constrain the epistemic breadth of the search process. The literature on multi-objective variants highlights that competing targets introduce additional layers of interaction, in which convergence along one objective may cascade restrictions onto others [12]. Systems-level insights emerge when considering how these interactions scale across compositional spaces, revealing patterns wherein early stabilization correlates with diminished capacity for holistic property optimization.

Uncertainty-aware extensions further nuance these dynamics by embedding probabilistic envelopes around predictions [13, 14]. Conceptually, these envelopes function as regulatory structures that modulate convergence thresholds. Yet their effectiveness hinges on the fidelity with which uncertainty reflects true epistemic gaps; misalignments can precipitate feedback loops that accelerate closure around approximate rather than comprehensive representations.

Graph representations and equivariant architectures

Graph neural networks and equivariant message-passing models have introduced powerful abstractions for encoding atomic interactions within materials systems [15-17]. These architectures facilitate transfer of learned representations across chemical domains, conceptually expanding the scope of guided search. Nevertheless, when integrated into optimization loops, their convergence behaviors reflect the interplay between local structural motifs and global landscape features. Interpretive analysis suggests that high-fidelity local representations may inadvertently hasten convergence by providing compelling but partial signals, thereby influencing higher-order exploration decisions [18, 19].

Performance assessments of such models underscore trade-offs between accuracy and computational overhead, framing cost structures as embedded in the very representations that enable guidance [20]. When early convergence truncates the sampling of diverse graph configurations, the interpretive consequence is a narrowing of the conceptual space available for deriving generalizable insights into material stability or functionality.

Active learning, transfer mechanisms, and systems-level considerations

Transfer learning strategies, including shotgun approaches, enable leveraging prior knowledge to bootstrap search in novel domains [21, 22]. These mechanisms interact with convergence dynamics by providing initialization anchors that may predispose trajectories toward rapid stabilization. Analytically, this predisposition highlights feedback structures wherein transferred representations reinforce certain subspaces at the expense of others, generating epistemic costs measured in terms of foregone diversity.

Broader reviews of machine learning in energy materials and interatomic potentials emphasize the importance of maintaining open search horizons to sustain innovation potential [23-25]. When viewed through a systems lens, early convergence emerges as a modulator of these horizons, interacting with algorithmic and representational components to shape overall discovery ecosystems. Insights from constraint-aware and human-in-the-loop variants further illustrate how external steering inputs can either mitigate or exacerbate convergence tendencies, depending on their alignment with underlying uncertainty structures [26, 27].

Collectively, these threads portray AI-guided search as a rich tapestry of interacting dynamics. Early convergence functions not as an isolated event but as an integrative phenomenon whose costs permeate epistemic, computational, and structural dimensions. The synthesis reveals recurring motifs of feedback amplification and trade-off propagation that warrant deeper conceptual articulation.

Proposed conceptual framework

The convergence cost cascade model

The Convergence Cost Cascade Model (CCCM) advances a systems-interpretive understanding of early convergence in AI-guided materials discovery by positioning it not as an isolated optimization artifact but as a cascading epistemic condition that progressively restructures the architecture of scientific search. Within this framework, convergence is not interpreted as a terminal event in algorithmic optimization but as an evolving structural phase that reconditions how knowledge is generated, represented, and propagated across discovery systems. The model, therefore, reframes convergence from a performance outcome into a constitutive force shaping the epistemic horizon of materials innovation.

At its foundation, the model conceptualizes AI-guided search as a recursive network composed of tightly coupled feedback relations linking surrogate modeling infrastructures, acquisition steering mechanisms, and representational updating processes. These three components operate as interdependent epistemic circuits rather than as sequential modules. Predictive surrogates generate probabilistic estimates of material properties, acquisition functions translate these estimates into sampling decisions, and representation systems encode the resulting knowledge into latent structural embeddings. Each iteration of search consequently modifies the informational substrate upon which subsequent exploration unfolds. Early convergence occurs when these recursive circuits undergo premature stabilization, leading to predictive confidence, sampling density, and representational encoding coalescing around localized performance optima before the broader design space has been sufficiently traversed.

Within this interpretive framing, convergence signifies a reduction in epistemic permeability rather than merely a narrowing of candidate selection. The accessible landscape of material possibilities becomes progressively filtered, not only in spatial terms but also in conceptual and representational terms. Configurations that fall outside early high-confidence clusters become increasingly difficult for the system to recognize, prioritize, or even meaningfully encode. The cascade is thus initiated when localized certainty outpaces exploratory diversity.

The first interpretive layer of the cascade is described as local epistemic closure. At this stage, uncertainty quantification architectures begin to densify confidence signals in early-sampled regions. Bayesian surrogate models, ensemble predictors, and evidential inference systems register declining posterior variance in localized neighborhoods of the design space. While such densification may appear operationally advantageous, CCCM interprets it as a form of epistemic contraction. The probabilistic field through which the system perceives novelty begins to flatten, diminishing gradient signals that would otherwise guide exploration toward uncharted configurations. Predictive certainty deepens but not broadens, producing a condition in which the system becomes locally over-certain yet globally under-informed. This closure does not halt exploration outright; rather, it subtly reshapes exploratory salience, rendering certain regions epistemically luminous while others fade into algorithmic obscurity.

As epistemic closure intensifies, its effects are absorbed into the AI system's structural substrate, giving rise to a second cascade layer: structural reinforcement. Representation architectures such as graph neural networks, equivariant transformers, and crystal graph encoders embed observed configurations into latent manifolds that structure future predictions. When training exposure is already convergence-biased, these embeddings develop anisotropic geometries that privilege sampled material families. Latent clustering becomes denser around canonical chemistries and crystallographic motifs, while representational curvature toward unexplored domains flattens. The model's internal ontology of materials thus becomes selectively compressed, encoding not the full diversity of materials space but a convergence-filtered projection of it. Epistemic closure is thereby transformed into architectural memory, stabilizing narrowed discovery pathways within the geometry of predictive inference itself.

The third interpretive layer, systemic propagation, extends the cascade beyond model internals into the broader ecosystem of AI-guided discovery. Constrained representations influence candidate generation systems, autonomous experimentation pipelines, high-throughput simulation prioritization, and even human interpretive reasoning. Downstream discovery trajectories inherit embedded biases originating in early convergence dynamics. The system begins to reproduce familiar material families, reinforcing canonical design logics while attenuating the emergence of paradigm-deviant candidates. Innovation scope contracts, hypothesis diversity diminishes, and the conceptual boundaries of materials search become historically path-dependent. CCCM emphasizes that these systemic consequences are not

additive degradations but multiplicative propagations. Constraints imposed at earlier layers amplify as they traverse institutional, computational, and epistemic infrastructures, producing compounded contraction of discovery horizons.

A defining property of the cascade is its recursive amplification architecture. Confidence signals generated within localized regions feed acquisition steering mechanisms, biasing sampling decisions toward known optima. These biased samples retrain surrogate predictors, further tightening uncertainty envelopes. Representational embeddings update accordingly, reinforcing structural emphasis on converged domains. Each cycle intensifies the attractor dynamics of convergence, transforming early stabilization into self-reinforcing epistemic gravity. Exploration signals do not disappear; rather, they become progressively attenuated, overwhelmed by the coherence-reinforcing logic of feedback coupling.

Acquisition steering logics occupy a pivotal modulatory position within this system. Their encoded reward structures determine whether convergence cascades are dampened or accelerated. When acquisition functions prioritize immediate expected improvement, greedy exploitation, or deterministic confidence weighting, they effectively lower the threshold at which convergence becomes self-reinforcing. Exploration is not eliminated; it is systematically deprioritized. Conversely, acquisition logics incorporating diversity regularization, entropy weighting, or novelty sensitivity can slow the initiation of cascades by preserving epistemic heterogeneity within sampling trajectories. CCCM therefore situates acquisition design not merely as a technical optimization problem but as an epistemic governance mechanism shaping the topology of discovery itself.

The cascade gives rise to a temporal trade-off dynamic between short-term efficiency and long-horizon innovation capacity. Early convergence often yields rapid gains in predictive accuracy, reduced experimental burden, and accelerated optimization cycles. However, these immediate efficiencies incur deferred epistemic costs. Generalization robustness weakens, unconventional material regimes remain unexplored, and the representational elasticity required for paradigm innovation diminishes. CCCM interprets this asymmetry as a structural property of AI-guided search systems, in which efficiency benefits are measurable in real time while epistemic losses accumulate diffusely over extended discovery timelines.

Central to the framework is the concept of a cascade initiation threshold. This threshold represents the inflection point at which feedback tightening becomes self-reinforcing rather than reversible. Its position is conditioned by factors such as initial dataset diversity, surrogate inductive bias, exploration weighting within acquisition functions, representation plasticity, and noise tolerance in uncertainty estimation. Before threshold crossing, convergence remains adjustable through parameter modulation. Beyond it, reversal requires architectural intervention, as convergence dynamics become embedded within representational and inferential substrates.

From a systems perspective, CCCM reframes convergence costs as multiplicative rather than localized. Constraints arising from uncertainty collapse propagate into the representation geometry, then into systemic discovery infrastructures. Even modest early closure can therefore yield disproportionate long-term contraction of epistemic possibility. This multiplicativity explains why minor acquisition biases or dataset imbalances can ultimately produce large-scale discovery blind spots.

The framework further introduces an ethical and epistemic reasoning dimension. AI systems guiding materials discovery do not merely optimize candidate selection; they shape the narrative contours of scientific possibility. Early convergence determines which materials are rendered visible, fundable, manufacturable, and industrially scalable. Entire classes of unconventional materials may remain algorithmically marginalized, not because of a lack of value but because of structural invisibility within converged search architectures. CCCM thus frames convergence costs as socio-technical and epistemic phenomena, implicating governance responsibilities in transparency of acquisition, exploration, auditing, and diversity reporting within AI discovery systems **(Table 1)**.

Table 1. Interpretive layers and cascading costs within the convergence cost cascade model

Cascade layer	Convergence mechanism in AI-guided search	Propagated constraint	Resulting epistemic cost
Local epistemic closure	Rapid uncertainty tightening around early high-confidence clusters	Reduced exploratory signal diversity	Truncated mapping of materials design space
Structural reinforcement	Embedding of narrowed configurations within graph/equivariant representations	Representational anisotropy toward sampled motifs	Constrained structure–property generalization
Systemic propagation	Inheritance of converged priors across downstream discovery systems	Innovation pathway contraction	Long-horizon limitation on materials novelty and knowledge breadth

In this reconceptualization, convergence is no longer treated as a universal optimization objective. Instead, it is interpreted as a structural phase transition within epistemic search dynamics, one that must be timed, regulated, and contextually interpreted rather than maximized indiscriminately. The goal is not to eliminate convergence but to understand its cascade properties and manage its initiation relative to discovery maturity.

Figure 1 visualizes the Convergence Cost Cascade Model, a directed graph depicting how early algorithmic convergence, through recursive feedback and structural reinforcement, propagates across representational and systemic layers to constrain the scope of innovation and contract epistemic boundaries.

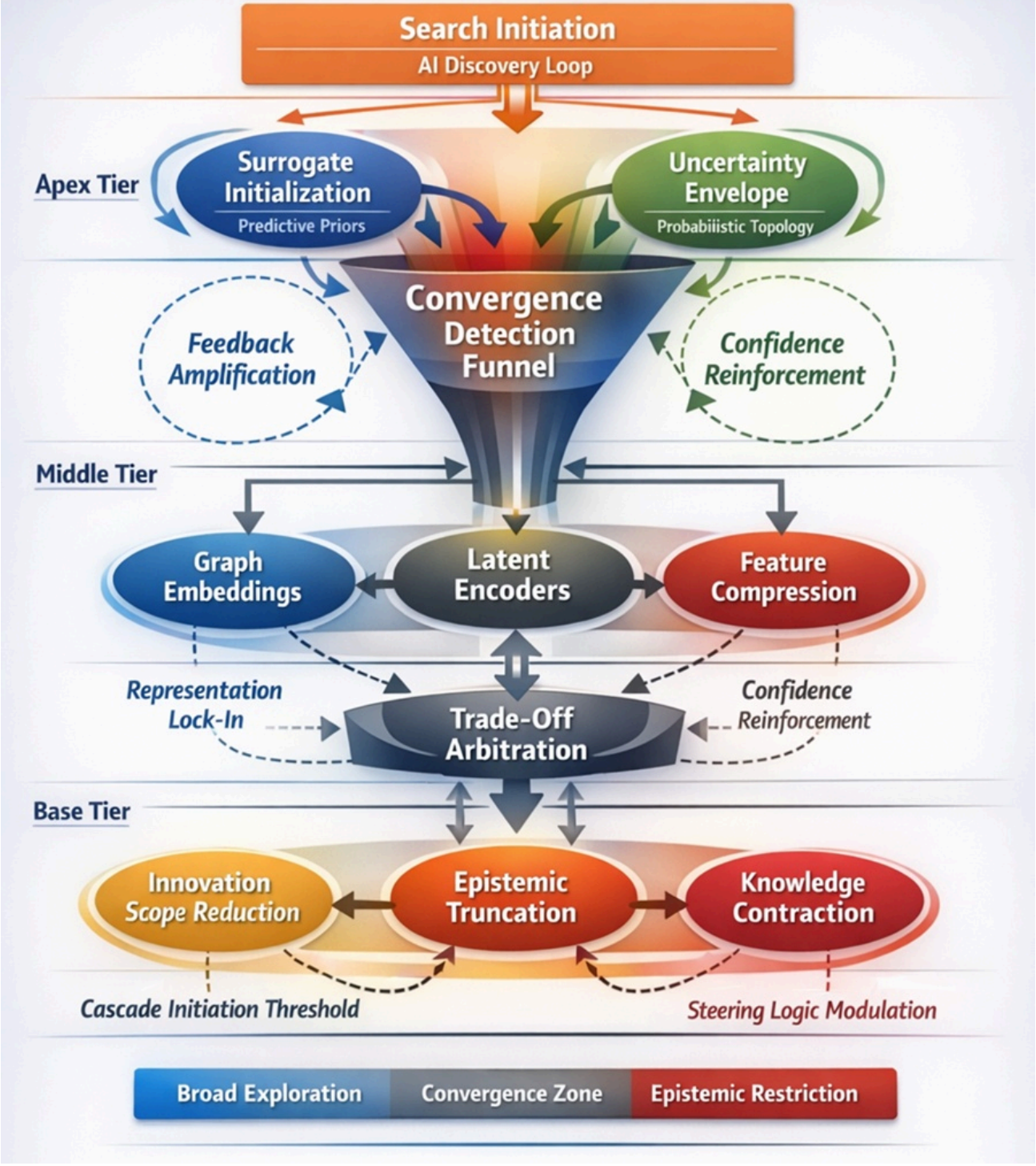

Figure 1. The convergence cost cascade model: a vertically stratified systems map illustrating the cascade from localized algorithmic convergence to systemic epistemic constraint propagation in AI-guided materials discovery.

Analytical implications

The Convergence Cost Cascade Model offers interpretive depth into the layered consequences of early stabilization within AI-guided materials search. At the initial layer of local epistemic closure, surrogate models and uncertainty quantification mechanisms interact to rapidly tighten the confidence boundaries around sampled clusters. This dynamic,

when unchecked, curtails the propagation of alternative signals, leading to interpretive outcomes where the conceptual representation of the design space becomes disproportionately anchored to initial coherence rather than comprehensive coverage. Consequently, downstream acquisition decisions operate under inherited constraints, amplifying the preference for refinement over diversification and embedding a structural bias toward incremental rather than transformative insights.

In the structural reinforcement layer, graph-based and equivariant representations are mediated by encoding the narrowed subspace into predictive embeddings. These embeddings then feed back into subsequent iterations, reinforcing localized motifs at the expense of broader chemical or configurational variability. Analytically, this interaction generates compounded interpretive costs: the fidelity of local atomic interactions, while advantageous for precision within the converged domain, simultaneously diminishes the model's capacity to generalize across underrepresented regions of the landscape. Trade-offs become evident here, as the efficiency gained from focused exploitation comes at the cost of a systemic reduction in the diversity of structure-property mappings that could otherwise emerge. Feedback amplification in these loops further entrenches the cascade, in which each cycle of updating increases the weight assigned to converged signals, progressively marginalizing outlier configurations that might challenge prevailing conceptual models of material behavior.

At the systemic propagation layer, these restrictions cascade into higher-order effects on innovation trajectories. Discovery processes inherit bounded epistemic horizons, wherein the range of explorable material families contracts, influencing not only immediate candidate identification but also the conceptual scaffolding available for future inquiries. Systems-level insights reveal multiplicative dynamics: epistemic truncation at lower-layer scales nonlinearly constrains the overall knowledge ecosystem and fosters path dependencies that favor certain innovation pathways while rendering others conceptually latent. For instance, when steering logics embedded in acquisition functions prioritize immediate expected gains, they modulate the cascade threshold in ways that favor short-term coherence, generating interpretive tensions between computational expediency and long-term robustness of understanding.

Ethical and epistemic reasoning underscores the responsibilities inherent in these structures. The model highlights how algorithmic preferences shape not merely search efficiency but the very character of generated knowledge, inviting reflection on the implicit values encoded in uncertainty handling and sampling criteria. By interpreting early convergence through cascading feedback, the framework illuminates hidden interdependencies wherein individual component choices reverberate across the entire discovery architecture, necessitating nuanced awareness of how such systems configure the boundaries of attainable material insights. These analytical dimensions collectively reposition convergence costs as constitutive features of guided search, rather than peripheral inefficiencies, emphasizing their role in defining the interpretive scope of materials innovation.

Results and Discussion

Integrating the Convergence Cost Cascade Model with established threads in probabilistic optimization and adaptive learning reveals deeper systemic patterns in AI-guided materials search. The interplay between surrogate initialization, uncertainty envelopes, and acquisition steering manifests as self-reinforcing architectures that prioritize stabilization over sustained openness. This integration underscores how early convergence functions as an emergent regulator of epistemic breadth, interacting with multi-objective trade-offs to produce interpretive trade-offs in which gains in targeted property optimization come at the expense of holistic landscape characterization. Literature on entropy-driven and constraint-aware approaches further contextualizes these patterns, portraying convergence not as an anomaly but as an intrinsic outcome of logics that reward local consistency amid high-dimensional complexity [25-29].

Epistemic reasoning within this context emphasizes the provisional nature of knowledge produced under constrained trajectories. When cascades propagate structural reinforcements from graph representations, the resulting conceptual models risk overgeneralization from partial samples, limiting the development of more inclusive frameworks for understanding material phenomena. Systems-level insights extend this to innovation ecosystems, where repeated emphasis on converged subspaces may channel collective efforts toward densely populated regions of design space

while underrepresenting sparse but potentially disruptive alternatives. Such dynamics invite consideration of feedback structures that could modulate cascade initiation, such as regularization mechanisms that maintain diversity thresholds without compromising core guidance functions [28-31].

Ethical dimensions arise from the steering influence of algorithmic components on discovery narratives. Deploying these systems entails implicit choices about which uncertainties merit attention and which pathways warrant pursuit, raising questions of responsibility in configuring the epistemic horizons of materials research. Interpretively, the framework encourages viewing these choices as shaping not only technical outcomes but also the broader contours of scientific imagination, wherein premature closure may curtail serendipitous intersections across compositional families. Trade-off propagation thus carries normative weight, highlighting tensions between accelerated identification of viable candidates and the cultivation of expansive conceptual repertoires essential for addressing multifaceted materials challenges.

Broader analytical implications extend to the reflexive calibration of guidance paradigms. By framing convergence costs through cascading layers, the model facilitates recognition of how representational fidelity and sampling preferences co-evolve to influence long-term trajectories. This perspective fosters interpretive humility, acknowledging that while guided search expands accessibility to complex spaces, it simultaneously embeds structural logics that warrant ongoing scrutiny. The discussion thereby synthesizes interaction dynamics across probabilistic, representational, and adaptive elements, portraying early convergence as a pivotal nexus where efficiency imperatives intersect with epistemic and systemic consequences [29-32].

Conclusion

The interpretive examination of early convergence in AI-guided materials search, articulated through the Convergence Cost Cascade Model, illuminates its embedded role in shaping discovery architectures. Through layered feedback structures and propagating trade-offs, premature stabilization emerges as a determinant of epistemic scope, structural reinforcement, and systemic innovation boundaries. These dynamics underscore the intricate steering logics that govern the interactions among uncertainty quantification, representational embeddings, and adaptive sampling, revealing how local closures cascade into broader constraints on knowledge diversity and conceptual generality.

By emphasizing analytical implications and integrative insights, this conceptual contribution reframes convergence costs as constitutive aspects of probabilistic guidance rather than ancillary considerations. The resulting perspective encourages sustained reflection on the epistemic responsibilities and ethical dimensions inherent in algorithmic design for materials innovation, highlighting the need for models attuned to the multiplicative effects of feedback amplification and horizon modulation. Ultimately, such interpretive engagement enriches understanding of the subtle architectures that configure trajectories in high-dimensional search, fostering awareness of the interplay between expediency and openness in advancing materials knowledge.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 18 Jul 2022 Revised: 21 Nov 2022 Accepted: 17 Dec 2022
Published online: 18 July 2023

References

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;11:5966.
https://doi.org/10.1038/s41467-020-19597-w.

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;10:156.
https://doi.org/10.1038/s41524-024-01326-2.

Khatamsaz D, Vela B, Singh P, Johnson DD, Allaire D, Arróyave R. Bayesian optimization with active learning of design constraints using an entropy-based approach. npj Comput Mater. 2023;9:49.
https://doi.org/10.1038/s41524-023-01006-7.

Biswas A, Liu Y, Creange N, Liu YC, Jesse S, Yang JC, et al. A dynamic bayesian optimized active recommender system for curiosity-driven partially human-in-the-loop automated experiments. npj Comput Mater. 2024;10:29.
https://doi.org/10.1038/s41524-023-01191-5.

Liang Q, Gongora AE, Ren Z, Tiihonen A, Liu Z, Sun S, et al. Benchmarking the performance of bayesian optimization across multiple experimental materials science domains. npj Comput Mater. 2021;7:188.
https://doi.org/10.1038/s41524-021-00656-9.

Pyzer-Knapp EO, Pitera JW, Staar PWJ, Takeda S, Laino T, Sanders DP, et al. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. npj Comput Mater. 2022;8:84.
https://doi.org/10.1038/s41524-022-00765-z.

Tian Y, Li T, Pang J, Zhou Y, Xue D, Ding X, et al. Materials design with target-oriented bayesian optimization. npj Comput Mater. 2025;11:209.
https://doi.org/10.1038/s41524-025-01704-4.

Khatamsaz D, Neuberger R, Roy AM, Zadeh SH, Otis R, Arróyave R. A physics-informed bayesian optimization approach for material design: application to NiTi shape memory alloys. npj Comput Mater. 2023;9:221.
https://doi.org/10.1038/s41524-023-01173-7.

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.

Jablonka KM, Jothiappan GM, Wang S, Smit B, Yoo B. Bias-free multiobjective active learning for materials design and discovery. Nat Commun. 2021;12:2312.
https://doi.org/10.1038/s41467-021-22437-0.

Batzner S, Musaelian A, Sun L, Geiger M, Mailoa JP, Kornbluth M, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. Nat Commun. 2022;13:2453.
https://doi.org/10.1038/s41467-022-29939-5.

Deng B, Zhong P, Jun K, Riebesell J, Han K, Bartel CJ, et al. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. Nat Mach Intell. 2023;5(9):1031-41.
https://doi.org/10.1038/s42256-023-00716-3.

Goodall REA, Lee AA. Predicting materials properties with little data using shotgun transfer learning. Nat Commun. 2020;11:4043.
https://doi.org/10.1038/s41467-020-17871-9.

Vandermause J, Torrisi SB, Batzner S, Xie Y, Sun L, Kolpak AM, et al. Active learning of interatomic potentials with bayesian optimization. npj Comput Mater. 2021;7:95.
https://doi.org/10.1038/s41524-021-00555-7.

Zeni C, Pinsard E, Sanna S, Martinelli L, Ceriotti M. Uncertainty-aware active learning for interatomic potentials. npj Comput Mater. 2023;9:157.
https://doi.org/10.1038/s41524-023-01100-w.

Rosen AS, Fung V, Huck P, O'Donnell CT, Horton MK, Truhlar DG, et al. High-throughput predictions of metal-organic framework electronic properties: theoretical challenges, graph neural networks, and data exploration. npj Comput Mater. 2022;8:112.
https://doi.org/10.1038/s41524-022-00796-6.

Kim Y, Kim E, Antono E, Meredig B, Ling J. Machine-learned metrics for predicting the likelihood of success in materials discovery. npj Comput Mater. 2020;6:131.
https://doi.org/10.1038/s41524-020-00405-0.

Wang AY, Kauwe SK, Murdock RJ, Sparks TD. Compositionally restricted attention-based network for materials property predictions. npj Comput Mater. 2021;7:77.
https://doi.org/10.1038/s41524-021-00545-9.

Schmidt J, Pettersson L, Verdozzi C, Botti S, Marques MAL. Crystal graph attention networks for the prediction of stable materials. Sci Adv. 2021;7(49):eabi7948.
https://doi.org/10.1126/sciadv.abi7948.

Chen C, Ong SP. A universal graph deep learning interatomic potential for the periodic table. Nat Comput Sci. 2022;2(11):718-28.
https://doi.org/10.1038/s43588-022-00349-3.

Schwalbe-Koda D, Tan AR, Xie Y, Wen J, Gómez-Bombarelli R. Differentiable sampling of molecular geometries with uncertainty-based adversarial attacks. Nat Commun. 2021;12:5104.
https://doi.org/10.1038/s41467-021-25342-2.

Tran K, Neiswanger W, Yoon J, Zhang Q, Xing E, Ulissi ZW. Methods for comparing uncertainty quantifications for material property predictions. Mach Learn Sci Technol. 2020;1:025006.

Wang Y, Xie T, France-Lanord A, Berkley A, Johnson JA, Shao-Horn Y, et al. Toward efficient discovery of new materials with targeted properties through active learning. npj Comput Mater. 2022;8:46.

https://doi.org/10.1038/s41524-022-00728-4.

Musaelian A, Batzner S, Johansson A, Sun L, Owen CJ, Kornbluth M, et al. Learning local equivariant representations for large-scale atomistic dynamics. Nat Commun. 2023;14:579.
https://doi.org/10.1038/s41467-023-36329-y.

Zuo Y, Chen C, Li X, Deng Z, Chen Y, Behler J, et al. Performance and cost assessment of machine learning interatomic potentials. J Phys Chem A. 2020;124(4):731-45.
https://doi.org/10.1021/acs.jpca.9b08723.

Chen C, Zuo Y, Ye W, Li X, Deng Z, Ong SP. A critical review of machine learning of energy materials. Adv Energy Mater. 2020;10(8):1903242.
https://doi.org/10.1002/aenm.201903242.

Solomou A, Zhao G, Boluki S, Joy JK, Liu X, Karaman I, et al. Multi-objective materials bayesian optimization with active learning of design constraints: design of ductile refractory multi-principal-element alloys. Acta Mater. 2022;239:118264.
https://doi.org/10.1016/j.actamat.2022.118264.

Hickman RJ, Parakh A, Cheng M, Tao H, Aspuru-Guzik A. AlphaFold and the future of structural biology. Nat Rev Mater. 2023;8(11):709-10.
https://doi.org/10.1038/s41578-023-00572-0.

Lee J, Seko A, Shitara K, Tanaka I. Prediction of lattice thermal conductivity from first-principles using machine learning. Phys Rev B. 2020;101(14):144105.
https://doi.org/10.1103/PhysRevB.101.144105.

Fung V, Zhang J, Juarez E, Sumpter BG. Benchmarking graph neural networks for materials science. npj Comput Mater. 2023;9:92.
https://doi.org/10.1038/s41524-023-01000-1.

Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Phys Rev Lett. 2020;120(14):145301.

Batatia I, Kovacs DP, Simm G, Ortner C, Csányi G. MACE: Higher-order equivariant message passing neural networks for fast and accurate force fields. Adv Neural Inf Process Syst. 2022;35:11423-36.