ORIGINAL RESEARCH **Open access**

Algorithmic Discovery vs. Scientific Discovery: A Conceptual Boundary for AI-Driven Materials Research

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

Abstract

The integration of artificial intelligence into materials research has transformed how chemical and structural spaces are explored, enabling algorithmic systems to generate and evaluate candidate materials at an unprecedented scale. While these approaches dramatically accelerate exploration, they operate under epistemic conditions that differ fundamentally from those of traditional scientific discovery. This conceptual manuscript articulates a boundary between *algorithmic discovery*—defined by probabilistic inference, large-scale search, and optimization within computational objectives—and *scientific discovery*, which emphasizes causal understanding, theoretical coherence, and explanatory integration. Rather than treating these modes as competing or hierarchical, the framework conceptualizes their relationship as a permeable boundary through which interaction, feedback, and epistemic governance occur. The analysis examines how algorithmic breadth and scientific depth are coordinated through steering mechanisms such as uncertainty awareness, constraint propagation, and selective interpretation. By foregrounding boundary dynamics, the manuscript clarifies how AI reshapes discovery not by replacing scientific reasoning but by reconfiguring the conditions under which explanation, validation, and legitimacy are achieved. The framework contributes a systems-level conceptual vocabulary for positioning AI as an augmentative instrument in materials research, preserving the epistemic integrity of scientific discovery while enabling scalable exploration beyond human cognitive limits.

Keywords Artificial intelligence, Materials discovery, Conceptual frameworks, Algorithmic discovery, Scientific discovery, Epistemic boundaries

*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

Materials science stands at the confluence of increasing computational power and data availability, where artificial intelligence (AI) and machine learning (ML) have reshaped the landscape of discovery processes [1, 2]. High-throughput computational screening, generative modeling, and autonomous experimentation systems now routinely propose novel compounds, predict properties, and optimize performance metrics at scales previously unattainable through conventional methods [3-5]. These developments have expanded the accessible materials space by orders of magnitude, with demonstrations of discovering millions of stable crystal structures and identifying candidates for applications ranging from energy storage to catalysis [1, 6].

Yet the accelerating pace of such outputs raises conceptual questions about the nature of discovery itself. Traditional scientific discovery in materials research has relied on iterative cycles of hypothesis formulation, mechanistic

interpretation, theoretical unification, and experimental validation, often guided by human intuition and domain-specific knowledge [7]. In contrast, algorithmic approaches prioritize exhaustive exploration, statistical inference, and optimization within defined objective functions, frequently without explicit reference to underlying causal mechanisms [8, 9]. This divergence is not merely methodological but epistemic: algorithmic systems excel in identifying correlations and generating candidates through pattern-based generalization, whereas scientific discovery seeks explanatory coherence and transferable understanding [10].

The literature reflects growing recognition of these distinctions. Efforts to scale graph neural networks for property prediction illustrate how deep learning can achieve broad generalization across chemical compositions. Yet, such models often remain opaque with respect to physical principles [1]. Similarly, active learning and closed-loop workflows integrate computational predictions with automated synthesis, creating data flywheels that amplify throughput but defer deeper interpretation to subsequent human analysis [2, 11]. Autonomous laboratories further exemplify this shift, where AI-driven decision-making guides robotic execution, yet the attribution of discovery remains contested [12].

These trends invite scrutiny of the boundary between algorithmic and scientific discovery. The boundary is not absolute but relational, defined by the interplay of scope and depth, speed and insight, and automation and agency. Algorithmic discovery operates effectively in regimes of high-dimensional search and probabilistic novelty, where the volume of candidates overwhelms human capacity [13]. Scientific discovery, however, thrives in regimes that require causal linkage, contextual integration, and the revision of conceptual frameworks [14]. The interaction between these modalities generates feedback structures that can either reinforce or challenge existing paradigms.

Clarifying this boundary holds analytical value for materials research. It illuminates trade-offs inherent in AI adoption—such as gains in efficiency against potential losses in explanatory power—and underscores steering logics that govern transitions across domains. For instance, algorithmic outputs can serve as starting points for scientific inquiry, while scientific constraints can impose physically motivated priors on algorithmic search [15]. Such dynamics suggest an integrative rather than a substitutive relationship, where the boundary functions as a permeable interface that facilitates mutual enrichment.

This manuscript develops a novel conceptual framework to articulate this boundary. Drawing on recent advances in AI-driven materials workflows, it synthesizes insights from computational scaling, generative modeling, and autonomous systems to examine interaction dynamics and epistemic implications. The framework avoids predictive or quasi-empirical claims, focusing instead on interpretive structures that reveal how algorithmic and scientific modalities coexist and interact within contemporary materials research.

Theoretical Background and Literature Synthesis

AI and machine learning foundations in materials science

Recent years have marked the maturation of machine learning techniques explicitly adapted to the structural and chemical complexity of materials systems. Graph neural networks, equivariant architectures, and symmetry-aware models now enable accurate property prediction across diverse chemical and crystallographic spaces, often without reliance on handcrafted descriptors [1, 16]. Scaling laws, analogous to those observed in large language models, indicate predictable improvements in model performance with increasing data volume and architectural capacity, allowing learned representations to generalize beyond narrowly sampled training regimes [1]. In some cases, models trained on compositional or stoichiometric inputs alone achieve competitive predictive accuracy, underscoring the power of statistical abstraction in navigating materials space [17].

These advances have been tightly coupled with high-throughput computation. Machine learning models are routinely embedded within screening pipelines, where active learning strategies iteratively refine predictions by prioritizing candidates associated with high uncertainty or expected information gain [11, 18]. Such workflows generate self-

reinforcing data cycles, in which predicted stability, synthesizability, or functional performance directs subsequent data acquisition, further expanding the searchable design space [19]. Collectively, these developments have shifted materials discovery toward regimes characterized by scale, iteration, and probabilistic inference.

Generative and autonomous approaches

Beyond predictive modeling, generative methods have emerged as central instruments for proposing novel material candidates. Variational autoencoders, generative adversarial networks, and diffusion-based models enable direct navigation of chemical and structural manifolds, supporting inverse design workflows in which desired properties guide candidate generation [20, 21]. These approaches move discovery beyond enumeration toward the synthesis of statistically plausible yet previously unexplored configurations.

Autonomous systems extend this logic by coupling generative models with robotic synthesis and characterization, forming closed-loop platforms that execute discovery cycles with minimal human intervention [12, 22]. In these settings, decision-making is increasingly delegated to algorithmic agents that balance exploration and exploitation based on learned objectives. While such systems dramatically increase throughput, they also intensify a shift toward data-centric discovery paradigms, where the volume and velocity of iteration dominate over serial hypothesis formulation and testing [2].

At the same time, generative and autonomous approaches surface persistent tensions. Although they excel at producing diverse and optimized candidates, they often lack intrinsic mechanisms for assessing causal plausibility, mechanistic stability, or theoretical consistency [23]. As a result, scientific interpretation is frequently deferred to post hoc analysis, reinforcing a separation between candidate generation and explanatory understanding.

Epistemic and conceptual perspectives

The growing influence of AI in materials research has prompted sustained engagement with its epistemic implications. A recurring theme in the literature concerns interpretability: high predictive accuracy does not necessarily translate into physical insight, particularly when models rely on latent representations that resist mechanistic interpretation [24]. Critical examinations of machine-learned stability predictions further demonstrate how extrapolation beyond training regimes can yield results that diverge from ground-truth physical principles [25].

Parallel discussions in broader scientific contexts emphasize that discovery cannot be reduced to pattern identification alone. Attribution of discovery depends on explanatory depth, theoretical embedding, and the capacity to integrate findings into coherent conceptual frameworks [11]. In materials science, these concerns manifest in debates over the role of human expertise in validating algorithmic outputs, imposing physical constraints, and determining when a computational proposal constitutes genuine scientific knowledge rather than a provisional statistical artifact [26].

Integration and feedback structures

Hybrid workflows exemplify how algorithmic and scientific modalities are increasingly interwoven. In these systems, AI accelerates exploration while human-guided reasoning provides direction, interpretation, and constraint [2, 27]. Bayesian optimization, uncertainty quantification, and active learning function as steering mechanisms that regulate the balance between exploration and exploitation, determining when algorithmic search should expand and when scientific consolidation should occur [18].

Additional integration occurs through data extraction and representation learning from unstructured sources, where knowledge graphs and multimodal embeddings link disparate experimental and computational findings into navigable information structures [28]. Across these workflows, discovery emerges not from the dominance of a single modality, but from feedback and constraint propagation across distinct epistemic roles.

Taken together, the literature depicts a discovery landscape in which algorithmic efficiency and scientific depth intersect without collapsing into one another. Instead, their interaction sustains a productive conceptual tension that shapes discovery trajectories in contemporary materials research. **Table 1** synthesizes epistemic distinctions across the literature, highlighting how algorithmic and scientific discovery differ in their goals, inference modes, and evaluation criteria while remaining structurally interdependent.

Table 1. Epistemic characteristics of algorithmic discovery and scientific discovery in materials research

Dimension	Algorithmic discovery	Scientific discovery
Primary epistemic goal	Efficient exploration of large possibility spaces	Generation of explanatory, generalizable knowledge
Dominant mode of inference	Inductive, statistical generalization	Deductive and abductive reasoning
Role of data	Central driver of inference and optimization	Evidential support within the theoretical context
Treatment of causality	Implicit or indirect; often correlational	Explicit; mechanistic, and theory-linked
Typical outputs	Ranked candidates, predicted properties, generative proposals	Mechanistic explanations, validated principles, revised models
Evaluation criteria	Predictive accuracy, novelty, optimization performance	Explanatory coherence, causal plausibility, and theoretical integration
Temporal position in workflows	Early-stage, exploratory phases	Later-stage, consolidative, and interpretive phases
Relationship to human agency	Delegates search and pattern recognition	Centers interpretation, judgment, and validation
Risk profile	Spurious correlations, unphysical artifacts	Over-constraint, reduced exploratory reach

Proposed conceptual framework

The proposed framework conceptualizes the boundary between algorithmic discovery and scientific discovery as a permeable epistemic interface rather than a rigid methodological divide. Algorithmic discovery operates within a computational possibility space, governed by data-driven inference, probabilistic generalization, and optimization under explicit objective functions. It is optimized for breadth, excelling in regimes where vast candidate enumeration, pattern recognition, and iterative refinement yield novelty through scale and speed.

Scientific discovery, by contrast, operates within an explanatory possibility space. It emphasizes causal linkage, theoretical unification, and contextual coherence, prioritizing depth over coverage. Scientific discovery transforms knowledge by embedding it within mechanistic narratives, revising conceptual frameworks, and assessing its generalizability beyond specific instances.

The boundary between these modalities is not fixed but dynamically enacted through trade-offs and steering logics. Algorithmic processes achieve high throughput by abstracting away mechanistic detail, relying on correlational patterns and learned representations [1, 16]. While this abstraction enables exploration beyond human intuition, it introduces epistemic distance from underlying physical principles. Scientific processes counterbalance this distance by reintroducing causal reasoning and explanatory constraints, often triggered by algorithmic anomalies, inconsistencies, or unexpected candidates that resist straightforward interpretation [9].

Interaction across the boundary is structured through bidirectional feedback. Algorithmic outputs—candidate structures, property predictions, or generative proposals—function as stimuli for scientific inquiry, prompting mechanistic analysis, experimental validation, or theoretical refinement. Conversely, scientific insights impose priors, constraints, and reinterpretive frames that redirect algorithmic search, narrowing computational possibility spaces toward physically meaningful regions [2, 11]. These feedback loops generate steering structures in which algorithmic breadth supplies raw novelty while scientific depth filters, integrates, and stabilizes understanding.

Trade-offs further define the boundary. Algorithmic discovery gains speed, scope, and serendipity at the expense of interpretability and causal grounding, whereas scientific discovery gains explanatory power and conceptual durability at the expense of scale and exploratory reach [24, 29]. Rather than treating these trade-offs as deficiencies to be eliminated, the framework views them as complementary tensions to be actively managed. The boundary thus modulates the allocation of epistemic resources between exploration and consolidation, enabling hybrid discovery trajectories that neither modality can achieve in isolation. **Figure 1** presents a textual schematic of the framework as two regions separated by a permeable boundary: an upper algorithmic discovery domain characterized by high-dimensional search, probabilistic generalization, generative proposal, data flywheel iteration, and optimization under objectives, and a lower scientific discovery domain characterized by causal mechanistic reasoning, theoretical unification, explanatory coherence, contextual integration, and epistemic agency. Bidirectional arrows crossing the boundary at multiple points indicate feedback flows (constraints and priors upward, candidates and anomalies downward). At the same time, a dashed permeability zone annotated with trade-offs (breadth vs. depth, speed vs. insight) and steering logics (exploration–consolidation modulation) emphasizes that integration occurs at the boundary without either domain subsuming the other.

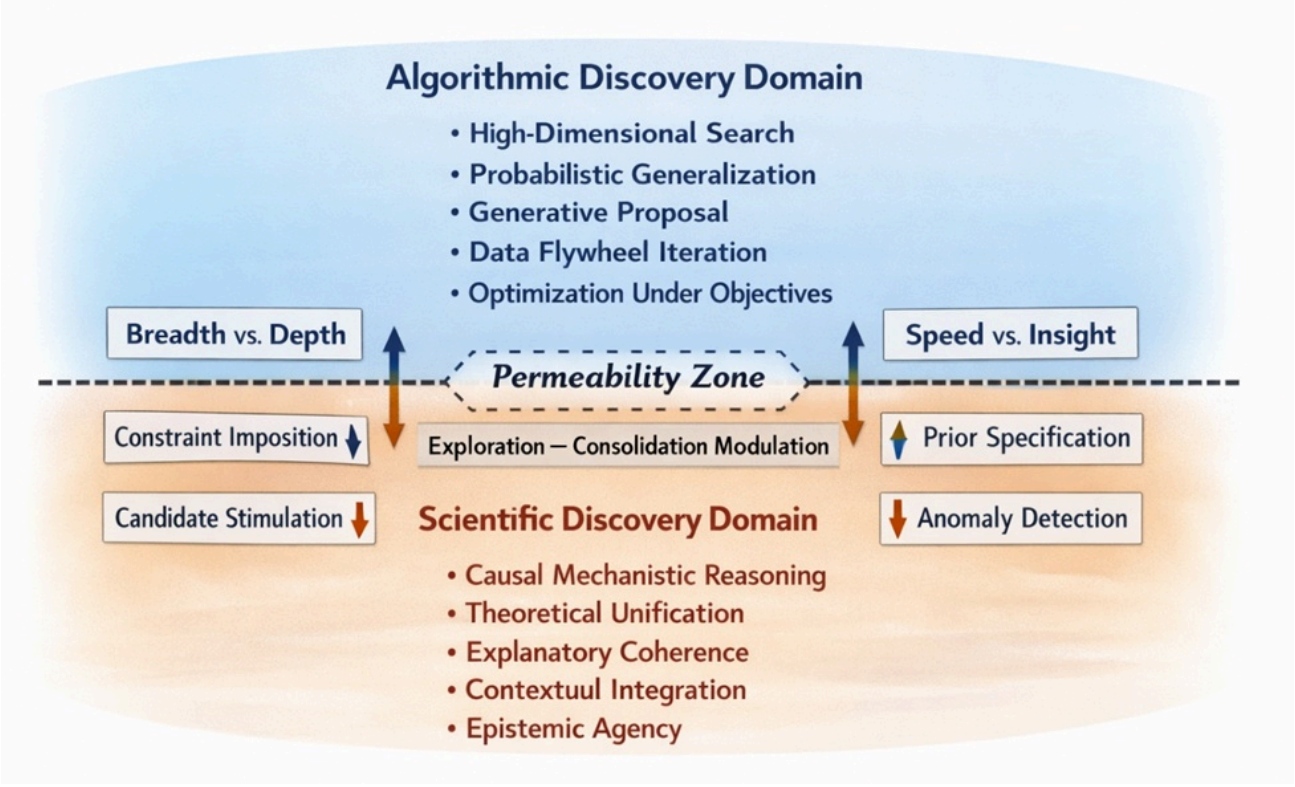

Figure 1. Boundary framework linking algorithmic and scientific discovery domains in AI-driven materials research.

Boundary governance and design implications for AI-Driven materials research

Reframing the relationship between algorithmic discovery and scientific discovery as a boundary condition rather than a methodological hierarchy has important consequences for how AI-driven materials research is designed, governed, and

evaluated. Rather than yielding narrow analytical implications, the framework foregrounds issues of epistemic governance—how discovery processes are structured, constrained, and legitimized across hybrid human–machine systems.

Boundary governance as an epistemic design problem

The first implication concerns governance at the discovery architecture level. When algorithmic and scientific discovery are treated as interchangeable outputs of a unified pipeline, the epistemic role of AI risks becoming overextended. The boundary framework instead clarifies that algorithmic systems govern search while scientific systems govern meaning. This distinction implies that governance should focus not on controlling algorithms in isolation, but on managing transitions across the boundary—specifically, when and how algorithmic outputs are elevated to scientific claims.

In practice, this shifts attention from performance metrics (e.g., prediction accuracy, novelty scores) to boundary criteria: What justifies moving from candidate generation to mechanistic interpretation? Which algorithmic outputs merit theoretical investment, and which should remain exploratory artifacts? These questions are not technical but epistemic, requiring explicit boundary rules embedded in workflow design. Without such rules, AI systems may accumulate vast catalogs of candidates without corresponding advances in scientific understanding, creating what may be termed discovery inflation—an expansion of outputs without commensurate explanatory consolidation.

Design implications for hybrid discovery workflows

The framework also reframes how hybrid workflows should be architected. Conventional AI-driven pipelines often emphasize linear progression: data → model → prediction → validation. The boundary perspective, in contrast, favors modular designs with explicit handoff points between algorithmic and scientific modalities. These handoff points—where algorithmic exploration pauses and scientific interpretation begins—are critical sites of epistemic decision-making.

Designing for such transitions implies that AI systems should not be optimized solely for throughput or novelty, but for interrogability. Algorithmic discovery systems that surface uncertainty structures, anomaly clusters, or representation shifts better support boundary crossings than systems optimized only for ranking or generation. Similarly, scientific workflows benefit from being designed to selectively absorb algorithmic outputs, treating them as prompts for explanation rather than conclusions. The implication is a move away from seamless automation toward deliberately interruptible systems that preserve space for human judgment and theoretical reflection.

Boundary sensitivity and the allocation of cognitive labor

A further implication concerns the allocation of cognitive labor between machines and human researchers. Algorithmic discovery redistributes labor by externalizing search and pattern recognition, while scientific discovery concentrates labor in interpretation, explanation, and synthesis. The boundary framework highlights that this redistribution is not neutral: excessive reliance on algorithmic breadth can shift scientific labor toward post-hoc rationalization, while excessive scientific constraint can reduce AI to a confirmatory tool. **Table 2** summarizes the functional roles played by the boundary between algorithmic and scientific discovery, illustrating how epistemic steering is enacted through bidirectional transitions rather than linear pipelines.

Table 2. Boundary functions and steering mechanisms linking algorithmic and scientific discovery

Boundary function	Description	Direction of influence	Typical mechanisms
Candidate filtration	Selection of algorithmic outputs for deeper scientific analysis	Algorithmic → Scientific	Uncertainty thresholds, anomaly detection, expert triage
Constraint	Injection of physical or theoretical	Scientific →	Physics-informed losses, symmetry

imposition	priors into algorithmic search	Algorithmic	constraints, domain priors
Epistemic validation	Determination of when outputs qualify as scientific knowledge	Scientific → Boundary	Mechanistic explanation, experimental corroboration
Exploration modulation	Regulation of breadth versus depth across discovery phases	Bidirectional	Active learning, exploration–exploitation balance
Paradigm challenge	Identification of outputs that conflict with established theory	Algorithmic → Scientific	Outlier analysis, counter-example detection
Knowledge consolidation	Integration of findings into coherent theoretical structures	Scientific → Knowledge base	Model revision, conceptual unification

These boundary functions underscore that governance in AI-driven materials research operates through selective transitions rather than unilateral control of either algorithmic or scientific components.

Boundary sensitivity—the ability to recognize when exploration should yield to explanation, and when explanation should give way to renewed exploration—thus becomes a core epistemic skill. This sensitivity cannot be automated; it must be cultivated through training, institutional norms, and shared conceptual frameworks. From this perspective, AI literacy in materials science is not merely technical proficiency but an understanding of where algorithmic discovery ends and scientific responsibility begins.

Implications for scientific legitimacy and attribution

The framework also carries implications for how discovery is attributed and legitimized. In AI-driven contexts, claims of discovery often blur the line between algorithmic suggestion and scientific confirmation. By maintaining a clear boundary, the framework supports a more precise attribution logic: algorithmic systems propose, scientific systems dispose. Discovery, in the full scientific sense, occurs only when algorithmic outputs are integrated into coherent explanatory structures.

This distinction matters for publication practices, credit assignment, and epistemic accountability. Treating algorithmic outputs as discoveries in their own right risks diluting scientific standards of explanation, while denying their epistemic contribution understates their transformative role. The boundary framework avoids both extremes by situating discovery as a distributed process with differentiated roles rather than a singular event.

Long-Term trajectories and paradigm evolution

Finally, the boundary perspective reframes how AI may influence long-term paradigm development in materials science. Algorithmic discovery tends to favor dense exploration of existing representational spaces, potentially reinforcing dominant paradigms through data amplification. Scientific discovery, by contrast, enables paradigm revision by reinterpreting anomalies and integrating disparate findings into new conceptual schemes.

The boundary functions as the site where these forces interact. When properly governed, it allows algorithmic anomalies to challenge established theories, enabling paradigm evolution rather than entrenchment. When poorly governed, it can either suppress novelty through premature constraint or overwhelm theory through uncontrolled expansion. The implication is that paradigm change in AI-augmented materials science will depend less on algorithmic sophistication than on how effectively the boundary between algorithmic and scientific discovery is maintained and navigated.

Results and Discussion

The framework developed herein deliberately refrains from prescriptive recommendations or operational directives. Instead, it clarifies the conceptual architecture through which AI-driven materials research currently unfolds,

foregrounding the boundary between algorithmic discovery and scientific discovery as a central organizing structure. This boundary is not merely a descriptive distinction but a locus of epistemic tension, where different virtues of knowledge production—scalability, speed, and probabilistic novelty on one side; explanatory depth, causal coherence, and theoretical integration on the other—are continuously negotiated [1, 2]. By making this boundary explicit, the framework provides a vocabulary for interpreting contemporary discovery practices without reducing them to either technical optimization or philosophical abstraction.

A key insight of the framework is that interaction across this boundary is neither symmetrical nor temporally fixed. Algorithmic discovery tends to dominate early-stage exploration, where the primary objective is expansive coverage of high-dimensional chemical or structural spaces [13, 20]. In these regimes, probabilistic generalization and large-scale candidate generation outperform human intuition, enabling forms of exploration that would otherwise be infeasible. As algorithmic outputs accumulate, however, the epistemic demands of discovery shift. The boundary becomes the site where scientific reasoning is progressively infused—through mechanistic interpretation, contextualization, and theoretical filtering—transforming candidate abundance into structured understanding [11, 18]. This transition is not automatic; it is mediated by steering mechanisms such as uncertainty quantification, active learning queries, and constraint propagation, which regulate when exploration yields to explanation and when renewed exploration is warranted [15].

The discussion also highlights the importance of boundary maintenance in preventing epistemic drift. In the absence of robust boundary conditions, AI-driven workflows risk pathological extremes. Over-optimization within algorithmic regimes may entrench spurious correlations, reinforce data-driven biases, or produce artifacts that are statistically compelling yet physically ungrounded [23, 25]. Conversely, premature or overly rigid imposition of scientific constraints can suppress genuine novelty, limiting algorithmic systems to rediscovering known regimes rather than probing the margins of existing theory. The permeable nature of the boundary mitigates these risks by enabling bidirectional correction: algorithmic outputs challenge scientific assumptions through anomalies and unexpected candidates, while scientific insights recalibrate algorithmic objectives by imposing revised priors, representations, or evaluation criteria [9, 27].

Importantly, the framework reframes debates about automation and human agency. Rather than treating AI as either a substitute for or a tool subordinate to human reasoning, the boundary perspective positions discovery as a distributed process with differentiated epistemic roles. Algorithmic systems externalize search and pattern recognition, while scientific systems retain responsibility for explanation, meaning, and legitimacy. This division of labor does not diminish human agency; instead, it relocates it to the boundary, where decisions about interpretation, validation, and theoretical integration are made. From this vantage point, human expertise is not displaced by AI but rendered more consequential, as it governs the transitions that convert algorithmic outputs into scientific knowledge.

In broader philosophical terms, the distinction articulated here resonates with longstanding discussions on scientific progress and modes of inference. Algorithmic discovery aligns with inductive, data-intensive traditions that emphasize pattern accumulation and statistical regularity. In contrast, scientific discovery retains deductive and abductive elements that anchor claims in causal narratives and theoretical frameworks [3, 10]. The contribution of this manuscript lies in showing that these modes need not compete for epistemic primacy. Instead, their interaction—structured through a consciously maintained boundary—expands the space of possible discoveries while preserving the standards by which discoveries are understood and justified.

Taken together, the discussion situates the proposed framework as an interpretive lens rather than a procedural blueprint. Its value lies in clarifying how AI reshapes discovery not by replacing scientific reasoning, but by reconfiguring the conditions under which reasoning operates. By attending to the boundary between algorithmic and scientific discovery, materials research can harness the generative power of AI while maintaining continuity with the explanatory aims that define science itself.

Conclusion

This manuscript has advanced a conceptual framework that delineates a boundary between algorithmic discovery and scientific discovery in AI-driven materials research. By characterizing their distinct epistemic modalities and examining how they interact through feedback, steering mechanisms, and selective interpretation, the framework clarifies how discovery unfolds across hybrid human–machine systems. Algorithmic processes contribute scale, speed, and probabilistic novelty, enabling expansive exploration of high-dimensional material spaces. Scientific processes confer depth, coherence, and causal intelligibility, transforming algorithmic outputs into durable knowledge. The boundary between these modalities, conceived as a permeable interface rather than a rigid divide, enables mutual calibration without epistemic subsumption.

Attending to this boundary is essential as materials research becomes increasingly data-intensive and automated. It highlights that the central challenge posed by AI is not technical performance alone, but the governance of transitions from search to explanation, from candidate generation to scientific understanding. By reframing discovery as a distributed process with differentiated epistemic roles, the framework preserves the centrality of human explanatory agency while recognizing the transformative capacity of algorithmic systems.

Beyond materials science, the conceptual structure developed here offers a transferable lens for analyzing AI-driven inquiry in other domains where data-driven exploration intersects with theory-driven understanding. As artificial intelligence continues to reshape knowledge production, maintaining clarity about the boundary between algorithmic generation and scientific discovery will remain critical for sustaining the explanatory aspirations that define science itself.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 21 Jun 2024 Revised: 25 Jul 2024 Accepted: 21 Aug 2024
Published online: 18 January 2025

References

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624(7990):80-5.
https://doi.org/10.1038/s41586-023-06735-9.

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

Goodall REA, Lee AA. Predicting materials properties without crystal structure: Deep representation learning from stoichiometry. Nat Commun. 2020;11(1):6280.
https://doi.org/10.1038/s41467-020-19964-7.

Choudhary K, DeCost B, Chen C, Jain A, Tavazza F, Cohn R, et al. Recent advances and applications of deep learning methods in materials science. npj Comput Mater. 2022;8(1):59.
https://doi.org/10.1038/s41524-022-00734-6.

Dan Y, Zhao Y, Li X, Xu G, Hu J, Hu J. Small data machine learning in materials science. npj Comput Mater. 2023;9(1):1-12.

Oviedo F, Ferres JL, Buonassisi T, Butler KT. Interpretable and explainable machine learning for materials science and chemistry. Acc Mater Res. 2022;3(9):889-903.

Xie T, Grossman JC. Machine learning for high-entropy alloys. Annu Rev Mater Res. 2021;51:339-65.

Zhang Y, Ling C. Transfer learning for small datasets in materials science. npj Comput Mater. 2021;7(1):1-10.

Ramprasad R, Batra R, Pilania G, Chen C, Ward L, Jain A. Machine learning in materials science. Nat Rev Mater. 2021;6(5):387-9.

Jablonka KM, Ongari D, Moosavi SM, Smit B. Big-data science in porous materials: The case of metal-organic frameworks. Chem Rev. 2020;120(16):7931-81.

Boiko DA, MacKnight R, Kline B, Košmrlj A, White AD. Autonomous chemical research with large language models. Nature. 2023;624(7992):570-8.

Zitnick L, Das A, Kolluri S, Shoaib M, Shoaib M, Zitnick L, et al. The open catalyst 2020 (oc20) dataset and competitions. J Chem Inf Model. 2021;61(12):5865-76.

Elton DC, Boukouvalas Z, Fuge MD, Chung PW. Applying machine learning to materials science. MRS Bull. 2020;45(5):374-80.

Meredig B. From screening to discovery: Rethinking the ai-driven materials workflow. Matter. 2020;3(3):559-61.

Chen C, Ong SP. Opportunities for next-generation machine learning in materials science. Nat Rev Mater. 2022;7(11):858-70.

Jain A, Hautier G, Ong SP, Persson K. New opportunities for high-throughput materials discovery from the materials project database. APL Mater. 2020;8(8):080401.

Lookman T, Balachandran P, Xue D, Yuan R, Guo H. Adaptive strategies for materials design using active learning. MRS Bull. 2020;45(5):381-9.

Pilania G, Gubernatis JE, Lookman T. Multi-fidelity machine learning models for accurate bandgap predictions of inorganic solids. Comput Mater Sci. 2020;173:109457.

Dan Y, Zhao Y, Li X, Hu J. Generative models for materials inverse design. Nat Rev Mater. 2021;6(5):387-9.

Pollice R, dos Passos Gomes G, Aldeghi M, Hickman RJ, Krenn M, Lavigne C, et al. Data-driven strategies for accelerated materials design. Acc Chem Res. 2021;54(4):849-60.

De Meulemeester H, De Moor B. Unsupervised embeddings for categorical variables. In: 2020 International Joint Conference on Neural Networks (IJCNN); 2020; IEEE. p. 1-8.

Oviedo F. Knowledge-integrated machine learning for materials: Lessons from catalysis. Nat Rev Mater. 2023;8:123-35.

Oviedo F, Ferres JL, Butler KT. Explainable machine learning in materials science. npj Comput Mater. 2022;8(1):204.

Chen C, Ye W, Ong SP. Accelerating the prediction of stable materials with machine learning. Nat Comput Sci. 2023;3(11):949-58.

Bai Y, Hu J, Chen C. Density of states prediction for materials discovery via contrastive learning. Nat Commun. 2022;13(1):925.

Lookman T, Balachandran P, Xue D, Yuan R, Guo H. Active learning in materials discovery. Annu Rev Mater Res. 2021;51:339-65.

Jablonka KM, Moosavi SM, Ongari D, Smit B. Data extraction for materials science. Patterns. 2022;3(5):100473.

Elton DC, Boukouvalas Z, Fuge MD, Chung PW. Machine learning for materials science. MRS Bull. 2021;46(5):374-80.

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624(7990):80-5.