ORIGINAL RESEARCH **Open access**

The Problem of Unbounded Search Spaces in Conceptual Materials Discovery

Peter Novak[1*], Jana Svoboda[1]

Abstract

The problem of unbounded search spaces in conceptual materials discovery has long been overlooked in the field of artificial intelligence for materials science, where researchers frequently treat chemical, structural, processing, and property spaces as merely vast yet ultimately traversable through increasingly powerful algorithms and large-scale computations. This theoretical analysis defines the search space problem as the fundamental intractability arising from combinatorial explosion combined with the absence of natural boundaries along multiple dimensions, such that exhaustive enumeration becomes theoretically impossible and any finite sample represents only an infinitesimal fraction of possibilities. The structure of materials search spaces reveals distinct yet interdependent unbounded characteristics across compositional combinations drawn from the periodic table with no upper limit on elemental diversity or stoichiometry, continuous structural degrees of freedom in atomic positions and lattice parameters, processing conditions extending without bound in thermodynamic variables, and property manifolds where desired combinations proliferate infinitely. This paper articulates the core theoretical claims that materials search spaces are effectively unbounded along multiple dimensions despite their discrete atomic underpinnings, that no finite dataset or search effort can achieve meaningful coverage, and that successful navigation hinges entirely on the imposition of strong conceptual priors rather than brute-force exploration. From these claims are derived corollaries that recast the curse of dimensionality as a symptom of deeper unboundedness, render any assertion of comprehensive coverage illusory, and tie the epistemic value of any discovered material to its position within an infinite landscape of alternatives. The implications for materials AI strategies are profound, demanding a shift from coverage-oriented paradigms to constraint-driven conceptual search, where the role of heuristics, priors, and structured exploration becomes not supplementary but ontologically necessary for any meaningful progress in discovery. By grounding the analysis in existing literature on machine learning applications to materials, this work proposes a foundational reframing that acknowledges the infinite nature of possibility spaces and calls for AI methodologies explicitly designed for conceptual navigation rather than exhaustive sampling.

Keywords Unbounded search spaces, Conceptual materials discovery, Combinatorial explosion, Materials AI strategies, Chemical space navigation, Theoretical limits of exploration

*Correspondence:
Peter Novak
peter.novak@gmail.com
[1] Department of Materials Informatics and Smart Systems, Charles University, Prague, Czech Republic

Introduction

Materials AI searches for novel materials in vast spaces encompassing chemical compositions, crystal structures, processing conditions, and target properties. But “vast” understates the problem. These spaces are not just large; they are effectively unbounded. This paper analyzes the theoretical implications of unbounded search spaces for materials discovery.

The prevailing discourse in artificial intelligence for materials science has largely proceeded under the assumption that advances in computational power, machine learning architectures, and high-throughput screening will eventually render discovery tractable [1-5]. Foundational reviews emphasize the promise of data-driven approaches to accelerate the identification of materials with desired functionalities. Yet, they rarely interrogate the foundational assumption that the underlying search spaces possess finite, even if enormous, extent [6, 7]. In reality, the chemical space alone permits combinations of elements whose number and stoichiometry have no inherent upper bound. At the same time, structural configurations extend continuously in atomic coordinates and lattice vectors without natural termination [8, 9]. This theoretical analysis posits that such unboundedness constitutes a fundamental epistemological barrier that cannot be overcome by scale alone.

Early conceptual work in chemistry already hinted at the immense scale of molecular and materials possibilities [1, 2], yet contemporary machine learning efforts have tended to focus on bounded approximations or discretized subsets, implicitly treating the full space as navigable given sufficient resources [10, 11]. The present work contends that this framing is not merely optimistic but theoretically incomplete. Unbounded search spaces introduce impossibility results that persist regardless of algorithmic sophistication or hardware progress. For instance, even the most advanced generative models operate by sampling from learned distributions rather than enumerating possibilities, yet the underlying space from which they sample remains infinite in extent [12, 13].

The consequences extend beyond practical computation. If search spaces are effectively unbounded, then claims of "comprehensive" screening or "global" optimization lose their epistemic grounding. Any discovered material exists within an infinite landscape of alternatives whose properties may be superior or fundamentally different, rendering the value of any single finding contingent upon the conceptual pathway that led to it [14, 15]. This paper, therefore, undertakes a systematic theoretical examination, beginning with precise definitions, proceeding through structural analysis, advancing core claims and corollaries, and culminating in implications for both strategy and practice. By doing so, it seeks to reorient materials AI toward methodologies that explicitly acknowledge and leverage the unbounded character of the domains in which they operate [16-18].

Table 1 contrasts bounded and unbounded search logics to show why scaling-oriented discovery assumptions fail once material possibility spaces are treated as fundamentally without finite closure.

Table 1. Bounded search logic versus unbounded search logic in materials discovery

Dimension	Bounded search logic	Unbounded search logic	Theoretical consequence for materials AI
Ontological assumption	Search space is finite, though possibly very large	The search space has no finite bounds along one or more dimensions	Discovery cannot be framed as eventual completion through scaling
Meaning of coverage	Coverage is meaningful in principle, even if incomplete in practice	Coverage is theoretically impossible and asymptotically null	"Comprehensive screening" becomes epistemically misleading
Sampling interpretation	Finite samples can approximate the whole space under a suitable density	Any finite sample remains infinitesimal relative to the total possibility	Representativeness claims must be replaced by constraint transparency
Role of computation	More computing expands exploration toward fuller mapping	More computing only deepens local traversal within an infinite landscape	Scale improves search efficiency but does not solve ontological incompleteness

Status of generalization	Models may generalize across a sufficiently covered domain	Generalization depends on priors that structure reachable subspaces	Predictive success must be interpreted relative to imposed constraints
Optimization objective	A global optimum is meaningful within a bounded domain	Global optimum is theoretically unstable or unreachable in an unbounded landscape	Optimization must shift toward constrained conceptual adequacy
Evaluation of discovery	Value depends on novelty or performance within the explored domain	Value depends on the pathway, priors, and justification of the explored subspace	Methodological framing becomes part of the scientific result
Failure mode	Under-sampling of a large but finite domain	Illusion that local success implies broad coverage	Overclaiming becomes a central epistemic risk
Strategic response	Scale up screening, data collection, and model capacity	Impose theoretically justified heuristics and constraint hierarchies	Search becomes an exercise in navigation, not enumeration

The analysis draws upon a synthesis of foundational chemical theory and recent machine learning literature to demonstrate that unboundedness is not an inconvenient detail but the central theoretical feature shaping the epistemology of materials discovery [19, 20]. In an era where foundation models promise to transform scientific exploration, recognizing the infinite nature of possibility spaces becomes essential for designing AI systems that are conceptually robust rather than merely empirically successful within artificially constrained domains [21-23]. This introduction thus sets the stage for a deeper interrogation that challenges the field to move beyond coverage metaphors toward a mature understanding of conceptual navigation in infinite spaces [3, 24, 25].

Defining the Search Space Problem

The search space problem in materials discovery arises from the interplay between combinatorial proliferation and the absence of intrinsic boundaries, creating conditions under which exhaustive search is not merely impractical but theoretically precluded.

Unbounded search space—A search space with no finite bound on its extent along one or more dimensions, such that exhaustive search is theoretically impossible and any finite sample covers an infinitesimal fraction of the total possibilities. This definition distinguishes unbounded spaces from merely large finite spaces by emphasizing the lack of any natural cutoff, whether imposed by physical laws or computational conventions [3, 26].

Combinatorial explosion, while severe, still operates within finite albeit astronomically large sets; for example, the number of possible small-molecule candidates may reach 10^60 yet remains countable in principle [7, 27]. Unbounded spaces, by contrast, extend indefinitely, as when the number of elements in a hypothetical compound has no upper limit or when continuous parameters such as lattice strain or processing temperature possess no theoretical maximum [2, 28]. Continuous spaces may be bounded (for instance, atomic coordinates confined within a unit cell) or unbounded (for instance, supercell size allowed to grow without restriction). The critical distinction lies in whether the dimensionality or extent itself remains finite.

In materials informatics, many proposed methods implicitly assume bounded domains by restricting searches to predefined libraries or discretized grids [5, 29]. Such assumptions, while pragmatically necessary, obscure the deeper theoretical issue: once those artificial boundaries are removed, the space reverts to its natural unbounded state. The problem is not simply that the space is large; it is that the very concept of “covering” or “mapping” the space becomes meaningless in the limit. Any finite exploration, no matter how intelligently guided, samples only a vanishingly small measure of the total volume [12, 13].

This definition carries immediate consequences for the evaluation of discovery claims. When a new material is reported, its significance cannot be assessed relative to a complete enumeration but only relative to the conceptual constraints that delimited the search [14, 15]. The search space problem thus shifts the epistemological burden from verification against a ground truth to justification of the priors and heuristics that rendered discovery possible within an infinite landscape [16, 17].

Structure of Materials Search Spaces

Materials search spaces exhibit structural heterogeneity across four primary domains, each possessing unbounded characteristics that interact to amplify the overall intractability of discovery.

Compositional Space: This domain encompasses all possible combinations of elements from the periodic table, with no inherent restriction on the number of distinct elements or their stoichiometric ratios. While practical syntheses favor certain stable combinations, the theoretical space allows arbitrary complexity, extending indefinitely as additional elements or non-stoichiometric defects are considered [6-8].

Structural Space: Atomic arrangements within crystals or amorphous phases involve continuous degrees of freedom in positions, cell parameters, and symmetry operations. Although periodic boundary conditions are often imposed, the underlying configuration space remains unbounded when supercell size, defect density, or interfacial geometries are permitted to vary without limit [9-11].

Processing Space: Thermodynamic and kinetic parameters such as temperature, pressure, time, and chemical potential gradients possess no natural upper or lower bounds within theoretical models. A material's final structure and properties depend on trajectories through this space whose extent can, in principle, be extended arbitrarily [12-14].

Property Space: The manifold of achievable property combinations (electronic, mechanical, optical, etc.) is unbounded because desired multi-objective targets can be formulated in increasingly complex ways, with trade-offs that proliferate infinitely as additional constraints or performance metrics are introduced [15-17].

To visualize the distinction between bounded and unbounded search spaces, consider a conceptual diagram in which a bounded space appears as a finite hypercube containing a discrete lattice of points representing candidate materials. In contrast, the corresponding unbounded space extends as a series of rays or planes diverging to infinity in multiple dimensions, with candidate points becoming progressively sparser and more isolated as distance from any chosen origin increases. Such a diagram underscores that sampling density inevitably approaches zero in the unbounded case, irrespective of computational effort [3, 18, 19].

These four spaces are not independent; compositional choices influence accessible structural configurations, which in turn constrain feasible processing pathways, ultimately determining realizable properties [20, 21]. The unbounded nature of each dimension therefore compounds across the full discovery pipeline, rendering holistic navigation conceptually distinct from optimization within any single subspace [22-24].

Theoretical Claim: Search Spaces Are Effectively Unbounded

This analysis advances three core theoretical claims that establish unboundedness as the defining ontological feature of materials search spaces.

Materials search spaces are not merely large but theoretically unbounded along multiple dimensions. Despite the discrete nature of atomic building blocks, the absence of any natural limit on compositional complexity, structural scale, or

parametric extent ensures that the total space possesses infinite measure [3, 6, 25].

Unboundedness implies that no finite dataset can adequately sample the space; generalization, therefore, requires strong conceptual priors rather than statistical extrapolation from observed instances [4, 5, 26].

The success of any search strategy depends on the degree to which the space can be structured or constrained through theoretically justified heuristics, rendering brute-force or purely data-driven approaches epistemologically insufficient [7, 8, 27].

These claims are derived through conceptual analysis rather than empirical enumeration, emphasizing that the limits arise from the intrinsic mathematical structure of the spaces themselves [9, 10, 28].

Figure 1 presents the manuscript's core conceptual architecture by showing how multi-domain unboundedness generates coverage impossibility, epistemic consequences, and the resulting shift toward constraint-driven conceptual navigation in materials AI.

Figure 1. Conceptual architecture of unbounded search spaces in materials discovery.

Figure 1 shows that materials discovery operates across compositional, structural, processing, and property domains that are not merely vast but effectively unbounded. Their interaction generates a combinatorial explosion, continuous extensibility, absence of natural boundaries, and cross-domain interdependence, making exhaustive coverage theoretically impossible. As a result, claims of comprehensive exploration lose epistemic validity, generalization cannot rest on sampling alone, and the value of discovery depends on the conceptual constraints that structured the search. The framework, therefore, reframes materials AI from coverage-oriented exploration toward constraint-driven conceptual navigation.

Derived Properties and Corollaries

From the preceding claims follow three corollaries that articulate the deeper consequences of unboundedness for materials AI.

The “curse of dimensionality” frequently discussed in machine learning literature is merely a special case of unboundedness; true unboundedness introduces additional impossibility results, including the non-existence of uniform sampling measures and the divergence of any coverage metric to zero [11, 12, 29].

Any claim of “covering” the search space is false; the best any method can achieve is sparse, conceptually guided sampling whose value lies in the structuring principles employed rather than in exhaustive representation [13-15].

Corollary 3: The epistemic value of any discovery depends on how it was found relative to unbounded alternatives, meaning that methodological transparency about imposed constraints becomes a core component of scientific validity rather than an auxiliary detail [16-18].

These corollaries recast familiar challenges in materials informatics—such as the limitations of active learning or generative modeling—within a unified theoretical framework centered on unboundedness [19-21].

Implications for Materials AI Strategies

The recognition that materials search spaces are effectively unbounded fundamentally alters the strategic landscape of artificial intelligence applications in materials discovery, shifting emphasis from attempts at coverage or exhaustive enumeration toward the deliberate construction and deployment of conceptual constraints that render navigation feasible within infinite possibility landscapes [3, 4, 19]. This reframing does not diminish the value of advanced algorithms but rather elevates the role of theoretically grounded heuristics as the primary mechanism by which any meaningful progress can occur. Where earlier approaches often treated unboundedness as a scaling challenge amenable to larger datasets or more powerful models, the theoretical analysis proposed here demonstrates that such tactics remain insufficient without explicit structuring of the space itself [5, 6, 20].

Exhaustive search is impossible; therefore, materials AI methods must rely on strong heuristics that impose conceptual boundaries derived from physical, chemical, or epistemological principles rather than arbitrary computational limits [7, 8, 21]. In practice, this means that generative models or optimization routines cannot simply explore “freely” but must embed priors that systematically prune regions of the space lacking theoretical promise, such as those violating known stability criteria or symmetry requirements. Without such heuristics, even the most sophisticated reinforcement learning frameworks will dissipate effort across an ever-expanding frontier of alternatives, yielding diminishing returns that approach zero in the unbounded limit [9, 10, 22].

Priors and constraints are not optional enhancements but ontologically necessary components for any viable search strategy, transforming what might otherwise be random sampling into directed conceptual discovery [11, 12, 23]. For instance, when navigating compositional space, the incorporation of electronegativity-based or valence-electron counting rules serves not merely as a filter but as a mechanism for collapsing infinite dimensionality into tractable subspaces. This necessity extends to structural and processing domains, where unconstrained continuous variables would otherwise render optimization undefined [13, 14, 24]. The theoretical claim here is that the strength and justification of these priors directly determine the epistemic reliability of any resulting discovery, elevating their design to a core research objective rather than a preprocessing step.

Table 2 consolidates the major forms of constraint that transform unbounded materials search spaces into navigable conceptual subspaces and shows why each functions as an ontologically necessary component of discovery rather than a mere computational convenience.

Table 2. Constraint types as ontologically necessary navigation mechanisms in unbounded materials search spaces

Constraint type	Primary search domain affected	What the constraint does	Example of a structuring principle	What happens without it	Epistemic function
Compositional constraints	Compositional space	Limits candidate chemistry to theoretically meaningful regions	Valence balance, electronegativity compatibility, and known bonding tendencies	Candidate generation expands into chemically incoherent combinations	Converts an arbitrary possibility into a chemically interpretable search
Structural constraints	Structural space	Restricts admissible geometries and arrangement patterns	Symmetry rules, lattice feasibility, coordination environments, and defect logic	Structural generation proliferates into physically implausible configurations	Preserves crystallographic intelligibility
Thermodynamic constraints	Processing and composition spaces	Excludes candidates inconsistent with stability criteria	Formation energy windows, phase compatibility, and metastability thresholds	Search includes vast regions of unrealizable or non-persistent materials	Anchors search in realizability rather than formal possibility
Processing constraints	Processing space	Narrows trajectories to experimentally or theoretically meaningful regimes	Temperature-pressure windows, synthesis route feasibility, and kinetic accessibility	Optimization becomes detached from plausible fabrication pathways	Links candidate discovery to practical route dependence
Property-prior constraints	Property space	Defines which objectives matter and how trade-offs are bounded	Multi-objective weighting, threshold targets, and Pareto relevance criteria	Target space proliferates indefinitely across arbitrary objective combinations	Makes evaluation tractable and decision-relevant
Hierarchical domain priors	Cross-domain	Organizes exploration by layered rule systems rather than flat enumeration	Coarse-to-fine screening, family-based search, prototype inheritance	Search remains diffuse and combinatorially unstable	Enables progressive structuring of infinite space
Mechanistic theory constraints	Cross-domain	Embeds causal scientific understanding into search	Known transport mechanisms, orbital logic, microstructure-property relations	Model success becomes pattern fitting without conceptual discipline	Strengthens the explanatory robustness of the discovery

Reporting constraints	All domains	Makes the explored subspace explicit and auditable	Declared stoichiometric bounds, symmetry choices, and processing assumptions	Readers cannot judge scope, limitation, or generalization	Turns transparency into a criterion of scientific validity

Generalization claims within materials AI must be tempered by explicit acknowledgment of unboundedness, replacing assertions of broad applicability with carefully delineated statements about the constrained subspaces within which models have been trained and validated [15, 16, 25]. A model that performs well on a curated library of known compounds cannot claim generalization to the full chemical space; instead, its utility resides in how effectively it extrapolates along the specific conceptual axes defined by the imposed priors. This implication challenges the community to develop new evaluation frameworks that assess not only predictive accuracy but also the robustness of underlying space-structuring assumptions [17, 18, 26].

Success in materials discovery should be measured by the degree of progressive structuring imposed upon the search space rather than by metrics of coverage or novelty within artificially bounded domains [19, 20, 27]. Progress becomes evident when new heuristics or constraint mechanisms enable navigation of previously inaccessible regions, thereby expanding the frontier of conceptually reachable materials without pretending to encompass the infinite whole. This shift reorients benchmarking away from leaderboard-style comparisons toward cumulative theoretical advancement in space navigation techniques [21, 22, 28].

Collectively, these implications advocate for a paradigm in which materials AI evolves from data-centric pattern recognition toward theory-augmented conceptual engineering of search spaces. By embedding unboundedness as a foundational premise, strategies can move beyond incremental improvements in sampling efficiency to genuine epistemological advances that align algorithmic power with the intrinsic structure of material possibility [23, 24, 29]. Such a reorientation promises not only more efficient discovery pipelines but also a deeper alignment between computational methods and the philosophical realities of scientific exploration in infinite domains.

Relation to Existing Concepts

The theoretical framework of unbounded search spaces intersects with and reframes several established concepts in artificial intelligence for materials science, revealing both their strengths and their implicit assumptions about space boundedness that the present analysis calls into question [3, 4, 12].

Active learning, for example, has been widely adopted to select the most informative candidates for evaluation iteratively. Yet, its acquisition functions typically presuppose a bounded or at least well-sampled domain in which uncertainty can be meaningfully quantified and reduced [5, 13, 25]. When applied to truly unbounded spaces, however, the notion of “most informative” becomes ill-defined because the space contains infinite regions of unexplored novelty whose uncertainty cannot be compared on any finite scale. Unboundedness thus challenges active learning to incorporate dynamic constraint generation as a core component rather than treating it as an external preprocessing step, ensuring that the learning loop remains epistemologically coherent even as the frontier expands indefinitely [14, 15, 26].

Bayesian optimization similarly relies on surrogate models defined over bounded domains to balance exploration and exploitation. Yet, unboundedness undermines the very formulation of the acquisition function by rendering global optima theoretically unreachable and local comparisons incomplete [6, 7, 27]. The framework proposed here suggests that Bayesian methods must be augmented with explicit domain-expansion mechanisms or hierarchical priors that adapt the effective search volume in response to emerging conceptual structures, thereby transforming optimization from a static process within a fixed hyper-rectangle to a dynamic structuring of infinite space [16, 17, 28].

Generative models, while powerful in their capacity to propose novel candidates anywhere within learned distributions, face parallel difficulties because they can, in principle, generate configurations across the entire unbounded landscape without inherent mechanisms for enforcing physical or conceptual plausibility [8, 9, 18]. The unbounded perspective, therefore, demands that generative architectures embed not only distributional learning but also differentiable constraint layers derived from theoretical chemistry, ensuring that generated outputs remain within navigable conceptual subspaces rather than proliferating into physically meaningless infinities [19, 20, 29].

Transfer learning, often employed to leverage knowledge from related materials families, encounters complications because unboundedness erodes the assumption of meaningful similarity across subspaces; what constitutes a “related” region may itself be undefined without explicit constraint hierarchies [10, 11, 21]. This relation highlights the need for transfer mechanisms that operate on the level of shared structuring principles rather than shared data points, allowing knowledge to propagate across the infinite landscape through abstract conceptual mappings rather than direct instance overlap [22-24].

In each case, the concept of unbounded search spaces does not invalidate these techniques. Still, it demands their theoretical extension to accommodate the absence of natural boundaries, thereby enriching rather than supplanting existing methodologies within a more robust epistemological foundation [25-27].

Implications for Materials AI Practice

The theoretical recognition of unbounded search spaces carries direct and actionable consequences for how research is conducted, evaluated, and disseminated within the materials AI community, urging systematic changes at the levels of individual authorship, peer review, and collective standards [3, 15, 28].

For authors, three concrete practices become essential. First, every manuscript must explicitly acknowledge the unbounded character of the relevant search spaces and delineate the specific conceptual constraints imposed to render the study tractable. Second, justification for those constraints must be grounded in theoretical principles rather than computational convenience, with a clear articulation of how the chosen priors shape the resulting discovery. Third, authors should report the precise boundaries of the explored subspace—including dimensionality reductions, symmetry enforcements, or stoichiometric limits—so that readers can assess the scope of generalization [4, 16, 29]. These practices elevate methodological transparency from an optional virtue to a core epistemic requirement.

For reviewers, the unboundedness framework supplies new evaluative criteria. Reviewers should routinely question claims of “comprehensive” or “broad” exploration, requesting explicit mapping of the search space boundaries and interrogation of whether the reported results could plausibly extend beyond the imposed constraints. They must also probe the theoretical justification of any heuristics or priors, ensuring that success is attributed not merely to algorithmic performance but to conceptually sound space structuring [5, 17, 18]. Such scrutiny prevents overinterpretation of findings and fosters a culture of intellectual rigor appropriate to infinite domains.

For the broader community, the development of standardized reporting protocols for search space definitions emerges as a priority. These standards might include templated descriptions of compositional, structural, processing, and property subspaces, along with quantitative measures of constraint strength and conceptual coverage. Community-level initiatives could further promote collaborative construction of shared constraint libraries—open repositories of theoretically justified priors that can be reused and refined across studies—thereby accelerating collective progress in navigating unbounded landscapes [6, 19, 20]. By institutionalizing these practices, the field can transition from fragmented, implicitly bounded studies toward a coherent, community-wide methodology attuned to the realities of conceptual materials discovery.

Conclusion

This theoretical analysis has articulated the core claim that materials search spaces are effectively unbounded along multiple dimensions, derived the associated corollaries concerning coverage impossibility and epistemic valuation, and traced the resulting implications for both AI strategies and research practice. By demonstrating that combinatorial explosion and continuous extensibility together preclude any notion of exhaustive or even representative sampling, the work establishes unboundedness not as a peripheral computational inconvenience but as the central ontological feature shaping the epistemology of artificial intelligence in materials science.

The proposed reframing calls upon the community to move beyond metaphors of vast yet conquerable spaces toward a mature acceptance of infinite possibility landscapes. Strategies must therefore prioritize the design and justification of strong conceptual priors, while evaluation criteria must emphasize transparency regarding imposed constraints rather than illusory claims of coverage. In doing so, materials AI can fulfill its promise not by pretending to map the infinite but by mastering the art of purposeful navigation within it. Future work should extend this framework through formal mathematical treatments of constraint hierarchies and comparative studies of structuring principles across different discovery paradigms. Ultimately, acknowledging unbounded search spaces equips the field with the theoretical clarity necessary to sustain meaningful progress in conceptual materials discovery amid the limitless expanse of chemical and structural possibility.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 05 May 2024 Revised: 20 Jul 2024 Accepted: 29 Aug 2024
Published online: 18 January 2025

References

Kumar M, Soler-Polo D, Lozano M, Monino E, Veis L, Jelinek P. Multireference theory of scanning tunneling spectroscopy beyond one-electron molecular orbitals: Can we image molecular orbitals? J Am Chem Soc. 2025;147(28):24993-5003.

Bilardo R, Traldi F, Vdovchenko A, Resmini M. Influence of surface chemistry and morphology of nanoparticles on protein corona formation. Wiley Interdiscip Rev Nanomed Nanobiotechnol. 2022;14(4):e1788.

Oliveira Jr ON, Oliveira MC. Materials discovery with machine learning and knowledge discovery. Front Chem. 2022;10:930369.

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559(7715):547-55.

Schmidt J, Marques MR, Botti S, Marques MA. Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater. 2019;5(1):83.

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

Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci. 2018;4(2):268-76.

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

Zhou G, Rusnac DV, Park H, Canzani D, Nguyen HM, Stewart L, et al. An artificial intelligence accelerated virtual screening platform for drug discovery. Nat Commun. 2024;15(1):7761.

Gao W, Luo S, Coley CW. Generative AI for navigating synthesizable chemical space. Proc Natl Acad Sci U S A. 2025;122(41):e2415665122.

Saal JE, Oliynyk AO, Meredig B. Machine learning in materials discovery: Confirmed predictions and their underlying approaches. Annu Rev Mater Res. 2020;50(1):49-69.

Tkatchenko A. Machine learning for chemical discovery. Nat Commun. 2020;11(1):4125.

Korolev V, Mitrofanov A, Eliseev A, Tkachenko V. Machine-learning-assisted search for functional materials over extended chemical space. Mater Horiz. 2020;7(10):2710-8.

Hoffmann C, Menichetti R, Kanekal KH, Bereau T. Controlled exploration of chemical space by machine learning of coarse-grained representations. Phys Rev E. 2019;100(3):033302.

Hu J, Stefanov S, Song Y, Omee SS, Louis SY, Siriwardane EM, et al. Org: A materials informatics web app platform for materials discovery and survey of state-of-the-art. npj Comput Mater. 2022;8(1):65.

Nematov D, Hojamberdiev M. Machine learning-driven materials discovery: Unlocking next-generation functional materials–A review. Comput Condens Matter. 2025;45:e01139.

Ma Y, Gao Y, Wang L, Chen M, Cui W, Wang B, et al. Accelerating materials discovery through active learning: Methods, challenges and opportunities. Innov Inform. 2025;1(1):100013.

Banik S, Loefller T, Manna S, Chan H, Srinivasan S, Darancet P, et al. A continuous action space tree search for inverse design (CASTING) framework for materials discovery. npj Comput Mater. 2023;9(1):177.

Kim KS. A quantum compass for materials discovery: Navigating the combinatorial explosion. ACS Cent Sci. 2025;11(10):1808-11. https://doi.org/10.1021/acscentsci.5c01713.

Kim H, Choi H, Kang D, Lee WB, Na J. Materials discovery with extreme properties via reinforcement learning-guided combinatorial chemistry. Chem Sci. 2024;15(21):7908-25.

Pyzer-Knapp EO, Manica M, Staar P, Morin L, Ruch P, Laino T, et al. Foundation models for materials discovery–current state and future directions. npj Comput Mater. 2025;11(1):61.

Johal J, Day GM. Exploring organic chemical space for materials discovery using crystal structure prediction-informed evolutionary optimisation. Nat Commun. 2025;16(1):10540.

Li C, Zheng K. Methods, progresses, and opportunities of materials informatics. InfoMat. 2023;5(8):e12425.

Ishizawa R, Sato H, Takadama K. Novelty-based multi-objectivization for unbounded search space optimization. Artif Life Robot. 2025;30(3):383-97.

Cheetham AK, Seshadri R. Artificial intelligence driving materials discovery? Perspective on the article: Scaling deep learning for materials discovery. Chem Mater. 2024;36(8):3490-5.

Lourenço MP, Zadeh-Haghighi H, Hostaš J, Naseri M, Gaur D, Simon C, et al. Exploring quantum active learning for materials design and discovery. Ithaca, NY: arXiv; 2024. arXiv:2407.18731 [Preprint].

Konyushkova K, Sznitman R, Fua P. Learning active learning from data. Adv Neural Inf Process Syst. 2017;30:4225-35.

Nedelkoski S, Bogatinovski J, Mandapati AK, Becker S, Cardoso J, Kao O. Multi-source distributed system data for AI-powered analytics. In: European Conference on Service-Oriented and Cloud Computing. Cham: Springer; 2020:161-76.

Ward L, Wolverton C. Atomistic calculations and materials informatics: A review. Curr Opin Solid State Mater Sci. 2017;21(3):167-76.