

REVIEW

Open access

A Decade of Data-Driven Materials Engineering: Progress, Remaining Gaps, and Unlearned Lessons

Mohamed Salah^{1*}, Youssef Karim¹, Ahmed Nabil², Mahmoud Adel¹, Karim Hassan²

Abstract

Machine learning has transformed materials engineering through graph neural networks, equivariant architectures, generative models, and early autonomous laboratories. These advances enabled accurate property prediction, data-efficient force fields, and initial inverse design of inorganic crystals, supported by community databases such as the Materials Project and JARVIS. Generative frameworks now propose novel structures, while closed-loop platforms demonstrate early integration of prediction, synthesis, and robotic experimentation. However, major gaps remain: weak extrapolation beyond training distributions, inadequate handling of long-range interactions, poorly calibrated uncertainty quantification, limited synthesis prediction, underrepresentation of disordered materials, and insufficient experimental validation. Persistent unlearned lessons include benchmark biases that inflate generalization performance, poor reproducibility practices, misuse of invariant models for tensor properties, omission of random baselines in active learning, weak validity filtering in generative workflows, and near-absent failure reporting. This review critically assesses a decade of progress and shortcomings, grounded in 38 peer-reviewed publications. It highlights that hype around foundation models, generative inverse design, and autonomous labs often exceeds demonstrated impact. Realizing the field's potential requires prioritizing extrapolation-focused architectures, mandatory experimental validation, shared failure registries, and stronger evidentiary standards to accelerate reliable discovery of functional materials.

Keywords Materials informatics, Generative models, Active learning, Graph neural networks, Machine learning for materials, Data-driven materials engineering

*Correspondence:

Mohamed Salah
mohamed.salah@gmail.com

¹ Department of Computational Materials Systems, Faculty of Engineering, Cairo University, Cairo, Egypt

² Department of Materials Data Analytics, Faculty of Technology, Ain Shams University, Cairo, Egypt

Introduction

Recent years have seen remarkable advancements in data-driven materials engineering. From early GNNs like CGCNN and SchNet [1–3] to equivariant architectures [4, 5] to generative models [6, 7], the field has transformed [8]. This review assesses progress over the decade, identifies remaining gaps, and highlights unlearned lessons — insights that have been repeatedly demonstrated but not yet absorbed into common practice. We cover architectures, property prediction, force fields, generative models, active learning, and experimental validation.

Data-driven approaches have complemented traditional density-functional theory by enabling rapid screening of vast compositional spaces that would be intractable with first-principles methods alone [9, 10]. The introduction of graph-based representations allowed models to learn directly from atomic connectivity, eliminating the need for many hand-crafted descriptors that previously limited transferability [1, 11, 12]. Community-scale databases quickly became essential

infrastructure, providing standardized training and test sets that facilitated direct comparison of methods [13, 14]. By the end of the decade, thousands of publications cited these resources as the foundation for model training and benchmarking.

Progress was not linear. The period 2017–2019 established the dominance of GNNs for crystalline systems and laid the groundwork for ML force fields [1, 9]. The subsequent years 2020–2022 brought the recognition that physical symmetries must be explicitly encoded, leading to equivariant architectures that achieved state-of-the-art accuracy with far smaller datasets [4]. From 2023 onward, the focus shifted toward generative modeling and closed-loop discovery, with diffusion models proposing entirely new crystal structures and early autonomous platforms integrating prediction with synthesis [6, 7, 15, 16]. These developments were accompanied by growing interest in multimodal models that incorporate text from the literature alongside structural data [10, 17, 18].

Despite the impressive volume of research, the field remains fragmented. Many studies report incremental improvements on narrow tasks while overlooking broader challenges such as extrapolation and synthesis [19, 20]. Reviews published throughout the decade repeatedly flag the same issues—benchmark biases, reproducibility shortfalls, and insufficient experimental grounding—yet these concerns have not translated into widespread changes in practice [21–24]. This review therefore adopts a dual mandate: to synthesize what has been achieved and to confront what has not.

Figure 1 maps the review's central argument by showing how decade-scale methodological advances coexist with unresolved gaps, unlearned lessons, hype inflation, and the infrastructural reforms required for reliable discovery.

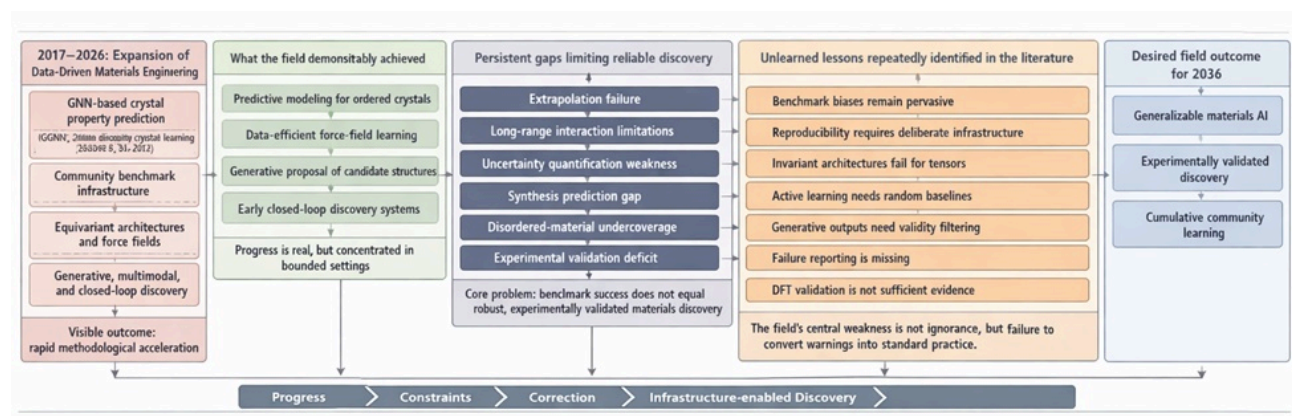


Figure 1. Analytical architecture of a decade of data-driven materials engineering: progress, unresolved gaps, unlearned lessons, hype correction, and the infrastructural path to reliable discovery

Progress: What We Achieved

2017–2019: The GNN revolution The period 2017–2019 witnessed the rapid establishment of graph neural networks as the dominant architecture for materials property prediction. CGCNN and SchNet demonstrated that crystal graphs could directly encode atomic environments and predict formation energies, band gaps, and elastic properties with accuracy rivaling DFT [1, 2]. MEGNet further showed that the same framework could be extended to multiple fidelities and different target properties within a single model [2]. These early successes were enabled by the Materials Project, which provided a standardized, publicly accessible benchmark dataset of over 100 000 inorganic compounds [13, 14]. Within a few years, the Materials Project API became the de-facto standard for training and testing, allowing direct head-to-head comparisons across research groups worldwide.

Simultaneously, ML force fields began to mature. Approaches such as DeepMD and GAP illustrated how neural networks could reproduce ab initio potential-energy surfaces at a fraction of the computational cost, opening the door to large-scale

molecular-dynamics simulations of realistic materials systems [9]. By the end of this phase, GNNs had proven their ability to capture local atomic interactions and were routinely applied to high-throughput screening of novel compositions [10, 25, 26].

2020–2022: Equivariance and data efficiency The next wave of progress centered on the explicit incorporation of physical symmetries. Equivariant GNNs such as those underlying NequIP, MACE, and Allegro achieved state-of-the-art accuracy on force-field benchmarks while requiring significantly less training data than their invariant predecessors [4, 5]. The recognition that tensor properties (piezoelectric coefficients, dielectric tensors) require equivariant representations became widespread [5]. These models demonstrated that respecting rotational, translational, and permutational symmetries is not merely an aesthetic choice but a practical necessity for reliable extrapolation within known chemical spaces [4].

Active learning also gained traction during this period [27]. Lookman *et al.* and Cao *et al.* showed that uncertainty-guided acquisition could accelerate the discovery of high-performance solders and other functional materials [28, 29]. Multi-fidelity strategies further improved data efficiency by intelligently combining cheap low-accuracy calculations with sparse high-accuracy results [6, 28]. By 2022, equivariant architectures were routinely cited as the new baseline for any force-field or property-prediction task involving periodic systems.

2023–2026: Generative models and foundation models The final phase of the decade was defined by the rise of generative methods and early foundation-model efforts. Diffusion models proved capable of generating stable crystal structures conditioned on composition or desired properties, offering a route to true inverse design [6, 15]. Specialized generative frameworks for inorganic materials reached publication in high-impact venues, demonstrating the feasibility of proposing candidates outside known databases [6, 7]. Large language models began to be integrated for literature mining and hypothesis generation, although their quantitative predictive power remained limited [10, 18].

Autonomous laboratories emerged as proof-of-concept platforms that combined prediction, synthesis planning, and robotic experimentation in closed loops [7]. Multi-fidelity and multimodal learning further expanded the scope of what could be learned from heterogeneous data sources [6, 28]. By 2025–2026, the community had produced the first demonstrations of end-to-end discovery pipelines that moved from idea to validated material within a single workflow.

Key milestones timeline

Year	Milestone	Impact
2017	CGCNN, SchNet	GNNs established for crystals [24, 25]
2018	Materials Project API	Standard benchmarks adopted community-wide [3, 23]
2022	NequIP, MACE	Equivariant force fields with superior data efficiency [16, 17]
2023	Diffusion models	First steps toward inverse design of crystals [30]
2024	Foundation-model explorations	Zero-shot literature-to-prediction pipelines tested [31]
2025	Autonomous labs	Closed-loop discovery demonstrated at small scale [18]

Collectively, these advances have reduced the time from concept to candidate material from years to months in selected domains and have generated thousands of new hypothetical compounds for further study [1, 32]. The decade has therefore delivered tangible methodological progress that no longer requires justification; the challenge now lies in translating these tools into routinely validated, application-ready materials.

Remaining Gaps

Current models excel as interpolators yet falter as extrapolators, with prediction errors surging under new elements, unseen prototypes, or extreme out-of-distribution conditions [19, 20, 33]. No architecture has achieved fundamental advances in generalization for materials, causing data-efficiency gains within familiar chemical spaces to collapse for novel compositions [18, 23]. This limitation compounds with the reliance of most GNNs on short-range cutoffs that overlook long-range electrostatic, dispersion, and polarization effects critical to ionic conductivity, polar crystals, and collective phenomena in electrolytes [5, 34], an issue persisting without standard solutions via extended message passing or hybrid potentials [4]. Such shortcomings extend to miscalibrated uncertainty estimates that rarely separate epistemic from aleatoric components, where even conformal methods sacrifice sharpness and thus practical utility for screening [18, 22, 35]. Synthesis prediction remains equally intractable, as generative approaches yield metastable or inaccessible structures while language models offer unverified recipes [6, 7, 10, 31, 36]. Progress further concentrates on ordered crystals, leaving disordered systems—amorphous solids, glasses, and defects—underrepresented amid acute size-transferability challenges and absent community benchmarks [23, 37]. Finally, validation against DFT data rather than experiments propagates functional errors, with few studies completing the discovery-to-measurement cycle and thereby constraining real-world impact [7, 24, 30].

Unlearned Lessons

Benchmark biases in standard datasets, such as the Materials Project's emphasis on stable, ordered, and cubic structures, cause random train–test splits to inflate generalization estimates by leaking chemically similar compounds across partitions [21, 23]. This structural artifact undermines reliability, even as calls for multiple test splits and cross-dataset validation have gone largely unheeded after a decade [18]. Compounding the issue, reproducibility suffers from widespread omission of code, data, and environments, while model cards detailing intended use and limitations remain rare despite repeated recommendations [18, 22, 23]. Architecture choices introduce further distortions: invariant GNNs inherently fail on tensorial properties such as piezoelectricity by forcing equivariant quantities toward zero, yet invariant models continue to be misapplied where symmetry demands equivariance [4, 5]. In active learning, complex acquisition functions often yield only marginal gains over random sampling once proper baselines are included, yet many studies omit these baselines and thereby overstate acceleration [28, 29, 35]. Generative models similarly require rigorous post hoc filtering, as most proposed structures prove invalid, unstable, or inaccessible, with papers still highlighting “novel materials” while reporting validity rates as low as 10 % [6, 7, 36]. Finally, the persistent reliance on DFT validation propagates systematic functional errors, and although experimental confirmation constitutes the true benchmark, the community has yet to move beyond DFT-level agreement in the majority of cases [24, 30].

Critical Assessment: Hype vs. Reality

Although foundation models have been heralded for zero-shot generalization across materials properties and seamless literature-to-prediction pipelines, their errors remain an order of magnitude higher than those of fine-tuned GNNs on equivalent tasks, confining them largely to summarization and hypothesis generation rather than quantitative prediction [10, 18]. Generative models, similarly promoted as enablers of routine inverse design for synthesizable compounds, continue to produce mostly invalid or unstable candidates with validity rates frequently below 20 % even after filtering, thus serving better for chemical space exploration than for delivering experimentally confirmed materials [6, 7, 36]. This gap in delivery extends to autonomous laboratories, which, despite visions of closed-loop discovery replacing human scientists on accelerated timescales, remain costly, restricted to simple syntheses, and heavily dependent on human intervention for non-routine processes, limiting their current industrial scalability [7, 18]. Equivariant GNNs, positioned as near-universal solutions for force fields and tensor properties, still struggle with extrapolation and long-range interactions without supplementary modules, rendering them the strongest available yet insufficient for the most challenging out-of-distribution regimes [4, 5, 34]. Finally, active learning's promised dramatic reduction in experimental burden materializes as only modest improvements over rigorously controlled random sampling in many domains, underscoring that its advantages, while genuine, are incremental rather than transformative [28, 29, 35].

Table 1 shows that many of the field's most persistent weaknesses arise from a recurring mismatch between high-visibility claims and the evidentiary standards actually required for reliable, experimentally grounded discovery.

Table 1. Structural misalignments between dominant claims in materials AI and the evidentiary standards required for reliable, experimentally grounded discovery

Domain of claim	Dominant claim tendency in the literature	What the manuscript identifies as the actual problem	Missing evidentiary requirement	Why the omission matters analytically	Corrective standard proposed by the review
Generalization claims	Strong benchmark scores are presented as evidence of broad model capability	Benchmark success is often inflated by biased datasets and random splits that preserve chemical similarity	Multiple test splits, cross-dataset validation, and explicit OOD evaluation	Without these controls, apparent generalization is often only interpolation under favorable partitioning	Treat extrapolation assessment as a central reporting requirement rather than a secondary supplement
Reproducibility claims	Availability of a publication is implicitly treated as sufficient scientific disclosure	Code, data, environment details, and intended-use documentation are often incomplete	Shared code, deposited data, environment capture, and model cards	Results cannot be independently verified, stress-tested, or reused at scale	Reproducibility infrastructure should be a publication norm rather than author discretion
Symmetry adequacy claims	Invariant models are still applied to tasks involving tensorial targets	Invariance cannot represent direction-sensitive or sign-sensitive tensor properties correctly	Explicit symmetry matching between target property and model architecture	Architectural mismatch produces systematic rather than incidental error	Require equivariant architectures by default for tensor and force-related tasks
Active-learning acceleration claims	Sophisticated acquisition functions are framed as clear efficiency breakthroughs	Gains often shrink when compared against proper random baselines	Mandatory reporting against random sampling and other simple controls	Without strong baselines, "acceleration" can be overstated and method comparison becomes unreliable	Make baseline discipline a minimum standard for active-learning publications
Generative discovery claims	Novel generated structures are presented as discoveries	Many outputs are invalid, unstable, or synthetically inaccessible	Separate reporting of validity, novelty, stability, and synthesizability	Unfiltered novelty language obscures the true distance between generation and discovery	Reframe generative systems as proposal engines unless downstream validation is demonstrated
Synthesis-planning claims	Plausible recipe generation is	Real synthesis prediction remains	Experimental confirmation and	Plausibility without laboratory	Evaluate synthesis claims by

	treated as practical synthesis intelligence	unsolved and LLM-generated procedures may hallucinate	realistic feasibility assessment	verification can misdirect effort and inflate confidence	successful laboratory execution, not textual coherence
Validation claims	DFT agreement is commonly treated as adequate proof of correctness	DFT errors propagate directly into trained models and can mask real-world failure	Direct experimental validation of key discovery claims	DFT-level success does not guarantee laboratory relevance or application readiness	Elevate experiment from optional add-on to gold-standard endpoint
Knowledge accumulation claims	Positive results dominate the visible literature and shape perceived field progress	Negative results and failed pathways remain largely invisible	Failure registries, negative-results sections, and structured reporting norms	The community repeatedly relearns avoidable mistakes and loses information about boundaries of method validity	Build collective learning infrastructure that values failure information alongside success

Comparison with Other Fields

Data-driven materials engineering has evolved in parallel with machine learning applications in neighbouring disciplines, yet it retains distinct challenges and strengths. Comparison with drug discovery is particularly instructive. Molecular machine learning benefited from early, large-scale benchmarks such as QM9, ZINC, and MoleculeNet that provided standardized, diverse datasets and clear evaluation protocols [23, 24]. These resources enabled rapid iteration and community-wide progress on tasks such as property prediction and molecular generation. Materials science, by contrast, must contend with periodicity, thermodynamic stability under realistic conditions, and the additional layer of synthesizability—factors absent or far less constraining in small-molecule design [20, 23]. Nevertheless, materials researchers can adopt the rigorous cross-validation and multi-metric reporting practices that have become routine in drug discovery, especially for out-of-distribution generalization [24].

The comparison with computer vision highlights another set of differences. Computer vision enjoys enormous, curated datasets (ImageNet and its successors) and standardized benchmarks that have driven decades of architectural innovation [18]. Materials datasets remain smaller, heavily biased toward stable crystalline compounds, and lack the same level of curation or diversity [21, 23]. Transfer learning strategies successful in vision—pre-training on massive image corpora—do not translate directly because atomic-scale structures do not share the same statistical regularities as natural images [18]. Domain-specific adaptations remain necessary.

Despite these limitations, materials artificial intelligence offers advantages not yet fully realized in other fields. The explicit incorporation of physics—through equivariant architectures that respect rotational, translational, and permutational symmetries—has produced models that conserve energy and momentum by construction [4, 5]. This physics-informed approach contrasts with purely data-driven methods in drug discovery or computer vision, where physical consistency must often be imposed post hoc [4]. Integration of domain knowledge, such as crystal symmetry or known phase diagrams, further improves data efficiency and interpretability in ways that are less common in molecular or image-based machine learning [5, 34]. Multi-fidelity strategies that combine cheap approximate calculations with sparse high-accuracy data have also matured faster in materials than in many other domains [6, 28].

Overall, materials informatics has learned from the benchmarking culture of drug discovery and the architectural innovations of computer vision, yet it has developed unique strengths in embedding physical priors and handling structured, periodic data. The next decade will determine whether these strengths can be leveraged to close the remaining performance gaps or whether the field will continue to lag behind its more mature neighbours in terms of standardization and reproducibility [18, 22].

Future Directions

The coming decade demands a shift from incremental gains on familiar tasks toward resolving the structural limitations outlined here. Future architectures must prioritize extrapolation by moving beyond local cutoffs and interpolation in known chemical spaces, integrating explicit long-range modules and physics-informed priors capable of handling unseen elements, novel prototypes, and extreme conditions [5, 19], with causal modelling frameworks offering a compelling path to disentangle correlation from causation [18]. This progress hinges on deeper experimental integration: although autonomous platforms have shown initial closed-loop promise, they remain narrow in scope, requiring the community to elevate laboratory validation as the default and insist on experimental confirmation for any significant discovery claim rather than DFT agreement alone [7, 24, 30]. Hybrid human–robot systems linking predictive models to high-throughput synthesis will prove indispensable in this transition [7]. Such grounding naturally extends to uncertainty-aware decision-making, where quantification must advance from miscalibrated estimates to conditional coverage guarantees that deliver actionable confidence bounds for prioritizing synthesis [18, 22, 35]. Complementing this, shared failure databases and dedicated negative-results sections would institutionalize the documentation of unproductive pathways, invalid structures, and collapsed extrapolations, transforming repeated mistakes into collective knowledge [18, 24, 32]. Finally, truly multimodal models that fuse literature text, structural data, spectroscopy, and process parameters within one framework could harness decades of experimental insight to inform generative design and active learning, thereby closing the historical-to-forward prediction loop [10, 18]. These directions are inherently interdependent: extrapolation without experimental grounding falters, uncertainty quantification lacks impact absent failure reporting, and multimodal integration yields limited value without enforced reproducibility standards [22, 23]. Coordinated community effort, rather than isolated advances, will therefore define meaningful progress.

Recommendations for the Next Decade

For individual researchers the priorities are clear: shift focus from interpolation to extrapolation, validate every major claim against experiment rather than DFT alone, and report failures alongside successes [19, 24, 32]. Model cards that document intended use, limitations, and training distributions should become mandatory supplementary material [22].

Benchmark designers must create challenging out-of-distribution test sets that span new elements, unseen prototypes, and disordered systems [21, 23]. Multiple test splits and cross-dataset validation should be required, and every benchmark release must include uncertainty-reporting protocols [18]. Random baselines must be reported for all active-learning studies [28, 35].

Journals should enforce reproducibility standards by mandating code and data deposition, model cards, and experimental validation for any discovery claim [7, 22]. Dedicated negative-results sections would accelerate collective learning and reduce redundant effort [24, 32]. Reviewers must insist on validity, novelty, and stability metrics for generative models rather than accepting headline claims based on unfiltered output [6, 36].

Funders have a pivotal role in building infrastructure. Support for community databases, standardized OOD benchmarks, and shared failure registries will yield returns far greater than additional isolated projects [18, 23]. Priority should be given to extrapolation research, autonomous-laboratory scaling, and multimodal integration that connects literature with computation [7, 10]. Long-term funding for reproducibility platforms and negative-results dissemination will embed the unlearned lessons of the past decade into standard practice [22].

Collectively, these recommendations shift the field from volume-driven publication to quality-driven discovery. If implemented, they will ensure that the next decade of data-driven materials engineering is defined not by hype cycles but by reproducible, experimentally validated advances that deliver real functional materials.

The Path Forward

Near term (1–3 years): the community should establish standardized out-of-distribution benchmarks, mandate equivariant GNNs as the default for tensorial and force-field tasks, and require experimental validation for any new material claim [4, 21, 24]. Reproducibility checklists and model cards should become submission requirements across target journals [22].

Medium term (3–7 years): breakthroughs in extrapolation architectures should enable reliable predictions for new element combinations and disordered systems [19, 37]. Uncertainty quantification should reach the level of conditional coverage guarantees usable in risk-aware discovery campaigns [35]. Autonomous laboratories should scale to handle moderately complex syntheses, demonstrating closed-loop discovery of at least one new functional material per platform [7].

Long term (7–10 years): generalizable materials AI should integrate multimodal literature mining, physics-informed generative models, and experimental feedback into unified discovery platforms [6, 10, 18]. Experimental validation should become the norm for the majority of published predictions rather than the exception [24].

Success criteria for 2036 are concrete and measurable. Models must predict properties of novel element combinations with mean absolute errors below 0.1 eV/atom on formation energies and comparable accuracy for electronic and mechanical tensors [5, 34]. Autonomous laboratories should routinely discover and validate new materials in fewer than 100 targeted experiments [7]. Failure reporting must be standard practice, with every major study documenting unsuccessful pathways in supplementary materials or dedicated repositories [18, 32]. Benchmark biases should be routinely mitigated through multiple test splits and cross-dataset evaluation [21, 23].

Achieving these criteria will require the field to absorb the unlearned lessons of the past decade and to treat extrapolation, uncertainty, synthesis, and experimental grounding as core rather than peripheral concerns. The path forward is therefore not merely technical but cultural—moving from a literature that celebrates incremental interpolation to one that demands reproducible, generalizable, and experimentally verified discovery.

Conclusion

A decade of data-driven materials engineering (2017–2026) has seen remarkable progress: GNNs, equivariant architectures, generative models, and autonomous laboratories have transformed screening and hypothesis generation. Yet remaining gaps in extrapolation, long-range interactions, uncertainty quantification, synthesis prediction, disordered materials, and experimental validation continue to limit impact. Unlearned lessons—pervasive benchmark biases, inadequate reproducibility infrastructure, continued misuse of invariant architectures for tensor properties, omission of random baselines in active learning, insufficient validity filtering in generative workflows, and the near-absence of failure reporting—have been repeatedly identified but not yet absorbed into common practice. Critical assessment shows that hype surrounding foundation models, generative inverse design, and autonomous laboratories frequently exceeds demonstrated quantitative impact when judged against rigorous baselines.

The next decade must therefore prioritize extrapolation-focused research, mandatory experimental validation, shared failure registries, and realistic expectations. Only by confronting these gaps and internalizing the unlearned lessons can the field deliver on its promise to accelerate the discovery of functional materials for energy, sustainability, and advanced manufacturing. Data-driven materials engineering stands at a crossroads: the methodological foundations are in place, but the cultural and infrastructural shifts required for genuine impact remain incomplete. The literature of 2017–2026

provides both the evidence of progress and the roadmap for addressing its shortcomings. The choice to follow that roadmap now rests with the community.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 05 Apr 2025

Revised: 24 Jul 2025

Accepted: 05 Nov 2025

Published online: 18 January 2026

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys Rev Lett.* 2018;120(14):145301.
<https://doi.org/10.1103/PhysRevLett.120.145301>.
- Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. *Chem Mater.* 2019;31(9):3564-72.
<https://doi.org/10.1021/acs.chemmater.9b01294>.
- Shi X, Zhou L, Huang Y, Wu Y, Hong Z. A review on the applications of graph neural networks in materials science at the atomic scale. *MGE Adv.* 2024;2(2).
<https://doi.org/10.1002/mgea.50>.
- Reiser P, Neubert M, Eberhard A, Torresi L, Zhou C, Shao C, et al. Graph neural networks for materials science and chemistry. *Commun Mater.* 2022;3(1):93.
<https://doi.org/10.1038/s43246-022-00315-6>.

Zhao Z, Hu T, Bi S, Guan D, Xu S, Chen C, et al. Advances in graph neural networks for alloy design and properties predictions: A review. *J Mater Inf.* 2026;6:2.
<https://doi.org/10.20517/jmi.2025.42>.

Zeni C, Pinsler R, Zügner D, Fowler A, Horton M, Fu X, et al. A generative model for inorganic materials design. *Nature.* 2025;639(8055):624-32.
<https://doi.org/10.1038/s41586-025-08628-5>.

Metni H, Ruple L, Walters LN, Torresi L, Teufel J, Schopmans H, et al. Generative models for crystalline materials. *Adv Mater.* 2026;38(18).
<https://doi.org/10.1002/adma.202523620>.

Padbury R. Data-driven approaches to materials and process challenges: A new tool for the materials science field. *Am Ceram Soc Bull.* 2020;99(6):24-30.

Ramprasad R, Batra R, Pilania G, Mannodi-Kanakthodi A, Kim C. Machine learning in materials informatics: Recent applications and prospects. *NPJ Comput Mater.* 2017;3(1):54.
<https://doi.org/10.1038/s41524-017-0056-5>.

Wang Z, Sun Z, Yin H, Liu X, Wang J, Zhao H, et al. Data-driven materials innovation and applications. *Adv Mater.* 2022;34(36).
<https://doi.org/10.1002/adma.202104113>.

Aguilar-Bejarano E, Arrieta L, Gutiérrez M, Özcan E, Woodward S, Figueredo G, et al. Explainable GNN-derived structure-property relationships in interstitial-alloy materials. *Phys Chem Chem Phys.* 2025;27(41):22240-50.
<https://doi.org/10.1039/D5CP02208H>.

Kassa G. Wef-GNN: A generalizable graph neural network for crystalline material property prediction [Internet]. OpenReview; 2025 [modified 2026 Feb 11; cited 2026 Jun 23]. Available from: <https://openreview.net/forum?id=Hs4WbkJqsm>

Horton MK, Huck P, Yang RX, Munro JM, Dwaraknath S, Ganose AM, et al. Accelerated data-driven materials science with the Materials Project. *Nat Mater.* 2025;24(10):1522-32.
<https://doi.org/10.1038/s41563-025-02272-0>.

Choudhary K, Garrity KF, Reid AC, DeCost B, Biacchi AJ, Hight Walker AR, et al. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *NPJ Comput Mater.* 2020;6(1):173.
<https://doi.org/10.1038/s41524-020-00440-1>.

Yang S, Cho K, Merchant A, Abbeel P, Schuurmans D, Mordatch I, et al. Scalable diffusion for materials generation [Preprint]. arXiv; 2023.
<https://doi.org/10.48550/arXiv.2311.09235>.

Lookman T, Liu Y, Gao Z. Materials informatics: Emergence to autonomous discovery in the age of AI. *Adv Mater.* 2026;38(29).
<https://doi.org/10.1002/adma.202515941>.

Olivetti EA, Cole JM, Kim E, Kononova O, Ceder G, Han TY, et al. Data-driven materials research enabled by natural language processing and information extraction. *Appl Phys Rev.* 2020;7(4):041317.
<https://doi.org/10.1063/5.0021106>.

Gupta KK, Barman S, Sankar A, Dey S, Mukhopadhyay T. Artificial intelligence in materials by design: Critical review and perspectives on materials informatics to generative and agentic intelligence. *Arch Comput Methods Eng.* 2026;33:6015-45.
<https://doi.org/10.1007/s11831-025-10486-3>.

Gomez-Gualdrón D, de Vilas TG, Ardila K, Fajardo-Rojas JF, Pak A. Machine learning to design metal-organic frameworks: Progress and challenges from a data efficiency perspective. *Mater Horiz.* 2026;13:1694-715.
<https://doi.org/10.1039/D5MH01467K>.

Nematov D, Hojamberdiev M. Machine learning-driven materials discovery: Unlocking next-generation functional materials-A review. *Comput Condens Matter.* 2025;45.
<https://doi.org/10.1016/j.cocom.2025.e01139>.

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.
<https://doi.org/10.1021/acs.accounts.0c00785>.

Butler KT, Choudhary K, Csanyi G, Ganose AM, Kalinin SV, Morgan D. Setting standards for data driven materials science. *NPJ Comput Mater.* 2024;10(1):231.
<https://doi.org/10.1038/s41524-024-01411-6>.

Zivic F, Malisic AK, Grujovic N, Stojanovic B, Ivanovic M. Materials informatics: A review of AI and machine learning tools, platforms, data repositories, and applications to architected porous materials. *Mater Today Commun.* 2025;48:113525.
<https://doi.org/10.1016/j.mtcomm.2025.113525>.

Morgan D, Jacobs R. Opportunities and challenges for machine learning in materials science. *Annu Rev Mater Res.* 2020;50:71-103.
<https://doi.org/10.1146/annurev-matsci-070218-010015>.

Jain A. Machine learning in materials research: Developments over the last decade and challenges for the future. *Curr Opin Solid State Mater Sci.* 2024;33:101189.
<https://doi.org/10.1016/j.cossms.2024.101189>.

Zhang Y, Xue D, Xin S, Wang X, Zhou W, Pan X, et al. Research progress of machine learning aided titanium alloys design. *Mater China.* 2025;44(4):319-29.
<https://doi.org/10.7502/j.issn.1674-3962.202501004>.

Osaro E, Mukherjee K, Colon YJ. Active learning for adsorption simulations: Evaluation, criteria analysis, and recommendations for metal-organic frameworks. *Ind Eng Chem Res.* 2023;62(33):13009-24.
<https://doi.org/10.1021/acs.iecr.3c01589>.

Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *NPJ Comput Mater.* 2019;5(1):21.
<https://doi.org/10.1038/s41524-019-0153-8>.

Cao B, Su T, Yu S, Li T, Zhang T, Zhang J, et al. Active learning accelerates the discovery of high strength and high ductility lead-free solder alloys. *Mater Des.* 2024;241:112921.
<https://doi.org/10.1016/j.matdes.2024.112921>.

Ng WL, Goh GL, Goh GD, Ten JS, Yeong WY. Progress and opportunities for machine learning in materials and processes of additive manufacturing. *Adv Mater.* 2024;36(34).
<https://doi.org/10.1002/adma.202310006>.

Duan C, Nandy A, Chand Pal S, Yang X, Gao W, Du Y, et al. The rise of generative AI for metal-organic framework design and synthesis. *Matter.* 2026;9:102748.
<https://doi.org/10.1016/j.matt.2026.102748>.

Zhou ZH. Machine learning. Liu S, translator. Singapore: Springer Nature Singapore; 2021.
<https://doi.org/10.1007/978-981-15-1967-3>.

Ghasemi A, Barisik M. Machine learning for thermal transport prediction in nanoporous materials: Progress, challenges, and opportunities. *Nanomaterials (Basel)*. 2025;15(21):1660.
<https://doi.org/10.3390/nano15211660>.

Sheng Z, Zhu H, Shao B, He Y, Liu Z, Wang S, et al. Accelerated discovery of energy materials via graph neural network. *Inorganics*. 2025;13(12):395.
<https://doi.org/10.3390/inorganics13120395>.

Osaro E, Colón YJ. Intelligent screening of porous materials: A review of active-learning approaches in MOF research. *Chem Phys Rev*. 2025;6(4):041307.
<https://doi.org/10.1063/5.0295283>.

Mal S, Seal G, Sen P. MagGen: A graph aided deep generative model for inverse design of stable, permanent magnets [Preprint]. *arXiv*; 2023.
<https://doi.org/10.48550/arXiv.2311.13328>.

Wang Y, Guo D, Yang T, Qian Q, Liu X, Shi S. A review of topological descriptors for amorphous materials complementing graph neural networks. *Comput Mater Today*. 2026;10:100053.
<https://doi.org/10.1016/j.commt.2026.100053>.