Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

REVIEW

View Article Online
View Journal

Cite this: DOI: 10.1039/d6dd00132g

Achieving a scalable machine learning workflow for crystal structure discovery with experimental validation

Danila Shiryaev,[a] Emil I. Jaffal,[bc] Sangjoon Lee,[d] Balaranjan Selvaratnam [c] and Anton O. Oliynyk *[bc]

Machine learning (ML) has become a central component of data-driven materials discovery, yet its practical impact remains heavily dependent on how these predictions are translated into experimentally-realizable outcomes. In this review, we examine ML-guided crystal structure discovery through the lens of recommendations as well as unconstrained generation, to emphasize interpretable workflows embedding chemical intuition, physical constraints, and experimental validation. Surveying standalone ML, hybrid ML-DFT, and machine-learned interatomic potential (MLIP) approaches, we highlight how constrained design spaces, data preprocessing, and validation strategies shape novel discovery success. Drawing on our own experimentally-validated case studies, ranging from supervised to unsupervised learning, as well as recommendation-type explorations, we outline the shift towards interpretable and explainable ML models that guide synthetic decisions, reveal trends that were previously difficult to identify, confirm established patterns, and uncover new ones. Collectively, we highlight the results of interpretable ML, which is more effective when deployed within experimental workflows to bridge learning and chemistry, to enable a reliable discovery pathway to solid state materials.

Received 22nd March 2026
Accepted 22nd May 2026

DOI: 10.1039/d6dd00132g

rsc.li/digitaldiscovery

1 Introduction

1.1 Forward and inverse paradigms in data-driven materials discovery

Historically, the discovery of functional solid-state materials has relied on hypothesis-driven experimentation grounded in chemical intuition, structure–property analogies, and systematic exploration of phase diagrams. Many core functional materials, including the permanent magnet $Nd_2Fe_{14}B$,[1] the ferroelectric perovskite $BaTiO_3$,[2] the thermoelectric alloy Bi_2Te_3,[3] and the cuprate superconductor $YBa_2Cu_3O_{7-\delta}$,[4] were identified through iterative synthesis and characterization cycles guided by empirical trends rather than explicit predictive frameworks. Such traditional approaches emphasize local exploration within chemically plausible spaces and depend heavily on accumulated experimental knowledge. While this paradigm seems slow and expensive, the byproduct of the discovery is the formalized extracted knowledge, which acts as the foundation for future discoveries. Data-driven strategies aim to formalize and accelerate the discovery process by treating existing experimental and computational results as a navigational map of materials space. Machine learning models can extract latent regularities from prior discoveries and use them to prioritize promising candidates, thereby transforming empirical insight into a scalable, quantitative guide for exploration. At the same time, knowledge extraction widely remains out of the focus, or simply inaccessible given the black-box nature of the models.

Data-driven materials discovery in the solid state is commonly organized around two complementary paradigms: *forward prediction*[5,6] and *inverse design*.[7,8] This distinction reflects how machine learning is deployed within a discovery workflow rather than differences in algorithmic form. Forward approaches use machine learning (ML) to approximate mappings from composition or structure to properties, while inverse approaches frame discovery as a decision-making problem aimed at identifying candidates that meet predefined targets under physical and experimental constraints.[9,10] These two paradigms and their points of interaction across structure, property, and performance objectives are schematically summarized in Fig. 1.

In the forward paradigm, supervised ML models are trained on large, static datasets, most often derived from density functional theory (DFT), to predict properties such as formation energy,[11,12] elastic moduli,[13–15] transport coefficients,[16–20] or

[a]PhD-Track Program in Advanced Materials, Institut Polytechnique de Paris, Palaiseau 91120, France
[b]Ph.D. Program in Chemistry, The Graduate Center of the City University of New York, New York, NY 10016, USA. E-mail: anton.oliynyk@hunter.cuny.edu
[c]Department of Chemistry, Hunter College, City University of New York, New York, NY 10065, USA
[d]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA

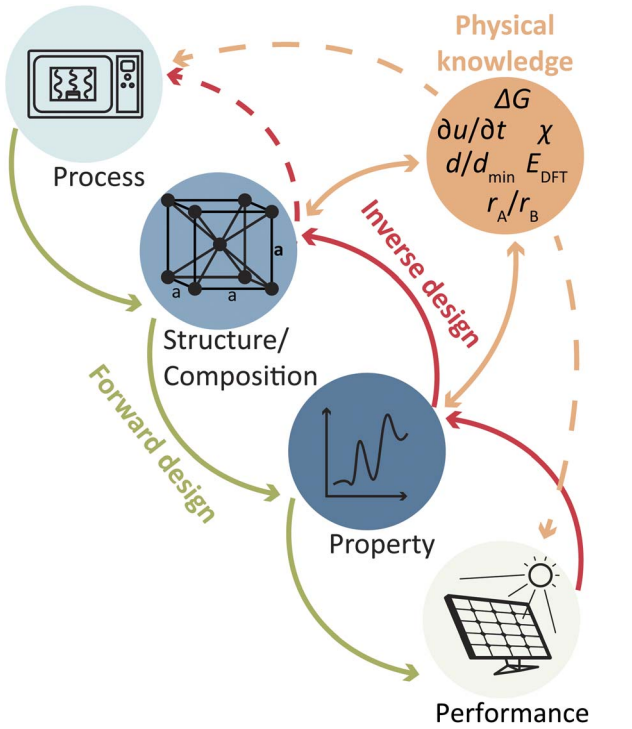

Fig. 1 Forward and inverse paradigms for data-driven materials discovery.

magnetic ordering temperatures.[21–23] Given the consistency of the data (*e.g.*, DFT), these reliable models enable rapid screening and ranking within predefined chemical or structural spaces and are widely used as surrogates for first principles calculations in high-throughput studies.[24,25] Forward prediction is therefore most effective when the hypothesis space is well-defined; however, commonly, the validation remains computational rather than experimental.[26–30]

Inverse design, as the name suggests, inverts this logic by starting from a desired property or performance objective and iteratively narrowing the search space through ranked recommendations, which should enable faster material targeting. In this setting, ML models are embedded within optimization or feedback loops, such as Bayesian optimization, active learning, or evolutionary search, where their primary role is to prioritize candidates under uncertainty and limited validation budgets.[10,31,32] When the exploration space is large, uncertainty-driven exploration can mitigate biases inherent to supervised learning workflows. These include dataset selection bias, the overrepresentation of well-studied chemistries, and the exclusion of negative or failed experimental outcomes, which collectively restrict exploration to historically sampled regions of materials space.[33] The success of inverse workflows is evaluated not by predictive accuracy alone, but by their ability to convert ranked candidates into experimentally confirmed structures or material properties.[34,35]

While forward prediction and inverse design provide the two principal operational modes of machine-learning-driven discovery, several recent studies have proposed a complementary paradigm referred to as physics-informed machine learning (PIML).[36,37] In this framework, machine learning models are explicitly constrained by physical knowledge, which may include governing equations,[38] as well as structural,[39] compositional,[40] chemical,[41] and physical[42] constraints that define feasible regions of materials space. Rather than relying solely on correlations learned from data, physics-informed approaches incorporate domain knowledge directly into model architectures, descriptors, or training objectives to guide learning toward physically admissible solutions. Such constraints can take many forms, ranging from conservation laws and differential equations to empirical chemical rules or stability criteria derived from materials thermodynamics. The common objective of these approaches is to improve generalization and reduce the risk of discovering candidates that are mathematically plausible yet physically unrealistic. In addition to academic implementations, physics-informed and ML-integrated materials modeling frameworks are increasingly available through commercial platforms, such as Matlantis,[43] QuantumATK,[44] Schrödinger Materials Science Suite[45] and Citrine Platform,[46] which integrate machine learning with first-principles or physics-based simulations to enable scalable and experimentally-relevant materials discovery workflows.

In the context of materials discovery, this integration of physical knowledge is particularly important because the accessible design space is strongly restricted by factors that are difficult to capture purely from data.[47] Structural stability, phase competition, chemical compatibility, and synthesis pathways impose constraints that may not be explicitly represented in available datasets but nevertheless determine whether a predicted material can exist and be experimentally realized. In practice, inverse design strategies often rely on predefined candidate spaces derived from materials databases, enumerated substitutions, or structure prototypes, which constrain exploration to chemically plausible regions of materials space.[48,49] Other approaches attempt to learn the underlying distribution of materials and sample new compositions or structures from continuous latent representations.[32,34,50] Regardless of the specific strategy, the feasible design space of solid-state materials remains strongly limited by phenomena such as polymorphism,[51–53] metastability,[54] and synthesis pathways,[55] which ultimately determine whether a predicted compound can be experimentally realized.

For this reason, many recent materials discovery workflows increasingly combine machine-learning models with physical knowledge to guide exploration toward experimentally meaningful regions of materials space.[7,56] In such approaches, physical constraints act as an additional source of inductive bias that complements both forward prediction and inverse design strategies. While forward models enable scalable property prediction and inverse workflows support targeted search within large chemical spaces, physics-informed approaches help ensure that the resulting candidates remain consistent with known physical and chemical principles. Together, these strategies form a complementary toolkit for translating

machine-learning predictions into experimentally realizable materials discoveries.

Large language models (LLMs) represent an increasingly prominent yet distinct branch of AI-driven materials discovery,[57] with applications spanning automated literature mining and dataset construction,[58] autonomous chemical research and reaction planning,[59] text-guided crystal structure generation,[60] multimodal property prediction from atomic structure and natural language,[61] and closed-loop inverse design workflows for functional materials such as perovskites.[62] While these approaches demonstrate considerable promise in accelerating data aggregation and hypothesis generation, they operate through fundamentally different mechanisms than the constraint-driven, experimentally validated workflows that form the focus of this review, and are therefore not examined in detail here.

In this review, we demonstrate that the most successful machine-learning strategies for materials discovery are not those that generate unconstrained candidate structures, but those that operate within chemically and physically constrained regions of materials space and provide interpretable recommendations that guide experimental exploration. To illustrate this perspective, we examine different classes of recent machine-learning approaches – from data-driven predictive models to hybrid ML–DFT workflows and machine-learned interatomic potentials, and analyze how these frameworks incorporate physical knowledge, handle experimental constraints, and ultimately enable experimentally validated materials discovery. We define successful ML-driven materials discovery as experimentally validated outcomes, including (i) successful synthesis of the predicted phase, (ii) agreement between predicted and measured properties when applicable, and (iii) reproducibility of the reported synthesis pathway, which is frequently not demonstrated in existing studies. These criteria distinguish computational predictions from experimentally actionable discoveries.

1.2 Data sources and preprocessing for solid state learning

The effectiveness of machine learning guided materials discovery is constrained by the scope and consistency of available data. In this context, consistency refers to methodological uniformity in how data are generated and represented, traceable provenance, and reproducibility across calculations or measurements, all of which determine whether datasets can be reliably combined for model training. Traditional experimental data sources remain central to solid-state informatics, which ultimately should result in experimental validation, closing the loop of the iterative experimental process. Established repositories such as the Inorganic Crystal Structure Database (ICSD),[63] Cambridge Crystallographic Data Centre (CCDC),[64] the International Centre for Diffraction Data (ICDD),[65] and the Pearson Crystal Database (PCD)[66] provide experimentally-determined crystal structures, compositions, diffraction patterns, and related reference properties. In addition, many research groups maintain locally curated laboratory datasets derived from synthesis campaigns and characterization workflows, which often capture practical constraints and negative results not represented in public archives.[67]

In contrast, training data in modern machine-learning workflows are predominantly based on high-throughput DFT repositories,[68,69] with the Open Quantum Materials Database (OQMD)[70] and Materials Project[71] as widely used examples. These resources provide formation energies and relaxed crystal structures at scale, which makes them practical foundations for stability screening and for training structure-based models. In parallel, expert-curated approaches such as ME-AI[72] demonstrate that interpretable descriptor discovery can be achieved using small, measurement-based datasets, highlighting a complementary data paradigm to large DFT-dominated repositories.

Interoperability across repositories has become critical as workflows increasingly combine prediction, screening, and validation. The OPTIMADE[73] provides a standardized application programming interface for querying structures and metadata across independent databases. AFLOW[74] provides complementary infrastructure for symmetry normalization, prototype assignment, and standardized descriptor generation. Together, these tools reduce ambiguity during data aggregation and enable reproducible cross database screening.

Pre-processing choices strongly influence which physical distinctions a model can learn, and which biases it inherits. Structural standardization, symmetry reduction, and removal of duplicate or near-duplicate entries are essential when repositories contain many closely related variants of the same prototype family. Several discovery workflows therefore enforce explicit space group and Wyckoff constraints during structure handling to align candidate evaluation with physically meaningful degrees of freedom.[75] A recurring limitation of large DFT repositories is the disconnect between zero-Kelvin stability labels and experimental realizability. Many datasets emphasize formation energy and convex hull stability, while information on metastability, kinetics, and processing-dependent phase selection is sparse. This gap has motivated synthesis-aware strategies that move beyond the static stability screening based solely on zero-Kelvin thermodynamic criteria, including autonomous experimentation[35] and temperature-dependent phase stability prediction.[76]

The authors would like to note that it is often that constrained ML workflows may outperform unconstrained generation, with a particular focus on obeying general chemical physics. Recent work by Amazon's Science team goes into great-depth about the enforcement of more elementary physics-based laws as it relates to deep-learning methods at the industrial level. One case[77] handled is on enforcing physics through boundary conditions, exhibiting "a 20-fold performance improvement over previous operator models." In another case[78] dealing with mass conservation, they found that their constrained model "outperforms other ML-based approaches that do not guarantee volume conservation" and had the "smallest relative L2 errors across various values of viscosity ν, by enforcing the exact boundary values."

Prior academic-based ML-guided discovery studies[79] demonstrate that an ML-guided search for Li superionic

conductors was 2.7 times more likely to identify fast ion conductors than random exploration (unconstrained generation), and achieved at least a 44-fold improvement in the log-average room-temperature ionic conductivity of selected candidates. The model additionally achieved an F1 score roughly 3.5 times greater than random selection, and substantially outperformed human expert screening in both predictive accuracy and speed. These are only some collective results representing the benefit of constraining workflows to maintain scientific rigor across models, ensuring that chemically-informed strategies improve efficiency and still drive exploration across broad and diverse materials spaces.

In tandem, data sources and preprocessing choices define the effective design space explored by ML models. In experimentally-oriented discovery, data handling functions as an active decision layer that shapes candidate ranking and validation, rather than as a neutral technical step.[80–82]

2 Machine learning strategies for materials discovery

2.1 ML approaches without first principles coupling

Standalone ML approaches in materials science aim to predict or propose materials without explicit coupling to first principles calculations during inference. In solid state discovery, these methods are typically trained on existing experimental databases and operate either at the level of property prediction, structural classification, or candidate recommendation. Their appeal lies in computational efficiency and scalability, but their effectiveness depends strongly on how closely the learned task

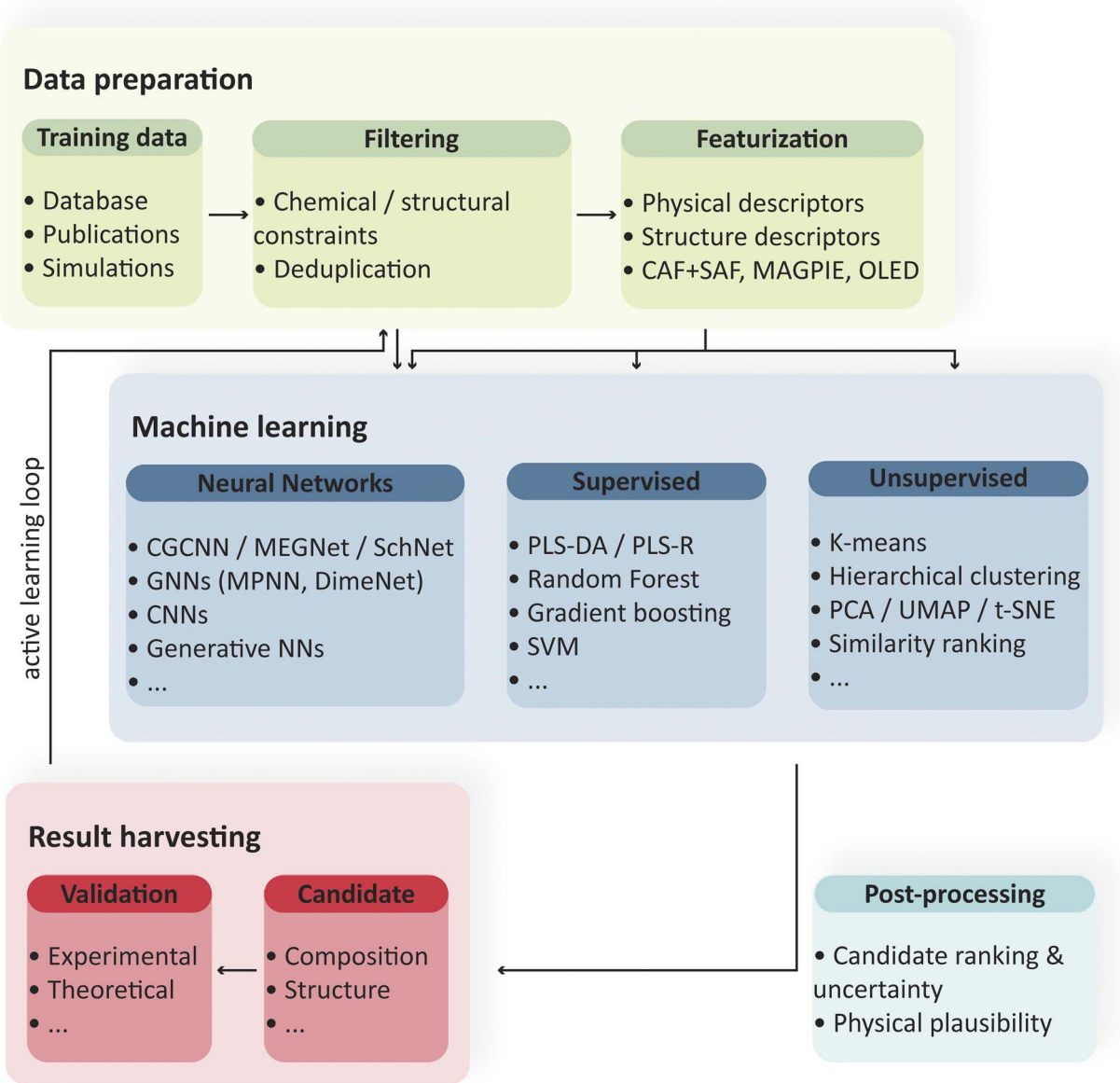

Fig. 2 Generalized workflow of machine-learning-driven materials discovery.

aligns with experimentally meaningful decision boundaries and how individual steps, from data curation to candidate validation, are combined in practice (Fig. 2).

One common class of ML approaches focuses on property prediction from composition or structure using supervised learning. Structure-aware neural networks such as crystal graph convolutional neural network (CGCNN) variants predict formation energies, elastic properties, or transport-related quantities directly from relaxed crystal structures, enabling rapid ranking of known or hypothetical materials within fixed design spaces.[83,84] Composition-based models are also widely used, particularly when structural information is unavailable, and have demonstrated success in narrowly defined application targets such as magnetocaloric screening and thermoelectric property mapping.[85,86]

Another class of approaches applies generative or sequence-based models to propose new crystal structures. Previous work demonstrated that deep generative models can reproduce statistical features of crystallographic data without explicit physical constraints.[34] More recent autoregressive and diffusion-based models[87–89] extend this idea by generating full crystal structures conditioned on symmetry or composition, but validation typically relies on downstream stability filtering rather than intrinsic guarantees of synthesizability. As a result, these models function more reliably as proposal engines than as standalone discovery tools.

2.2 Recent experimentally-validated ML works

The experimental realization of new materials is a challenging endeavor, as it requires domain knowledge and, often, several experiments. This can be rationalized by looking at the number of reported compounds ($\sim 10^6$) that are predicted to be stable by first principles calculations *versus* the number of experimentally realized compounds ($\sim 10^5$).[90] Further, failed experiments are seldom reported. For instance, it has been estimated that, in the search of a catalyst for the Haber–Bosch process, more than 6500 experiments were performed with over 2500 catalysts.[91] However, the data from the failed/non-optimal reactions are valuable in discovering new materials.[67] Another bias in chemical data is the anthropogenic bias in chemical space selection for exploration, reagent choices, and synthetic method selections which can concentrate data on small chemical space, while leaving other spaces sparse. Studies show that this biases selection does not correlate with the success of the reaction, and random explorations are better in mapping the synthesis parameter space to experimental outcomes.[92]

Recommendation-focused ML approaches avoid unconstrained generation by ranking candidates drawn from predefined and physically meaningful spaces. A representative example of such a constrained, recommendation-driven workflow is shown in Fig. 3, which summarizes results from the study of Fe_2P-type magnetocaloric compounds[93–95] where supervised learning is embedded within an iterative pipeline linking literature-derived data, compositional screening, and experimental validation. In this work, the authors constructed a curated experimental dataset comprising 603 samples collected from published studies and new measurements, of which 558 remained after filtering and deduplication. The feature space was intentionally restricted to chemical composition and annealing conditions, reflecting the most consistently reported experimental descriptors. A feed-forward neural network with a single hidden layer of 256 neurons was trained to predict the magnetic transition temperature; the resulting agreement between predicted and experimental values across the full dataset is shown in Fig. 3a, with the model achieving a mean absolute error of approximately 20 K and $R^2 = 0.89$ on held-out data. The trained model was subsequently used to map low-temperature regions of the Mn–Fe–P–Si compositional space, with Fig. 3b presenting the predicted transition temperature landscape for the base Mn–Fe–P–Si system together with experimental compositions overlaid for direct validation. Fig. 3c presents the predicted transition temperature landscape for Co-substituted compositions; however, experimental transition temperatures are not overlaid in this panel,

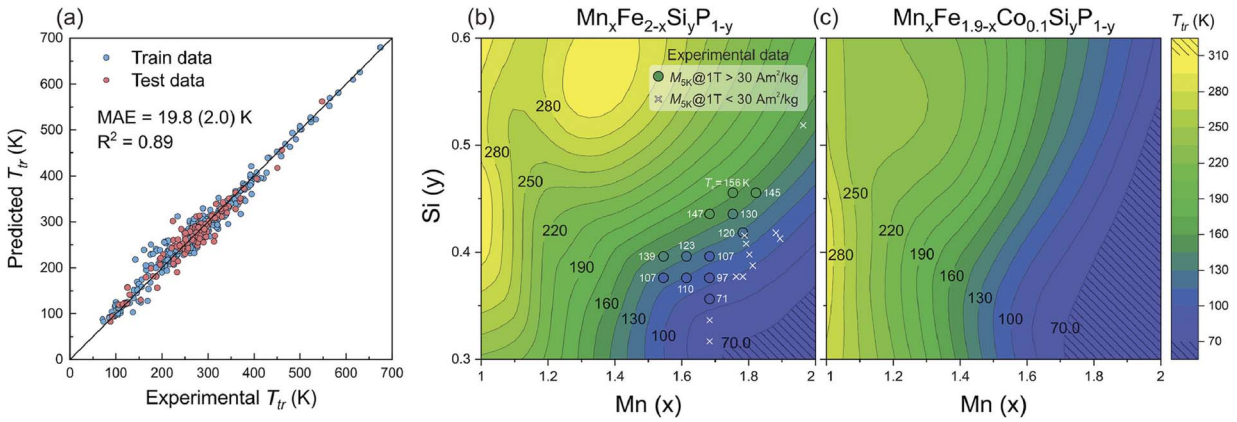

Fig. 3 Machine-learning prediction of transition temperatures in Fe_2P-type magnetocaloric compounds. (a) Comparison between experimentally measured transition temperatures and values predicted by the trained neural-network model, illustrating predictive accuracy, (b) predicted transition-temperature map for the $Mn_xFe_{2-x}Si_yP_{1-y}$ compositional space under fixed annealing conditions, with experimental samples overlaid for validation and (c) corresponding prediction map for Co-substituted compositions, showing the expansion of the low-temperature region upon partial Fe–Co substitution. Adapted with permission from ref. 93. Copyright 2022 Elsevier.

Table 1 Summary of experimentally validated machine-learning predictions (without first-principles calculations) from 2020–2025

Material class	Application	Predicted properties	Initial training dataset	ML/AI algorithm	Constraints	# Candidates evaluated	Year	Ref.
Co-based superalloys	Structural	γ' solvus temperature, γ' fraction	Literature (417 entries)	SVM, RF, GBoost, XGB, KNN	Thermodynamic	6 experiments	2020	106
High-entropy alloys (HEA)	Magnetism	Phase formation probability	4 HEA datasets (~2000 entries)	Evidence-based recommender	None	1 experiment	2021	76
Transition-metal borides/carbides	Mechanical	Formation energy, elastic moduli, stability	Materials Project (12 277 entries)	MEGNet GNN + Bayesian optimization	Structural Thermodynamic	2 experiments	2021	34
MnZnSb intermetallics	Magnetism	Curie temperature	ICSD, PCD (~200 000 entries)	RF Regressor	Structural	4 experiments	2021	107
FCC high-entropy alloys (Ni-based)	Mechanical	Precipitate fraction, morphology	Ni-based superalloy database (not specified)	Artificial NN	Thermodynamic Kinetic	1 experiment	2021	108
Fe–Co–B intermetallics	Magnetism	Formation energy, magnetic anisotropy	Materials Project (28 046 entries)	CGCNN (1G/2G) + AGA	Structural Thermodynamic	1 experiment	2022	75
High-entropy alloys	Mechanical	Hardness	Literature (370 entries)	SVM + SHAP	Chemical	3 experiments	2022	105
Half-Heusler materials	Energy	Thermoelectric suitability	Materials Project (456 entries)	Unsupervised clustering	Physical	2 experiments	2022	10
Fe_2P-type intermetallics	Magnetism	Transition temperature	Literature (603 entries)	GB/SVRs, artificial NNs	Structural Physical	8 experiments	2022	93
SnSe materials	Energy	Thermal conductivity	In-house dataset (776 entries)	XGBoost, RF, SVR, k-NN	Chemical	7 experiments	2023	109
Sn–Ag–Cu–Bi–In–Ti alloys	Mechanical	Strength, ductility	Experimental data + active learning (27 entries)	Active learning with GPR, Bayesian optimization (UCB)	Thermodynamic Physical	1 experiment	2024	110
B2 multi-principal intermetallics	Structural	Phase stability, mechanical performance	In-house dataset (1251 entries)	CVAE + physics-informed ANN	Thermodynamic Physical	3 experiments	2025	111
Laves phases (AB_2)	Magnetism	Curie temperature, phase stability	In-house dataset (2060 entries)	RF, GB, NN	Thermodynamic Chemical	9 experiments	2025	94
Novel intermetallics	Structural	Structure-type, latent variables	AB_3 CIFs, PCD (2366 entries)	PLS-DA, SVM, XGBoost	Structural Chemical	1 experiment	2025	112
Novel intermetallics	Structural, energy	Structure-type, thermoelectric properties	Multiple datasets (191 236 entries)	PLS-DA	Structural Chemical	1 experiment	2025	113

despite the fact that selected compositions were synthesized and characterized in subsequent analysis. Rather than generating unconstrained candidates, the workflow therefore operates as a recommendation engine over a physically admissible design space, with model performance assessed by its ability to identify experimentally realizable compositions.

Evidence-based recommenders and matrix factorization methods have been applied to alloy discovery where labelled data are sparse, enabling prioritization of compositions likely to form targeted phases and supporting direct experimental validation.[96–98] Similar logic underpins unsupervised and similarity based workflows that progressively narrow candidate sets based on proximity to known functional materials, leading to experimentally confirmed discoveries without explicit property regression.[99]

Across the literature, ML-only approaches succeed when the prediction target is conservative and closely tied to available data, such as phase existence or relative ranking within a restricted family. Failures are most common when models extrapolate beyond the structural or chemical support of the training set or when generated candidates lack an explicit pathway to synthesis.[100–103] These limitations propose the integration of first-principles information or explicit physical constraints, which is addressed in the following section.

Table 1 highlights the recent cases where ML driven ranking directly informed experimental synthesis and characterization, particularly focused on alloys and intermetallic systems. The most common application areas we bring forth here are structural performance, energy-related properties, and especially magnetism (*e.g.*, type of magnetism and transition temperature), as key metrics can be often represented by a single scalar value which is straightforward to model.

Predicted properties are generally framed within established physical theories; phase stability is often linked to formation energy and electronic structure, magnetic ordering to some form of exchange interactions, mechanical strength to microstructure, and so forth. Training data sets are often quite small, and this has been known as a problem in the materials informatics community for some time (compared to big data scaling to millions of data points). Experimental entries are even more limited, consisting of a couple hundred entries, whereas computational databases such as the Materials Project can have many thousands. In terms of the models used, there is quite a good diversity, spanning from support vector machine (SVM), neural network (NN), and ensemble tree models such as random forest (RF), gradient boosting (GB), and extreme gradient boosting (XGBoost). With regard to the data for analysis and model development, although open availability and machine-readable reporting are on the rise thanks to the recent efforts such as findable, accessible, interoperable and reusable (FAIR) standards,[104] the bulk of the data is still behind paywalls.

Across the analysed literature, only a small fraction of studies report direct experimental validation of machine-learning predictions, and even among those, negative results are seldom disclosed. This positive-outcome reporting bias, however, is commonplace across academia in general, and this bias hampers the development of models to navigate the experimental synthesis landscape. Another important aspect of progress in machine-learning-driven materials chemistry is the availability of data and code. This availability is highly inconsistent across the studies listed in Table 1. While a small fraction of studies make their data and code readily accessible, in most cases they are only available upon request. While most of the studies presented in Table 1 report only successfully synthesized compounds with predicted properties, some report experimental validation failures as well. In a study on ML-guided design of high-entropy alloys with enhanced hardness,[105] two out of three recommended compositions failed to meet their hardness targets upon synthesis, with prediction errors of 58.2% and 39.1% attributed to extrapolation beyond the training domain and insufficient data in the target compositional region, respectively. In a study on accelerated design of Co-based superalloys,[106] four out of six ML-filtered candidates were found upon synthesis to contain unwanted secondary phases beyond the desired $\gamma + \gamma'$ two-phase microstructure. In a study combining Bayesian optimization and graph deep learning for the discovery of ultra-incompressible hard materials,[34] six out of eight candidate compositions attempted *via* spark plasma sintering yielded multiphase products rather than the predicted single-phase compounds.

Most experimentally validated studies incorporate explicit domain constraints within the machine-learning workflow. In Fe_2P-type magnetocaloric materials, the search is restricted to a single structural family and a curated dataset of 558 experimentally derived compositions, following filtering from 603 reported samples.[10] The model is then applied to map transition temperatures within this predefined compositional space, rather than exploring unconstrained chemistries. In high-entropy alloy design, compositions are constrained to physically meaningful ranges of 5–35 at% per element and restricted to a limited set of alloying systems.[105] The feature space is further reduced from 142 candidate descriptors to 5 key variables, and candidate generation is guided using inverse projection and high-throughput screening to target high-performance regions.

2.3 Hybrid DFT and ML discovery pipelines

Hybrid discovery pipelines integrate ML with DFT to balance physical fidelity and scalability. In these workflows, DFT provides reference energetics and properties, while machine learning supplies surrogate models that enable ranking across large candidate spaces before targeted first principles validation. Structure-based neural networks trained on first principles datasets are widely used to predict formation energy, stability metrics, and selected functional properties, enabling systematic prescreening prior to explicit DFT relaxation and evaluation.[83,114]

A dominant hybrid pattern uses machine learning as an energy predictor embedded within structure search or optimization loops. Graph-based models approximate total energies and guide evolutionary or adaptive genetic algorithms to explore large structure spaces efficiently, with explicit DFT calculations reserved for a small subset of low-energy

Table 2 Summary of machine learning predictions embedded with first-principles calculations from 2020–2025

Material class	Application	Predicted properties	Computational details	Design space	# Candidates evaluated	Year	Ref.
Pnictides, chalcogenides	Thermoelectrics	Power factor ($S^2\sigma$)	DFT dataset (~1600 compounds) + active learning gradient boosting regressor	Binary and ternary diamond-like compounds	Computational	2020	119
Mn–Fe–P–Si intermetallics	Magnetism	Curie temperature (T_C), ΔS_m, ΔT_{hys}	Feed-forward neural network (trained on experimental datasets: $n(T_C) = 503$, $n(\Delta T_{hys}) = 465$, $n(\Delta S_m) = 660$) + DFT validation trends	Mn–Fe–P–Si composition + processing parameters	Computational	2022	95
High-entropy alloys	Low thermal expansion alloys	Thermal expansion coefficient	Active learning loop: autoencoder generative model + ensemble regression + DFT + CALPHAD	Fe–Ni–Co–Cr–Cu high-entropy alloys	17 experiments	2022	48
Laves phases	Magnetism	Formation enthalpy	Descriptor-based ML (SISSO symbolic regression) trained on >600 DFT-calculated rare-earth compounds + high-throughput screening + experimental XRPD validation	$RE(TM_1TM_2)$ compositions	3 experiments	2022	120
Metal oxides, phosphates	Autonomous solid-state synthesis	Reaction yield, phase purity, phase fractions	NLP-based recipe generator + ML temperature model + active learning (ARROWS) + DFT convex-hull screening	58 predicted targets	41 experiments	2023	35
Inorganic crystalline materials	Multiple purposes	Formation energy, decomposition energy (DFT stability)	Large-scale GNNs (GNoME)+ active learning + DFT relaxation (VASP)	>10^9 generated; 2.2M predicted stable; 381k on convex hull	Cross-validation with literature (736 structures)	2023	83
Intermetallic nanoparticles	Oxygen reduction reaction (ORR)	Ordering energy, thermodynamic stability, ORR activity	Active-learning Gaussian process regression trained on DFT formation energies + DFT slab calculations for adsorption energetics	Ternary Pt_2CoM intermetallics (M = 16 candidate elements)	2 experiments	2024	121
Inorganic crystalline materials	Multiple purposes	Stability (DFT hull), magnetic density, symmetry, mechanical and electronic properties	Diffusion-based generative model (MatterGen) + adapter fine-tuning for property constraints + MLFF pre-filtering + DFT relaxation and validation	607k structures; full periodic table	1 experiment	2025	117
Rutile-type binary oxides	Oxygen evolution reaction (OER)	OER overpotential (DFT), free-energy barriers	DFT + Gaussian-process Bayesian optimization	66 binary oxides (6 hosts × 11 dopants)	1 experiment	2025	122

candidates.[114,115] Bayesian optimization under symmetry and Wyckoff constraints further replace direct structural relaxation, allowing scalable evaluation of hypothetical crystals while preserving crystallographic constraints.[75] Several compounds identified using these approaches were subsequently synthesized and structurally validated.

Hybrid workflows are also used to overcome the limitations of static zero-Kelvin stability screening. Machine learning potentials trained on first principles data enable finite temperature sampling, configurational averaging, and free energy estimation, allowing prediction of synthesis relevant stability windows inaccessible to conventional DFT screening.[76,116] These approaches have been applied to complex alloys and ceramics, where entropic stabilization and disorder play a decisive role in phase formation.

Representative examples where temperature-dependent predictions aligned with experimental phase formation are included in Table 2. In contrast to Table 1, which emphasizes ML-guided experimental validation without first-principles calculations, Table 2 highlights workflows in which ML is tightly integrated with DFT, including cases of high-throughput studies, and in some cases, generative modelling. While both tables span roughly the same applications (energy, magnetism, structural prediction) as well as with similar materials classes (intermetallics, alloys), the key distinction here is within the scaling potential, with considerably more hypothetical structures to be screened computationally with DFT.[83,117]

Subsequently, the studies outlined in Table 2 operate over vastly larger design spaces, combining quite unique pieces (such as active learning, DFT relaxation, adsorption-energy calculations) as integral parts of the workflow to refine predictions before validation. In turn, this allows the exploration of new chemical spaces that have not been included in the original training data. Algorithmically, there is much overlap with Table 1, particularly with the use of gradient boosted models, Gaussian processes (GP), and NN. Notable differences however include shifts toward symbolic regressions such as sure independence screening and sparsifying operator (SISSO), graph NNs, and closed-loop discovery frameworks.[118] Their success reflects the incorporation of physical knowledge, conservative energetic ordering, and iterative validation, illustrating how physics-informed machine-learning frameworks can bridge the gap between computational prediction and experimental realization.

Another major class of hybrid pipelines combines machine learning screening with adaptive feedback between computation and experiment. In these closed-loop workflows, ML models rank candidates, DFT refines stability and property predictions, and experimental outcomes are used to iteratively update the ranking policy. Such closed-loop discovery schemes are demonstrated in the active-learning framework developed for high-entropy Invar alloys,[48] where generative sampling, DFT calculations, thermodynamic modeling, and experimental measurements were integrated into a continuous discovery loop. Starting from a sparse database of roughly 699 known compounds, the workflow iteratively proposed candidates, screened them using physics-informed descriptors, and

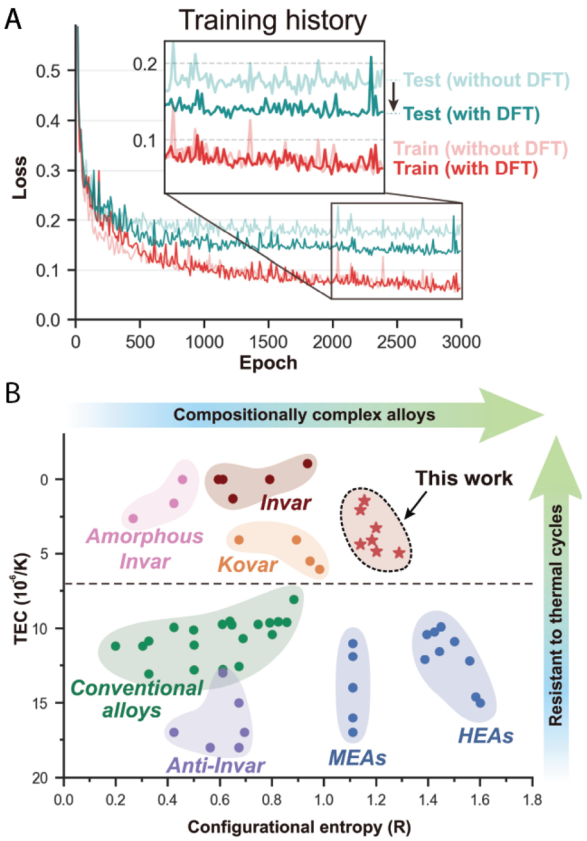

Fig. 4 Closed-loop machine-learning discovery of high-entropy Invar alloys. (A) Training and testing history of the regression model with and without physics-based descriptors, demonstrating that inclusion of DFT- and CALPHAD-derived quantities reduces prediction error and stabilizes learning under sparse data and (B) property landscape showing the thermal expansion coefficient as a function of configurational entropy for known alloys and compositions discovered in this work, highlighting the identification of alloys combining Invar-level thermal expansion with high entropy. Adapted with permission from ref. 48. Copyright 2022 American Association for the Advancement of Science (AAAS).

experimentally validated selected alloys. The inclusion of DFT- and CALPHAD-derived descriptors significantly improved predictive performance, as shown in Fig. 4a, which compares the model training and testing history with and without physics-based descriptors, reducing the testing error from about 19% to 14% compared with composition-only learning. This closed-loop strategy enabled the identification of new high-entropy alloys combining very low thermal expansion coefficients ($\approx 2 \times 10^{-6}$ K^{-1}) with high configurational entropy, illustrated in Fig. 4b, which maps the thermal expansion coefficient against configurational entropy for both known alloys and compositions discovered through the active learning loop, thereby demonstrating how adaptive ML–DFT pipelines can progressively narrow vast composition spaces under real experimental constraints. This strategy has enabled discovery of new intermetallic and magnetic compounds by progressively narrowing the design space under real validation constraints.[31]

Across the literature, hybrid DFT and machine learning pipelines consistently outperform standalone ML approaches

when experimental validation is required. Their success reflects the use of conservative physical constraints, consistent energetic ordering, and iterative validation rather than unconstrained exploration.

2.4 Machine learning interatomic potentials as engines for structure and phase exploration

Machine-learned interatomic potentials (MLIPs) represent another pathway toward physics-informed machine learning, where ML models are trained to reproduce first-principles potential energy surfaces and can therefore be used as simulation engines for exploring phase stability, kinetics, and structural transformations. MLIPs reposition machine learning from a predictive layer to a simulation engine for atomistic exploration. Rather than targeting isolated properties, these models learn potential energy surfaces from large and internally consistent first principles datasets, enabling structural relaxation, defect sampling, phase transformation analysis, and kinetic pathway exploration at scales inaccessible to direct density functional theory. This paradigm is illustrated by the low-rank potential developed for the high-entropy carbide $TiZrNbHfTaC_5$,[86] trained on DFT energies of supercells and parameterized with roughly 600 independent parameters, achieving validation errors of about 9 meV per atom. Coupled with canonical Monte Carlo simulations of supercells containing up to 16 000 atoms, the model enabled temperature-dependent sampling of configurational equilibria across the range 500–2000 K. The simulations predicted phase segregation at low temperature and entropy-stabilized single-phase behavior at high temperature, as illustrated in Fig. 5, which contrasts the spontaneous chemical segregation into multiple carbide-rich domains observed at 500 K (Fig. 5a) with the homogeneous elemental distribution recovered at 2000 K (Fig. 5b), with the layer-resolved concentration profiles in panels (c) and (d) quantifying the suppression of compositional fluctuations upon heating, directly guiding experimental synthesis conditions. The formation of the predicted $TiZrNbHfTaC_5$ high-entropy carbide was experimentally validated through electric arc synthesis, with a single-phase material obtained at 2000 K and confirmed by X-ray diffraction and microscopy analysis.

Comprehensive reviews position ML interatomic potentials as reactive force fields that support recommendation of low energy structures, metastable phases, diffusion pathways, and deformation mechanisms across metals, oxides, and complex functional materials.[123,124] Universal and transferable graph

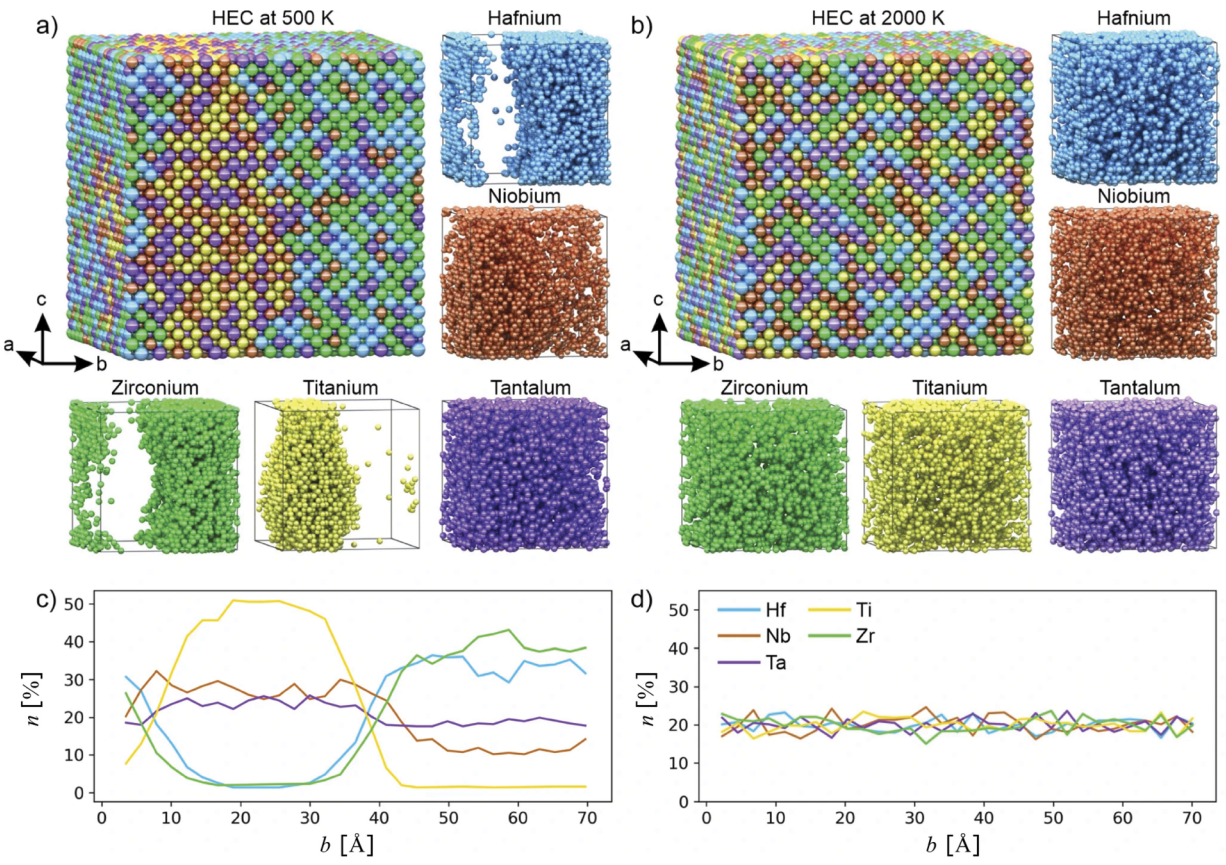

Fig. 5 Temperature-dependent configurational stability of the $TiZrNbHfTaC_5$ high-entropy carbide predicted using a machine-learned interatomic potential. (a) Atomic distribution in a simulated supercell at 500 K, showing spontaneous chemical segregation into multiple carbide-rich domains, (b) corresponding supercell at 2000 K, exhibiting a homogeneous single-phase solid solution stabilized by configurational entropy and (c and d) layer-resolved concentration profiles of metallic species along the simulation cell, quantifying compositional fluctuations in the segregated state and their suppression in the high-temperature phase. Adapted with permission from ref. 86. Copyright 2023 Springer Nature.

based architectures further expand this role by enabling chemically diverse simulations without refitting for each new system, allowing large scale structure exploration prior to selective first principles verification.[125]

Beyond structural stability, interatomic potentials enable mechanism level recommendation by resolving atomistic processes that govern macroscopic behaviour. Graph-neural-network-based potentials trained on energies and forces reproduce experimentally-measured lattice parameters, stacking fault energies, and elastic responses, enabling molecular dynamics simulations that reveal dominant deformation and transport mechanisms in chemically complex alloys.[126] Data-efficient formulations demonstrate that carefully chosen low-dimensional descriptors can outperform highly expressive representations when training data are limited.[127] For example, in a five-component refractory HEA dataset comprising ~2.9k structures (~1.5×10^5 atomic environments) spanning the full composition range of Mo–Nb–Ta–V–W alloys, models based on simple two- and three-body, and density descriptors achieved ~2–3 meV per atom accuracy while reaching acceptable predictive performance using only 25–35% of the training data-roughly a two-to-three-fold improvement in data efficiency relative to SOAP-based representations,[128] obtained through a local expansion of the Gaussian-smeared atomic density using orthonormal radial basis functions combined with spherical harmonics. These models can additionally be tabulated to enable molecular dynamics simulations that are two to three orders of magnitude faster, making large-scale simulations of multicomponent alloys computationally tractable. This behaviour reflects the fact that descriptor dimensionality controls both data requirements and sampling efficiency in compositionally complex systems. These studies demonstrate that ML interatomic potentials do not merely accelerate existing workflows but expose which atomic scale mechanisms control phase stability, strength, and ductility.

Interatomic potentials also enable thermodynamic recommendation beyond zero-Kelvin stability. Structure agnostic training combined with free energy integration allows direct prediction of temperature composition phase diagrams without assuming known crystal structures, producing phase boundaries consistent with experimental observations under processing relevant conditions.[129] In contrast to generative or purely data-driven recommendation approaches, ML interatomic potentials operate through physically grounded energy landscapes, providing a bridge between first principles theory and experimentally meaningful phase behavior. Although distinct from high-throughput screening databases[71] and end-to-end generative design models,[117] interatomic potentials define a complementary paradigm in which discovery emerges from exhaustive and physically-constrained exploration rather than direct proposal of candidate materials.

2.5 Experimental validation, practical shortcuts, and hidden assumptions

Experimental validation is the defining criterion that separates computational demonstration from materials discovery. Across ML-guided studies, validation spans a wide spectrum, ranging from confirmation of predicted crystal structures and phase stability to indirect verification through property trends or partial synthesis success. Many report validation at the level of structural existence or phase formation rather than quantitative agreement of target properties, reflecting the practical constraints of synthesis and characterization.[9,12]

A recurring shortcut in experimentally successful workflows is restriction of the design space prior to learning or screening. Rather than exploring unconstrained chemical space, many studies limit candidates to known prototypes, fixed stoichiometries, or narrow composition ranges, implicitly encoding chemical intuition into the workflow. This strategy reduces extrapolation risk[48] and increases experimental conversion rates, but also biases discovery toward incremental extensions of known structure families. Such constraints are rarely framed as assumptions, yet they play a decisive role in reported success rates.

The most explicit realization of tightly constrained, validation-centered discovery workflows is the A-Lab autonomous platform,[35] which integrates DFT-based stability screening, machine-learning-guided synthesis planning, robotic execution, and active learning within a closed experimental loop. Starting from targets predicted to lie on or near the convex hull (<10 meV per atom), the system autonomously proposed synthesis recipes, executed them using robotic powder processing and furnace operations, and evaluated products *via* automated X-ray diffraction analysis. Over 17 days of continuous operation, the platform performed 355 synthesis experiments and successfully realized 41 of 58 target compounds, corresponding to a reported success rate of 71% (Fig. 6). However, synthesis outcomes plotted against predicted decomposition energies show little systematic correlation between thermodynamic stability and experimental success. Rather than functioning as a reliable predictor of synthesizability, zero-Kelvin stability appears primarily as a coarse prioritization heuristic, with outcomes strongly mediated by precursor selection, reaction pathways, and kinetic accessibility.

This disconnect reflects a broader assumption embedded in many ML-guided discovery workflows: zero-Kelvin thermodynamic stability can serve as a proxy for synthesizability. Convex-hull distance is therefore widely used to rank candidates, yet numerous experimentally realized materials are metastable, kinetically trapped, or stabilized only under specific processing conditions. Stability metrics thus capture only one dimension of experimental feasibility while neglecting reaction pathways, competing intermediates, and practical constraints of precursor chemistry. Recognizing this limitation, several recent studies[81,82,130,131] incorporate auxiliary recommendation layers, including synthesizability classifiers, synthesis-aware ranking schemes, or heuristic filters derived from experimental data, to bridge the gap between thermodynamic prediction and laboratory realization. These approaches acknowledge that reliable experimental targeting requires integrating thermodynamic, kinetic, and procedural knowledge rather than relying on stability prediction alone.

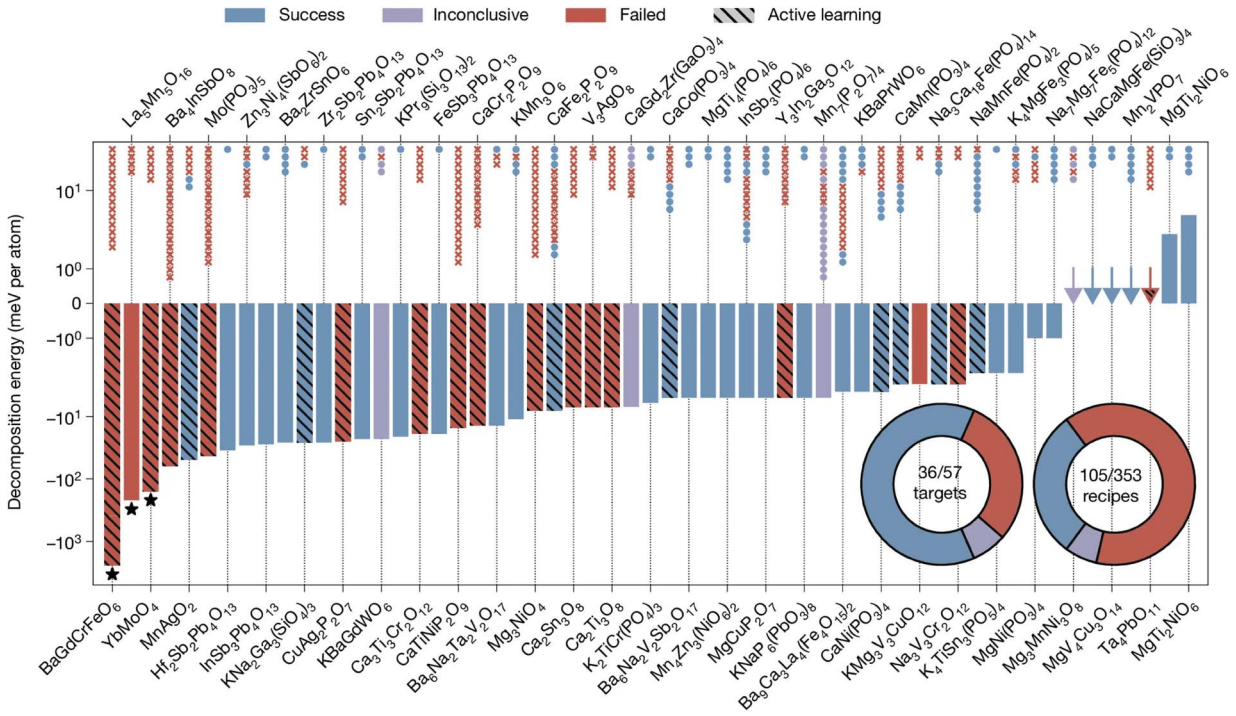

Fig. 6 Experimental outcomes of A-Lab synthesis campaigns plotted against predicted decomposition energies. Each bar represents a target compound, coloured by successful realization, failure, or optimization *via* active learning, while overlaid markers indicate outcomes of individual synthesis recipes. The inset charts summarize the fractions of successful targets and successful recipes. The distribution demonstrates that experimental success is only weakly correlated with zero-temperature thermodynamic stability, highlighting the importance of kinetic factors and synthesis pathway design in autonomous materials discovery. Adapted with permission from ref. 35. Copyright 2023 Springer Nature.

Finally, validation itself is often selective rather than exhaustive. Experimental follow-up typically focuses on top ranked candidates, while lower ranked or ambiguous predictions remain untested, limiting the ability to assess false negative rates. Closed-loop workflows partially address this limitation by iteratively updating ranking policies based on experimental feedback, but they remain resource intensive and are still applied to narrowly defined systems. Taken together, these observations indicate that experimental validation in ML-guided discovery is inseparable from workflow design choices, shortcuts, and assumptions that must be made explicit when assessing reported success.

Establishing validation strategies and incorporating physical constraints addresses how models should be evaluated and guided. However, an equally critical challenge lies in defining which properties are most meaningful to predict and measure, and how these properties are connected to experimental conditions and materials realization.

3 Approaches with a focus on interpretability/explainability

To address this challenge, ML-guided discovery can focus less on algorithmic novelty and more on model interpretability (model's inherent knowledge extraction) and explainability (*post hoc* knowledge extraction method) with direct chemistry and physics-informed insight, providing some correlation with experimental results. The following case studies reflect how we progressively moved from using ML as a diagnostic tool to deploying it as a recommendation engine directly embedded in experimental decision-making. Together, they illustrate the strengths, limitations, and practical design choices that shape real-world ML-driven discovery workflows.

3.1 Recommendations as a test set for ML models

3.1.1 Classifying binary equiatomic *AB* structures[132] (RhCd). An earlier effort to apply ML in the crystal structure classification was with an equiatomic binary *AB* study, in which crystallographic data was extracted from the Pearson Crystal Database[66] in tandem with the ASM Alloy Phase Diagram Database[133] (excluding hydrogen, noble gases, and heavy radioactive and actinide elements, and having a strict 1 : 1 ratio). 974 compounds composed a dataset of experimentally confirmed compounds to exist under ambient temperatures and pressures, crystallizing in 107 unique structure types. Although many of these structure types are rare, 706 adopt one of seven common structure types, with each having at least 30 representatives (Fig. 7a). For modelling, the data were represented in a 974 × 56 matrix of elemental descriptors, and only the 706 compounds belonging to these seven structure types were retained to ensure statistical reliability. Both PLS-DA and SVM techniques were used to train models on this dataset.

Through a procedure of forward selection and backward elimination for iterative feature selection, our 56 initial

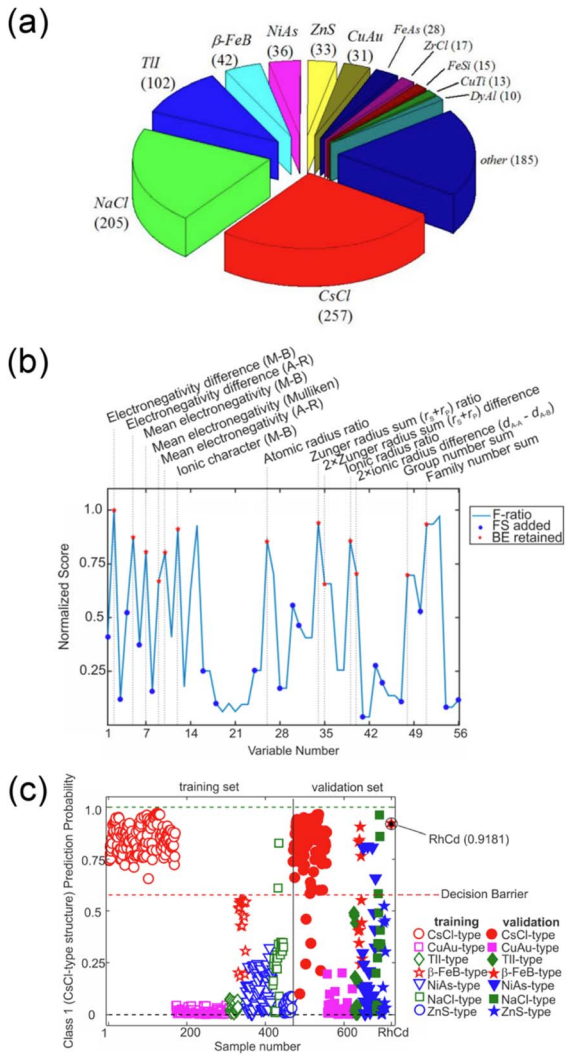

Fig. 7 (a) Structure types adopted by binary AB phases. (b) Fisher ratio scores for variables selected in the CR-FS procedure. (c) Predicted probability for CsCl-type structures using a machine-learning model based on SVM. Adapted with permission from ref. 101. Copyright 2016 American Chemical Society.

Table 3 PLS-DA and SVM results

Model	Sensitivity (%)	Specificity (%)	Accuracy (%)
PLS-DA	96.5	66.0	77.1
SVM	94.2	92.7	93.2

descriptors were filtered down to 31, in an unbiased manner (highlighted in Fig. 7b). Model results are displayed in Table 3. The SVM model, in conjunction with our feature selection process, gave excellent performance for classifying the crystal structures of *AB* compounds. To validate the predictive results, RhCd was synthesized, which at the time of discovery was the first new binary *AB* compound to be discovered in over 15 years. The synthetic procedure was done by annealing a stoichiometric ratio of the elements at 800 °C after sintering and characterizing it through X-ray diffraction. Among the list of candidates with high prediction probability (Fig. 7c), the RhCd candidate (91.8% probability) was chosen based on the straightforward synthesis procedure and that underexplored systems contained precious metals.

The recommendation for the experimental validation in this study is based on the test set, where the prediction probability is the only factor driving the recommendation. The feature selection, even on the relatively small feature set, improves the model performance and enhances the explainability.

3.1.2 Interpreting experimental outcomes of ML prediction in ternary *ABC* phases. A similar approach was later done for equiatomic ternary phases, with 1037 individual compounds that adopt the most populated ternary 1:1:1 structure types (TiNiSi-, ZrNiAl-, PbFCl-, LiGaGe-, YPtAs-, UGeTe-, and LaPtSi-type) and validated using a held-out set of 519 compounds.[134] The same feature selection as before was applied, and only SVM was used here due to its earlier success. The final model prediction sensitivity, specificity, and accuracy were 97.3%, 93.9%, and 96.9%, respectively, establishing the previous reliability of our binary model for predicting the crystal structure given only its composition. It is important to note two things regarding the ternary phases as compared to the binary phases. Firstly, extracting data from the literature, including ternary phase diagrams, is considerably more difficult and time-consuming than for the binary compounds. Secondly, the problem of polymorphism, in which more than one structure forms under different conditions, presents an additional challenge for the interpretation of the results.

For example, among equiatomic *ABC* phases, polymorphism between the orthorhombic TiNiSi-type or the hexagonal ZrNiAl-type structures is prevalent and especially common for metal-rich phosphides M_2P.[135] Analysing 19 compositions that are experimentally reported in both structure types, the model correctly identifies with high confidence (>0.7) the low-temperature polymorph from its high-temperature form. Interestingly, the class probability plot (Fig. 8a), especially in the validation subset, has datapoints misclassified or with the probabilities close to the decision barrier. Typically, this indicates a lower model quality, but with the chemistry knowledge behind the datapoint, it reveals an interesting phenomenon. ML suggested that certain compositions cannot be clearly differentiated and lie in a region of *confusion* (confidence score ranging between 0.3 and 0.7) (Fig. 8b), suggesting both polymorphs may be observed in a single sample at varying synthetic conditions.

At first, these intermediate probability values, lying in this *confused* region, may seem uninformative or even indicative of poor model performance. In reality, they reflect cases where experimental reports show that closely related synthesis conditions can yield two different polymorphs. This is unknown to the software engineer who works outside of the traditional chemistry realm; thus, we think that it is of the utmost importance for software to be not only validated, but also curated within each domain. We cannot further stress that chemical intuition can only come from being part of the field. In such polymorphic systems, one phase is likely thermodynamically

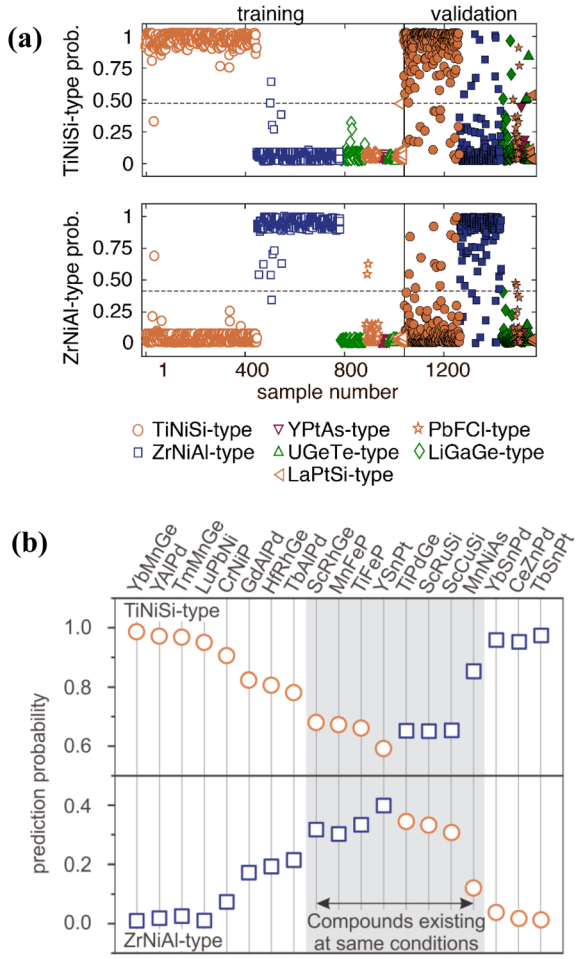

Fig. 8 (a) Predicted probability for TiNiSi-type (upper panel) and ZrNiAl-type (lower panel) structures based on an SVM model, with a dashed line marking the decision barrier. (b) Predicted probability for 19 compounds adopting both TiNiSi- and ZrNiAl-type structures. Adapted with permission from ref. 134. Copyright 2017 American Chemical Society.

favored. While the other persists in a metastable form, its transformation is suppressed. Remarkably, the model appears to encode this subtle energetic competition, capturing behavior that is often difficult to resolve even with first-principles approaches. A striking example is TiFeP, which exhibits one of the most uncertain predictions, 60% likelihood for the TiNiSi-type structure and 40% for the ZrNiAl-type structure. Consistent with this ambiguity, the literature reports conflicting structural assignments. A reinvestigation confirms that TiFeP readily forms as a two-phase mixture, even after extended annealing, likely due to a kinetically hindered phase transformation resulting in the coexistence of both phases rather than simple aforementioned model confusion.

Of a similar flavor, recent work has extended the ability of actually predicting synthesizability from composition alone to the level of specific crystal structures, directly addressing the previous polymorph problem.[136] By fine-tuning large language models (LLMs), which have thus far not been highlighted in this text, on text-based descriptions from CIFs, researchers framed synthesizability as a positive-unlabelled (PU) learning task using experimentally realized *versus* hypothetical structures. Models incorporating these explicitly structural descriptions significantly outperformed stoichiometry-only approaches and were competitive with, or better than, the bespoke GNNs. Even greater gains were achieved by using LLM-derived embeddings as structural representations for a dedicated PU classifier, suggesting that the text-based encodings here capture symmetry, coordination, and environmental information that traditional graph constructions may incompletely represent, particularly for the structurally complex systems where polymorphism is prevalent.[137]

The logic behind the LLM is quite simple and intuitive, but also interacts with the user's prompt by directly addressing why or why not a particular polymorph is synthesizable. Using GPT-4o[138] and associated probabilities, returning "*true*" to a question can be traced back to a principle common in textbooks. For example, the LLM can respond to a query with "*If a new compound exhibits uniform bond lengths, then it likely has high internal strain, indicating that the compound is unstable and unsynthesizable.*" Other reasons such as thermodynamic stability alone account for a minority of predictions, reinforcing the experimental reality that energy-above-hull criteria are insufficient proxies for synthetic accessibility. Sensitivity analyses done further show that small perturbations to atomic positions dramatically alter predicted synthesizability, underscoring the delicate structural balance governing polymorph selection. In this way, intermediate probabilities and structural uncertainty, such as those observed in our TiNiSi/ZrNiAl case, may reflect genuine thermodynamic–kinetic competition rather than model confusion, highlighting how interpretable frameworks can augment chemical intuition in understanding and predicting polymorphic crystal structures.

Shifting back our focus to structure prediction of polymorphs, the ParetoCSP2 work did not treat polymorph prediction as a simple global-minimum search problem,[139] but instead explicitly interrogated why conventional algorithms fail to recover experimentally-known competing phases. The authors demonstrated here that single-objective evolutionary searches tend to prematurely converge into one symmetry basin, which ends up suppressing structurally distinct but energetically-close polymorphs. By reformulating structure prediction as a multi-objective optimization problem, balancing energy with symmetry diversity, the method preserves multiple structural lineages throughout the search. This design makes the algorithm's behavior interpretable at the population level: one can directly monitor how different symmetry classes/space-group representations are maintained, suppressed, or amplified over generations, rather than observing a black-box collapse into a single dominant structure type.

The study done by the authors showed that recovery of experimentally-reported polymorphs correlates with sustained representation of distinct space groups during optimization, rather than strictly with lowest-energy ranking. In several benchmark systems containing multiple Materials Project[71] polymorphs with identical cell sizes, higher-energy but

experimentally realized structures only appeared when symmetry diversity constraints were enforced.

This ultimately provided some mechanistic insight into polymorph competition, particularly that structural accessibility emerges from the interplay between near-degenerate energy basins and search-path diversity, rather than absolute energetic ordering alone. Much like the *region of confusion* between TiNiSi- and ZrNiAl-type phases, ParetoCSP2 makes visible the existence of parallel low-energy funnels and demonstrates that algorithmic diversity control can reveal genuine thermodynamic–kinetic competition, offering a more interpretable picture of why multiple structure types coexist within a single chemical system.

3.2 Elemental reactivity maps as a constrained recommendation framework for ternary materials discovery

Predicting which elemental combinations are likely to yield stable compounds, prior to any consideration of composition or crystal structure, represents a practically powerful but methodologically underexplored entry point into materials discovery. This approach directly addresses the combinatorial explosion of candidate ternary systems by framing elemental reactivity as a binary classification task, deliberately avoiding the more complex and computationally demanding problem of predicting specific compositions or crystal structures.[140] The central methodological challenge was the positive-unlabeled (PU) nature of available data: unreported elemental combinations cannot be assumed to be non-reactive, as they may simply remain unexplored experimentally. To construct reliable negative training data, a kNN-based similarity scoring procedure using cosine similarity between crystallographically motivated elemental descriptors was developed. Unlabeled element sets (UESs) with the lowest average similarity to known reactive element sets (RESs) were designated as reliable non-reactive element sets (RNRESs), with four models trained at negative data thresholds of $x = 25$, 50, 75, and 100%.

Training data were sourced from the Materials Project (queried April 2018, energy above hull = 0), yielding 19 835 stable compounds mapped onto 14 099 RESs as positive data, against 71 301 UESs drawn from 85 400 total single, binary, and ternary combinations among 80 laboratory-accessible elements. The machine learning model was a permutation-invariant neural network in which descriptors for each of the three constituent elements were independently processed by a shared two-layer sub-network, then max-pooled into a single order-invariant representation before a final classification layer with sigmoid output. The permutation invariance and use of coordination-geometry-derived elemental descriptors, rather than learned embeddings, make the approach interpretable by design: the model captures chemically meaningful elemental features rather than arbitrary statistical correlations, and the predicted reactivity scores are visualized as two-dimensional heatmaps across the full periodic table for any target element. As illustrated in Fig. 9 for Al-containing ternary systems, these maps allow researchers to directly identify element set combinations with high predicted reactivity but no reported compounds, with known Materials Project and ICSD entries overlaid as reference points.

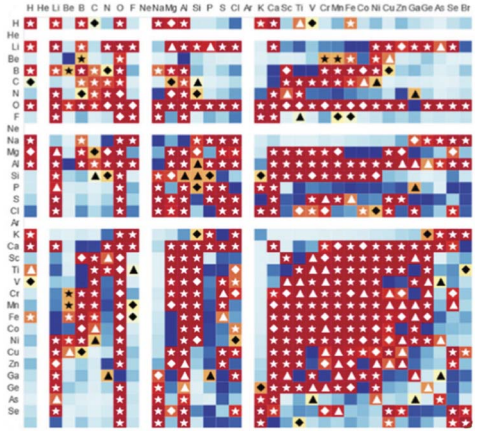

Fig. 9 Elemental reactivity map for Al-containing ternary systems predicted by Model [75%], visualized as a two-dimensional heatmap where deeper blue indicates higher predicted reactivity. Red markers denote element sets with known compounds reported in the Materials Project (stars) or ICSD (diamonds and triangles), while unmarked dark blue squares identify unreported but highly reactive element combinations as candidates for novel materials discovery. The Co–Al–Ge system, highlighted as an unexplored high-reactivity candidate, was subsequently confirmed through synthesis of two novel ternary compounds. Adapted with permission from ref. 140. Copyright 2025 American Chemical Society.

Validation against an independent external dataset of 5852 ICSD entries absent from the Materials Project confirmed that predicted reactivity scores correlate with experimental synthesizability. Model [25%] recovered 91% of high-quality ICSD entries as reactive, while Model [75%] improved precision, reaching 24% true positive rate among all predicted reactive UESs compared to 15% for Model [25%], demonstrating a controllable precision-recall tradeoff governed by the negative data threshold. Experimental validation targeted the Co–Al–Ge ternary system, identified as a high-reactivity UES (score 0.987 under Model [75%]) with no prior Materials Project entry. Simple solid-state synthesis yielded two novel compounds, B20-type $Co_4Ge_{3.19}Al_{0.81}$ and Ni_2Al_3-like $Co_2Al_{1.26}Ge_{1.74}$, confirmed by single-crystal X-ray analysis, a result that would likely have been missed under composition- or structure-constrained search strategies.

3.3 Unsupervised ML approaches ($TbIr_3$)

Unsupervised ML approaches are the rarest kind in experimentally-validated ML works, especially in the solid state domain. Recently, the prediction and synthesis of novel $TbIr_3$ ($PuNi_3$-type) was reported.[112] The dataset for this was extracted from Pearson's Crystal Database,[66] and encompassed 2366 entries with 97 total features (44 compositional, 53 structural). The features were generated from custom-made Python code, namely, the composition analyzer featurizer (CAF)[141] and structure analyzer featurizer (SAF),[142] with more structural features added from *cifkit*.[143] To enable broad, reproducible exploration of these data, we developed a Python-based

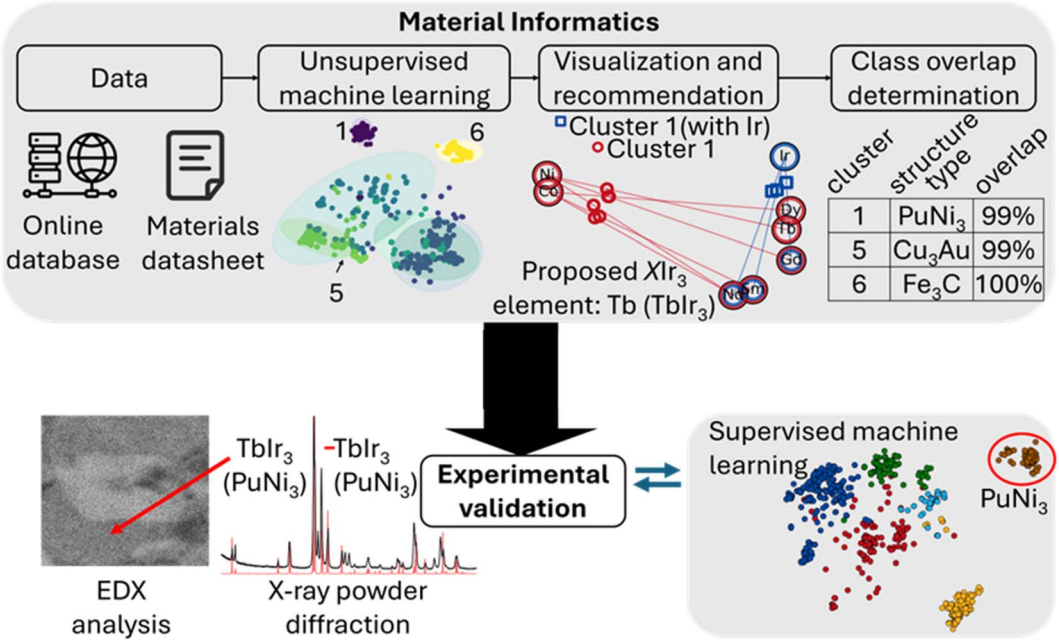

Fig. 10 Application of unsupervised ML for novel structure prediction, followed by supervised ML confirmation, and experimental validation. Adapted with permission from ref. 112. Copyright 2025 American Chemical Society.

workflow, *Automated Machine Learning Workflow for Excel Data*,[144] that supports both unsupervised and supervised analyses from Excel-structured inputs. For visualization and recommendation purposes, we developed *Structure-Type Explorer (STEx)*,[145] for exploring structure types in a user-defined chemical space. The recommendation engine is an augmentation of chemical intuition that provides a scored list of candidates to expand the structural series, wherein the highest score would be the most promising candidate. A workflow of the process is presented below in Fig. 10.

In the unsupervised track, PCA was applied for dimensional reduction, followed by clustering (K-means, DBSCAN, hierarchical), using cluster separability and visualization (*e.g.*, PCA projections and confidence ellipses) to guide parameter selection. This approach identified well-segregated clusters that could be interpreted *post hoc* as structure families; notably, the PLS-DA scatterplot correctly assigned the $TbIr_3$ compound (CIF input obtained from Rietveld refinement) to the $PuNi_3$-type structure cluster (Fig. 11), demonstrating the success of the recommendation engine and validity of the supervised ML approach. The recommendation engine was then used to propose chemically adjacent candidates for exploratory synthesis, leading to $TbIr_3$ as the top-ranked extension of AIr_3 (where A = Y, La–Nd, Sm, and Gd). Experimental validation *via* arc-melting and annealing (which was independently reproduced in two laboratories) followed by PXRD and Pawley/Rietveld refinement confirmed the presence of $TbIr_3$ in the $PuNi_3$-type structure, while also revealing that multiphase mixtures are difficult to avoid, consistent with kinetically suppressed transformations in this composition range as with the previous ternary problem. Finally, a supervised validation step using the same feature representation showed near-perfect agreement with the experimental assignment, with PLS-DA, SVM, and XGBoost achieving quite high accuracies (Table 4), supporting the robustness of the structure-type classification and the predictive value of combining unsupervised discovery with supervised confirmation. A compilation of related unsupervised works for a variety of intermetallic-related outcomes

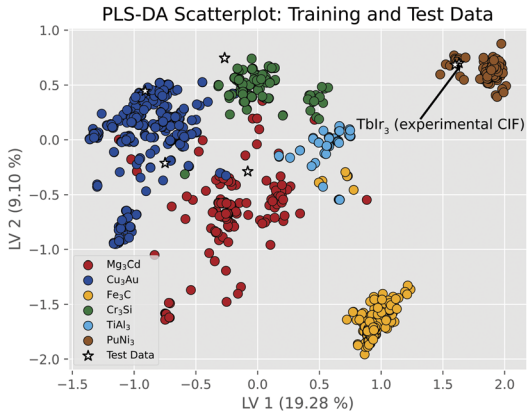

Fig. 11 PLS-DA plot of intermetallic AB_3 structure type classification.

Table 4 PLS-DA, SVM, and XGBoost results

Model	Accuracy (%)
PLS-DA	96.7
SVM	99.7
XGBoost	99.9

Table 5 Summary of some unsupervised works

Goal	Type	Ref.
Reveal separability of 7 intermetallic structure classes in latent space	Manifold learning (t-SNE) on learned structural embeddings	147
Identify materials families in Cu–S space, link clusters to structure and properties	Wasserstein distance + clustering	148
Determine crystal structures directly from powder XRD without prototype bias	Symmetry-constrained model search guided by pattern similarity	149
Decompose high-throughput XRD datasets into phase end-members	Non-negative Matrix Factorization (NMF)	150
Discover distinct atomic environments (bulk, defects, surfaces) without labels	Manifold learning + clustering on local descriptors	151
Explore vast superalloy design space to sample promising candidates	Clustering + similarity-based exploration	152
Organize Li-containing compounds *via* diffraction-derived representations	Unsupervised learning on digital XRD features	153
Identify structural patterns correlated with superconductivity	Descriptor-based clustering/representation learning (SOAP)	154

(*i.e.*, structure clustering, design of materials) is provided below in Table 5. A particular value of the unsupervised models is its ability to provide counter-intuitive recommendations. In the case of $TbIr_3$ prediction, suggesting a different structure from two recently reported polymorphs[146] is an excellent case of the bias, which researchers might have against exploring well-studied domains.

3.4 Recommendation engines with exploratory purposes ($Gd_{10}RuCd_3$)

A similar recommendation engine (workflow presented in Fig. 12) was used to predict novel intermetallic $Gd_{10}RuCd_3$ ($Y_{10}RuCd_3$-type),[113] combining strong neutron absorption (99.96% percentile performance) with negative thermal expansion, a candidate for neutron management and criticality control. This method was particularly helpful as traditional experimental trial-and-error is quite costly with *RE* elements. Using crystallographic site parsing from CIFs, each unique atomic site in a given structure type was treated as a class and analyzed using PLS-DA. The resulting latent variables define a chemically meaningful projection space in which both known and hypothetical elements can be positioned according to their similarity. This enables direct visualization of chemical trends and systematic ranking of unexplored substitutions. Unlike black-box recommendation systems, our framework, like our models, is fully interpretable, with elemental contributions traced to underlying physical and chemical descriptors such as electronegativity, valence, radii, cohesive energy, bulk modulus, *etc.*

In contrast to the earlier *STEx* implementation, which constructed a global PCA chemical space from the full elemental descriptor set, the present workflow first derives a site-resolved PLS-DA space from elements experimentally observed in each crystallographic position of the target structure type. The remaining elements of the periodic table are then projected into this latent space using the extracted variable loadings, enabling expansion of candidate substitutions while preserving structure-type-specific chemical constraints.

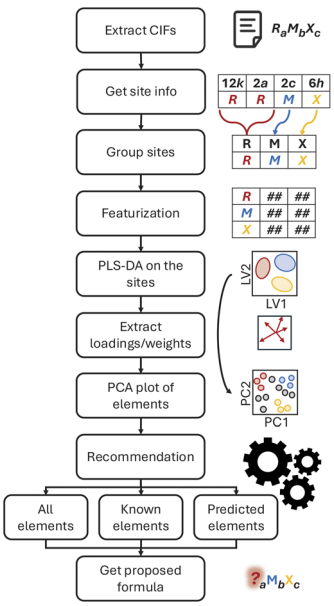

Fig. 12 Recommendation engine workflow. Adapted with permission from ref. 113. Copyright 2025 American Chemical Society.

A key development here was the implementation of three complementary explainable recommendation modes: (i) an unrestricted search that identifies elements closest to any previously occupied site, (ii) a conservative mode restricted to elements already known to occupy that crystallographic site in related compounds, and (iii) a cluster-based method that limits candidates to the PLS-DA-defined chemical space of the target structure family. Types of recommendations are displayed in Fig. 13. The final candidate is selected using a weighted consensus score across all three methods. Applied to the $RE_{10}RuCd_3$ structure type, this strategy identified $Gd_{10}RuCd_3$ as the top-ranked unexplored composition, despite the complete absence of any prior reports in the Gd–Ru–Cd system in the most recent version of Pearson's Crystal Database. Subsequent

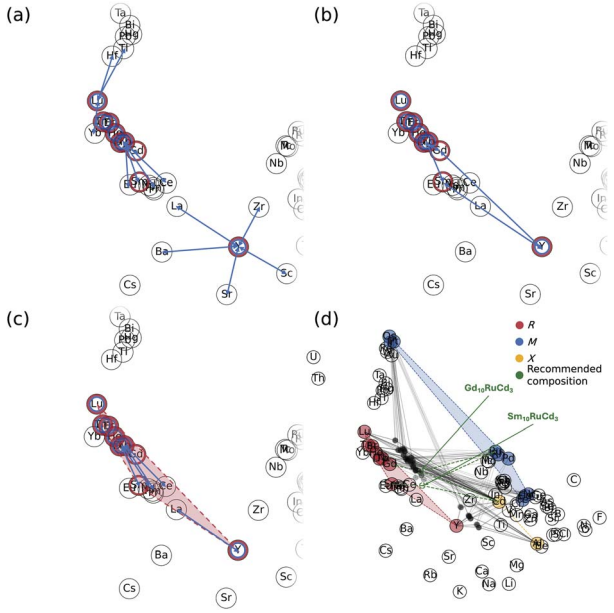

Fig. 13 Types of recommendations: (a) unrestricted method, proximity to all elements, (b) conservative method, proximity to the elements observed in a particular site in a particular structure, (c) cluster method, proximity to elements within the cluster area, (d) combined score result plotted as a recommended compound. Adapted with permission from ref. 113. Copyright 2025 American Chemical Society.

synthesis and PXRD/SCXRD confirmed the structure, validating the predictive capability of the engine. This compound also displayed unusually low thermal conductivity.

3.5 Constraint-aware, interpretable chemical recommendation framework (*CRAFT*)

Building on the explainable recommendation strategies demonstrated for $Gd_{10}RuCd_3$, *STEx* was cited, although at that point in time, it was renamed to *CRAFT* (Chemical Recommendation & Analysis for Future Targets) as a constraint-aware, interpretable framework for proposing new compound compositions within user-defined chemical and physical boundaries.[155] In the work with $Gd_{10}RuCd_3$, *STEx* was implemented and refined to *CRAFT*, as a Python-based workflow with a *Jupyter* notebook interface enabling the same site-resolved analysis and substitution-driven discovery using chemically meaningful descriptors. Particular differences from the $TbIr_3$ recommendation to the newly-developed $Gd_{10}RuCd_3$ include a total of three recommendations rather than just one. The core engine applies PLS-DA to site-specific elemental classes derived from crystallographic parsing, followed by PCA on the learned loadings to construct a low-dimensional chemical similarity space. Within this space, both known and hypothetical elements are projected, enabling visualization of trends and systematic ranking of candidate substitutions.

The updated iteration of *CRAFT* herein includes fixed-site substitution, alternative fixed-site combinations, and exploratory searches within the learned chemical manifold, with optional constraints such as charge balance, thus applied as *post hoc* filters removing chemically implausible candidates without altering the underlying model logic. This framework represents a generalizable, interpretable approach to chemical recommendation that augments intuition while remaining compatible with experimental feasibility and domain-imposed constraints.

3.6 Targeted discovery of layered ternary antimonides *via* formation energy screening and interpretable structural classification

A representative example of combining DFT-based prescreening with an interpretable ML structural classifier is the discovery of ternary antimonides A–Al–Sb (A = Rb or Cs),[156] targeting compounds with specific structural motifs relevant to thermoelectric applications. Formation energies for several hundred hypothetical compositions $Rb_xAl_ySb_z$ and $Cs_xAl_ySb_z$ were estimated from composition alone using an online model trained on DFT calculations from the OQMD,[70] identifying $RbAlSb_2$, $CsAlSb_2$, $Rb_2Al_2Sb_3$, and $Cs_2Al_2Sb_3$ as priority targets. To predict structural motifs – clathrate, layer, channel, or network, machine learning models were then trained on 119 experimentally known ternary Rb- and Cs-containing Zintl phases and polar intermetallics drawn from Pearson's Crystal Data,[66] using 23 features combining compositional subscripts with elemental properties of the M and X components (electronegativity, covalent and metallic radii, atomic number, Mendeleev number, period, group, melting point, and boiling point).

Given the small dataset size, conventional algorithms such as support vector classifier and random forest were benchmarked but performed considerably more poorly than the sure independence screening and sparsifying operator (SISSO) method, which constructs low-dimensional descriptors by combining primary features through algebraic operators across up to 10^{10} candidate combinations. Classification was achieved in two stages: a coarse separation of open structures (clathrates, layers, channels) from denser 3D networks, followed by finer discrimination among the open classes. A key strength of the SISSO approach is its inherent interpretability – the resulting descriptors take the form of explicit mathematical expressions involving physically meaningful quantities such as differences in electronegativity or melting and boiling points, allowing the structure maps to be read and rationalized in chemical terms rather than treated as black-box outputs.

Experimental validation confirmed three of four predicted targets: $RbAlSb_2$, $CsAlSb_2$ (both adopting layered tetragonal structures in space group $P4_2/nmc$), and $Rb_2Al_2Sb_3$ (monoclinic, $Na_2Al_2Sb_3$-type, $P2_1/c$), while $Cs_2Al_2Sb_3$ could not be synthesized despite multiple attempts at varying temperatures. All confirmed structures were layered, in agreement with the ML predictions. Reinvestigation of the related compound $Cs_2In_2Sb_3$ additionally revealed a low thermal conductivity of 0.64 W m^{-1} K^{-1} and p-type semiconducting behavior with the Seebeck coefficient reaching 253 μV K^{-1} at 300 K, demonstrating how ML-guided structural targeting can simultaneously accelerate property discovery in underexplored compositional spaces.

3.7 Inverse design of FeNiCrCoCu MPEAs with explainable AI

Integrating surrogate machine learning models with evolutionary optimization algorithms into a closed inverse design loop represents a scalable alternative to exhaustive first-principles or molecular dynamics (MD) searches for alloy composition discovery. This study[157] developed a four-stage computational workflow targeting FeNiCrCoCu multi-principal element alloys (MPEAs) with high bulk modulus and unstable stacking fault energy (USFE), two mechanical descriptors governing hardness and resistance to nanoscale plastic deformation. Rather than sampling compositions uniformly, the workflow used Particle Swarm Optimization (PSO) guided MD simulations in the first stage to generate an enriched, targeted training dataset, concentrating data density in compositionally relevant regions of high bulk modulus and USFE before any machine learning model was trained.

The training dataset was generated entirely from MD simulations using an embedded atom method (EAM) force field on FCC systems of 4000 atoms, with PSO iteratively adjusting elemental concentrations of Cu, Cr, Co, and Fe within 5 to 35 at% each, with Ni as the balancing element. After 137 PSO optimization cycles, a total of 19 728 compositions comprising 59 184 individual structures were generated. Two separate surrogate models were developed on this data. A two-layer stacked ensemble machine learning (SEML) model, combining Bayesian Ridge regression, Stochastic Gradient Descent regression, and a Multilayer Perceptron in its first layer, with a second-layer MLP aggregating their outputs, was trained on elemental concentrations as input features to predict USFE, achieving $R^2 = 0.87$ and MAE below 2.5 mJ m^{-2}. For bulk modulus prediction, a one-dimensional convolutional neural network (1D CNN) was developed, taking as input a spatially ordered 4000-atom array encoding element identity and local neighbour relationships across 20 atomic layers, achieving $R^2 \geq 0.98$ and MSE ≤ 0.16 GPa^2 on both Top10K and Top25K subsets. The structural encoding of the atom array was designed to preserve local chemical short-range order (CSRO), enabling the CNN to capture neighbourhood-level compositional effects rather than relying solely on mean-field concentration descriptors. These two surrogate models were then integrated with three independent optimization algorithms, namely PSO, a Genetic Algorithm (GA), and a reinforcement learning method using temporal difference (TD) learning with SARSA, each exploring the compositional design space over 200 iterations with 128 candidates per iteration.

Explainability was addressed through SHapley Additive exPlanations (SHAP) analysis applied separately to both models, providing *post hoc* interpretation of feature contributions. As shown in Fig. 14a, SHAP analysis of the SEML model revealed a clear ranking of elemental importance for USFE prediction, with Fe and Ni identified as the dominant positive contributors (average absolute SHAP values of 4.09 and 2.98 mJ m^{-2}, respectively), consistent with their higher intrinsic USFE values of 485.91 and 409.29 mJ m^{-2} as calculated from the EAM potential for pure FCC structures. The Pearson correlation matrix in Fig. 14b further confirms these trends quantitatively, showing moderate positive correlations of Fe and Ni with USFE (0.55 and 0.44, respectively), while Cu, Cr, and Co displayed negative correlations of −0.21, −0.20, and −0.27, attributed to weaker metallic bonding and associated electron density changes at the stacking fault interface. For the CNN model, atom-wise SHAP values demonstrated that more than 99% of Cu and Cr atoms carried negative SHAP contributions to the bulk modulus, while over 70% of Fe, Ni, and Co atoms contributed positively. Critically, the CNN SHAP analysis further revealed that the sign of an element's contribution was not determined solely by its identity but also by its local chemical environment: Co and Ni atoms surrounded predominantly by same-species neighbours were more likely to show positive SHAP values, indicating that local clustering and CSRO directly modulate the effective mechanical contribution of individual atoms.

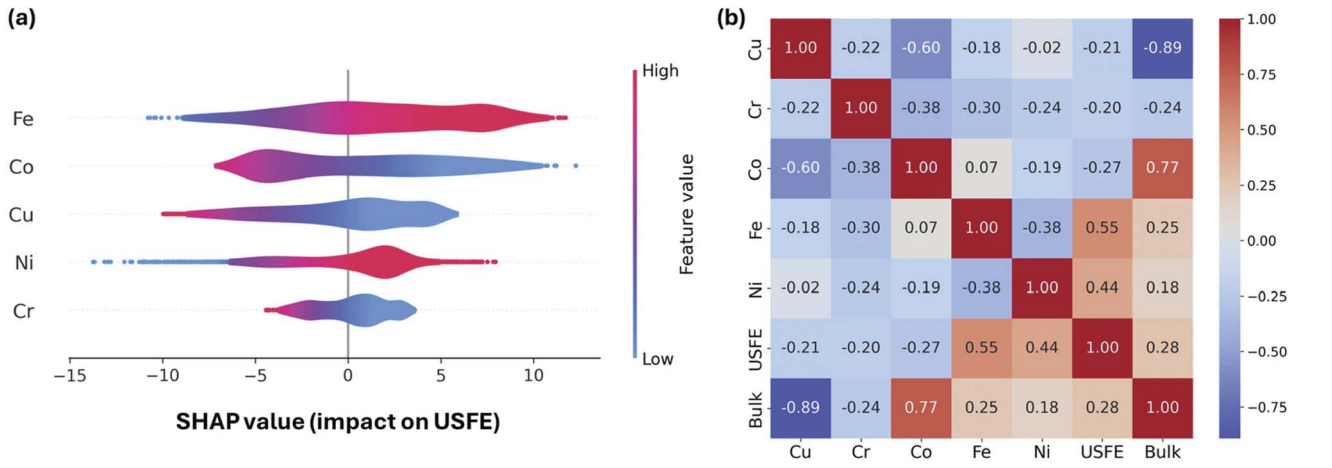

Fig. 14 SHAP analysis of the SEML model for USFE prediction. Panel (a) shows elemental feature importance ranked by the average absolute SHAP value, where red and blue indicate high and low elemental concentrations, respectively. Panel (b) presents the Pearson correlation matrix confirming that higher Fe and Ni concentrations positively correlate with USFE, while Cu, Cr, and Co show negative contributions. Adapted with permission from ref. 157. Copyright 2025 Springer Nature.

Five candidate compositions were selected for experimental synthesis by arc melting, three derived from the PSO-guided MD training data and two from the ML and DL optimization outputs. All five produced single-phase face-centred cubic structures confirmed by powder X-ray diffraction. Measured hardness for all new MPEAs exceeded 3.0 GPa compared to 1.49 GPa for the equimolar CuCrCoFeNi reference, and measured Young's moduli ranged from 179.86 to 197.74 GPa, in qualitative agreement with MD predictions despite known discrepancies arising from strain rate differences and force field limitations. The composition $Cu_{5.35}Cr_{5.37}Co_{29.55}Fe_{31.21}Ni_{28.52}$ achieved the highest experimental Young's modulus of 197.74 $\pm$ 6.03 GPa, consistent with the workflow's prediction that low Cu and Cr fractions favor enhanced stiffness.

4 Conclusions

A consistent theme emerges across these various ML-guided strategies, namely that successful materials discovery is driven less by unconstrained prediction accuracy and more by recommendation frameworks wherein the main driving forces are those respecting physical, chemical, and experimental constraints. Standalone ML approaches are effective when applied to conservative, well-defined tasks. Hybrid ML-DFT pipelines and MLIPs provide the physical grounding necessary for robust experimental validation. Additional studies demonstrate that interpretable, structure-aware models can capture subtle energetic competitions, reveal polymorphic uncertainty, and guide synthesis toward experimentally-realizable compounds, even in sparsely-explored chemical spaces. By explicitly encoding constraints, prioritizing explainability, and *closing the loop* with experiments, ML transitions from a descriptive tool to an active decision-making engine, which is increasingly being adopted at the industrial level. We propose that future progress in crystal structure discovery will not depend on ever more complex algorithms, but on the thoughtful integration of ML with chemical intuition, first-principles insight, and experimentally-informed validation strategies that will make discovery reproducible, interpretable, and actionable – particularly a combinatorial strategy of screening, and preliminary and concluding validation.

Author contributions

Anton Oliynyk: formal analysis, visualization, writing – review & editing, supervision. Danila Shiryaev: formal analysis, writing – original draft. Emil I. Jaffal: formal analysis, writing – original draft. Sangjoon Lee: writing – review & editing. Balaranjan Selvaratnam: review & editing.

Conflicts of interest

The authors report that there are no conflicts to declare.

Data availability

No new data was generated or used in the current review article.

Acknowledgements

AOO thanks Hunter College for the startup funds.

Notes and references

1 M. Sagawa, S. Fujimura, N. Togawa, H. Yamamoto and Y. Matsuura, *J. Appl. Phys.*, 1984, **55**, 2083–2087.
2 H. Megaw, *Nature*, 1945, **155**, 484–485.
3 H. J. Goldsmid, *Proc. Phys. Soc.*, 1958, **72**, 17–26.
4 J. G. Bednorz and K. A. Müller, *Z. Phys. B: Condens. Matter*, 1986, **64**, 189–193.
5 G. Huang, Y. Guo, Y. Chen and Z. Nie, *Materials*, 2023, **16**, 5977.
6 J. E. Gubernatis and T. Lookman, *Phys. Rev. Mater.*, 2018, **2**, 120301.
7 R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi and C. Kim, *npj Comput. Mater.*, 2017, **3**, 54.
8 S. V. Dudiy and A. Zunger, *Phys. Rev. Lett.*, 2006, **97**, 046401.
9 M. Kusaba, C. Liu and R. Yoshida, *Comput. Mater. Sci.*, 2022, **211**, 111496.
10 X. Jia, Y. Deng, X. Bao, H. Yao, S. Li, Z. Li, C. Chen, X. Wang, J. Mao, F. Cao, J. Sui, J. Wu, C. Wang, Q. Zhang and X. Liu, *npj Comput. Mater.*, 2022, **8**, 34.
11 C. J. Court and J. M. Cole, *npj Comput. Mater.*, 2020, **6**, 18.
12 A. Wang, Y. Yin, Z. Sun, G. Jin and C. Xin, *Adv. Theory Simul.*, 2024, **7**, 2400048.
13 N. Linton and D. S. Aidhy, *APL Mach. Learn.*, 2023, **1**, 016109.
14 Y.-J. Hu, G. Zhao, M. Zhang, B. Bin, T. Del Rose, Q. Zhao, Q. Zu, Y. Chen, X. Sun, M. De Jong and L. Qi, *npj Comput. Mater.*, 2020, **6**, 25.
15 V. Revi, S. Kasodariya, A. Talapatra, G. Pilania and A. Alankar, *Comput. Mater. Sci.*, 2021, **198**, 110671.
16 M. Mukherjee, S. Satsangi and A. K. Singh, *Chem. Mater.*, 2020, **32**, 6507–6514.
17 H. Wei, H. Bao and X. Ruan, *Int. J. Heat Mass Transfer*, 2020, **160**, 120176.
18 Q. Guo, J. Wu, Y. Yang, X. Liu, Z. Lan, J. Lin, M. Huang, Y. Wei, J. Dong, J. Jia and Y. Huang, *Research*, 2019, **2019**, 4049793.
19 M. W. Gaultois, A. O. Oliynyk, A. Mar, T. D. Sparks, G. J. Mulholland and B. Meredig, *APL Mater.*, 2016, **4**, 053213.
20 A. Furmanchuk, J. E. Saal, J. W. Doak, G. B. Olson, A. Choudhary and A. Agrawal, *J. Comput. Chem.*, 2018, **39**, 191–202.
21 S. G. Jung, G. Jung and J. M. Cole, *J. Chem. Inf. Model.*, 2024, **64**, 6388–6409.
22 C. M. Acosta, E. Ogoshi, J. A. Souza and G. M. Dalpian, *ACS Appl. Mater. Interfaces*, 2022, **14**, 9418–9432.
23 P. Singh, T. Del Rose, A. Palasyuk and Y. Mudryk, *Chem. Mater.*, 2023, **35**, 6304–6312.
24 F. Ren, L. Ward, T. Williams, K. J. Laws, C. Wolverton, J. Hattrick-Simpers and A. Mehta, *Sci. Adv.*, 2018, **4**, eaaq1566.

25 K. Choudhary, I. Kalish, R. Beams and F. Tavazza, *Sci. Rep.*, 2017, **7**, 5179.
26 L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, *npj Comput. Mater.*, 2016, **2**, 16028.
27 K. Kim, L. Ward, J. He, A. Krishna, A. Agrawal and C. Wolverton, *Phys. Rev. Mater.*, 2018, **2**, 123801.
28 G. Pilania, A. Mannodi-Kanakkithodi, B. P. Uberuaga, R. Ramprasad, J. E. Gubernatis and T. Lookman, *Sci. Rep.*, 2016, **6**, 19375.
29 W. Setyawan, R. M. Gaume, S. Lam, R. S. Feigelson and S. Curtarolo, *ACS Comb. Sci.*, 2011, **13**, 382–390.
30 Y. Zhuo, A. Mansouri Tehrani and J. Brgoch, *J. Phys. Chem. Lett.*, 2018, **9**, 1668–1673.
31 S. Kong, F. Ricci, D. Guevarra, J. B. Neaton, C. P. Gomes and J. M. Gregoire, *Nat. Commun.*, 2022, **13**, 949.
32 J. Noh, G. H. Gu, S. Kim and Y. Jung, *Chem. Sci.*, 2020, **11**, 4871–4881.
33 P. Ball, *Nature*, 2016, 19866.
34 Y. Zuo, M. Qin, C. Chen, W. Ye, X. Li, J. Luo and S. P. Ong, *Mater. Today*, 2021, **51**, 126–135.
35 N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng and G. Ceder, *Nature*, 2023, **624**, 86–91.
36 S. Liu, B. B. Kappes, B. Amin-Ahmadi, O. Benafan, X. Zhang and A. P. Stebner, *Appl. Mater. Today*, 2021, **22**, 100898.
37 B. Mortazavi, *Adv. Energy Mater.*, 2025, **15**, 2403876.
38 Y. Bar-Sinai, S. Hoyer, J. Hickey and M. P. Brenner, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 15344–15349.
39 T. Xie and J. C. Grossman, *Phys. Rev. Lett.*, 2018, **120**, 145301.
40 E. I. Jaffal, S. Lee, D. Shiryaev, A. Vtorov, N. K. Barua, H. Kleinke and A. O. Oliynyk, *Digital Discovery*, 2025, **4**, 548–560.
41 A. O. Oliynyk, E. Antono, T. D. Sparks, L. Ghadbeigi, M. W. Gaultois, B. Meredig and A. Mar, *Chem. Mater.*, 2016, **28**, 7324–7331.
42 A. Kumar and T. Echekki, *Data-Centric Eng.*, 2025, **6**, e34.
43 S. Takamoto, C. Shinagawa, D. Motoki, K. Nakago, W. Li, I. Kurata, T. Watanabe, Y. Yayama, H. Iriguchi, Y. Asano, T. Onodera, T. Ishii, T. Kudo, H. Ono, R. Sawada, R. Ishitani, M. Ong, T. Yamaguchi, T. Kataoka, A. Hayashi, N. Charoenphakdee and T. Ibuka, *Nat. Commun.*, 2022, **13**, 2991.
44 S. Smidstrup, T. Markussen, P. Vancraeyveld, J. Wellendorff, J. Schneider, T. Gunst, B. Verstichel, D. Stradi, P. A. Khomyakov, U. G. Vej-Hansen, M.-E. Lee, S. T. Chill, F. Rasmussen, G. Penazzi, F. Corsetti, A. Ojanperä, K. Jensen, M. L. N. Palsgaard, U. Martinez, A. Blom, M. Brandbyge and K. Stokbro, *J. Phys.: Condens. Matter*, 2020, **32**, 015901.
45 *Schrödinger Release 2026-2: Materials Science Suite*, Schrödinger, LLC, New York, NY, 2025.
46 J. Stuckner, S. M. Taheri-Mousavi and J. E. Saal, *AMP Tech. Artic.*, 2024, **182**, 14–20.
47 W. Sun, S. T. Dacek, S. P. Ong, G. Hautier, A. Jain, W. D. Richards, A. C. Gamst, K. A. Persson and G. Ceder, *Sci. Adv.*, 2016, **2**, e1600225.
48 Z. Rao, P.-Y. Tung, R. Xie, Y. Wei, H. Zhang, A. Ferrari, T. P. C. Klaver, F. Körmann, P. T. Sukumar, A. Kwiatkowski Da Silva, Y. Chen, Z. Li, D. Ponge, J. Neugebauer, O. Gutfleisch, S. Bauer and D. Raabe, *Science*, 2022, **378**, 78–85.
49 P. B. D. Castro, K. Terashima, T. D. Yamamoto, Z. Hou, S. Iwasaki, R. Matsumoto, S. Adachi, Y. Saito, P. Song, H. Takeya and Y. Takano, *NPG Asia Mater.*, 2020, **12**, 35.
50 B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science*, 2018, **361**, 360–365.
51 W. Huang, P. Martin and H. L. Zhuang, *Acta Mater.*, 2019, **169**, 225–236.
52 S. Y. Lee, S. Byeon, H. S. Kim, H. Jin and S. Lee, *Mater. Des.*, 2021, **197**, 109260.
53 Y. Li and W. Guo, *Phys. Rev. Mater.*, 2019, **3**, 095005.
54 M. Bianchini, J. Wang, R. J. Clément, B. Ouyang, P. Xiao, D. Kitchaev, T. Shi, Y. Zhang, Y. Wang, H. Kim, M. Zhang, J. Bai, F. Wang, W. Sun and G. Ceder, *Nat. Mater.*, 2020, **19**, 1088–1095.
55 H. Hayashi, K. Hayashi, K. Kouzai, A. Seko and I. Tanaka, *Chem. Mater.*, 2019, **31**, 9984–9992.
56 G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang and L. Yang, *Nat. Rev. Phys.*, 2021, **3**, 422–440.
57 F. Yang, W. Chen and J. D. Evans, *Digital Discovery*, 2026, **5**, 1981–1990.
58 P. Kalhor, N. Jung, S. Bräse, C. Wöll, M. Tsotsalas and P. Friederich, *Adv. Funct. Mater.*, 2024, **34**, 2302630.
59 D. A. Boiko, R. MacKnight, B. Kline and G. Gomes, *Nature*, 2023, **624**, 570–578.
60 H. Park, A. Onwuli and A. Walsh, *Nat. Commun.*, 2025, **16**, 4379.
61 Y. Tang, W. Xu, J. Cao, W. Gao, S. Farrell, B. Erichson, M. W. Mahoney, A. Nonaka and Z. J. Yao, *Nat. Mach. Intell.*, 2026, **8**, 588–601.
62 Z. Yang, Y. Ma, H. Liu and Y. Ji, *J. Alloys Compd.*, 2026, **1066**, 188359.
63 D. Zagorac, H. Müller, S. Ruehl, J. Zagorac and S. Rehme, *J. Appl. Crystallogr.*, 2019, **52**, 918–925.
64 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr., Sect. B*, 2016, **72**, 171–179.
65 S. N. Kabekkodu, J. Faber and T. Fawcett, *Acta Crystallogr., Sect. B*, 2002, **58**, 333–337.
66 P. Villars and K. Cenzual, *Pearson's Crystal Data: Crystal Structure Database for Inorganic Compounds*, ASM International, Materials Park, OH, USA, 2023.
67 P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, *Nature*, 2016, **533**, 73–76.
68 L. Talirz, S. Kumbhar, E. Passaro, A. V. Yakutovich, V. Granata, F. Gargiulo, M. Borelli, M. Uhrin, S. P. Huber, S. Zoupanos, C. S. Adorf, C. W. Andersen, O. Schütt, C. A. Pignedoli, D. Passerone, J. VandeVondele, T. C. Schulthess, B. Smit, G. Pizzi and N. Marzari, *Sci. Data*, 2020, **7**, 299.

69 M. Scheidgen, L. Himanen, A. N. Ladines, D. Sikter, M. Nakhaee, Á. Fekete, T. Chang, A. Golparvar, J. A. Márquez, S. Brockhauser, S. Brückner, L. M. Ghiringhelli, F. Dietrich, D. Lehmberg, T. Denell, A. Albino, H. Näsström, S. Shabih, F. Dobener, M. Kühbach, R. Mozumder, J. F. Rudzinski, N. Daelman, J. M. Pizarro, M. Kuban, C. Salazar, P. Ondračka, H.-J. Bungartz and C. Draxl, *J. Open Source Softw.*, 2023, **8**, 5388.
70 J. E. Saal, S. Kirklin, M. Aykol, B. Meredig and C. Wolverton, *JOM*, 2013, **65**, 1501–1509.
71 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.
72 Y. Liu, M. Jovanovic, K. Mallayya, W. J. Maddox, A. G. Wilson, S. Klemenz, L. M. Schoop and E.-A. Kim, *Commun. Mater.*, 2025, **6**, 212.
73 C. W. Andersen, R. Armiento, E. Blokhin, G. J. Conduit, S. Dwaraknath, M. L. Evans, Á. Fekete, A. Gopakumar, S. Gražulis, A. Merkys, F. Mohamed, C. Oses, G. Pizzi, G.-M. Rignanese, M. Scheidgen, L. Talirz, C. Toher, D. Winston, R. Aversa, K. Choudhary, P. Colinet, S. Curtarolo, D. Di Stefano, C. Draxl, S. Er, M. Esters, M. Fornari, M. Giantomassi, M. Govoni, G. Hautier, V. Hegde, M. K. Horton, P. Huck, G. Huhs, J. Hummelshøj, A. Kariryaa, B. Kozinsky, S. Kumbhar, M. Liu, N. Marzari, A. J. Morris, A. A. Mostofi, K. A. Persson, G. Petretto, T. Purcell, F. Ricci, F. Rose, M. Scheffler, D. Speckhard, M. Uhrin, A. Vaitkus, P. Villars, D. Waroquiers, C. Wolverton, M. Wu and X. Yang, *Sci. Data*, 2021, **8**, 217.
74 S. Curtarolo, W. Setyawan, G. L. W. Hart, M. Jahnatek, R. V. Chepulskii, H. T. Stokes, D. O. Demchenko and D. Morgan, *Comput. Mater. Sci.*, 2012, **58**, 218–226.
75 W. Xia, M. Sakurai, B. Balasubramanian, T. Liao, R. Wang, C. Zhang, H. Sun, K.-M. Ho, J. R. Chelikowsky, D. J. Sellmyer and C.-Z. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2204485119.
76 M.-Q. Ha, D.-N. Nguyen, V.-C. Nguyen, T. Nagata, T. Chikyow, H. Kino, T. Miyake, T. Denœux, V.-N. Huynh and H.-C. Dam, *Nat. Comput. Sci.*, 2021, **1**, 470–478.
77 N. Saad, G. Gupta, S. Alizadeh and D. C. Maddix, Guiding Continuous Operator Learning through Physics-Based Boundary Constraints, in *International Conference on Learning Representations*, ICLR, 2023.
78 D. Hansen, D. C. Maddix, S. Alizadeh, G. Gupta and M. W. Mahoney, *Phys. D*, 2024, **457**, 133952.
79 A. D. Sendek, E. D. Cubuk, E. R. Antoniuk, G. Cheon, Y. Cui and E. J. Reed, *Chem. Mater.*, 2019, **31**, 342–352.
80 S. M. Moosavi, K. M. Jablonka and B. Smit, *J. Am. Chem. Soc.*, 2020, **142**, 20273–20287.
81 H. Hayashi, A. Seko and I. Tanaka, *npj Comput. Mater.*, 2022, **8**, 217.
82 S. Back, A. Aspuru-Guzik, M. Ceriotti, G. Gryn'ova, B. Grzybowski, G. H. Gu, J. Hein, K. Hippalgaonkar, R. Hormázabal, Y. Jung, S. Kim, W. Y. Kim, S. M. Moosavi, J. Noh, C. Park, J. Schrier, P. Schwaller, K. Tsuda, T. Vegge, O. A. Von Lilienfeld and A. Walsh, *Digital Discovery*, 2024, **3**, 23–33.
83 A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, *Nature*, 2023, **624**, 80–85.
84 Y.-G. Yan, D. Lu and K. Wang, *Tungsten*, 2023, **5**, 32–49.
85 T. Zhu, R. He, S. Gong, T. Xie, P. Gorai, K. Nielsch and J. C. Grossman, *Energy Environ. Sci.*, 2021, **14**, 3559–3566.
86 A. Y. Pak, V. Sotskov, A. A. Gumovskaya, Y. Z. Vassilyeva, Z. S. Bolatova, Y. A. Kvashnina, G. Ya. Mamontov, A. V. Shapeev and A. G. Kvashnin, *npj Comput. Mater.*, 2023, **9**, 7.
87 M. Alverson, S. G. Baird, R. Murdock, E. Sin-Hang Ho, J. Johnson and T. D. Sparks, *Digital Discovery*, 2024, **3**, 62–80.
88 R. Chang, A. Guerra, N. Richardson, N. Zhan, S. Liu, A. Pak, R. Marr, A. M. Ganose, R. P. Adams and E. Ertekin, Crystal Generative Modeling with Explicit Autoregressive Conditional Likelihoods and Nontrivial Space Group Stabilizers, in *AI for Accelerated Materials Design*, ICLR, 2025.
89 L. M. Antunes, K. T. Butler and R. Grau-Crespo, *Nat. Commun.*, 2024, **15**, 10570.
90 K. Kovnir, *Chem. Mater.*, 2021, **33**, 4835–4841.
91 S. Chen, S. Perathoner, C. Ampelli and G. Centi, in *Horizons in Sustainable Industrial Chemistry and Catalysis*, ed. S. Albonetti, S. Perathoner and E. A. Quadrelli, Elsevier, 2019, vol. 178, pp. 31–46.
92 X. Jia, A. Lynch, Y. Huang, M. Danielson, I. Lang'at, A. Milder, A. E. Ruby, H. Wang, S. A. Friedler, A. J. Norquist and J. Schrier, *Nature*, 2019, **573**, 251–255.
93 J. Lai, A. Bolyachkin, N. Terada, S. Dieb, X. Tang, T. Ohkubo, H. Sepehri-Amin and K. Hono, *Acta Mater.*, 2022, **232**, 117942.
94 V. G. S. Lunde, A. B. Møller, B. G. F. Eggert, A. M. Döring, J.-C. Grivel, R. Bjørk, F. Veillon, K. Skokov, O. Gutfleisch, A. O. Sjåstad, B. C. Hauback and C. Frommen, *Acta Mater.*, 2025, **297**, 121282.
95 D. Tu, J. Yan, Y. Xie, J. Li, S. Feng, M. Xia, J. Li and A. P. Leung, *J. Mater. Sci. Technol.*, 2022, **96**, 241–247.
96 S. Swateelagna, M. Singh and M. R. Rahul, *Intermetallics*, 2024, **167**, 108198.
97 J. Luo, X. Liu, Q. Ma, C. Pei, H. Yao, J. Xiong and Q. Gao, *J. Mater. Inf.*, 2026, **6**, 5.
98 C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, L. Dai, T. Lookman and Y. Su, *Acta Mater.*, 2019, **170**, 109–117.
99 D. Marchand and W. A. Curtin, *Phys. Rev. Mater.*, 2022, **6**, 053803.
100 J. Yu, S. Xi, S. Pan, Y. Wang, Q. Peng, R. Shi, C. Wang and X. Liu, *J. Mater. Inf.*, 2021, **1**, 9.
101 A. O. Oliynyk and A. Mar, *Acc. Chem. Res.*, 2018, **51**, 59–68.
102 X. Zhong, B. Gallagher, S. Liu, B. Kailkhura, A. Hiszpanski and T. Y.-J. Han, *npj Comput. Mater.*, 2022, **8**, 204.
103 J. Xie, Y. Zhou, M. Faizan, Z. Li, T. Li, Y. Fu, X. Wang and L. Zhang, *Nat. Comput. Sci.*, 2024, **4**, 322–333.
104 M. Scheffler, M. Aeschlimann, M. Albrecht, T. Bereau, H.-J. Bungartz, C. Felser, M. Greiner, A. Groß, C. T. Koch,

K. Kremer, W. E. Nagel, M. Scheidgen, C. Wöll and C. Draxl, *Nature*, 2022, **604**, 635–642.
105 C. Yang, C. Ren, Y. Jia, G. Wang, M. Li and W. Lu, *Acta Mater.*, 2022, **222**, 117431.
106 J. Yu, C. Wang, Y. Chen, C. Wang and X. Liu, *Mater. Des.*, 2020, **195**, 108996.
107 P. A. E. Murgatroyd, K. Routledge, S. Durdy, M. W. Gaultois, T. W. Surta, M. S. Dyer, J. B. Claridge, S. N. Savvin, D. Pelloquin, S. Hébert and J. Alaria, *Adv. Funct. Mater.*, 2021, **31**, 2100108.
108 T. Zheng, X. Hu, F. He, Q. Wu, B. Han, D. Chen, J. Li, Z. Wang, J. Wang, J. Kai, Z. Xia and C. T. Liu, *J. Mater. Sci. Technol.*, 2021, **69**, 156–167.
109 N. K. Barua, A. Golabek, A. O. Oliynyk and H. Kleinke, *J. Mater. Chem. C*, 2023, **11**, 11643–11652.
110 B. Cao, T. Su, S. Yu, T. Li, T. Zhang, J. Zhang, Z. Dong and T.-Y. Zhang, *Mater. Des.*, 2024, **241**, 112921.
111 W. Zhao, Z. Chen, Y. Shang, Q. Wang, L. Wang, B. Liu, Y. Liu and Y. Yang, *npj Comput. Mater.*, 2025, **11**, 278.
112 S. S. Sethi, A. Dutta, E. I. Jaffal, N. Yadav, D. Shiryaev, B. Hoang, A. Machathi, S. Lee, K. Das, P. P. Jana and A. O. Oliynyk, *J. Am. Chem. Soc.*, 2025, **147**, 14739–14755.
113 B. Xhabrahimi, E. I. Jaffal, D. Shiryaev, N. K. Barua, M. Donohoe, N. Pozdnyakova, M. Ismail, B. Selvaratnam, E. Niknam, H. Kleinke and A. O. Oliynyk, *J. Am. Chem. Soc.*, 2025, **147**, 36589–36603.
114 B. Balasubramanian, M. Sakurai, C.-Z. Wang, X. Xu, K.-M. Ho, J. R. Chelikowsky and D. J. Sellmyer, *Mol. Syst. Des. Eng.*, 2020, **5**, 1098–1117.
115 G. Cheng, X.-G. Gong and W.-J. Yin, *Nat. Commun.*, 2022, **13**, 1492.
116 S. Kim, J. Noh, G. H. Gu, A. Aspuru-Guzik and Y. Jung, *ACS Cent. Sci.*, 2020, **6**, 1412–1420.
117 C. Zeni, R. Pinsler, D. Zügner, A. Fowler, M. Horton, X. Fu, Z. Wang, A. Shysheya, J. Crabbé, S. Ueda, R. Sordillo, L. Sun, J. Smith, B. Nguyen, H. Schulz, S. Lewis, C.-W. Huang, Z. Lu, Y. Zhou, H. Yang, H. Hao, J. Li, C. Yang, W. Li, R. Tomioka and T. Xie, *Nature*, 2025, **639**, 624–632.
118 T. Ueno, T. D. Rhone, Z. Hou, T. Mizoguchi and K. Tsuda, *Mater. Discovery*, 2016, **4**, 18–21.
119 Y. Sheng, Y. Wu, J. Yang, W. Lu, P. Villars and W. Zhang, *npj Comput. Mater.*, 2020, **6**, 171.
120 P. Singh, T. Del Rose, G. Vazquez, R. Arroyave and Y. Mudryk, *Acta Mater.*, 2022, **229**, 117759.
121 P. Yin, X. Niu, S.-B. Li, K. Chen, X. Zhang, M. Zuo, L. Zhang and H.-W. Liang, *Nat. Commun.*, 2024, **15**, 415.
122 X. Niu, Y. Chen, M. Sun, S. Nagao, Y. Aoki, Z. Niu and L. Zhang, *Sci. Adv.*, 2025, **11**, eadw0894.
123 V. Eyert, J. Wormald, W. A. Curtin and E. Wimmer, *J. Mater. Res.*, 2023, **38**, 5079–5094.
124 P. Friederich, F. Häse, J. Proppe and A. Aspuru-Guzik, *Nat. Mater.*, 2021, **20**, 750–761.
125 N. Fedik, R. Zubatyuk, M. Kulichenko, N. Lubbers, J. S. Smith, B. Nebgen, R. Messerly, Y. W. Li, A. I. Boldyrev, K. Barros, O. Isayev and S. Tretiak, *Nat. Rev. Chem.*, 2022, **6**, 653–672.
126 L. Wu and T. Li, *J. Mech. Phys. Solids*, 2024, **187**, 105639.
127 J. Byggmästar, K. Nordlund and F. Djurabekova, *Phys. Rev. Mater.*, 2022, **6**, 083801.
128 A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2013, **87**, 184115.
129 E. A. Antillon and N. Bernstein, *Phys. Rev. Mater.*, 2025, **9**, 083801.
130 A. Lee, S. Sarker, J. E. Saal, L. Ward, C. Borg, A. Mehta and C. Wolverton, *Commun. Mater.*, 2022, **3**, 73.
131 J. Jang, G. H. Gu, J. Noh, J. Kim and Y. Jung, *J. Am. Chem. Soc.*, 2020, **142**, 18836–18843.
132 A. O. Oliynyk, L. A. Adutwum, J. J. Harynuk and A. Mar, *Chem. Mater.*, 2016, **28**, 6672–6681.
133 P. Villars, H. Okamoto and K. Cenzual, *ASM Alloy Phase Diagrams Database*, ASM International, Materials Park, OH, USA, 2016.
134 A. O. Oliynyk, L. A. Adutwum, B. W. Rudyk, H. Pisavadia, S. Lotfi, V. Hlukhyy, J. J. Harynuk, A. Mar and J. Brgoch, *J. Am. Chem. Soc.*, 2017, **139**, 17870–17881.
135 Ya. F. Lomnytska and Yu. B. Kuz'ma, *J. Alloys Compd.*, 1998, **269**, 133–137.
136 S. Kim, J. Schrier and Y. Jung, *Angew. Chem., Int. Ed.*, 2025, **64**, e202423950.
137 A. Niyongabo Rubungo, C. Arnold, B. P. Rand and A. B. Dieng, *npj Comput. Mater.*, 2025, **11**, 186.
138 *OpenAI, ChatGPT, GPT-4o*, 2024, https://chat.openai.com.
139 S. S. Omee, L. Wei, S. Dey and J. Hu, *Adv. Sci.*, 2025, **12**, e10792.
140 Y. Inada, M. Fujioka, H. Morito, T. Sugahara, H. Yamane and Y. Katsura, *Chem. Mater.*, 2025, **37**, 2097–2105.
141 S. Lee and E. Jaffal, *Composition-Analyzer-Featurizer*, https://github.com/bobleesj/composition-analyzer-featurizer, accessed 2026-02-26.
142 S. Lee and E. Jaffal, *Structure-Analyzer-Featurizer*, https://github.com/bobleesj/structure-analyzer-featurizer, accessed 2026-02-26.
143 S. Lee and A. O. Oliynyk, *J. Open Source Softw.*, 2024, **9**, 7205.
144 A. Machathi, *Automated Machine-Learning Workflow for Excel Data*, 2024, https://github.com/AnirudhM2110/Automated-Machine-Learning-Workflow-for-Excel-Data, accessed 2026-02-26.
145 D. Shiryaev, *Structure Type Explorer (STEx)*, 2025, https://github.com/dshirya/structure-type-explorer, accessed 2026-02-26.
146 B. Mondal, S. Dan, S. Mondal, R. N. Bhowmik, R. Ranganathan and C. Mazumdar, *Phys. Chem. Chem. Phys.*, 2019, **21**, 16923–16936.
147 D. Zhu, M. Nie, H.-H. Wu, C. Shang, J. Zhu, X. Zhou, Y. Zhu, F. Wang, B. Wang, S. Wang, J. Gao, H. Zhao, C. Zhang and X. Mao, *J. Mater. Inf.*, 2025, **5**, 8.
148 S. Hao, T. Xia, R. Zhang and M. Guo, *Sci. Rep.*, 2024, **14**, 31602.
149 A. S. Parackal, R. E. A. Goodall, F. A. Faber and R. Armiento, *Phys. Rev. Mater.*, 2024, **8**, 103801.
150 V. Stanev, V. V. Vesselinov, A. G. Kusne, G. Antoszewski, I. Takeuchi and B. S. Alexandrov, *npj Comput. Mater.*, 2018, **4**, 43.

151 L. Kývala, P. Montero De Hijes and C. Dellago, *npj Comput. Mater.*, 2025, **11**, 50.
152 W. Liao, R. Yuan, X. Xue, J. Wang, J. Li and T. Lookman, *npj Comput. Mater.*, 2024, **10**, 171.
153 Y. Zhang, X. He, Z. Chen, Q. Bai, A. M. Nolan, C. A. Roberts, D. Banerjee, T. Matsunaga, Y. Mo and C. Ling, *Nat. Commun.*, 2019, **10**, 5260.
154 J. Zhang, Z. Zhu, X.-D. Xiang, K. Zhang, S. Huang, C. Zhong, H.-J. Qiu, K. Hu and X. Lin, *J. Phys. Chem. C*, 2022, **126**, 8922–8927.
155 D. Shiryaev, *CRAFT*, https://github.com/dshirya/CRAFT, accessed 2026-02-26.
156 B. Owens-Baird, V. Gvozdetskyi, A. Sarkar, B. Selvaratnam, A. Mar and K. Kovnir, *Chem. Mater.*, 2024, **36**, 6180–6192.
157 F. Wang, A. G. Iwanicki, A. T. Sose, L. A. Pressley, T. M. McQueen and S. A. Deshmukh, *npj Comput. Mater.*, 2025, **11**, 124.