nature computational science

Perspective

https://doi.org/10.1038/s43588-023-00412-7

Accelerating the design of compositionally complex materials via physics-informed artificial intelligence

Received: 11 August 2022

Accepted: 7 February 2023

Published online: 31 March 2023

Check for updates

Dierk Raabe ✉, Jaber Rezaei Mianroodi ✉ & Jörg Neugebauer ✉

The chemical space for designing materials is practically infinite. This makes disruptive progress by traditional physics-based modeling alone challenging. Yet, training data for identifying composition–structure–property relations by artificial intelligence are sparse. We discuss opportunities to discover new chemically complex materials by hybrid methods where physics laws are combined with artificial intelligence.

New materials are crucial in two aspects. On the one hand, they enable disruptive leaps in civilization. Examples are early ceramics for pottery; bronze for agriculture; steels for machinery; cement for buildings; aluminium for aviation; titanium for spaceships; rare earth elements for magnets; semiconductors for computer chips; platinum-group metals for catalysts; and polymers for packaging and medicine. On the other hand, material production is the largest single source of greenhouse gas emissions, energy consumption and environmental pollution, a fact that forces us to entirely rethink the way we produce, use and recycle them[1,2].

The drive towards ever-improving materials has led to their higher chemical complexity, as property improvement often requires tweaking the intrinsic and microstructure-dominated features by composition adjustment. Examples are chemically fine-tuned intermetallic phases in superalloys[3–5], complex precipitation pathways in high-performance aluminium alloys[6–8] or interfaces in advanced magnets[9,10]. Another challenge is the near atomic-scale blending of multiple elements in microelectronics, where the borders between product and material become blurred, such as in the 2 nm process in semiconductor manufacturing. Both trends enhance compositional complexity of materials and highly integrated systems: they are preconditions for advanced product properties and open doors to new solid-state phenomena[11–14]. Yet, chemistry never comes alone: compositional complexity of materials translates to their microstructure[15]. Changes in the chemical composition affect many defect features, often with an exponential dependence: examples include changes in the solute decoration state and energy of the defects, drag forces acting on them and the formation of new phases at defects. This means that changes in chemical complexity are linked to changes in microstructure complexity. The latter aspect is important because materials are practically never used in their thermodynamic equilibrium state, but in a transient state, equipped with a complex microstructure cosmos consisting of point defects, line defects (dislocations) and interfaces (Fig. 1).

The aim to predict structures and properties of materials in such a huge and uncharted chemical space with its multiple scales and vast cosmos of chemically decorated defects and resulting microstructure phenomena (Fig. 1) creates completely new types of challenges and opportunities that require disruptive changes in computational materials science[16–21]. Classical physics-based simulation approaches see their limits already in the first 'simple' steps, meaning, the quantitative prediction of high-dimensional phase equilibria (Fig. 1a)[22,23], not to mention the reliability of the simulations for the non-equilibrium microstructures and properties (Fig. 1b).

An additional constraint in this context is sustainability: while enhancing chemical complexity ever more is a tempting design route because it takes materials far into infinite composition and property spaces, it makes product disassembly and material recycling much harder, reducing the fitness of materials and products for a circular economy. Some high-entropy materials match or even outperform established materials[10,24,25], but they often contain environmentally questionable elements, impairing their environmental footprint. This means that computational design of chemically complex materials has to balance opportunities for new discoveries with responsible synthesis and use of elements.

In this Perspective, we discuss the challenges when designing compositionally complex materials, examine the limitations of conventional models and further discuss how artificial intelligence (AI) can help to guide the discovery of new chemically complex materials.

Physical and chemical complexity from massive chemical mixing

Materials have several dimensions of complexity, for example, in terms of their chemical composition, synthesis, processing and defects (Fig. 1).

Max-Planck-Institut für Eisenforschung, Düsseldorf, Germany. ✉e-mail: d.raabe@mpie.de; j.mianroodi@mpie.de; j.neugebauer@mpie.de

https://doi.org/10.1038/s43588-023-00412-7

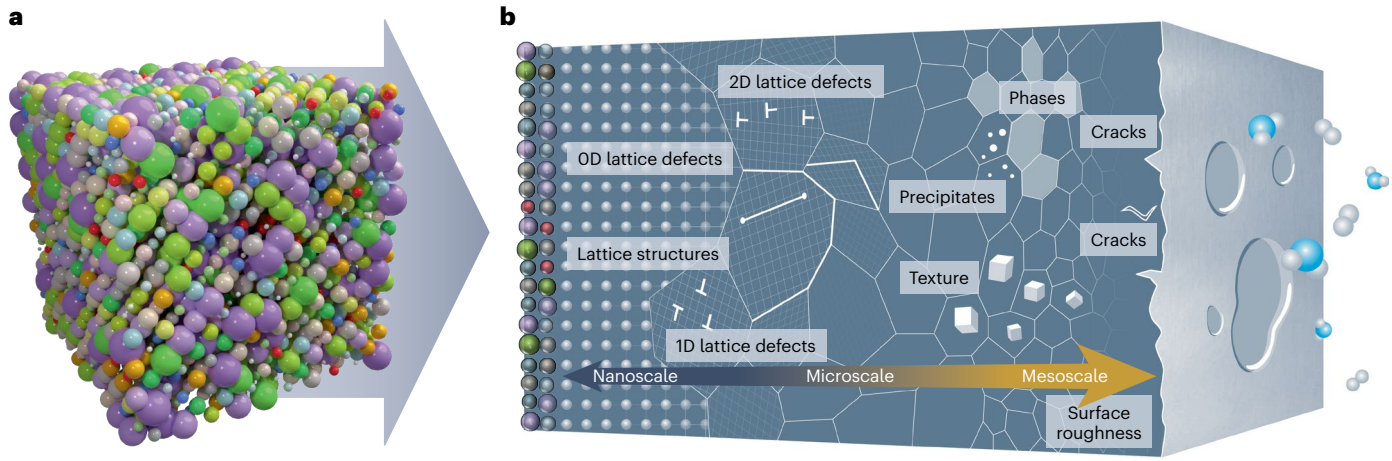

Fig. 1 | Chemical complexity influences microstructure complexity. a, View of a chemically complex material, where different types of atoms (indicated by the colored spheres) are mixed together in a crystalline solid solution, stabilized by configurational entropy. Computational studies on chemical complexity therefore often start with high-component thermodynamics and equilibrium phase diagrams. Mixing multiple elements in high fractions breaks the crystal symmetry locally, creates complex electronic states, leads to magnetic ordering effects, and introduces high distortions and potentially also short-range order, making such materials profoundly different from low-component alloys. **b**, Schematic view of lattice defect types in solids that are affected by compositional complexity: point defects (zero-dimensional (0D) defects), dislocations (one-dimensional (1D) defects) and interfaces (two-dimensional (2D) defects). The chemical and structural features of compositionally complex solid solutions affect the energies and kinetics of the lattice defects in them, thus altering the resulting microstructures that lend materials many of their properties such as strength or conductivity. Understanding and quantifying the link between chemical and structural complexity by using AI is one of the grand challenges for the design of compositionally complex solid solutions

Many features of materials can change substantially when multiple principal elements are blended together[12,26,27]. Table 1 compiles some of these aspects.

With only about 10^4 different types of synthetic materials used today, out of an almost infinite space of about 10^{80} potentially meaningful chemical combinations, we learn that we stand at only the beginning of materials research in all its compositional complexity[28,29]. Nowadays, we have reliable data only for low-component number phase diagrams and even many allegedly 'simple' binary phase equilibria are not well understood, for instance, when magnetism is involved and/or complex low-temperature phase transformations occur, where experimental access is limited by kinetics. Therefore, the chemistry of materials deserves particular attention[30], owing to its huge undiscovered space, profound influence on microstructure, properties and sustainability, and insufficient theoretical methods to study it for systems with more than three or four components.

Using more chemical elements in materials helps to realize so-far-inaccessible thermodynamic, kinetic, structural and functional features and also leads to new local chemical interactions with the materials' internal defects (its microstructure)[31–33], an approach that enables the chemical and structural modification of defect-related properties (Table 1)[26].

The use of multiple elements in materials allows for tuning not only their bulk properties but also their internal defects and surface features. The Gibbs adsorption isotherm makes a thermodynamic statement about the magnitude of defect stabilization when decorated by solutes in dilute cases, but it does not answer similar questions for multi-component systems when defects undergo complex atomic motif formation and chemical reconstructions.

A principle downside and topic that needs consideration in the context of chemically complex materials lies in their reduced recyclability and the accumulation of (potentially harmful) impurity elements in waste streams that enter new materials when recycled[1]. While the scientific questions arising from the presence of multiple recycling-related tramp elements in materials are in principle the same as those described in Table 1, life-cycle aspects must be considered as an additional set of constraints, rendering the task a system challenge.

Understanding complexity via physics-based modeling

The discovery of chemically complex materials often relies on trial and error, serendipity and phenomenological rules. Exceptions are tools that have heralded modern computational materials design such as density functional theory[34–37] and statistical thermodynamics[28,38–43].

Density functional theory simulations can be used to extract energetic and certain structural features of materials, also under consideration of thermal and magnetic effects, with high quantum mechanical fidelity. However, owing to the limited super-cell sizes, their ability to deal with the high-dimensional and symmetry-breaking chemical complexity of materials is limited and it is also doubtful that quantum computing will be capable of unleashing ab initio simulations for large complex systems in the near future (Table 1).

Tools such as molecular dynamics, in conjunction with accurate machine-learning-trained interatomic potentials, can be used to investigate larger systems that also involve defects, kinetics and chemical reactions in materials with moderately complex compositions[44–54]. Mean-field methods such as phase-field theory can use ab initio-informed energy landscapes and mobility coefficients to study microstructure formation[55,56], opening the door towards the simulation of complex multi-physics and processing-related problems where multiple chemical elements are involved[57]. Multi-physics and machine-learning-enhanced phase-field simulations are thus attractive for problems where several kinetic mechanisms compete in materials with complex chemical compositions and environments such as in corrosion, battery, catalysis and reduction processes[58–60]. Yet, these methods take a coarse-grained approximation of the energy associated with defects in terms of gradient terms, and their energetic fidelity depends on the quality of the approximation of the system energetics by the underlying Landau energy forms. On a more mean-field level, dislocation dynamics and crystal-plasticity models[61–65] can be used to relate such microstructures that evolve in chemically complex materials to properties[66]. Phase-field methods and crystal mechanical models are increasingly being combined in the same simulation tools[67,68] and AI methods are used to accelerate them[58,60,69–73].

Table 1 | Complexity of materials with multiple principal elements (for example, high-entropy alloys) or impurity elements (for example, recycled alloys) and the role of computational materials science

Challenges in materials design	Computational challenges and opportunities specific to multi-component and chemically complex materials
Characteristic bulk features • Up to ten alloying and/or impurity elements • Multiple (meta-)stable phases and phase transformations • Crystal and amorphous structures • Strong distortions and local crystalline symmetry-breaking • Chemical ordering • Segregation and heterogeneity across all scales • Challenging synthesis, fabrication and processing • Near-equilibrium or highly non-equilibrium synthesis pathways	**Challenges** • Prediction of bulk phase energies and kinetics • Prediction of short-range ordering • Insufficient databases and training data, that is, how to do AI with sparse data • Electronic and magnetic effects • Configurational, vibrational, magnetic and electronic entropy • Bridging discrete and continuum models • Macroscopic response from electronic and atomistic principles **Opportunities** • Extraction of thermodynamic data from literature by AI • Hybrid machine learning and active-learning methods for sparse data problems • AI for bridging scales and accelerated solvers • AI for bridging atomistic and continuum simulations • AI-guided automatic and even autonomous high-throughput simulations
Lattice defects • Solute-decorated point defects, line defects, interfaces and so on • Microstructure patterning and gradients	**Challenges** • Energies and kinetics of defects in chemical multi-component decorated state (due to the Gibbs adsorption isotherm) • From individual solute-decorated defects to patterning and gradients • Multi-physics interactions: chemistry, magnetism, mechanics • Process and property simulation **Opportunities** • Surrogating defect features and patterning with AI • Structure–chemistry–process–property linkage through AI • New routes for stabilizing nanocrystalline structures and chemically decorated defects
Life cycle, longevity, sustainability • Considering element scarcity, mining, refining, synthesis, processing and recycling • Corrosion and longevity in harsh environments	**Challenges** • Considering sustainability, material decay, harsh environments, element scarcity and life-cycle aspects in the design of chemically complex materials for a circular economy • Reconciliation of material design for advanced properties with the goal of improving material sustainability **Opportunities** • Holistic cradle-to-grave simulations and AI methods for materials, processes and recycling • High-component surface oxides and nitrides with better corrosion resistance • New materials with enhanced abrasion, fatigue and stress-corrosion resistance
Increased amount of data • Materials evolved in part over millennia • Materials design often done via trial and error • Dispersed knowledge (literature, patents, experience, communities)	**Challenges** • Systematic and quality-controlled collection of knowledge • No homogeneous or unified nomenclature, materials, processes and compositions **Opportunities** • Autonomous AI-based knowledge extraction from literature • Autonomous AI-based material and process design • AI guidance for material development

Many of the models sketched in this section use simplified approximations and rely on large numbers of variables and degrees of freedom when the material system and its processing grow in terms of chemical and structural complexity. The art of sensible modeling in this field thus lies in the task of making a model 'as simple as possible but not simpler'. Yet, where to draw the line in ignoring complexity is often not clear and not guided by crisp rules.

A reliable measure for 'permitted' simplification is to examine whether a simulation method is capable of reproducing the energetics of a system properly (at least in the targeted temperature and chemical potential regimes), such as matching portions of known phase diagrams, for instance. It is surprising how even frequently used and well-established bottom-up atomistic simulation methods based on using empirical potentials fail when it comes to reproducing Gibbs free-energy landscapes for multi-component systems. In the case of highly concentrated solid metallic solutions, this is probably due to their electronic and magnetic complexity as well as to short-range ordering phenomena[74–77]: when taking the well-known FeMnCrNiCo alloy as an example[13], it includes all the strongest ferromagnetic and antiferromagnetic elements in one (metastable) solid solution. This means that not only the electronic ground states must be correctly captured under consideration of all possible magnetic states but also all entropy contributions (mixing, phonon, electron, magnetic)[35,76]. Most conventional empirical potentials do not have the functional flexibility to properly reproduce all these effects.

Interesting entropy-related features have not only been studied for metals but also for compositionally complex and high-entropy ceramics[27,78–80]. At first view, these materials would have been traditionally assumed to be more enthalpy dominated rather than entropy dominated, due to the strong covalent and ionic bonds and localized electrons, compared with the delocalized valence electrons in metals resulting in weaker bonds that enable intense element mixing. However, it was found that interesting mixing effects can be triggered in multi-component ceramics by disorder-induced charge fluctuations[27]. It was also suggested that configurational entropy effects play a role in the deviation from ordering on the anion sublattice, an effect which increases the compositional space for ceramics substantially.

While ab initio methods are principally able to describe any material property, their huge computational costs limits them to rather small system sizes. Therefore, critical quantities are not or only very approximately accessible by direct ab initio calculations. For example, in contrast to the simulation of thermodynamic equilibrium conditions, the role of system dynamics and reactions turns out to be computationally very costly. These limitations apply particularly to chemically complex materials, where constrained or local equilibrium calculations only predict certain aspects of interest, such as thermal expansion, stacking fault energy or interfacial cohesion. Also, properties that are related to the microstructure of the materials, such as strength, toughness, ductility, magnetic dynamics, thermoelectricity, reactivity and so on, cannot be captured by ab initio simulations alone. For the

same reason, the chemical search space accessible to ab initio methods remains limited. This means that interesting and even unexpected material behavior arising from chemical complexity might occur in search regions of currently totally unchartered terrain.

Hence, a dogmatic decision to use the computational design of chemically complex materials based on either ab initio methods or mean-field continuum models alone seems to be futile. Furthermore, better chemical models or higher-fidelity homogenization methods are typically developed slowly as in practice the choices for better models that can deal with high-component chemical complexity are sparse. The expected progress in this field is limited due to the often too high computational efforts and/or missing data to fit potentials.

Physics-based modeling of descriptors

Classical physics-based models for chemically complex materials usually start with characterizing the thermodynamic stability of the system and identifying its stable phases. The thermodynamic description goes through a combination of experimental observation of phase stability, fed as coefficients into Gibbs energy approximations, or obtained as quantum-level calculations of atomic interactions and energies[81,82]. An accurate thermodynamic model is necessary to discover possible phases, ordering and certain microstructure phenomena. This step is challenging for chemically complex materials owing to the curse of dimensionality, due to the exploding number of possible combinations and multi-body interactions between the different atoms. Blends between established thermodynamic database approaches, AI and quantum computing are a promising avenue here.

Yet, materials are mostly not used in the equilibrium state, which makes an accurate description of kinetics necessary. In physics-based models, this is done by dynamic simulations, such as molecular dynamics, kinetic Monte Carlo methods or phase-field theory. An efficient approach to performing such forward simulations for chemically complex materials is to focus on computationally easy-to-calculate descriptors with high leverage on properties. An example from high-entropy alloys is to use density functional theory calculations of the stacking fault energy as the descriptor for phase stability, twinning and martensitic phase transformations[83–88]. All these energetic features have a high influence on the mechanical properties of metallic alloys and can thus be well coupled to corresponding mean-field simulations of the mechanical response[89–91]. This approach was, for example, successfully applied to design an alloy in which two high-entropy phases coexist and both of the coexisting metastable phases can undergo athermal phase transformations when mechanically loaded[92,93].

Successful identification and simulation of physics-based descriptors that correlate with microstructure evolution and properties are a promising approach that can be tackled not only with conventional models but also with AI[94]. When such descriptors capture the influence of chemical composition, they can be simulated starting from first principles and the resulting microstructure-dependent property predictions can be done by kinetic scale-hopping simulations that use them as central constitutive quantities, such as in analytical constitutive[79,95,96], phase-field[58–60,97,98] and crystal-plasticity[71–73,99–103] methods. The computational feasibility of these methods depends on the complexity of the energetic landscape and the kinetic rules, which again—in the context of compositionally complex materials—are expected to be also complex.

Descriptors are a powerful tool to explore the compositional space without having to resolve the energetic–kinetic–microstructure–property chain. However, their foundation is often phenomenologically based, which means that identifying appropriate descriptors requires careful and systematic experiments to identify the ones that reflect the bottleneck mechanisms of the specific problem addressed. In the future, it will be therefore important to have also methods available that allow to cover the entire simulation chain. This poses severe challenges for current computational methods when chemically complex materials are considered.

New approaches are thus needed that can cope with higher degrees of compositional, structural and property complexity and at the same time can consider larger length and longer timescales, such as AI or hybrid AI methods coupled with physics-based models. In the next section, we comment on several emerging methods in AI and their applicability to these challenges.

AI for compositionally complex materials

AI in materials science

AI-based methods offer solutions to tackle the curse of compositional, structural and defect dimensionality. We refer to AI methods here as predictive modeling approaches that can learn from data. Data that can be used in AI can be of diverse origin: it can come from existing results, classical physics-based simulations or all the hidden information that exists in the often dilute, ill-defined, sparse and unfiltered body of literature, which in itself can be heterogeneous in quality. Hence, AI can learn from sufficient amounts of training data (for instance, by using large-scale nonlinear regression via artificial neural networks or Bayesian optimization)[18,70,104–111]; by way of autonomous literature analysis (for instance, via natural language processing)[112–116]; from deep learning[69,98,117] (learning from example scenarios) or reinforcement learning[98,118] (where penalty and reward-driven algorithms are used in a modeled training universe in which an AI gains experience) to name but a few important variants used in materials science.

It is likely that the different AI approaches may be suitable for different types of problems in materials science. This means that, for instance, the recognition of certain defect, chemical or damage patterns in microscopic data (obtained from microscopy and spectroscopy) in materials or trend analyses of material properties on the basis of large sets of published training data may in many cases be better achieved by neural networks (feature recognition, segmentation and latent space analysis)[103,111,119], whereas systematic insights on existing synthesis procedures and processing methods and their respective influence on microstructures and properties might be better revealed by natural language processing[114,115]. However, for the latter approach to function, literature must have already reported the content of interest. For undiscovered materials, such as in the field of high-entropy alloys, text mining may not readily work. Yet, Pei et al.[116] recently showed that certain high-entropy alloys could have been discovered earlier if natural language processing had been used to screen the literature for element affinity in solid solutions. This means that it may not be possible to discover new materials directly by text mining, but it could point at compositional regions of interest.

Opportunities offered by AI in the context of compositionally complex materials lie particularly in the automated extraction of thermodynamic data from literature and ab initio data, hybrid machine learning methods for sparse data problems, more efficient bridging of multiple scales and multi-physics problems, the use of AI as accelerated solvers, the bridging of electronic, atomistic and continuum simulations, and automatic and even autonomous combinatorial and high-throughput simulations, guided and decided by AI[111]. In addition, the combination of physics-based and AI-based methods, namely, hybrid methods, are increasingly developed for tackling those challenges. Examples are machine learning methods in conjunction with thermodynamic concepts[117,120,121], using adaptive activation functions for physics-based neural networks[122,123], or employing kernels and operators with functional forms that come from physics[124]. These hybrid methods seem to be particularly promising for (chemically) complex materials[47,51,121,125].

Conventional AI models

In the context of composition-dependent microstructure features, AI methods can particularly help to surrogate chemical defect decoration features that stabilize, for instance, nanosized grains against

capillary-driven coarsening in high-temperature alloys[126,127]. Similar opportunities exist for the AI-based prediction of improved corrosion resistance, with respect to both the influence of the materials' average chemical composition on the protective features of oxide or nitride layers, and pitting corrosion, where local oxidative attack prevails[128–130]. Using AI for predicting chemical defect stabilization and surface modification thus opens a pathway to novel types of structure–chemistry–process–property linkage predictions for chemically complex materials.

Another promising avenue for using AI is the reconciliation of the rich property cosmos offered by the materials sector with the goal of improving its sustainability, one of the biggest challenges of our generation, as material production is the largest greenhouse gas emitter and energy consumer on the planet[1,131,132]. In this context, a specific downside of chemically complex materials lies in their reduced recyclability, use of valuable elements and the accumulation of (potentially harmful) impurity elements in waste streams that enter new materials when recycled[1]. In a circular economy, every atom must ideally re-enter the manufacturing chain and this should become a constraint in AI-based alloy design[2,133]. Also, high-entropy and similarly complex alloys not only enhance the number of chemical components used in them but also often contain rare elements. This means that research into complex chemical compositions of materials can be both a key enabler for advanced materials and a main future burden for the pursuit of a circular economy[134,135]. AI-based material development should therefore take constraints of sustainability into account, for example, by penalizing the use of harmful, rare or hard-to-recycle elements. Also, established life-cycle assessment methods could enter AI-based alloy development. Such a holistic and systemic approach to material development might be able to answer trade-off questions such as whether it is better to develop new materials with higher chemical complexity (for example, to reduce vehicle weight), use chemically lean materials (and heavier vehicles) or replace chemical complexity (a conserved quantity) by microstructural complexity (a non-conserved quantity), if at all possible[136].

Besides intentional elemental doping for targeted novel properties, there may be also unexpected opportunities lurking in chemically complex materials[137,138]. Examples are questions whether materials with an enhanced concentration of multiple elements might be more forgiving against unintentional impurities that enter through recycling, or whether they might be more resilient against harsh environmental attacks[1,139–141]. Such questions, where unintentional elemental intrusion poses new scientific challenges and opportunities, are currently emerging as a discipline referred to as the 'science of dirty alloys'[139].

Although there has been a sharp rise in the number of works on the application of AI in material science[48,71,142–144], these are not yet in all cases readily suited for applications related to compositionally complex materials along the challenging lines described above. Purely data-driven AI methods, although effective in finding patterns, do not automatically make discoveries outside of the training data. Yet, this feature is important for material design, where particularly outliers of compositional trajectories might be more promising than mere trend projection. For this to work, suitable mechanism- and microstructure-sensitive descriptors must be identified[145,146]. If descriptor-based approaches are not available for the targeted design tasks, physics-based AI coupled with domain knowledge seems to be a promising strategy (Fig. 2)[44,48,73,147]. Such coupled methods have the potential to largely accelerate physics calculations and fill the gaps in the often very sparse experimental data for compositionally complex materials[52,54].

Hybrid methods

Hybrids between physics models and AI, suited for compositionally complex materials, come in different forms. One type of approach pursues a purely 'physics based' AI. This has the advantage of working with established functional forms known from materials physics and can be seamlessly blended with existing simulation tools. The shortcoming of this approach, however, is exactly this, namely, their deliberate restriction to functional forms that are simpler (or maybe too simple) compared with those accessible in AI. This would be a step back in computational material discovery efforts, as it unnecessarily discards the existence or pertinence of more complex mathematical relations behind composition–structure–property relations of chemically complex materials. Yet, the many nonlinearities, mechanism interactions, microstructure features, phase states and high chemical dimensionality suggest that such complexity might be better captured when exploiting the advantages of traditional correlative AI, without imposing 'physics based' limits at an early stage of the modeling.

In contrast, the use of purely correlative and data-driven AI has the opposite problem, meaning, it works with a functional dimensionality that is way too large for typical datasets available in materials science. Using known physical relations (for instance, from thermodynamic databases, density functional theory calculations or chemical trends) allows to reduce dimensionality without restricting functional flexibility. Such hybrid approaches have been successfully developed and applied to discover novel compositionally complex alloys.

For example, an iterative hybrid active-learning AI model has recently been introduced that couples the advantages of physics-based and correlative AI approaches: Rao et al.[148] have suggested a fully integrated hybrid materials discovery strategy that combines AI, AI-guided experiments and density functional theory-based simulations of phase equilibria to solve the sparse data problem via an iterative approach in the form of an active-learning strategy. The approach was applied to accelerate the design of metallic high-entropy Invar alloys (materials with small thermal expansion) in the form of a closed-loop workflow, integrating active learning, thermodynamics simulations and experiments (Fig. 3). Foppa et al.[149] used symbolic regression guided by sure-independence-screening and the sparsifying-operator approach to derive analytical expressions for the prediction of descriptors such as the lattice constant and the cohesive energy for modeling the bulk modulus of perovskites. Kaufmann et al.[78] suggested an AI approach for the identification of thermodynamic and compositional features of multi-component disordered metal carbides, based on a random forests approach. They compared the values obtained by density functional theory simulations with those from machine learning predictions and identified the entropy-forming tendency of 70 new ceramic materials, some of which were then validated.

An issue in using AI for designing chemically complex materials is their lower accuracy compared with conventional methods and their limited extrapolation capability. However, even a relatively inaccurate approach that respects physical laws and considers domain knowledge in the form of a hybrid model can help to improve its predictive capability (Fig. 2). It can, for instance, include specific sustainability and/or thermodynamic constraints, allowing it to indicate interesting compositional directions for a more detailed search. This means that albeit we cannot hope to accurately screen 10^{80} possibilities for material discovery and design purposes given our limited databases in materials science, we can scan subsets of the chemical space with less precise but fast AI methods to propose candidates for more accurate computational and combinatorial investigation. Automated data mining tools and active learning become also more important in this context[104,150–154] (Fig. 3 and Table 1).

Identifying promising composition ranges where available data and literature reveal contradictions, unexpected outliers, or abrupt feature and property transitions for chemically complex materials are good starting points for such a divide-and-conquer strategy where hybrid and staggered approaches can be successful.

In this regard, access to data is crucial. This includes data quality, access to FAIR data (FAIR standing for findability, accessibility,

https://doi.org/10.1038/s43588-023-00412-7

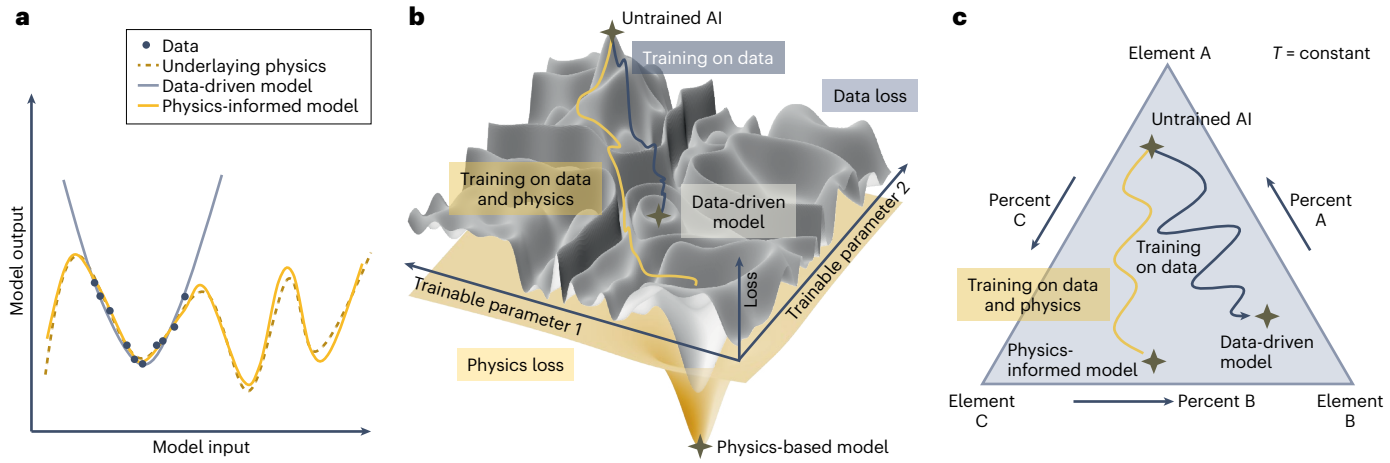

Fig. 2 | Combinations of AI and physics-based models for the simulation of compositionally complex materials. a, Example case where a limited number of sample data (black dots) from an underlying physical phenomenon (dashed line) are used to train an artificial neural network model. The model can be trained only on the data (gray solid line) or in a hybrid method with the data and aspects of the underlying physical laws (yellow solid line). The accuracy of the purely data-driven model is confined to regions where sufficient data were available for training, whereas the physics-based model is more reliable also outside of trained regions. **b**, Schematic diagram of the physics- and data-loss landscapes for a complex artificial neural network designed to surrogate a complex dataset. As the data are complex, the landscape of data loss (gray surface) on which the training process is taking place is also complex. The physical models that are typically included in the physics-based artificial neural network are typically a simplified description of the system, with a simpler loss landscape (yellow surface). A purely data-driven model trained solely on the gray surface may become trapped at a local minimum. A physics-based machine learning model that is trained on both loss landscapes is expected to find the global minimum of the loss landscape easier, as the optimization force from the physics loss is preventing the system from getting trapped in unphysical local minima. **c**, Schematic ternary phase diagram, where the strategy outlined in **a** and **b** has been implemented for the case of compositionally complex materials. In reality, such phase diagrams do not consist of three components, as shown here for the sake of simplicity, but may have up to ten and more axes. *T*, temperature.

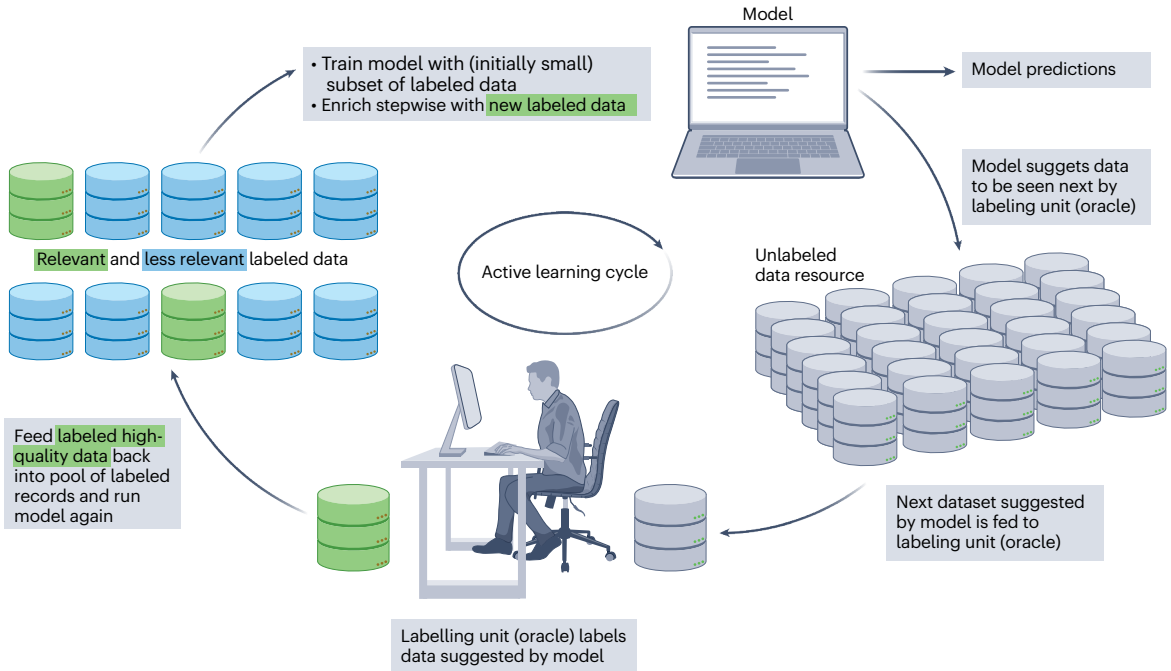

Fig. 3 | Active learning in materials science. Active learning is a powerful approach in which a key problem of AI, namely, the challenge of a limited availability of well-labeled data (one of the most expensive steps in AI) for training is tackled. This is done through iterative machine-assisted decision-making about data subsets to be labeled next, for the gradual enrichment of the initially sparse set of labeled data. The approach proceeds by first conventionally training an AI model with reliably labeled data (such as magnetic versus non-magnetic; face-centered cubic structure versus body-centered cubic structure; and so forth). These training data, however, form only a tiny subset of a much larger dataset of unlabeled data. The model then assists in suggesting those datasets that should be labeled by an expert, often called oracle (can be a human expert user, a physics model or an AI model), to be fed as additional labeled data into the next model iteration. This means that a machine learning model suggests those unlabeled data records that should be next selected for labeling. Thus, in active learning, the labeled dataset starts small and grows iteratively with each loop passed.

interoperability and reuse of digital assets)[155] and data sharing. This should include commercial data that are typically not disclosed by private companies. If manufacturers expect to be part of the coming digital revolution in materials science and profit from digitalized elements of a circular economy, they should develop strategies to share data on the materials they develop, process, use and recycle.

A more general challenge of current AI approaches to the simulation of chemically complex materials lies in the fact that they produce correlations, but not causality. AI has yet to prove that it can reveal causality, that is, so-far-undisclosed physics-based relationships between composition, microstructure and properties. Improving the latter without giving up the former is a pending quest in AI as will be addressed in more detail below. Some recent AI approaches[113,115,156,157] suggest that causality in material design and manufacturing could come to some extent from natural language processing. A pragmatic view of this could be that to predict something meaningful, all that is needed is correlation between input and output. Understanding causality may please the materials scientists but it does not necessarily improve the predictive power of the underlying methods. An example where AI produced results with high engineering significance and relevance for the case of the compositionally and structurally very complex superalloys was recently given by Conduit et al.[106]. In this study the authors trained an artificial neural network with experimental data and computational thermodynamic predictions, capable of discovering and optimizing novel variants of superalloys in a high-dimensional phase space considering multiple properties, including mechanical, physical and thermodynamic features, with a focus on the most critical aspects such as strength and corrosion resistance. This example showed that AI is capable of the training-based development and maturation of a very complicated material via correlation. For such advanced engineering materials as superalloys, this is in general hard to achieve with conventional forward physics models (that is, through causality alone), which as a rule cannot readily cope with such complex alloys under consideration of their manufacturing processes, microstructures, mechanical properties and oxidation response.

Although correlative (empirical) material discovery via AI has therefore clearly earned its place in materials science, pretending that a shortcut by traditional AI methods readily solves the dimensionality problems in material discovery is a promise that is unlikely to be held. The reason is that the required massive training data that are needed for identifying composition–structure–property relations in most cases simply do not exist for such large composition, processing and property spaces. This means that the success of AI in other fields such as language analysis and pattern recognition is not easy to translate to computational materials science, owing to a lack of data for training and validation (that is, data not used for training).

Outlook

A few specific key questions and opportunities can be filtered from the aspects discussed in the previous sections (Fig. 4).

(1) How can AI methods for such complex materials be trained with only sparse and often noisy data, a characteristic but problematic feature in the field of materials science? The best approach seems to be the use of (1) existing thermodynamic and kinetic databases; (2) combinatorial high-throughput experimental methods to produce massive material libraries in conjunction with automated characterization and active-learning-guided experiments[158–162]; (3) massive forward simulations; and/or (4) automated language processing and advanced word embedding methods[112,113,150], to provide better databases and training sets.

(2) How can AI methods for the design of chemically complex alloys be combined in a sensible manner with physics-based models? Promising approaches in this context are data-driven surrogating of physical models[58,163], physically informed neural networks[164,165] and operator learning[166,167]. These methods could be used to accelerate the calculation of material response and properties. Even purely data-driven models have been shown to be able to extrapolate beyond the training data in terms of microstructure, mechanics, damage and topology[70,71,104,143,168,169]. Classical computational methods can be partially replaced by AI methods. These include the application of machine learning for atomistic simulations[51,52,170,171], mesh-based models[172] and the phase-field method[173,174]. The application of these methods for the simulation of compositionally complex materials could allow in the near future the exploration of material candidate spaces that are orders of magnitude larger than what is currently computationally accessible.

(3) What are suited descriptors to link complex chemical composition to properties? While simulation methods to relate the compositions and microstructures of chemically complex materials to properties are fairly well established, the identification of likewise suited descriptors from machine learning methods or hybrid methods has not yet been well studied. A number of approaches to 'featurize' the composition and material data have been already proposed[175,176]. However, composition-to-property linkage requires much more information than only the composition. The heart of the problem is the disparate nature of the important data in the link from the composition to the property. These data could be qualitative (for example, corrosion resistance) or quantitative (for example, pitting potential), categorical (for example, alloy classes) or numerical (for example, carbon concentration), discrete (for example, the number of elements) or continuous (for example, concentration of elements), and all with possible spatial (for example, microstructure inhomogeneity) and temporal (for example, temperature changes) variations. Furthermore, setting a predetermined volume and type for this information is not pertinent. For example, describing the process route of making new material and evaluating its properties typically takes a full-length publication, which might not fit into a predetermined set of composition and processing descriptors. Here natural language processing could be useful to featurize publications as a whole and extract those descriptors that might be suitable for machine learning

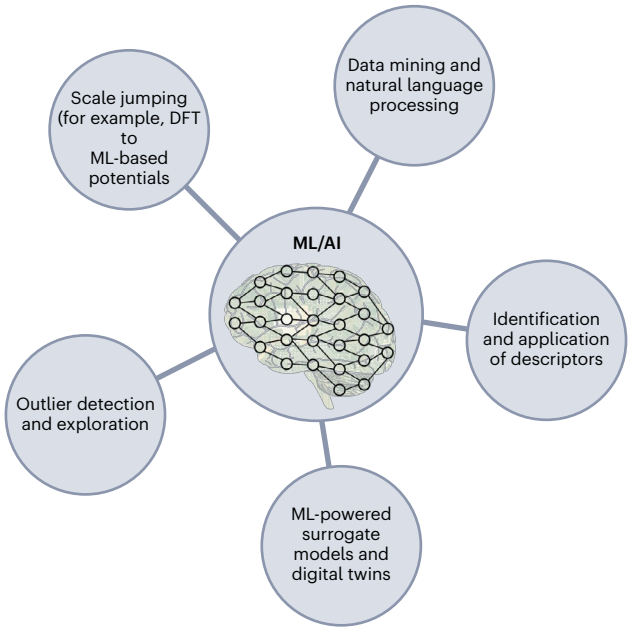

Fig. 4 | Different applications of data-driven and physics-informed AI for the modeling of compositionally complex materials. These include scale bridging through machine learning; text mining for data extraction from the vast trove of literature; identification and computation of easy-to-compute descriptors that strongly correlate with the targeted complex materials properties; surrogate modeling of more complex simulations to create digital twins that faithfully reproduce the response of the actual material; and material discovery through outlier exploration. ML, machine learning; DFT, density functional theory.

tasks in an automatic fashion. This approach might evolve into one of the pillars for an autonomous AI-based material discovery and design strategy, without much bias from an individual researcher, except, for defining, for instance, a specific desired target property.

(4) How can the vast existing phenomenological knowledge from the literature about complex materials be screened and exploited by automated data retrieval via using word embedding methods, considering practical limits such as copyright issues and restricted access, but also semantic challenges such as the very diverse nomenclatures and notations of materials and their compositions in the literature. Another important aspect is the heterogeneity in the information quality of papers regarding potentially relevant details, such as impurity content, processing, homogenization, segregation and microstructure features of chemically complex materials, aspects that are often not so well documented in papers. Here particularly subtle effects arising from elements such as hydrogen, nitrogen and carbon are important, for instance, in the case of metallic high-entropy alloys. Such information can turn out to be vital particularly when it comes to the identification of reliable composition–structure–property relationships.

(5) What are suitable methods to avoid overlooking interesting outliers and 'misfits' in the huge composition space (considering also the many other complexity dimensions discussed above)? This is an important question as it may in many cases be particularly the outliers that could be sources for real discoveries, while in many cases the observation of expected and plausible compositional trends that can in principle also be inferred by an educated guess and from existing theory might be less exciting research targets. This raises the question of what is the best hybrid method or machine learning approach to discover truly unknown and novel materials? A few examples of possible approaches of using machine learning for such material discovery pathways are already present in the literature[177,178].

(6) A more fundamental question of using AI methods in this field is how to use it for guiding material discovery, meaning an approach that actually helps decide about the specific research directions. AI can even make autonomous decisions about promising chemical subspaces to screen for the unknown. This is connected to the deep-learning and active-learning strategies presented above. What remains to be further explored here is to use AI to guide the search into composition subspaces entirely by itself. This would have the charm of eliminating the biased researcher from the picture, a factor that can otherwise lead to blind spots.

(7) A weak point of AI in compositionally complex materials design is insufficient consideration of the roles of unplanned elemental intrusion from synthesis and recycling as well as the associated compositional variations between charges. This will affect material design in the coming decades and must be considered when it comes to (unwanted) complex compositions. Also, constraints from life-cycle assessment should be included, to make AI more holistic and system-oriented and capable of targeting alloying, synthesis and processing pathways that are much more sustainable than today.

(8) The high-entropy alloying approach has meanwhile been extended from metallic systems, which are well suited for the formation of massive solid solutions, owing to the nature of the metallic bond and the associated relevance of the configurational entropy, to many other material classes. These include ceramics, low-dimensional materials, thin films, semiconductors and coatings as well as polymers. For these material classes, the role of (configurational) entropy is often much less relevant as a guiding parameter or material descriptor to material design. Instead, other descriptors need to be identified when designing non-metallic compositionally complex materials through simulations or AI.

(9) Irrespective of the many diverse AI approaches, no clear best-practice method trends are yet discernible. The reason is that it has often not yet been properly analyzed and understood why certain AI methods are actually suitable for addressing certain challenges associated with complex materials or not. There are many black-box successes where AI has led to very convincing results and improved material discovery efficiency, but the exact analysis of the reasons and success factors behind that are still pending. Deeper understanding of the reasons behind the efficiency of certain AI and hybrid methods might open interesting pathways to custom designing such modeling tools for future material design.

Despite the challenges and caveats above, we believe that AI will play a more important role in helping us understand and design compositionally complex materials, especially driven by the fast development of AI algorithms, the availability of high-quality materials datasets and high-performance computing resources.

References

1. Raabe, D., Tasan, C. C. & Olivetti, E. A. Strategies for improving the sustainability of structural metals. *Nature* **575**, 64–74 (2019).
2. Olivetti, E. A. & Cullen, J. M. Toward a sustainable materials system. *Science* **360**, 1396–1398 (2018).
3. Reed, R. C. *The Superalloys* (Cambridge Univ. Press, 2009); https://doi.org/10.1017/cbo9780511541285
4. Suzuki, A., Inui, H. & Pollock, T. M. L12-strengthened cobalt-base superalloys. *Annu. Rev. Mater. Res.* **45**, 345–368 (2015).
5. Sato, J. et al. Cobalt-base high-temperature alloys. *Science* **312**, 90–91 (2006).
6. Nicolas, M. & Deschamps, A. Characterisation and modelling of precipitate evolution in an Al–Zn–Mg alloy during non-isothermal heat treatments. *Acta Mater.* **51**, 6077–6094 (2003).
7. Dorin, T., Deschamps, A., Geuser, F., De & Sigli, C. Quantification and modelling of the microstructure/strength relationship by tailoring the morphological parameters of the T1 phase in an Al–Cu–Li alloy. *Acta Mater.* **75**, 134–146 (2014).
8. Zhao, H. et al. Hydrogen trapping and embrittlement in high-strength Al-alloys. *Nature* **602**, 437–441 (2022).
9. Gutfleisch, O. Controlling the properties of high energy density permanent magnetic materials by different processing routes. *J. Phys. D* **33**, R157–R172 (2000).
10. Han, L. et al. A mechanically strong and ductile soft magnet with extremely low coercivity. *Nature* **608**, 310–316 (2022).
11. Yeh, J. W. et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).
12. Yeh, J. W. in *High-Entropy Alloys: Fundamentals and Applications* (eds Gao, M. et al.) https://doi.org/10.1007/978-3-319-27013-5_1 (Springer, 2016).
13. Cantor, B., Chang, I. T. H., Knight, P. & Vincent, A. J. B. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* **375–377**, 213–218 (2004).
14. Cantor, B. Multicomponent and high entropy alloys. *Entropy* **16**, 4749–4768 (2014).
15. Zhou, X. et al. The hidden structure dependence of the chemical life of dislocations. *Sci. Adv.* **7**, 1–10 (2021).
16. Singh, R., Sharma, A., Singh, P., Balasubramanian, G. & Johnson, D. D. Accelerating computational modeling and design of high-entropy alloys. *Nat. Comput. Sci.* **1**, 54–61 (2021).
17. Kalidindi, S. R. *Hierarchical Materials Informatics* (Butterworth-Heineman, 2016).

18. Aykol, M. et al. Network analysis of synthesizable materials discovery. *Nat. Commun.* **10**, 2018 (2019).
19. Li, J. et al. Accelerated discovery of high-strength aluminum alloys by machine learning. *Commun. Mater.* **1**, 73 (2020).
20. Gubernatis, J. E. & Lookman, T. Machine learning in materials design and discovery: examples from the present and suggestions for the future. *Phys. Rev. Mater.* **2**, 120301 (2018).
21. Wen, C. et al. Machine learning assisted design of high entropy alloys with desired property. *Acta Mater.* **170**, 109–117 (2019).
22. Chang, Y. A. et al. Phase diagram calculation: past, present and future. *Prog. Mater. Sci.* https://doi.org/10.1016/S0079-6425(03)00025-2 (2004).
23. Schmid-Fetzer, R. Phase diagrams: the beginning of wisdom. *J. Phase Equilibria Diffus.* **35**, 735–760 (2014).
24. Kumari, P., Gupta, A. K., Mishra, R. K., Ahmad, M. S. & Shahi, R. R. A comprehensive review: recent progress on magnetic high entropy alloys and oxides. *J. Magn. Magn. Mater.* **554**, 169142 (2022).
25. Han, L. et al. Ultrastrong and ductile soft magnetic high-entropy alloys via coherent ordered nanoprecipitates. *Adv. Mater.* **33**, 2102139 (2021).
26. George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).
27. Oses, C., Toher, C. & Curtarolo, S. High-entropy ceramics. *Nat. Rev. Mater.* **5**, 295–309 (2020).
28. Murty, B. S., Yeh, J. W. & Ranganathan, S. *High Entropy Alloys* 57–76 (Butterworth-Heinemann, 2014); https://doi.org/10.1016/b978-0-12-800251-3.00004-3
29. Gorsse, S., Couzinié, J. P. & Miracle, D. B. From high-entropy alloys to complex concentrated alloys. *C. R. Phys.* **19**, 721–736 (2018).
30. Pei, Z., Yin, J., Hawk, J. A., Alman, D. E. & Gao, M. C. Machine-learning informed prediction of high-entropy solid solution formation: beyond the Hume–Rothery rules. *npj Comput. Mater.* **6**, 50 (2020).
31. Zhao, H. et al. Interplay of chemistry and faceting at grain boundaries in a model Al alloy. *Phys. Rev. Lett.* **124**, 106102 (2020).
32. Zhao, X., Chen, H., Wilson, N., Liu, Q. & Nie, J. F. Direct observation and impact of co-segregated atoms in magnesium having multiple alloying elements. *Nat. Commun.* **10**, 3243 (2019).
33. Raabe, D. et al. Grain boundary segregation engineering in metallic alloys: a pathway to the design of interfaces. *Curr. Opin. Solid State Mater. Sci.* **18**, 253–261 (2014).
34. Rao, Z. et al. Invar effects in FeNiCo medium entropy alloys: from an Invar treasure map to alloy design. *Intermetallics* **111**, 106520 (2019).
35. Wu, X. et al. Role of magnetic ordering for the design of quinary TWIP-TRIP high entropy alloys. *Phys. Rev. Mater.* **4**, 33601 (2020).
36. Counts, W. A., Friak, M., Raabe, D. & Neugebauer, J. Using ab initio calculations in designing bcc Mg–Li alloys for ultra-lightweight applications. *Acta Mater.* **57**, 69–76 (2009).
37. Grabowski, B., Ismer, L., Hickel, T. & Neugebauer, J. Ab initio up to the melting point: anharmonicity and vacancies in aluminum. *Phys. Rev. B* **79**, 134106 (2009).
38. Senkov, O. N., Miller, J. D., Miracle, D. B. & Woodward, C. Accelerated exploration of multi-principal element alloys for structural applications. *Calphad* **50**, 32–48 (2015).
39. Gorsse, S. & Senkov, O. N. About the reliability of CALPHAD predictions in multicomponent systems. *Entropy* **20**, 899 (2018).
40. Zhang, C. & Gao, M. C. in *High-Entropy Alloys: Fundamentals and Applications* (eds Gao, M. et al.) 399–444 (Springer, 2016); https://doi.org/10.1007/978-3-319-27013-5_12
41. Miracle, D. B. & Senkov, O. N. A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511 (2017).
42. Zhang, F. et al. An understanding of high entropy alloys from phase diagram calculations. *Calphad* **45**, 1–10 (2014).
43. Ma, D. et al. Phase stability of non-equiatomic CoCrFeMnNi high entropy alloys. *Acta Mater.* **98**, 288–296 (2015).
44. Grabowski, B. et al. Ab initio vibrational free energies including anharmonicity for multicomponent alloys. *npj Comput. Mater.* **5**, 80 (2019).
45. Kostiuchenko, T., Körmann, F., Neugebauer, J. & Shapeev, A. Impact of lattice relaxations on phase transitions in a high-entropy alloy studied by machine-learning potentials. *npj Comput. Mater.* **5**, 55 (2019).
46. Husic, B. E. et al. Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys.* **153**, 194101 (2020).
47. Zhou, Z. et al. Machine learning guided appraisal and exploration of phase design for high entropy alloys. *npj Comput. Mater.* **5**, 128Z (2019).
48. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
49. Noé, F., Tkatchenko, A., Müller, K. R. & Clementi, C. Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.* **71**, 361–390 (2020).
50. Gubaev, K. et al. Finite-temperature interplay of structural stability, chemical complexity, and elastic properties of bcc multicomponent alloys from ab initio trained machine-learning potentials. *Phys. Rev. Mater.* **5**, 073801 (2021).
51. Li, Z., Kermode, J. R. & De Vita, A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **114**, 096405 (2015).
52. Dhaliwal, G., Nair, P. B. & Singh, C. V. Machine learned interatomic potentials using random features. *npj Comput. Mater.* **8**, 7 (2022).
53. Wang, J. et al. Machine learning of coarse-grained molecular dynamics force fields. *ACS Cent. Sci.* **5**, 755–767 (2019).
54. Westermayr, J., Gastegger, M., Schütt, K. T. & Maurer, R. J. Perspective on integrating machine learning into computational chemistry and materials science. *J. Chem. Phys.* **154**, 230903 (2021).
55. Chen, L.-Q. Q. Phase-field models for microstructure evolution. *Annu. Rev. Mater. Sci.* **32**, 113–140 (2002).
56. Hu, S. Y. & Chen, L. Q. A phase-field model for evolving microstructures with strong elastic inhomogeneity. *Acta Mater.* **49**, 1879–1890 (2001).
57. Diehl, M. et al. Solving material mechanics and multiphysics problems of metals with complex microstructures using DAMASK—the Düsseldorf Advanced Material Simulation Kit. *Adv. Eng. Mater.* **22**, 1901044 (2020).
58. Montes de Oca Zapiain, D., Stewart, J. A. & Dingreville, R. Accelerating phase-field-based microstructure evolution predictions via surrogate models trained by machine learning methods. *npj Comput. Mater.* **7**, 3 (2021).
59. Teichert, G. H. & Garikipati, K. Machine learning materials physics: Surrogate optimization and multi-fidelity algorithms predict precipitate morphology in an alternative to phase field dynamics. *Comput. Methods Appl. Mech. Eng.* **344**, 666–693 (2019).
60. Peivaste, I. et al. Machine-learning-based surrogate modeling of microstructure evolution using phase-field. *Comput. Mater. Sci.* **214**, 111750 (2022).
61. Roters, F. et al. DAMASK—the Düsseldorf Advanced Material Simulation Kit for modeling multi-physics crystal plasticity, thermal, and damage phenomena from the single crystal up to the component scale. *Comput. Mater. Sci.* **158**, 420–478 (2019).
62. Roters, F. et al. Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: theory, experiments, applications. *Acta Mater.* **58**, 1152–1211 (2010).
63. Mika, D. P. & Dawson, P. R. Effects of grain interaction on deformation in polycrystals. *Mater. Sci. Eng. A* **257**, 62–76 (1998).

64. Beaudoin, A. J., Dawson, P. R., Mathur, K. K., Kocks, U. F. & Korzekwa, D. A. Application of polycrystal plasticity to sheet forming. *Comput. Methods Appl. Mech. Eng.* **117**, 49–70 (1994).
65. Kalidindi, S. R., Duvvuru, H. K. & Knezevic, M. Spectral calibration of crystal plasticity models. *Acta Mater.* **54**, 1795–1804 (2006).
66. Helm, D., Butz, A., Raabe, D. & Gumbsch, P. Microstructure-based description of the deformation of metals: theory and application. *JOM* **63**, 26–33 (2011).
67. Liu, C. et al. An integrated crystal plasticity-phase field model for spatially resolved twin nucleation, propagation, and growth in hexagonal materials. *Int. J. Plast.* **106**, 203–227 (2018).
68. Shanthraj, P., Svendsen, B., Sharma, L., Roters, F. & Raabe, D. Elasto-viscoplastic phase field modelling of anisotropic cleavage fracture. *J. Mech. Phys. Solids* **99**, 19–34 (2017).
69. Khorrami, M. S. et al. An artificial neural network for surrogate modeling of stress fields in viscoplastic polycrystalline materials. Preprint at https://arxiv.org/abs/2208.13490 (2022).
70. Fernández, M., Rezaei, S., Rezaei Mianroodi, J., Fritzen, F. & Reese, S. Application of artificial neural networks for the prediction of interface mechanics: a study on grain boundary constitutive behavior. *Adv. Model. Simul. Eng. Sci.* **7**, 1 (2020).
71. Mianroodi, J. R., H. Siboni, N. & Raabe, D. Teaching solid mechanics to artificial intelligence—a fast solver for heterogeneous materials. *npj Comput. Mater.* **7**, 99 (2021).
72. Ibragimova, O., Brahme, A., Muhammad, W., Lévesque, J. & Inal, K. A new ANN based crystal plasticity model for fcc materials and its application to non-monotonic strain paths. *Int. J. Plast.* **144**, 103059 (2021).
73. Mangal, A. & Holm, E. A. Applied machine learning to predict stress hotspots I: Face centered cubic materials. *Int. J. Plast.* **111**, 122–134 (2018).
74. Schneeweiss, O. et al. Magnetic properties of the CrMnFeCoNi high-entropy alloy. *Phys. Rev. B* **96**, 014437 (2017).
75. Oh, H. S. et al. Lattice distortions in the FeCoNiCrMn high entropy alloy studied by theory and experiment. *Entropy* **18**, 321 (2016).
76. Ma, D., Grabowski, B., Körmann, F., Neugebauer, J. & Raabe, D. Ab initio thermodynamics of the CoCrFeMnNi high entropy alloy: importance of entropy contributions beyond the configurational one. *Acta Mater.* **100**, 90–97 (2015).
77. Löffler, A. et al. Quaternary Al–Cu–Mg–Si *Q* phase: sample preparation, heat capacity measurement and first-principles calculations. *J. Phase Equilibria Diffus.* **37**, 119–126 (2016).
78. Kaufmann, K. et al. Discovery of high-entropy ceramics via machine learning. *npj Comput. Mater.* **6**, 42 (2020).
79. Sarker, P. et al. High-entropy high-hardness metal carbides discovered by entropy descriptors. *Nat. Commun.* **9**, 4980 (2018).
80. Kaufmann, K. et al. Crystal symmetry determination in electron diffraction using machine learning. *Science* **367**, 564–568 (2020).
81. Kaufmann, L. & Bernstein, H. *Computer Calculation of Phase Diagrams* (Academic Press, 1970).
82. Spencer, P. J. A brief history of CALPHAD. *Calphad* **32**, 1–8 (2008).
83. Sandlöbes, S. et al. The relation between ductility and stacking fault energies in Mg and Mg–Y alloys. *Acta Mater.* **60**, 3011–3021 (2012).
84. Lei, Z. et al. Enhanced strength and ductility in a high-entropy alloy via ordered oxygen complexes. *Nature* **563**, 546–550 (2018).
85. Güvenç, O., Roters, F., Hickel, T. & Bambach, M. ICME for crashworthiness of TWIP steels: from ab initio to the crash performance. *JOM* **67**, 120–128 (2015).
86. De Cooman, B. C., Estrin, Y. & Kim, S. K. Twinning-induced plasticity (TWIP) steels. *Acta Mater.* **142**, 283–362 (2018).
87. Wei, S., He, F. & Tasan, C. C. Metastability in high-entropy alloys: a review. *J. Mater. Res.* **33**, 2924–2937 (2018).
88. Raabe, D., Li, Z. & Ponge, D. Metastability alloy design. *MRS Bull.* **44**, 266–272 (2019).
89. Curtin, W. A., Olmsted, D. L. & Hector, L. G. A predictive mechanism for dynamic strain ageing in aluminium–magnesium alloys. *Nat. Mater.* **5**, 875–880 (2006).
90. George, E. P., Curtin, W. A. & Tasan, C. C. High entropy alloys: a focused review of mechanical properties and deformation mechanisms. *Acta Mater.* https://doi.org/10.1016/j.actamat.2019.12.015 (2019).
91. Varvenne, C., Leyson, G. P. M., Ghazisaeidi, M. & Curtin, W. A. Solute strengthening in random alloys. *Acta Mater.* **124**, 660–683 (2017).
92. Li, Z., Pradeep, K. G., Deng, Y., Raabe, D. & Tasan, C. C. Metastable high-entropy dual-phase alloys overcome the strength-ductility trade-off. *Nature* **534**, 227–230 (2016).
93. Li, Z., Tasan, C. C., Pradeep, K. G. & Raabe, D. A TRIP-assisted dual-phase high-entropy alloy: grain size and phase fraction effects on deformation behavior. *Acta Mater.* **131**, 323–335 (2017).
94. Pei, Z. et al. Rapid theory-guided prototyping of ductile Mg alloys: from binary to multi-component materials. *New J. Phys.* **17**, 93009 (2015).
95. Nikolov, S. et al. Revealing the design principles of high-performance biological composites using ab initio and multiscale simulations: The example of lobster cuticle. *Adv. Mater.* **22**, 519–526 (2010).
96. Pei, Z. et al. From generalized stacking fault energies to dislocation properties: five-energy-point approach and solid solution effects in magnesium. *Phys. Rev. B* **92**, 64107 (2015).
97. Li, Q. et al. Quantification of flexoelectricity in $PbTiO_3/SrTiO_3$ superlattice polar vortices using machine learning and phase-field modeling. *Nat. Commun.* **8**, 1468 (2017).
98. Mianroodi, J. R., Siboni, N. H. & Raabe, D. Computational discovery of energy-efficient heat treatment for microstructure design using deep reinforcement learning. Preprint at https://arxiv.org/abs/2209.11259 (2022).
99. Yuan, M., Paradiso, S., Meredig, B. & Niezgoda, S. R. Machine learning-based reduce order crystal plasticity modeling for ICME applications. *Integr. Mater. Manuf. Innov.* **7**, 214–230 (2018).
100. Sangid, M. D. Coupling in situ experiments and modeling—opportunities for data fusion, machine learning, and discovery of emergent behavior. *Curr. Opin. Solid State Mater. Sci.* **24**, 100797 (2020).
101. Saidi, P. et al. Deep learning and crystal plasticity: a preconditioning approach for accurate orientation evolution prediction. *Comput. Methods Appl. Mech. Eng.* **389**, 114392 (2022).
102. Salmenjoki, H., Alava, M. J. & Laurson, L. Machine learning plastic deformation of crystals. *Nat. Commun.* **9**, 5307 (2018).
103. Holm, E. A. et al. Overview: computer vision and machine learning for microstructural characterization and analysis. *Metall. Mater. Trans. A* **51**, 5985–5999 (2020).
104. Bock, F. E. et al. A review of the application of machine learning and data mining approaches in continuum materials mechanics. *Front. Mater.* **6**, 110 (2019).
105. Devi, M. A. et al. An informatic approach to predict the mechanical properties of aluminum alloys using machine learning techniques. In *Proc. International Conference on Smart Electronics and Communication*. 536–541 (2020); https://doi.org/10.1109/ICOSEC49089.2020.9215277
106. Conduit, B. D., Jones, N. G., Stone, H. J. & Conduit, G. J. Design of a nickel-base superalloy using a neural network. *Mater. Des.* **131**, 358–365 (2017).
107. Barnett, M. R. et al. A scrap-tolerant alloying concept based on high entropy alloys. *Acta Mater.* **200**, 735–744 (2020).
108. Bartók, A. P. et al. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **3**, e1701816 (2017).

109. Ganesh, M., Hawkins, S. C., Kordzakhia, N. & Unicomb, S. An efficient Bayesian neural network surrogate algorithm for shape detection. *ANZIAM J.* **62**, C112–C127 (2022).
110. Vahid, A. et al. New Bayesian-optimization-based design of high-strength 7xxx-series alloys from recycled aluminum. JOM https://doi.org/10.1007/s11837-018-2984-z (2018).
111. Aggarwal, C. C. et al. Multi-objective Bayesian materials discovery. *Comput. Mater. Sci.* **3**, 227–235 (2017).
112. Swain, M. C. & Cole, J. M. ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature. *J. Chem. Inf. Model.* **56**, 1894–1904 (2016).
113. Kim, E. et al. Materials synthesis insights from scientific literature via text extraction and machine learning. *Chem. Mater.* **29**, 9436–9444 (2017).
114. Mahbub, R. et al. Text mining for processing conditions of solid-state battery electrolyte. *Electrochem. Commun.* **121**, 106860 (2020).
115. Olivetti, E. A. et al. Data-driven materials research enabled by natural language processing and information extraction. *Appl. Phys. Rev.* **7**, 041317 (2020).
116. Pei, Z., Yin, J., Liaw, P. K. & Raabe, D. Toward the design of ultrahigh-entropy alloys via mining six million texts. *Nat. Commun.* **14**, 54 (2023).
117. Zhang, T. & Sun, S. Thermodynamics-informed neural network (TINN) for phase equilibrium calculations considering capillary pressure. *Energies* **14**, 7724 (2021).
118. Mnih, V. et al. Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
119. Muhammad, W., Brahme, A. P., Ibragimova, O., Kang, J. & Inal, K. A machine learning framework to predict local strain distribution and the evolution of plastic anisotropy & fracture in additively manufactured alloys. *Int. J. Plast.* **136**, 1–38 (2021).
120. Hernandez, Q., Badias, A., Chinesta, F. & Cueto, E. Thermodynamics-informed graph neural networks. *IEEE Trans. Artif. Intell.* **4581**, 1–1 (2022).
121. Ding, J. et al. Machine learning for molecular thermodynamics. *Chin. J. Chem. Eng.* **31**, 227–239 (2021).
122. Jagtap, A. D., Kawaguchi, K. & Karniadakis, G. E. Adaptive activation functions accelerate convergence in deep and physics-informed neural networks. *J. Comput. Phys.* **404**, 109136 (2020).
123. Jagtap, A. D., Kawaguchi, K. & Karniadakis, G. E. Locally adaptive activation functions with slope recovery for deep and physics-informed neural networks. *Proc. R. Soc. A* **476**, 20200334 (2020).
124. Rodriguez-Torrado, R. et al. Physics-informed attention-based neural network for hyperbolic partial differential equations: application to the Buckley–Leverett problem. *Sci. Rep.* **12**, 7557 (2022).
125. Zhang, Z. & Gu, G. X. Physics-informed deep learning for digital materials. *Theor. Appl. Mech. Lett.* **11**, 100220 (2021).
126. Wagih, M. & Schuh, C. A. Learning grain-boundary segregation: from first principles to polycrystals. *Phys. Rev. Lett.* **129**, 046102 (2022).
127. Wagih, M., Larsen, P. M. & Schuh, C. A. Learning grain boundary segregation energy spectra in polycrystals. *Nat. Commun.* **11**, 6376 (2020).
128. Galvão, T. L. P., Novell-Leruth, G., Kuznetsova, A., Tedim, J. & Gomes, J. R. B. Elucidating structure–property relationships in aluminum alloy corrosion inhibitors by machine learning. *J. Phys. Chem. C* **124**, 5624–5635 (2020).
129. Mangos, J. & Birbilis, N. Computational alloy design and discovery using machine learning. Preprint at https://arxiv.org/abs/2105.14806 (2021).
130. Sasidhar, K. N. et al. Deep learning framework for uncovering compositional and environmental contributions to pitting resistance in passivating alloys. *npj Mater. Degrad.* **6**, 71 (2022).
131. Gaustad, G., Olivetti, E. & Kirchain, R. Toward sustainable material usage: evaluating the importance of market motivated agency in modeling material flows. *Environ. Sci. Technol.* **45**, 4110–4117 (2011).
132. Kirchain, R. E., Gregory, J. R. & Olivetti, E. A. Environmental life-cycle assessment. *Nat. Mater.* **16**, 693–697 (2017).
133. Gaustad, G., Olivetti, E. & Kirchain, R. Design for recycling. *J. Ind. Ecol.* **14**, 286–308 (2010).
134. Daehn, K. E., Cabrera Serrenho, A. & Allwood, J. M. How will copper contamination constrain future global steel recycling? *Environ. Sci. Technol.* **51**, 6599–6606 (2017).
135. Allwood, J. M. et al. *Sustainable Materials: With Both Eyes Open* (UIT Cambridge, 2012).
136. Cann, J. L. et al. Sustainability through alloy design: challenges and opportunities. *Prog. Mater. Sci.* **117**, 100722 (2020).
137. Lederer, Y., Toher, C., Vecchio, K. S. & Curtarolo, S. The search for high entropy alloys: a high-throughput ab-initio approach. *Acta Mater.* **159**, 364–383 (2018).
138. Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
139. Raabe, D. et al. Making sustainable aluminum by recycling scrap: the science of 'dirty' alloys. *Prog. Mater. Sci.* **128**, 100947 (2022).
140. Hiraki, T. et al. Thermodynamic criteria for the removal of impurities from end-of-life magnesium alloys by evaporation and flux treatment. *Sci. Technol. Adv. Mater.* https://doi.org/10.1088/1468-6996/12/3/035003 (2011).
141. Stemper, L., Tunes, M. A., Tosone, R., Uggowitzer, P. J. & Pogatscher, S. On the potential of aluminum crossover alloys. *Prog. Mater. Sci.* https://doi.org/10.1016/j.pmatsci.2021.100873 (2021).
142. Draxl, C. & Scheffler, M. The NOMAD laboratory: from data sharing to artificial intelligence. *J. Phys. Mater.* **2**, 036001 (2019).
143. Mianroodi, J. R., Rezaei, S., Siboni, N. H., Xu, B.-X. & Raabe, D. Lossless multi-scale constitutive elastic relations with artificial intelligence. *npj Comput. Mater.* **8**, 67 (2021).
144. Dimiduk, D. M., Holm, E. A. & Niezgoda, S. R. Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. *Integr. Mater. Manuf. Innov.* **7**, 157–172 (2018).
145. Sandlöbes, S. et al. A rare-earth free magnesium alloy with improved intrinsic ductility. *Sci. Rep.* **7**, 10458 (2017).
146. Sandlöbes, S. et al. Ductility improvement of Mg alloys by solid solution: ab initio modeling, synthesis and mechanical properties. *Acta Mater.* **70**, 92–104 (2014).
147. Haghighat, E., Raissi, M., Moure, A., Gomez, H. & Juanes, R. A Physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. *Comput. Methods Appl. Mech. Eng.* **379**, 113741 (2021).
148. Rao, Z. et al. Machine learning-enabled high-entropy alloy discovery. *Science* **85**, 78–85 (2022).
149. Foppa, L., Purcell, T. A. R., Levchenko, S. V., Scheffler, M. & Ghringhelli, L. M. Hierarchical symbolic regression for identifying key physical parameters correlated with bulk properties of perovskites. *Phys. Rev. Lett.* **129**, 55301 (2022).
150. Court, C. J. & Cole, J. M. Auto-generated aterials database of Curie and Neél temperatures via semisupervised relationship extraction. *Sci. Data* **5**, 180111 (2018).
151. Katnagallu, S. et al. Advanced data mining in field ion microscopy. *Mater. Charact.* **146**, 307–318 (2018).
152. Wang, C., Fu, H., Jiang, L., Xue, D. & Xie, J. A property-oriented design strategy for high performance copper alloys via machine learning. *npj Comput. Mater.* **5**, 87 (2019).
153. Yang, Z. et al. Deep learning approaches for mining structure-property linkages in high contrast composites from simulation datasets. *Comput. Mater. Sci.* **151**, 278–287 (2018).

154. Yang, Z. et al. Establishing structure-property localization linkages for elastic deformation of three-dimensional high contrast composites using deep learning approaches. *Acta Mater.* **166**, 335–345 (2019).
155. Wilkinson, M. D. et al. Comment: The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
156. Kajikawa, Y., Sugiyama, Y., Mima, H. & Matsushima, K. Causal knowledge extraction by natural language processing in material science: a case study in chemical vapor deposition. *Data Sci. J.* **5**, 108–118 (2006).
157. Cuomo, S. et al. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *J. Sci. Comput.* **92**, 88 (2022).
158. Cui, J. et al. Combinatorial search of thermoelastic shape-memory alloys with extremely small hysteresis width. *Nat. Mater.* **5**, 286–290 (2006).
159. Li, Z., Ludwig, A., Savan, A., Springer, H. & Raabe, D. Combinatorial metallurgical synthesis and processing of high-entropy alloys. *J. Mater. Res.* **33**, 3156–3169 (2018).
160. Löffler, T. et al. Discovery of a multinary noble metal-free oxygen reduction catalyst. *Adv. Energy Mater.* **8**, 1802269 (2018).
161. Raabe, D. et al. Ab initio-guided design of twinning-induced plasticity steels. *MRS Bull.* **41**, 320–325 (2016).
162. Gebhardt, T., Music, D., Takahashi, T. & Schneider, J. M. Combinatorial thin film materials science: from alloy discovery and optimization to alloy design. *Thin Solid Films* **520**, 5491–5499 (2012).
163. Mohammadzadeh, S. & Lejeune, E. Predicting mechanically driven full-field quantities of interest with deep learning-based metamodels. *Extrem. Mech. Lett.* **50**, 101566 (2022).
164. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
165. Pun, G. P. P., Batra, R., Ramprasad, R. & Mishin, Y. Physically informed artificial neural networks for atomistic modeling of materials. *Nat. Commun.* **10**, 2339 (2019).
166. Wang, S., Wang, H. & Perdikaris, P. Learning the solution operator of parametric partial differential equations with physics-informed DeepONets. *Sci. Adv.* **7**, eabi8605 (2021).
167. You, H., Zhang, Q., Ross, C. J., Lee, C.-H. & Yu, Y. Learning deep implicit Fourier neural operators (IFNOs) with applications to heterogeneous material modeling. *Comput. Methods Appl. Mech. Eng.* https://doi.org/10.1016/j.cma.2022.115296 (2022).
168. Guo, K., Yang, Z., Yu, C. H. & Buehler, M. J. Artificial intelligence and machine learning in design of mechanical materials. *Mater. Horiz.* **8**, 1153–1172 (2021).
169. Abueidda, D. W., Lu, Q. & Koric, S. Meshless physics-informed deep learning method for three-dimensional solid mechanics. *Int. J. Numer. Methods Eng.* **122**, 7182–7201 (2021).
170. Winkler, L., Müller, K. R. & Sauceda, H. E. High-fidelity molecular dynamics trajectory reconstruction with bi-directional neural networks. *Mach. Learn. Sci. Technol.* **3**, 025011 (2022).
171. Riniker, S., Wang, S., Bleiziffer, P., Böselt, L. & Esposito, C. Machine learning with and for molecular dynamics simulations. *Chimia* **73**, 1024–1027 (2019).
172. Pfaff, T., Fortunato, M., Sanchez-Gonzalez, A. & Battaglia, P. W. Learning mesh-based simulation with graph networks. Preprint at https://arxiv.org/abs/2010.03409 (2020).
173. Wight, C. L. & Zhao, J. Solving Allen–Cahn and Cahn–Hilliard equations using the adaptive physics informed neural networks. *Commun. Comput. Phys.* **29**, 930–954 (2021).
174. Attari, V. et al. Uncertainty propagation in a multiscale CALPHAD-reinforced elastochemical phase-field model. *Acta Mater.* **183**, 452–470 (2020).
175. Gierlich, C. & Palkovits, S. Featurizing chemistry for machine learning—methods and a coded example. *Curr. Opin. Chem. Eng.* **37**, 100840 (2022).
176. Kalidindi, S. R. Feature engineering of material structure for AI-based materials knowledge systems. *J. Appl. Phys.* **128**, 41103 (2020).
177. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **5**, 83 (2019).
178. Ojih, J., Al-Fahdi, M., Rodriguez, A. D. & Choudhary, K. Efficiently searching extreme mechanical properties via boundless objective-free exploration and minimal first-principles calculations. *npj Comput. Mater.* **8**, 143 (2022).

Acknowledgements

The authors are grateful for financial support by the BIGmax research network of the Max-Planck Society (https://www.bigmax.mpg.de/).

Author contributions

All authors contributed equally to the article, in terms of topical lines, discussion and writing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Dierk Raabe, Jaber Rezaei Mianroodi or Jörg Neugebauer.

Peer review information *Nature Computational Science* thanks Stefano Curtarolo, Peng-Fei Guan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Jie Pan, in collaboration with the *Nature Computational Science* team.