communications materials

REVIEW ARTICLE

https://doi.org/10.1038/s43246-021-00209-z

OPEN

Artificial intelligence for search and discovery of quantum materials

Valentin Stanev $^{1,2 \Join}$, Kamal Choudhary³, Aaron Gilad Kusne^{1,3}, Johnpierre Paglione 2,4,5 & Ichiro Takeuchi^{1,2}

Artificial intelligence and machine learning are becoming indispensable tools in many areas of physics, including astrophysics, particle physics, and climate science. In the arena of quantum materials, the rise of new experimental and computational techniques has increased the volume and the speed with which data are collected, and artificial intelligence is poised to impact the exploration of new materials such as superconductors, spin liquids, and topological insulators. This review outlines how the use of data-driven approaches is changing the landscape of quantum materials research. From rapid construction and analysis of computational and experimental databases to implementing physical models as pathfinding guidelines for autonomous experiments, we show that artificial intelligence is already well on its way to becoming the lynchpin in the search and discovery of quantum materials.

n the last decade, a broad new topic has emerged at the forefront of condensed matter physics. It includes recently discovered exotic systems such as two-dimensional (2D) materials¹, topological insulators², topological superconductors³, and Weyl semimetals⁴ alongside several materials that have been baffling researchers for many years—unconventional superconductors^{5–7} and quantum spin liquids⁸—in the loosely-defined family of "quantum materials"⁹. The unifying theme connecting these seemingly disparate systems is the idea of emergence—they cannot be understood as simple ensembles of noninteracting fermions or bosons in "free" space; rather, their unique properties spring out of the collective behavior of a macroscopic number of interacting quantum particles. These materials host various exotic excitations, e.g., "relativistic" fermions in Dirac materials, Majorana bound states in topological superconductors, and skyrmions in chiral magnets¹⁰, which not only enrich our fundamental understanding of matter but also hold the promise of "quantum" leaps in a multitude of technological arenas. Because of its high carrier mobility, graphene stands to change the future of high-speed electronics and communication devices¹¹. Topological insulators, which exhibit quantized anomalous quantum Hall effect^{12–14}, possess spin-momentum-locked conduction channels that are immune to backscattering and can be used to make perfect spin-filtering devices¹⁵.

But arguably the most exciting prospect is the use of topological superconductors in faulttolerant quantum computing¹⁶. Quantum information processing based on manipulating topologically protected Majorana states provides an inherent cover against the threat of decoherence—the Achilles heel of most embodiments of qubits. Great strides are being made toward realizing topological superconducting qubits using existing materials (e.g., by inducing superconductivity in topological insulators through the proximity effect)^{17,18}; however, there is very little progress in systematically identifying many true "intrinsic" topological superconductors.

In general, the search for new quantum materials presents an immense scientific challenge. The rich tapestries of quantum physics, which make them so exciting are also the very reasons



¹Department of Materials Science and Engineering, University of Maryland, College Park, MD 20742, USA. ² Maryland Quantum Materials Center, University of Maryland, College Park, MD 20742, USA. ³ National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. ⁴ Department of Physics, University of Maryland, College Park, MD 20742, USA. ⁵ The Canadian Institute for Advanced Research, Toronto, ON M5G 1Z, Canada. ⁵ email: vstanev@umd.edu

which render them so difficult to decipher. The notorious case in point is the superconductivity in cuprate materials, whose fundamental mechanism still eludes us after three decades of concerted research efforts^{5,6,19}. For some strongly correlated quantum materials such as spin liquids even the theoretical language necessary to adequately describe them is yet to be fully developed. When there is no guidance from rigorous theory, we are often left with serendipity to work with while exploring new compounds. There is, however, a glimmer of hope in an unusual form: research tools using Artificial Intelligence (AI) are coming to the rescue.

The explosive increase in computational power and volume of accessible data is fueling the growth of AI as it transforms many aspects of our everyday lives with emerging technologies such as self-driving vehicles and Internet of Things devices. AI, especially its subfield of Machine Learning (ML), is also becoming a breadand-butter analysis tool, with important applications in virtually all sciences, from cancer research²⁰ to astrophysics²¹. ML has already made its mark on condensed matter and materials physics by providing a robust platform for extracting a parsimonious description of a materials system from experimental and computational data^{22–28}. When the task at hand is the discovery of new quantum materials, then one obvious strategy is to unleash ML on information obtained from existing compounds to make predictions on compositions and collective behaviors of their atomic constituents.

In this review, we discuss the ongoing research spearheading the effort to turn AI into a ubiquitous and indispensable tool for the study of quantum materials. The complexities and rich physics of these materials make them simultaneously one of the most promising targets and a particularly challenging subject for the emerging AI-enabled methods. Only exhaustive highdimensional data encompassing a diverse range of physical properties can help scientists develop a comprehensive understanding of materials such as unconventional superconductors, with complex physics often rooted in strong electron-electron interactions, or graphene superlattices, hosting a multitude of competing phases²⁹. Thus, data need to be fused from disparate sources with vastly different modalities and veracities: theoretical models and first-principles computational tools, various experimental measurements, and results from previous studies. This review provides a vision for the future where ML algorithms are commonly proliferated for the creation, analysis, and visualization of such high-dimensional heterogeneous data collections. The focus of the article is superconductivity, which is one of the most intensely pursued topics in all of condensed matter physics: the continued research effort, combined with its relatively long history, has led to significant accumulation of computational and experimental data-a prerequisite for the use of ML tools. However, the tools being developed to study superconductors have the potential to dramatically increase the efficiency of the experimental and computational studies of all quantum materials, representing a veritable breakthrough in the exploration of these complex physical systems. (For some common AI terminology and recent examples of ML applied to quantum materials problems see Fig. 1 and Box 1.) Still, this vision is not without its share of potential pitfalls. In the sections below, we discuss the benefits and challenges associated with AI-based approaches to novel quantum materials.

AI for computational quantum materials

In application after application, it has been demonstrated that ML can turn complex multidimensional data into actionable knowledge (see, for example, refs. ^{30–32}). The success of ML methods, however, depends crucially on access to prodigious amounts of

high-quality data. Not surprisingly, in physical sciences, these algorithms were first adopted in fields with abundant and highly standardized measurements. In fact, the vast amounts of data generated in particle physics have for decades required the use of automated algorithms for data analysis^{21,33}. The importance of ML in this area is exemplified by the central role it played in the discovery of the Higgs boson at the Large Hadron Collider at CERN³³.

When it comes to materials data, sophisticated computational first-principles tools are the order of the day, and enormous databases with millions of computed property entries are now readily available at everyone's fingertips³⁴⁻³⁸. In the realm of quantum materials, however, the situation is less auspicious. The most widely used computational method for simulating materials properties is Density Functional Theory (DFT), which has served as the backbone of materials by design in areas such as energy and magnetic materials (see, for example, refs. ³⁹⁻⁴¹). Yet, DFT is built on several commonly used but essentially uncontrolled approximations. While the trends of the physical properties (e.g., as a function of composition) calculated by conventional DFT are generally correct, the veracity of the calculated parameters depends significantly on computational details. For quantum materials, calculated properties can become even more tenuous and inaccurate, due for example, to self-interaction error originating in electron-electron interactions, spin-orbit coupling, magnetism, etc. This severely limits the utility of conventional DFT for entire classes of quantum materials. A very promising emerging research area addresses this challenge by using ML methods to systematically improve the approximations made in DFT calculations^{42–45}. AI methods are also being used to accelerate high-level methods such as Quantum Monte Carlo (QMC)^{46,47} and Dynamical Mean Field Theory (DMFT)⁴⁸, which are more accurate than DFT but can be orders-of-magnitude slower and higher in computational cost.

These exciting developments notwithstanding, for now, conventional DFT remains the workhorse of computational materials science. It has been shown that the calculations can be combined with ML algorithms to efficiently and accurately predict materials properties such as melting temperature, bandgap, shear modulus, and heat capacity⁴⁹⁻⁵¹. DFT methods were also at the center of some of the early data-driven approaches to quantum materials, especially in the search for new superconductors. The discovery of superconducting materials has always been driven by serendipity and individual researchers' intuition. However, the complex physics of cuprates and other strongly correlated materials (for example organic charge-transfer salts, exhibiting a variety of exotic phases: unconventional superconductivity, Mott insulator, spin liquid, etc.⁵²) has forced scientists to explore more advanced datacentric approaches. In ref. 53, a series of DFT calculations of the band structure of hole-doped cuprates were compared, in order to identify physical parameters governing superconductivity in these materials. The authors of ref. 54 went further and proposed several distinct characteristics of the band structure of cuprates as key precursors of their superconductivity. These characteristics were then used as criteria to filter through the calculated electronic structures of all materials in Inorganic Crystal Structure Database (ICSD), containing hundreds of thousands of existing crystal structures, and over 100 materials were identified as potential high-temperature superconductors. A similar approach was used recently in ref. 55, where p-terphenyl-an organic material that shows hints of a possible superconducting transition at ≈123 K when doped with potassium-was used as a prototype. A database containing the electronic structure of more than 10,000 organic crystals was surveyed for compounds with a similar density of states (DOS)⁵⁶, and as a result, 15 compounds were proposed as candidate superconductors.

Type of Data	Type of Machine Learning	Type of Algorithm	Examples
Structured (Tabular) Data Sources: experimental and/or computational databases.	Supervised Learning Uses labeled data. Divided into classification (target properties with discrete categories) and regression (target prop- erties are continuous)	Parametric models Fixed complexity (e.g. analytic) models Examples: Linear and Logistic regression Pros and Cons: Interpretable, but not flexible	Ref. 74: Trained gradient boosting model to predict the topology of a material based on composition and crystal symmetry.
		Non-parametric models Complexity grows with the data Examples: Random Forest, Gradient Boosting	Ref. 88: Used random forest to predict superconducting T_c based on composition.
Examples: AFLOW ³⁴ , Materials Project ³⁵ , JARVIS ³⁸ , ICSD ⁸⁶ , Pauling, Eilo ⁹¹		Pros and Cons: Powerful, but not easily interpreted Deep Learning models:	Ref. 81: Trained gradient boosting classifier on DFT data to predict stability of potential 2D materials
Unstructured Data	erties are continuous)	Can learn features from both structured and unstructured data Example: Convolutional Neural Network (CNN) Pros and Cons: Very powerful, but require a lot of training data	Ref. 107: Used CNN on SISTM image data to find the order parameter type in cuprate materials.
Sources: images, spectra, text Examples: Imaging/spectroscopic experiments ^{107,110} , scientific articles ¹⁰⁰	Unsupervised Learning Used to discover latent structure in unlabeled data	Dimensionality Reduction Project high-dimensional data onto low- dimensional space preserving latent structures	Ref. 110: Applied CNN to XAS spectral data to find potential topological materials.
		Examples: PCA, NMF, t-SNE Clustering Find intrinsic groups in unlabeled data Examples: K-Means, Hierarchical Clustering	Ref. 87: Used NN model to predict crystal structure from composition. Cluste- red model representations to find potential topolo- gical materials.

Fig. 1 Commonly used machine-learning tasks and algorithms. The figure shows the most common types of machine-learning tasks and algorithms, as well as recent examples of these algorithms applied to quantum materials problems. See Box 1 for the definitions of some key machine-learning terms.

Box 1 Key terms

Artificial intelligence: "Intelligence" demonstrated by machines able to perceive their environment and take autonomous actions to achieve a goal. Machine learning: Branch of AI, a method for automated model building. Design of systems that can learn from data and make decisions with minimal human intervention.

Supervised learning: Subfield of ML, aimed at creating predictive models based on both input and output (target) data. Subdivides into: Classification: Modeling target data with discrete categories.

Regression: Modeling continuous target data.

Commonly used algorithms include Linear Regression, Support Vector Machines, Decision Trees, Gaussian Processes, Neural Networks.

Unsupervised learning: Subfield of ML, aimed at discovering internal representations from input data only. The two most widely used subfields are: **Dimensionality reduction**: Creating low-dimensional projections of high-dimensional data.

Clustering: Finding groups of related data points.

Commonly used algorithms include K-means Clustering, Hierarchical Clustering, Hidden Markov Models.

Active learning (optimal experimental design): Subfield of ML, designing algorithms able to interactively query an information source to obtain new outputs. Query strategies are typically some mix of exploration (aiming to maximize new information) and exploitation (aiming to optimize a property).

The studies mentioned above combine DFT calculations with heuristic criteria to search for novel superconductors. Such criteria unavoidably suffer from various biases and can be difficult to generalize. A much more rigorous approach is offered by ML methods, which provide a systematic way to extract important predictors of materials properties from complex highdimensional data. As an example, an investigation in 2015 used structural and electronic properties data from the AFLOW Online Repositories to define and calculate various fingerprints that encode materials' electronic band structures, as well as crystallographic and constitutional information⁵⁷. These fingerprints were then used to create both regression and classification ML models for the critical temperature of several hundred superconductors. The classification model separating the materials into two groups based on their critical temperature showed good predictive performance. Such models can be used to screen hundreds of thousands of existing and potential materials stored in computational databases.

In addition to directly predicting a property of interest (such as T_c) from existing DFT data, ML can be used to accelerate the intermediate computational steps. For example, in conventional superconductors, the coupling between the electrons and the

lattice vibrations creates an effective attractive electron–electron interaction leading to Cooper pair formation. However, an ab initio calculation of electron-phonon coupling strength requires knowledge of properties such as the phonon DOS, which are computationally expensive, especially for complex materials. In a recent publication, it was shown that a neural network model can be trained to create high-fidelity phonon DOS predictions. The authors utilized phonon DOS calculations of ~1500 crystalline solids as a training dataset⁵⁸. The predictions of the model were able to capture the main features of phonon DOS, even for crystalline solids with elements not present in the training set. Thus, by reducing the computational cost of ab initio methods, the use of ML can significantly enhance the throughput of computational materials screening.

In yet another example of incorporating AI into DFT for superconductivity, ML feature selection methods have been utilized to derive a simple analytical expression improving the Allen-Dynes approximation⁵⁹. The Allen-Dynes analytical formula is based on the Eliashberg-Migdal theory and is commonly used to predict the critical temperature of electron-phonon paired superconductors, reducing the number of required DFT calculations. The increase of the accuracy achieved in the new formula

can improve the computational pipeline, which in turn can lead to the discovery of novel superconductors faster.

In contrast to superconductors, the concept of materials with intrinsic topologically nontrivial states is relatively new^{2,3}. The current flurry of activities was initially spearheaded by a theoretical prediction⁶⁰ and the subsequent experimental confirmation of the existence of topological insulators^{61,62}—materials with insulating bulk but protected metallic surface states. Despite the significant efforts in the last decade, however, only a handful of distinct topological insulators, as well as other topological materials such as Dirac and Weyl semimetals^{4,63–65} are known, and given that many of them are difficult to synthesize and/or suffer from deleterious properties—such as the presence of bulk trivial states undermining the surface states, the community is always on the lookout for new topological materials.

To address this issue, several groups have developed algorithms such as symmetry indicators and spin-orbit spillage that use materials chemistry and symmetry in combination with electronic structures to calculate its topological properties, opening the door to automated screening for new topological insulators and semimetals⁶⁶⁻⁶⁸. Based on these ideas, several searches through ICSD generated extensive lists of candidates corresponding to various types of topological materials⁶⁹⁻⁷³. Such developments make the field a fertile ground for the application of ML methods (for an example see Fig. 2). A very recent work constructed a gradient boosting model (a particular type of powerful supervised learning algorithm) that can predict the topology of a given known material based only on "coarse-grained" chemical composition and crystal symmetry predictors with an accuracy of 90%⁷⁴. Similar to superconductivity, such models can be used to accelerate the search for novel materials by providing fast and efficient means to predict the possible topological nature of a given candidate.

Another class of quantum materials systems is 2D materials¹. Since the breakthrough discovery of graphene in 2004, many other 2D materials families (such as monolayers of hexagonal boron nitride, silicene, germanene, stanene, phosphorene, and transition metal dichalcogenides) have been synthesized¹. They have been the focus of a concerted investigation, both due to their rich potential in technologies such as electronics, sensing, and energy storage, and because they offer an entirely new avenue to

explore the interplay between particle-particle interactions, band structure, and constrained dimensionality (leading to effects such as the long-sought Wigner crystal⁷⁵). The search for new 2D materials and a systematic comparison of their properties is still in its infancy. High-throughput DFT has been used in the last several years to compile publicly available databases of potential 2D materials⁷⁶⁻⁷⁹. In ref. ⁸⁰, a JARVIS-DFT dataset containing results of 2D and 3D DFT calculations was used to develop handcrafted structural descriptors as inputs into gradient boosting models to predict properties including exfoliation energies, formation energies, and bandgaps. From the synthesis point of view, the exfoliation energy is particularly crucial for 2D materials design, yet its DFT is prohibitively expensive. These models were then used to discover exfoliable materials satisfying specific property requirements. In ref.⁸¹, the Computational 2D Materials Database (C2DB) was used to create an ML model (again based on gradient boosting) utilizing composition and structural predictors to classify potential 2D materials as having low, medium, or high stability. The model was used to discover potential 2D materials suitable for photoelectrocatalytic water splitting.

Beyond monolayers, there is a vast configuration space of possible van der Waals heterostructures, which can be formed by stacking different 2D layers. Such hybrid structures add additional degrees of freedom to their functionalities, creating 2D and 3D materials with untold tunability. Because DFT is too time-consuming to explore more than a small fraction of all possible combinations, incorporating ML in the process provides a practical alternative. Recent work has trained several ML models for predicting the interlayer distance and bandgap of bilayer heterostructures⁸². The models showed good accuracy on both tasks and were used to predict the properties of nearly 1500 hypothetical bilayer structures based on less than 300 DFT calculations. This underscores the promise of combining DFT and ML for rapid computational screening to identify new hybrid heterostructures with interesting and desirable properties.

Applying ML to experimental databases: making predictions from known materials

With the ability to quickly recognize patterns in large collections of quantitative information, applications of AI approaches in



Fig. 2 Example of Density Functional Theory used to predict topological properties of materials. Periodic table trend showing the percentage chance (color coded) that a compound containing any given element can be topologically nontrivial based on spin-orbit spillage approach, based on spillage data for 4835 nonmagnetic materials (reproduced from ref. ⁶⁸ under the terms of the Creative Commons CC BY license). The spillage is a machine-learnable quantity, and an accurate classification model predicting it (utilizing gradient boosting decision trees) was developed in ref. ⁷³.



Fig. 3 Examples of machine-learning methods applied to experimental data. a Three clusters (shown with blue, green, and red), representing groups of related materials, containing topologically nontrivial materials. The materials representations are extracted from a neural network model predicting the Bravais lattice from composition—for details, see ref. ⁸⁷. The clusters can be used to search for new topological materials. **b** The measured vs predicted $\ln(T_c)$ of various superconductors based on a random forest model presented in ref. ⁹¹. The same model can predict T_c of several distinct superconducting classes (blue markers: low- T_c materials; green markers: iron-based superconductors; red markers: cuprate superconductors).

experimental materials data are becoming just as prevalent. One natural task would be to apply them to databases amassing experimentally known compounds and to build models for making predictions. In the study of superconductors-a field with more than a century of extensive research history-there are large volumes of accumulated information. Sometimes even a small fraction of this data can be used to discover important trends. In an early pioneering work, Villars and Phillips demonstrated that using just three stoichiometric descriptors (associated with elemental makeups), the 60 known superconductors at the time (in the 80 s) with $T_c > 10$ K can all be clustered in three islands⁸³. Based on this observation the authors made predictions for potential high-temperature superconductors^{83,84}. In another early work, Hirsch used statistical methods to look for correlations between normal state properties and T_c of the metallic elements in the first six rows of the periodic table⁸⁵.

With time, more and more of the experimental data collected by generations of scientists are becoming easily accessible to researchers. Thanks to tedious curation efforts over decades by researchers at several institutes around the world, there are large databases with compiled experimental report entries such as the Phase Equilibria Diagram and the ICSD, arguably one of the largest compendiums of experimental materials data⁸⁶. Because of their comprehensive and exhaustive nature, such databases have become the de-facto standard-bearers for materials exploration and development, and they often serve as a starting point for building ML models (as already discussed above for use in conjunction with purely computational approaches). While these databases rarely contain functional properties, phase formation and stability as well as phase diagrams can be used as blueprints for navigating the materials exploration process. For example, a recent work used ICSD data to train a neural network to predict crystal structure information⁸⁷. The network's activations maps were then used to group materials according to their compositional and structural similarity, providing lists of materials potentially sharing properties with known superconductors or topological insulators (see Fig. 3a)

Unfortunately, when it comes to experimental databases of known quantum materials, usually there are very few entries in the datasets. In fact, one is often hard-pressed to find extensive databases of functional materials in general. A rare exception is the MatNavi database, managed by the National Institute of Materials Science (NIMS) in Japan. It is an experimental database encompassing well-curated information on materials with a variety of functional properties. Transcribing published journal results into formatted databases can be a massive undertaking, which NIMS materials data scientists have managed to do for decades. Some of the collected information is already being used for data-driven materials exploration^{88,89}. The Pauling File also contains up to tens of thousands of individual composition entries (and corresponding specific physical property quantities) diligently entered over decades⁹⁰.

This increase of the available experimental data, together with the surge of popularity of AI topics and the appearance of general-purpose ML libraries, led to a recent groundswell of activities introducing sophisticated ML methods in the study of superconductivity⁹¹⁻⁹⁶. In one notable example⁹¹, Stanev et al. considered more than 16,000 different compositions extracted from the MatNavi SuperCon database, which contains an exhaustive list of known superconductors as well as many "closely-related" materials varying only by small changes in stoichiometry. Compared to the early exercise by Villars and Philips in the 80 s, the orders-of-magnitude increase in the number of data points has led to the possibility to create a robust ML pipeline. The regression models developed to predict the values of T_c for different superconducting families used over one hundred stoichiometric descriptors, demonstrated strong predictive power and high accuracy, and offered valuable insights into the origins of superconductivity mechanisms in different materials groups (see Fig. 3b). The models also demonstrated an important limitation of ML in failing to extrapolate to materials families not included in the training set. A pipeline was then created to search for potential new superconductors among the roughly 110,000 different compositions contained in ICSD, which resulted in predictions of possible T_c 's above 20 K in 35 known compounds that had previously not been tested for superconductivity.

One interesting finding from this pipeline is that most of the newly identified possible superconductors possess a flat/nearlyflat band just below the Fermi energy, leading to an increase in the electronic DOS. Such enhanced DOS has long been considered a promising way to boost T_c . However, a very recent study using experimental T_c and DFT-based DOS calculations failed to discover a strong and consistent correlation between superconductivity and peaks in the electronic DOS⁹⁷. Yet, the strongest conclusion from the latter study was on "...the restrictions that the current availability and organization of materials data place on reliable machine-learning and data-based experimentation," underscoring the need for a more systematic approach to collecting and organizing quantum materials data.

This, however, is far easier said than done: the experimental exploration of quantum materials is a remarkably diverse and distributed endeavor, and it is not uncommon for tens or even hundreds of research groups to study the same material, with each group applying its unique toolbox and interrogation techniques. Also, many experiments are extremely resource-intensive and can only produce limited datasets. Combining numerous small datasets covering the same material can lead to issues with data consistency and incomplete metadata. Furthermore, the quantum materials community has yet to fully embrace the open data paradigm, and a significant fraction of the collected data are only published in unstructured form (i.e., text and images) and are not made easily available to other researchers. (Significant fraction of the collected information-especially "negative" results-is never published. For a rare example of work reporting both negative and positive results see ref. 98).

Thus, if one wants to construct an experimental database of quantum materials, the researcher may have to resort to the laborious manual extraction of data points from published articles. The significant effort required by this process at least partially explains the scarcity of databases in the field. However, emerging AI-driven automatic generation of databases can provide an alternative to the slow and tedious process. One example is the recently created database of almost 40,000 Curie and Néel phase-transition temperatures of magnetic materials⁹⁹. It is fully auto-generated and completely open-source. The database was produced using Natural Language Processing (NLP) and related ML methods, applied to the texts of published chemistry and physics articles. The method is quite precise, with the accuracy of the extracted transition temperatures reaching 82%. In a very recent extension of this work, the same authors were able to reconstruct the phase diagrams of wellknown magnetic and superconducting compounds, again using text data¹⁰⁰. They also demonstrated that it is possible to predict the phase-transition temperatures of compounds not present in the database. (A yet another recent effort to organize various experimental results into a curated database focuses on digitization of plots extracted from published papers¹⁰¹.)

Extracting latent knowledge from materials characterization data

In an ironic twist, in some areas of experimental materials physics, the issue is not the scarcity but the overabundance of data. In fact, negotiating the large quantities of data churned out in realtime by modern experimental labs is a significant emerging challenge. Continual improvements in materials characterization instrumentation coupled with the ever-increasing might of readily available computers for data-acquisition and memory storage have created the capacity to collect data on an unprecedented scale and with high speed. Through advances in scanning transmission electron microscopy, it is now possible to determine atomic column positions with picometer level precision (see, for example, refs. ^{102,103}). Spectroscopy and scanning probe measurements can provide detailed information about materials properties including electronic structure and symmetry of an order parameter^{104,105}. Pump-probe techniques permit the study of highly excited and out-of-equilibrium states and phases, sometimes revealing hidden tendencies¹⁰⁶.

With these techniques, even a single measurement of one material can generate large volumes of high-dimensional data necessitating the use of sophisticated statistical methods. In these instances, ML algorithms are implemented for uncovering physics working behind the scene, which affects the collective behavior of materials at an atomic scale. In one recent example, a neural network model was used to analyze Spectroscopic Imaging Scanning Tunneling Microscopy (SISTM) images of the CuO₂ planes, where transport takes place in high-temperature superconducting cuprates¹⁰⁷. The model was designed to recognize different symmetry-breaking ordered electronic states. It was trained on a set of artificial images, each generated to represent one of four distinct states, differing by their fundamental wavevector. Various forms of heterogeneity, intrinsic disorder, and topological defects were added to these images to mimic the experimental data. The model was then applied to experimental images from carrier-doped cuprates. Analyzing the noisy and complex data, the model was able to discover the existence of a translational-symmetry-breaking ordered state. The presence of this particular ordered state has important implications for theories aiming to explain the mysterious pseudogap phase of these materials.

In another notable example, researchers applied an ML method to the Angle-Resolved Photoemission Spectroscopy (ARPES) data of optimally doped and under-doped cuprates¹⁰⁸. The use of a restricted Boltzmann machine model allowed the recovery of a "hidden" feature in the spectra: prominent peak structures present both in the normal and anomalous self-energies of the singleparticle spectral function. These peaks structures cancel each other in the total self-energy and were only discovered by a model respecting physical constraints such as causality (encoded in the Kramers-Kronig relation). The use of ML allowed researchers to solve a non-linear underdetermined problem, and to obtain important information directly from experimental data, without the use of any theoretical model. The results clarified the role of the energy dissipation and quantum entanglement in the superconducting phase and provide a way to finally identify the boson responsible for the pairing in cuprates.

Beyond superconductivity, ML was recently applied to neutron scattering data of a frustrated magnet, exhibiting a complex phase diagram, and used to extract model Hamiltonians and to identify different magnetic regimes¹⁰⁹. An autoencoder—dimensionality reduction architecture based on a neural network—was trained to create a compressed representation of three-dimensional diffuse scattering, over a wide range of spin Hamiltonians. The autoencoder was able to find optimal Hamiltonian parameters matching observed scattering and heat capacity signatures. The autoencoder was also able to categorize different magnetic behaviors and eliminate background noise and artifacts in raw data. Thus, ML can augment many traditional diffraction and inelastic neutron scattering data analysis tools, which are often time-consuming and error-prone.

Another very recent work reported a convolutional-neuralnetwork-based classifier designed to distinguish topological from trivial materials and trained on X-ray Absorption Spectroscopy (XAS) data¹¹⁰. The model showed high accuracy in distinguishing the different classes of materials, demonstrating the potential of ML methods in recognizing topological character embedded in complex spectral features. XAS is a widely used materials characterization technique, and the ability to decipher the topological character of material from XAS signatures can substantially simplify and expedite the experimental identification of topologically nontrivial materials.



Fig. 4 Accelerating experimental discovery processes. a High-throughput methods can be used to perform validation of predicted materials. Visualization of resistivity vs temperature curves measured at different parts of a Fe-B composition spread. The horizontal range of each curve covers 2–300 K. The resistivity range for each curve is color coded (reproduced from ref. ¹⁴¹ under the terms of the Creative Commons CC BY license); **b** On-the-fly analysis of synchrotron diffraction data¹²⁴. A snapshot photo from the experiment. The upper screen shows the scanning stage with a thin-film library inside a beamline hutch. The bottom screen is a laptop, where unsupervised ML of diffraction patterns is carried out after each measurement. The same setup has also been used to carry out active-learning-based autonomous experiments.

Autonomous materials laboratories

Armed with predictions of potential quantum materials generated by theory or the ML techniques discussed above, experiments can be designed for systematic synthesis and characterization of novel compounds. However, even with the predictions providing blueprints, one often has to negotiate potentially large multidimensional parameter space as in any materials exploration and optimization task. One proven platform to effectively host such experiments is the high-throughput approach, which permits the interrogation of libraries that contain hundreds or even thousands of different compounds¹¹¹. Over the years combinatorial approaches have been effectively used to explore new superconductors¹¹² and map their compositional phase diagrams (for an example demonstrating the ability of this method to discover superconducting compositions see Fig. 4a). It has also been used to optimize the synthesis of a topological Kondo insulator SmB_6 in a thin-film form¹¹³.

A significant hurdle in the more widespread use of highthroughput methods is the need for special characterization tools, often required for the exploration of many classes of quantum materials. For instance, for detecting topological states, one typically turns to angle-resolved photoemission spectroscopy (ARPES) to look for their signature in electronic structure¹¹⁴. Antiferromagnetism—whose role in the superconductivity of several classes of materials is hotly debated—is best probed using neutron scattering¹¹⁵. X-ray magnetic circular dichroism has been used to investigate the electronic structure of quantum spin liquid α -RuCl₃¹¹⁶ These advanced techniques require unique synchrotron or neutron facilities, and their implementation in the screening of combinatorial libraries has been limited to date^{117,118}

The high-throughput approach generates large highdimensional datasets, requiring analysis techniques that can rapidly turn raw data into knowledge with limited or no human supervision¹¹⁹. This community adopted dimensionality reduction and data mining techniques early on^{120–123}, gradually creating a diverse ML toolbox for the rapid digestion of combinatorial data^{124–130}. One very common task is to quickly group (cluster) measurements from different points of a combinatorial library, and ML algorithms have been routinely applied to a large number of X-ray diffraction patterns to delineate structural phases and rapidly construct a composition-structure relationship^{120,124,129,130}. The authors of ref. ¹²⁴ took this idea a step further and used a comprehensive ML algorithm for on-the-fly analysis of synchrotron diffraction data from combinatorial libraries to facilitate a search for rare-earth-free permanent magnets (see Fig. 4b). Unsupervised ML for rapid data reduction is now commonly being applied to high-throughput characterization data from a variety of spectroscopic techniques, including Raman spectroscopy¹³¹, Time-of-Flight Secondary Ion Mass Spectrometry¹³², and X-ray photoelectron spectroscopy¹³³.

This close integration of experimental tools and ML can be considered as the first step towards an emerging AI-driven paradigm for materials exploration. It relies on active learning-a branch of AI dedicated to providing a systematic and rigorous approach for identifying the best experiment to perform to achieve an objective. This can be either finding the shortest path toward a material that optimizes some desired properties or identifying a series of experiments that maximizes knowledge of the explored space. Recently, materials physicists have begun to capitalize on active learning to accelerate experimental research¹³⁴. For example, active learning has been used to advise experimentalists on the next best experiment to perform in the search for various functional materials^{135–137}. Another work already demonstrated the potential of these methods in the field of quantum materials: an active-learning framework designed to discover the material with the highest T_c was evaluated on about 600 known superconductors¹³⁸. The framework did significantly better than pure random guessing, highlighting the impact it can have.

The active-learning algorithms guiding the exploration can indeed decisively outperform exhaustive experimental approaches, leading to a similar amount of actionable knowledge for only a fraction of the time and resources. This has profound implications: with a significant reduction in the number of iterative experimental runs, it may now be possible to incorporate



Fig. 5 Schematic for the autonomous experimental systems of the future. The AI in charge of the autonomous lab will have full control over the synthesis, characterization, and modeling tools. It will extract prior knowledge from structured databases, publications, and human guidance, do experiments and computations and be able to distill its findings in a succinct mathematical form. Such autonomous laboratories are already being implemented in other areas of materials science¹⁴². (Icons used in the figure are from INCORS).

active-learning-guided tools at synchrotron and neutron facilities for quick screening of quantum materials libraries in closed-loop cycles.

However, the examples of active learning discussed above only provided support for an important but limited set of questions, leaving the majority of the experimental design, execution, and analysis to human experts. This situation is rapidly changing, and truly autonomous experimental systems are beginning to arrive on the scene. In one study, an autonomous system was used to control synthesis parameters to optimize carbon nanotube growth rate¹³⁹. The Autonomous Research System (ARES) platform utilized an ML-based system linked to an automated growth reactor with in situ characterization to learn the optimum growth conditions.

Autonomous materials exploration can be particularly effective when coupled with high-throughput experimentation. In a recent study, a real-time closed-loop, autonomous system for materials exploration and optimization (CAMEO) was demonstrated on a combinatorial library.¹⁴⁰ The goal of CAMEO is to first map the structural phase diagram across the compositional landscape and then use the phase map as a blueprint to optimize a physical property of interest. It was used to quickly find the optimum composition of phase change memory materials with the largest bandgap difference between amorphous and crystalline states within a Ge-Sb-Te thin-film composition spread. To achieve this, analysis of ellipsometry measurements from the library and remotely-controlled synchrotron diffraction were carried out simultaneously. Combining these two sources of information, CAMEO discovered a novel composition with phase change performance superior to known materials in the field, for only about one-tenth of the time required to measure every point in the library.

Even more revolutionary AI-enabled systems, designed to help human scientists in optimizing every step of the research process, are already being considered and developed. These systems will coordinate a host of computational and experimental probes to help efficiently search a vast and mostly unexplored compositional space. The AI "brain" of the system will have access to all relevant information—results of past experiments, computations and theory, knowledge of a wide range of materials synthesis techniques, and measurement instrumentation. It will use this information to advise researchers which experiments to perform to maximize the knowledge gain and drastically accelerate progress. In its ultimate form, fully autonomous laboratories controlled by AI in command of various synthesis and characterization will orchestrate entire experimental campaigns, update their knowledge, and continue to explore until the desired goal is reached (see Fig. 5 for a schematic). These AI "scientists", dedicated to the exploration of quantum materials, can help us finally solve some of the most enduring mysteries of physics.

Outlook

As physicists focus more and more on materials with unusual properties grounded in many-body quantum mechanics, their research methods have to evolve. The ongoing revolution in AI is a great opportunity for the condensed matter physics community, and ML tools are starting to play an important role in the study of quantum materials. From accelerating the first-principles computational tools to helping analyze high-dimensional experimental data, these tools have already left their mark on the field. But even more exciting developments are gradually turning into reality. Researchers are working on automated research systems controlled by AI-guided robots. These systems will allow scientists to build penetrating multidimensional pictures of these complex materials, and at the same time accelerate the search and discovery process.

Apart from reshaping the experimental process, the use of such autonomous systems will reduce the many superfluous hurdles and dramatically lower the effort and time needed for running an experiment. This will allow scientists to focus on the most challenging and important parts of the research process, while also making science more "democratic" and equitable.

Received: 5 July 2021; Accepted: 12 September 2021; Published online: 13 October 2021

References

- Novoselov, K. S., Mishchenko, A., Carvalho, A. & Castro Neto, A. H. 2D materials and van der Waals heterostructures. *Science* 353, aac9439 (2016).
- Hasan, M. Z. & Kane, C. L. Colloquium: topological insulators. *Rev. Mod. Phys.* 82, 3045–3067 (2010).
- Qi, X.-L. & Zhang, S.-C. Topological insulators and superconductors. *Rev.* Mod. Phys. 83, 1057–1110 (2011).

- 4. Armitage, N. P., Mele, E. J. & Vishwanath, A. Weyl and Dirac semimetals in three-dimensional solids. *Rev. Mod. Phys.* **90**, 015001 (2018).
- Chu, C. W., Deng, L. Z. & Lv, B. Hole-doped cuprate high temperature superconductors. *Phys. C Supercond. Appl.* 514, 290–313 (2015).
- Armitage, N. P., Fournier, P. & Greene, R. L. Progress and perspectives on electron-doped cuprates. *Rev. Mod. Phys.* 82, 2421–2487 (2010).
- Paglione, J. & Greene, R. L. High-temperature superconductivity in iron-based materials. Nat. Phys. 6, 645 (2010).
- Savary, L. & Balents, L. Quantum spin liquids: a review. Rep. Prog. Phys. 80, 016502 (2016).
- 9. The rise of quantum materials. Nat. Phys. 12, 105 (2016).
- 10. Mühlbauer, S. et al. Skyrmion lattice in a chiral magnet. *Science* **323**, 915 (2009).
- 11. Ponomarenko, L. A. et al. Chaotic dirac billiard in graphene quantum dots. *Science* **320**, 356 (2008).
- 12. Chang, C.-Z. et al. Experimental observation of the quantum anomalous Hall effect in a magnetic topological insulator. *Science* **340**, 167 (2013).
- Kou, X. et al. Scale-invariant quantum anomalous Hall effect in magnetic topological insulators beyond the two-dimensional limit. *Phys. Rev. Lett.* 113, 137201 (2014).
- Checkelsky, J. G. et al. Trajectory of the anomalous Hall effect towards the quantized state in a ferromagnetic topological insulator. *Nat. Phys.* 10, 731–736 (2014).
- Fan, Y. et al. Magnetization switching through giant spin-orbit torque in a magnetically doped topological insulator heterostructure. *Nat. Mater.* 13, 699 (2014).
- Nayak, C., Simon, S. H., Stern, A., Freedman, M. & Das Sarma, S. Non-Abelian anyons and topological quantum computation. *Rev. Mod. Phys.* 80, 1083–1159 (2008).
- 17. Zhang, H. et al. Quantized Majorana conductance. Nature 556, 74 (2018).
- 18. Lee, S. et al. Perfect Andreev reflection due to the Klein paradox in a topological superconducting state. *Nature* **570**, 344–348 (2019).
- Proust, C. & Taillefer, L. The remarkable underlying ground states of cuprate superconductors. Annu. Rev. Condens. Matter Phys. 10, 409–429 (2019).
- 20. Marx, V. The big challenges of big data. Nature 498, 255 (2013).
- Kremer, J., Stensbo-Smidt, K., Gieseke, F., Pedersen, K. S. & Igel, C. Big universe, big data: machine learning and image analysis for astronomy. *IEEE Intell. Syst.* 32, 16–22 (2017).
- Jain, A., Hautier, G., Ong, S. P. & Persson, K. New opportunities for materials informatics: resources and data mining techniques for uncovering hidden relationships. *J. Mater. Res.* 31, 977–994 (2016).
- Agrawal, A. & Choudhary, A. Perspective: Materials informatics and big data: realization of the "fourth paradigm" of science in materials science. *APL Mater.* 4, 053208 (2016).
- Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* 3, 54 (2017).
- Mueller, T., Kusne, A. G. & Ramprasad, R. Machine learning in materials science. Reviews in Computational Chemistry. Vol. 29, ch. 4, pp. 186–273 (2016).
- Carrasquilla, J. & Melko, R. G. Machine learning phases of matter. Nat. Phys. 13, 431 (2017).
- Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* 559, 547–555 (2018).
- Rickman, J. M., Lookman, T. & Kalinin, S. V. Materials informatics: from the atomic-level to the continuum. *Acta Mater.* 168, 473–510 (2019).
- 29. Cao, Y. et al. Unconventional superconductivity in magic-angle graphene superlattices. *Nature* **556**, 43–50 (2018).
- McKinney, S. M. et al. International evaluation of an AI system for breast cancer screening. *Nature* 577, 89–94 (2020).
- 31. Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
- 32. Ho, D. Artificial intelligence in cancer therapy. Science 367, 982 (2020).
- Radovic, A. et al. Machine learning at the energy and intensity frontiers of particle physics. *Nature* 560, 41–48 (2018).
- 34. Curtarolo, S. et al. AFLOW: An automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* 58, 218–226 (2012).
- Jain, A. et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. APL Mater. 1, 011002 (2013).
- Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput Density Functional Theory: the open quantum materials database (OQMD). *JOM* 65, 1501–1509 (2013).
- 37. Draxl, C. & Scheffler, M. The NOMAD laboratory: from data sharing to artificial intelligence. J. Phys. Mater. 2, 036001 (2019).
- Choudhary, K. et al. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj Comput. Mater.* 6, 173 (2020).
- Zhang, Y. et al. Unsupervised discovery of solid-state lithium ion conductors. Nat. Commun. 10, 5260 (2019).

- Biswas, A. et al. Designed materials with the giant magnetocaloric effect near room temperature. Acta Mater. 180, 341–348 (2019).
- 41. Kim, J. R. et al. Stabilizing hidden room-temperature ferroelectricity via a metastable atomic distortion pattern. *Nat. Commun.* **11**, 4944 (2020).
- Snyder, J. C., Rupp, M., Hansen, K., Müller, K.-R. & Burke, K. Finding density functionals with machine learning. *Phys. Rev. Lett.* 108, 253002 (2012).
- Wellendorff, J. et al. Density functionals for surface science: exchangecorrelation model development with Bayesian error estimation. *Phys. Rev. B* 85, 235149 (2012).
- Suzuki, Y., Nagai, R. & Haruyama, J. Machine learning exchange-correlation potential in time-dependent density-functional theory. *Phys. Rev. A* 101, 050501 (2020).
- Yu, M., Yang, S., Wu, C. & Marom, N. Machine learning the Hubbard U parameter in DFT+U using Bayesian optimization. *npj Comput. Mater.* 6, 180 (2020).
- Liu, J., Qi, Y., Meng, Z. Y. & Fu, L. Self-learning Monte Carlo method. *Phys. Rev. B* 95, 041101 (2017).
- Huang, L. & Wang, L. Accelerated Monte Carlo simulations with restricted Boltzmann machines. *Phys. Rev. B* 95, 035105 (2017).
- Arsenault, L.-F., Lopez-Bezanilla, A., von Lilienfeld, O. A. & Millis, A. J. Machine learning for many-body physics: the case of the Anderson impurity model. *Phys. Rev. B* **90**, 155136 (2014).
- Seko, A., Maekawa, T., Tsuda, K. & Tanaka, I. Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids. *Phys. Rev. B* 89, 054303 (2014).
- Pilania, G. et al. Machine learning bandgaps of double perovskites. *Sci. Rep.* 6, 19375 (2016).
- 51. Isayev, O. et al. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **8**, 15679 (2017).
- Powell, B. J. & McKenzie, R. H. Quantum frustration in organic Mott insulators: from spin liquids to unconventional superconductors. *Rep. Prog. Phys.* 74, 056501 (2011).
- Pavarini, E., Dasgupta, I., Saha-Dasgupta, T., Jepsen, O. & Andersen, O. K. Band-structure trend in hole-doped cuprates and correlation with T_{cmax}. Phys. Rev. Lett. 87, 047003 (2001).
- Klintenberg, M. & Eriksson, O. Possible high-temperature superconductors predicted from electronic structure and data-filtering algorithms. *Comput. Mater. Sci.* 67, 282–286 (2013).
- Geilhufe, R. M., Borysov, S. S., Kalpakchi, D. & Balatsky, A. V. Towards novel organic high-T_c superconductors: data mining using density of states similarity search. *Phys. Rev. Mater.* 2, 024802 (2018).
- Borysov, S. S., Geilhufe, R. M. & Balatsky, A. V. Organic materials database: an open-access online database for data mining. *PLoS ONE* 12, e0171501 (2017).
- Isayev, O. et al. Materials cartography: representing and mining materials space using structural and electronic fingerprints. *Chem. Mater.* 27, 735–743 (2015).
- Chen, Z. et al. Direct prediction of phonon density of states with Euclidean neural networks. Adv. Sci. 8, 202004214 (2021).
- Xie, S. R., Stewart, G. R., Hamlin, J. J., Hirschfeld, P. J. & Hennig, R. G. Functional form of the superconducting critical temperature from machine learning. *Phys. Rev. B* 100, 174513 (2019).
- Bernevig, B. A., Hughes, T. L. & Zhang, S.-C. Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science* 314, 1757 (2006).
- 61. König, M. et al. Quantum spin Hall insulator state in HgTe quantum wells. *Science* **318**, 766 (2007).
- Hsieh, D. et al. A topological Dirac insulator in a quantum spin Hall phase. Nature 452, 970 (2008).
- Wan, X., Turner, A. M., Vishwanath, A. & Savrasov, S. Y. Topological semimetal and Fermi-arc surface states in the electronic structure of pyrochlore iridates. *Phys. Rev. B* 83, 205101 (2011).
- 64. Wang, Z. et al. Dirac semimetal and topological phase transitions in A_3 Bi (A = Na, K, Rb). *Phys. Rev. B* **85**, 195320 (2012).
- Young, S. M. et al. Dirac semimetal in three dimensions. *Phys. Rev. Lett.* 108, 140405 (2012).
- 66. Bradlyn, B. et al. Topological quantum chemistry. Nature 547, 298 (2017).
- Po, H. C., Vishwanath, A. & Watanabe, H. Symmetry-based indicators of band topology in the 230 space groups. *Nat. Commun.* 8, 50 (2017).
- Choudhary, K., Garrity, K. F. & Tavazza, F. High-throughput discovery of topologically non-trivial materials using spin-orbit spillage. *Sci. Rep.* 9, 8534 (2019).
- Tang, F., Po, H. C., Vishwanath, A. & Wan, X. Comprehensive search for topological materials using symmetry indicators. *Nature* 566, 486–489 (2019).
- Vergniory, M. G. et al. A complete catalogue of high-quality topological materials. *Nature* 566, 480–485 (2019).
- Zhang, T. et al. Catalogue of topological electronic materials. *Nature* 566, 475–479 (2019).

- 72. Frey, N. C. et al. High-throughput search for magnetic and topological order in transition metal oxides. *Sci. Adv.* **6**, eabd1076 (2020).
- Choudhary, K., Garrity, K. F., Ghimire, N. J., Anand, N. & Tavazza, F. Highthroughput search for magnetic topological materials using spin-orbit spillage, machine-learning and experiments. *Phys. Rev. B* 103, 155131 (2021).
- 74. Claussen, N., Bernevig, B. A. & Regnault, N. Detection of topological materials with machine learning. *Phys. Rev. B* **101**, 245117 (2020).
- 75. Zhou, Y. et al. Bilayer Wigner crystals in a transition metal dichalcogenide heterostructure. *Nature* **595**, 48–52 (2021).
- Choudhary, K., Kalish, I., Beams, R. & Tavazza, F. High-throughput identification and characterization of two-dimensional materials using Density Functional Theory. *Sci. Rep.* 7, 5179 (2017).
- Mounet, N. et al. Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. *Nat. Nanotechnol.* 13, 246–252 (2018).
- Haastrup, S. et al. The Computational 2D Materials Database: highthroughput modeling and discovery of atomically thin crystals. 2D Mater. 5, 042002 (2018).
- Zhou, J. et al. 2DMatPedia, an open computational database of twodimensional materials from top-down and bottom-up approaches. *Sci. Data* 6, 86 (2019).
- Choudhary, K., DeCost, B. & Tavazza, F. Machine learning with force-fieldinspired descriptors for materials: fast screening and mapping energy landscape. *Phys. Rev. Mater.* 2, 083801 (2018).
- Schleder, G. R., Acosta, C. M. & Fazzio, A. Exploring two-dimensional materials thermodynamic stability via machine learning. ACS Appl. Mater. Interfaces 12, 20149–20157 (2020).
- Tawfik, S. A. et al. Efficient prediction of structural and electronic properties of hybrid 2D materials using complementary DFT and machine learning approaches. *Adv. Theory Simul.* 2, 1800128 (2019).
- Villars, P. & Phillips, J. C. Quantum structural diagrams and high-T_c superconductivity. *Phys. Rev. B* 37, 2345–2348 (1988).
- Rabe, K. M., Phillips, J. C., Villars, P. & Brown, I. D. Global multinary structural chemistry of stable quasicrystals, high-T_c ferroelectrics, and high-T_c superconductors. *Phys. Rev. B* 45, 7650–7676 (1992).
- Hirsch, J. E. Correlations between normal-state properties and superconductivity. *Phys. Rev. B* 55, 9007–9024 (1997).
- Hellenbrandt, M. The inorganic crystal structure database (ICSD)—present and future. Crystallogr. Rev. 10, 17–22 (2004).
- Liang, H., Stanev, V., Gilad Kusne, A. & Takeuchi, I. CRYSPNet: Crystal structure predictions via neural network. *Phys. Rev. Mater.* 4, 123802 (2020).
- Castro, P. B. D. et al. Machine-learning-guided discovery of the gigantic magnetocaloric effect in HoB2 near the hydrogen liquefaction temperature. *NPG Asia Mater.* 12, 35 (2020).
- Matsumoto, R. et al. Crystal growth, structural analysis, and pressure-induced superconductivity in a AgIn5Se8 single crystal explored by a data-driven approach. *Inorg. Chem.* 59, 325–331 (2020).
- Villars, P., Cenzual, K., Gladyshevskii, R. & Iwata, S. Pauling File: toward a holistic view. In *Materials Informatics* (eds Isayev, O., Tropsha, A. & Curtarolo, S.) https://doi.org/10.1002/9783527802265.ch3 (2019).
- 91. Stanev, V. et al Machine learning modeling of superconducting critical temperature. *npj Comput. Mater.* **4**, 1–14 (2018).
- 92. Hamidieh, K. A data-driven statistical model for predicting the critical temperature of a superconductor. *Comput. Mater. Sci.* **154**, 346–354 (2018).
- 93. Zeng, S. et al. Atom table convolutional neural networks for an accurate prediction of compounds properties. *npj Comput. Mater.* **5**, 84 (2019).
- Matsumoto, K. & Horide, T. An acceleration search method of higher *Tc* superconductors by a machine learning algorithm. *Appl. Phys. Express* 12, 073003 (2019).
- Liu, Z.-L., Kang, P., Zhu, Y., Liu, L. & Guo, H. Material informatics for layered high-T_C superconductors. *APL Mater.* 8, 061104 (2020).
- Goodall, R. E. A., Zhu, B., MacManus-Driscoll, J. L. & Lee, A. A. Materials informatics reveals unexplored structure space in cuprate superconductors. *Adv. Funct. Mater.* 2104696 (2021).
- Yang, Z. S., Ferrenti, A. M. & Cava, R. J. Testing whether flat bands in the calculated electronic density of states are good predictors of superconducting materials. J. Phys. Chem. Solids 151, 109912 (2021).
- Hosono, H. et al. Exploration of new superconductors and functional materials, and fabrication of superconducting tapes and wires of iron pnictides. *Sci. Technol. Adv. Mater.* 16, 033503 (2015).
- Court, C. J. & Cole, J. M. Auto-generated materials database of Curie and Néel temperatures via semi-supervised relationship extraction. *Sci. Data* 5, 180111 (2018).
- 100. Court, C. J. & Cole, J. M. Magnetic and superconducting phase diagrams and transition temperatures predicted using text mining and machine learning. *npj Comput. Mater.* 6, 18 (2020).
- 101. Katsura, Y. et al. Data-driven analysis of electron relaxation times in PbTetype thermoelectric materials. Sci. Technol. Adv. Mater. 20, 511–520 (2019).

- 102. Muller, D. A. Structure and bonding at the atomic scale by scanning transmission electron microscopy. *Nat. Mater.* **8**, 263 (2009).
- Nelson, C. T. et al. Exploring physics of ferroelectric domain walls via Bayesian analysis of atomically resolved STEM data. *Nat. Commun.* 11, 6361 (2020).
- 104. Lu, D. et al. Angle-resolved photoemission studies of quantum materials. Annu. Rev. Condens. Matter Phys. 3, 129–167 (2012).
- 105. Kumigashira, H. et al. In situ photoemission spectroscopic study on La1-xSrxMnO3 thin films grown by combinatorial laser-MBE. J. Electr. Spectr. Relat. Phenom. 136, 31-36 (2004).
- Fausti, D. et al. Light-induced superconductivity in a stripe-ordered cuprate. Science 331, 189 (2011).
- Zhang, Y. et al. Machine learning in electronic-quantum-matter imaging experiments. *Nature* 570, 484–490 (2019).
- 108. Yamaji, Y., Yoshida, T., Fujimori, A. & Imada, M. Hidden self-energies as origin of cuprate superconductivity revealed by machine learning. *arXiv e-prints* **2019**, arXiv:1903.08060.
- 109. Samarakoon, A. M. et al. Machine-learning-assisted insight into spin ice Dy2Ti2O7. Nat. Commun. 11, 892 (2020).
- 110. Andrejevic, N., Andrejevic, J., Rycroft, C. H. & Li, M. Machine learning spectral indicators of topology. *arXiv e-prints* **2020**, arXiv:2003.00994.
- Koinuma, H. & Takeuchi, I. Combinatorial solid-state chemistry of inorganic materials. Nat. Mater. 3, 429 (2004).
- Yuan, J., Stanev, V., Gao, C., Takeuchi, I. & Jin, K. Recent advances in highthroughput superconductivity research. *Supercond. Sci. Technol.* 32, 123001 (2019).
- 113. Yong, J. et al. Robust topological surface state in Kondo insulator SmB6 thin films. *Appl. Phys. Lett.* **105**, 222403 (2014).
- Chen, Y. L. et al. Experimental realization of a three-dimensional topological insulator, Bi2Te3. *Science* 325, 178 (2009).
- 115. Qiu, Y. et al. Spin gap and resonance at the nesting wave vector in superconducting FeSe_{0.4} Te_{0.6}. *Phys. Rev. Lett.* **103**, 067008 (2009).
- 116. Agrestini, S. et al. Electronically highly cubic conditions for Ru in α-RuCl₃. *Phys. Rev. B* 96, 161107 (2017).
- 117. Horiba, K. et al. A high-resolution synchrotron-radiation angle-resolved photoemission spectrometer with in situ oxide thin film growth capability. *Rev. Sci. Instrum.* 74, 3406–3412 (2003).
- 118. Snow, R. J., Bhatkar, H., N'Diaye, A. T., Arenholz, E. & Idzerda, Y. U. Large moments in bcc FexCoyMnz ternary alloy thin films. *Appl. Phys. Lett.* **112**, 072403 (2018).
- 119. Hattrick-Simpers, J. R., Gregoire, J. M. & Kusne, A. G. Perspective: Composition-structure-property mapping in high-throughput experiments: turning data into knowledge. *APL Mater.* 4, 053211 (2016).
- Long, C. J. et al. Rapid structural mapping of ternary metallic alloy systems using the combinatorial approach and cluster analysis. *Rev. Sci. Instrum.* 78, 072217 (2007).
- 121. Baumes, L. A., Moliner, M., Nicoloyannis, N. & Corma, A. A reliable methodology for high throughput identification of a mixture of crystallographic phases from powder X-ray diffraction data. *CrystEngComm* 10, 1321–1324 (2008).
- 122. Long, C. J., Bunker, D., Li, X., Karen, V. L. & Takeuchi, I. Rapid identification of structural phases in combinatorial thin-film libraries using x-ray diffraction and non-negative matrix factorization. *Rev. Sci. Instrum.* 80, 103902 (2009).
- 123. LeBras, R. et al. in Constraint Reasoning and Kernel Clustering for Pattern Decomposition with Scaling, Principles and Practice of Constraint Programming, (ed. Lee, J.) 508–522 (Springer Berlin Heidelberg, 2011).
- Kusne, A. G. et al. On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets. *Sci. Rep.* 4, 6367 (2014).
- 125. Ermon, S. et al Pattern decomposition with complex combinatorial constraints: application to materials discovery. in *Proc. Twenty-Ninth AAAI Conference on Artificial Intelligence*, 636–643 (AAAI Press, 2015).
- 126. Kusne, A. G., Keller, D., Anderson, A., Zaban, A. & Takeuchi, I. Highthroughput determination of structural phase diagram and constituent phases using GRENDEL. *Nanotechnology* 26, 444002 (2015).
- Bunn, J. K., Hu, J. & Hattrick-Simpers, J. R. Semi-supervised approach to phase identification from combinatorial sample diffraction patterns. *JOM* 68, 2116–2125 (2016).
- Suram, S. K. et al. Automated phase mapping with AgileFD and its application to light absorber discovery in the V–Mn–Nb oxide system. ACS Combin. Sci. 19, 37–46 (2017).
- 129. Iwasaki, Y., Kusne, A. G. & Takeuchi, I. Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries. *npj Comput. Mater.* 3, 4 (2017).
- 130. Stanev, V. et al. Unsupervised phase mapping of X-ray diffraction data by nonnegative matrix factorization integrated with custom clustering. *npj Comput. Mater.* 4, 43 (2018).

- 131. Kan, D., Long, C. J., Steinmetz, C., Lofland, S. E. & Takeuchi, I. Combinatorial search of structural transitions: Systematic investigation of morphotropic phase boundaries in chemically substituted BiFeO3. J. Mater. Res. 27, 2691–2704 (2012).
- 132. Dell'Anna, R. et al. Data analysis in combinatorial experiments: applying supervised principal component technique to investigate the relationship between ToF-SIMS spectra and the composition distribution of ternary metallic alloy thin films. *QSAR Combin. Sci.* **27**, 171–178 (2008).
- 133. Chincarini, A., Gemme, G., Parodi, R. & Antoine, C. Statistical approach to XPS analysis: application to niobium surface treatment (SRF2001: Proceedings of the 10th workshop on RF superconductivity (KEK-PROC--2003-2), 2001, Tsukuba, Japan), 382–386 (High Energy Accelerator Research Organization, Tsukuba, Ibaraki (Japan), 2003).
- 134. Lookman, T., Balachandran, P. V., Xue, D. & Yuan, R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Comput. Mater.* 5, 21 (2019).
- 135. Xue, D. et al. Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* 7, 11241 (2016).
- 136. Yuan, R. et al. Accelerated discovery of large electrostrains in BaTiO3-based piezoelectrics using active learning. Adv. Mater. 30, 1702884 (2018).
- 137. Balachandran, P. V., Kowalski, B., Sehirlioglu, A. & Lookman, T. Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning. *Nat. Commun.* 9, 1668 (2018).
- Meredig, B. et al. Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery. *Mol. Syst. Des. Eng.* 3, 819–825 (2018).
- 139. Nikolaev, P. et al. Autonomy in materials research: a case study in carbon nanotube growth. *npj Comput. Mater.* **2**, 16031 (2016).
- 140. Kusne, A. G. et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **11**, 5966 (2020).
- 141. Jin, K. et al. Combinatorial search of superconductivity in Fe-B composition spreads. APL Mater. 1, 042101 (2013).
- 142. MacLeod, B. P. et al. Self-driving laboratory for accelerated discovery of thinfilm materials. *Sci. Adv.* **6**, eaaz8867 (2020).

Acknowledgements

We acknowledge James Warren for many useful discussions. Some of the ideas in this review were borne out of discussions we had at the 2018 Workshop on Machine Learning Quantum Materials, supported by the Gordon and Betty Moore Foundation's EPiQS Initiative and the Maryland Quantum Materials Center. This work was also supported by

ONR N00014-13-1-0635, ONR N00014-15-1-2222, AFOSR no. FA9550-14-10332, NIST through 60NANB19D027 and 70NANB19H117, and the Gordon and Betty Moore Foundation's EPiQS Initiative through Grant No. GBMF9071.

Author contributions

All authors (V.S., K.C., A.K., J.P. and I.T.) contributed to the design of the review. All authors wrote the text of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Valentin Stanev.

Peer review information *Communications Materials* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: John Plummer.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/ licenses/by/4.0/.

© The Author(s) 2021