



### Cite as

Nano-Micro Lett.

(2026) 18:109

## Artificial Intelligence Empowered New Materials: Discovery, Synthesis, Prediction to Validation

Received: 5 June 2025

Accepted: 9 September 2025

© The Author(s) 2026

### HIGHLIGHTS

- A comprehensive review focused on the recent advancement of artificial intelligence (AI) powered materials research from various aspects, including material discovery, synthesis, prediction and validation, is presented.
- The design strategies for the enhanced performance of AI for materials can be implemented from various procedures for cognizance of existing materials and discovery of novel materials with the data processing, algorithm design and automated laboratory construction included.
- A broad outlook on the future considerations of the AI systems for material is proposed.

**ABSTRACT** Recent years have witnessed the significant breakthrough in the field of new materials discovery brought about by the artificial intelligence (AI). AI has successfully been applied for predicting the formability, revealing the properties, and guiding the experimental synthesis of materials. Rapid progress has been made in the integration of increasing database and improved computing power. Though some reviews present the development from their unique aspects, reviews from the view of how AI empowered both discovery of new materials and cognition of existing materials that covers the completed contents with two synergistical aspects are few. Here, the newest development is systematically reviewed in the field of AI empowered materials, reflecting advanced design of the intelligent systems for discovery, synthesis, prediction and validation of materials. First, background and mechanisms are briefed, after which the design for the AI systems with data, machine learning and automated laboratory included is illustrated. Next, strategies are summarized to obtain the AI systems for materials with improved performance which comprehensively cover the aspects from the in-depth cognizance of existing material and the rapid discovery of new materials, and then, the design thought for future AI systems in material science is pointed out. Finally, some perspectives are put forward.

**KEYWORDS** Artificial intelligence; Material discovery and cognition; Design tactics; Review and perspective

✉ Hong Fu, hfu@edu.hk; Bingang Xu, ttxubg@polyu.edu.hk

<sup>1</sup> Nanotechnology Center, School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong 999077, People's Republic of China

<sup>2</sup> Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong 999077, People's Republic of China

<sup>3</sup> State Key Laboratory for Powder Metallurgy, Central South University, Changsha 410083, People's Republic of China

<sup>4</sup> School of Chemical Engineering and Technology, Sun Yat-Sen University, Zhuhai 519082, People's Republic of China

<sup>5</sup> Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, People's Republic of China

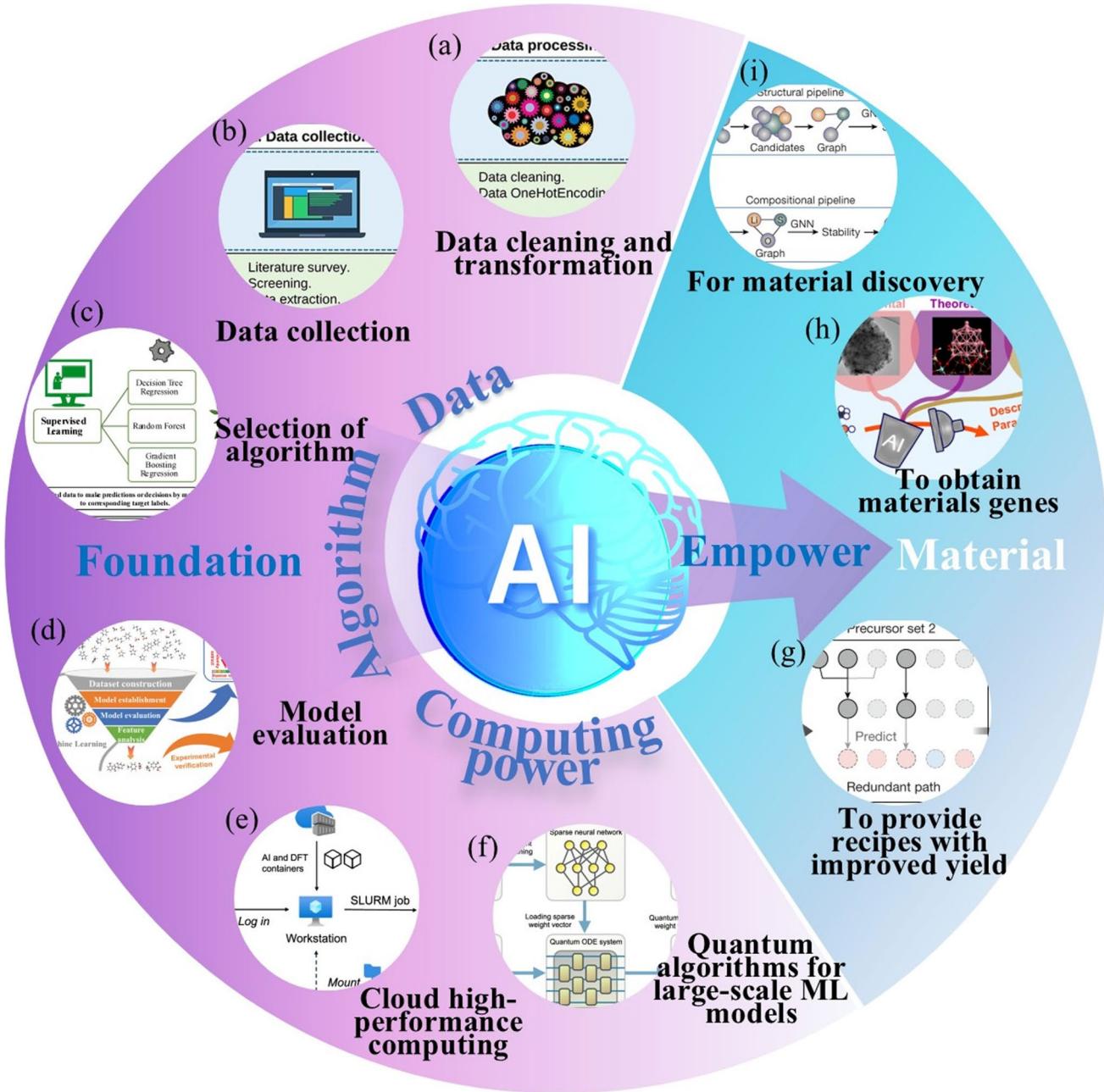
## 1 Introduction

The discovery and application of advanced materials and devices have promoted humans to combat the major global challenges [1–9]. Artificial intelligence (AI) has proved to be powerful tools for new material discovery [10–18], device performance prediction [19–24], and system performance improvements [25–32], and the emergent predictive capability has been verified with the assistant of increasing data, advanced algorithms and improved computing power [33–41] (Fig. 1). In particular, many novel information processing systems are developed, which will facilitate the progress made in the material science [42–46]. At the meantime, the rapid progress in the field of functional materials and devices has proposed high demand for AI [8, 47–50]. Novel methods for generating diverse candidate structures can be created, which can improve the efficiency of material discovery to a large extend [51, 52]. A large number of novel structures can be discovered by AI, many of which are beyond what human intuition can reach. Furthermore, as to the cognition of the existing materials, it is possible for AI to map the relationships between their structures and properties so as to make the prediction for previously uncharacterized properties [53–59] and device performances [60–72]. AI can also meet the challenge in illustrating the relationship between the physical properties of the stimuli in the external environment and their perceptual signals [73]. AI which can overcome the shortcomings of traditional trial-and-error method in material discovery and cognition has found its wide applications in many advanced functional materials [74–86], like two-dimensional perovskites [87], multicomponent oxides [88], nanomaterial [89], and silicon-oxygen compounds [90], which has prompted the development of many domains, such as information processing, clean energy harvesting, and catalysis discovery [51, 91–95].

The experimental synthesis of materials is also facilitated greatly by AI since the data-driven techniques, especially machine learning (ML), are managed to find the structure-property relationships of the materials, indicating the types of materials that are more feasible to be prepared, which used to be very difficult and time-consuming for humans to find the suitable methods and prepare new materials [87]. Moreover, efficient synthesis recipes can also be offered with the assistance from AI, which can simplify the manufacturing of complex materials and accelerate the synthesis of

theoretically predicted materials to a large extend [88]. High throughput and reproducibility can be realized at the same time by the robotic laboratories, making the exploration on the large-scale hypotheses to be rapid and reproducible [88].

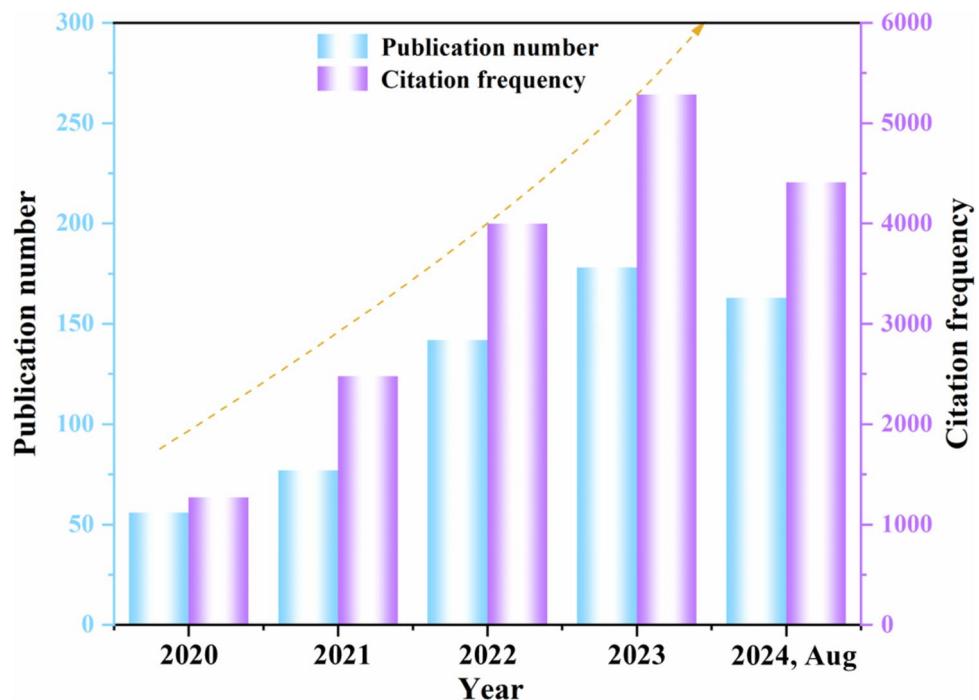
Recent years have witnessed a rapid development of the AI for materials science. The discovery of 2.2 million structures below the current convex hull has been realized with the efficiency of materials discovery promoted by an order of magnitude, among which many have been beyond the previous human chemical intuition [51]. A principal odor map has been developed which can make odor quality prediction for previously uncharacterized odorants [53]. The accurate and fast structure-adsorption prediction has been made by DeepSorption, a spatial atom interaction learning network [102]. The structural information has been provided for the disordered silicon at very-high pressure of up to 20 GPa via atomistic ML models, offering the predictions for the material systems even under experimentally challenging conditions [103]. As to the AI-assisted material synthesis, a universal framework has been developed for the preparation of two-dimensional perovskites with the ability of increasing the success rate of the synthesis feasibility by a factor of four compared to the traditional methods, which can be used in the typical laboratory [87]. It is noticeable that an autonomous laboratory has been successfully developed in order to achieve the accelerated synthesis of novel materials, which was managed to realize 41 novel compounds from a set of 58 targets under continuous operation of over 17 days [101]. As a result, many original works of high quality have been published with the citation frequency growing sharply over time (Fig. 2). Tactics have been developed for the AI empowered materials from many aspects, including synthesis, discovery, prediction, and variation, to realize the large-scale exploration, high throughput, and accelerated material discovery, which is demonstrated in Fig. 3. Several reviews relevant to the AI for material science are reported, and each of them has its own emphasis, with how AI promotes the membrane design [104], catalyst exploration [105], and development of other functional materials [106] included. Besides, other reviews provide us with the inspirations from other useful aspects, like the importance of interpretable ML for materials [107]. However, reviews from the view of how AI empowered both discovery of new materials and cognition of existing materials that covers the completed contents with these two synergistical aspects of cognition and discovery are few.



**Fig. 1** Overview of AI for materials with the data, algorithm, and computing power as fundamentals to support materials discovery and cognizance. **a** and **b** Reproduced with permission from Ref. [96]. Copyright 2022, Elsevier. **c** Reproduced under the terms of the CC-BY license [97]. Copyright 2024, The Authors, published by Wiley. **d** Reproduced with permission from Ref. [98]. Copyright 2024, Wiley-VCH GmbH. **e** Reproduced with permission from Ref. [99]. Copyright 2024, American Chemical Society. **f** Reproduced under the terms of the CC-BY license [100]. Copyright 2024, The Authors, published by Nature. **g** Reproduced under the terms of the CC-BY license [101]. Copyright 2023, The Authors, published by Nature. **h** Reproduced under the terms of the CC-BY license [91]. Copyright 2024, The Authors, published by American Chemical Society. **i** Reproduced under the terms of the CC-BY license [51]. Copyright 2023, The Authors, published by Nature

To be specific, the basic background of AI systems powered material research was introduced first, and then, the latest development in regard to the data collection and processing, the algorithm selection, and the automated laboratory

design for the AI systems applied in material science were demonstrated. Some important factors which should be under consideration when designing the advanced AI systems were discussed, including the strategies of how to



**Fig. 2** Publication number and citation frequency of the work focused on the artificial intelligence empowered materials discovery and prediction during the recent five years

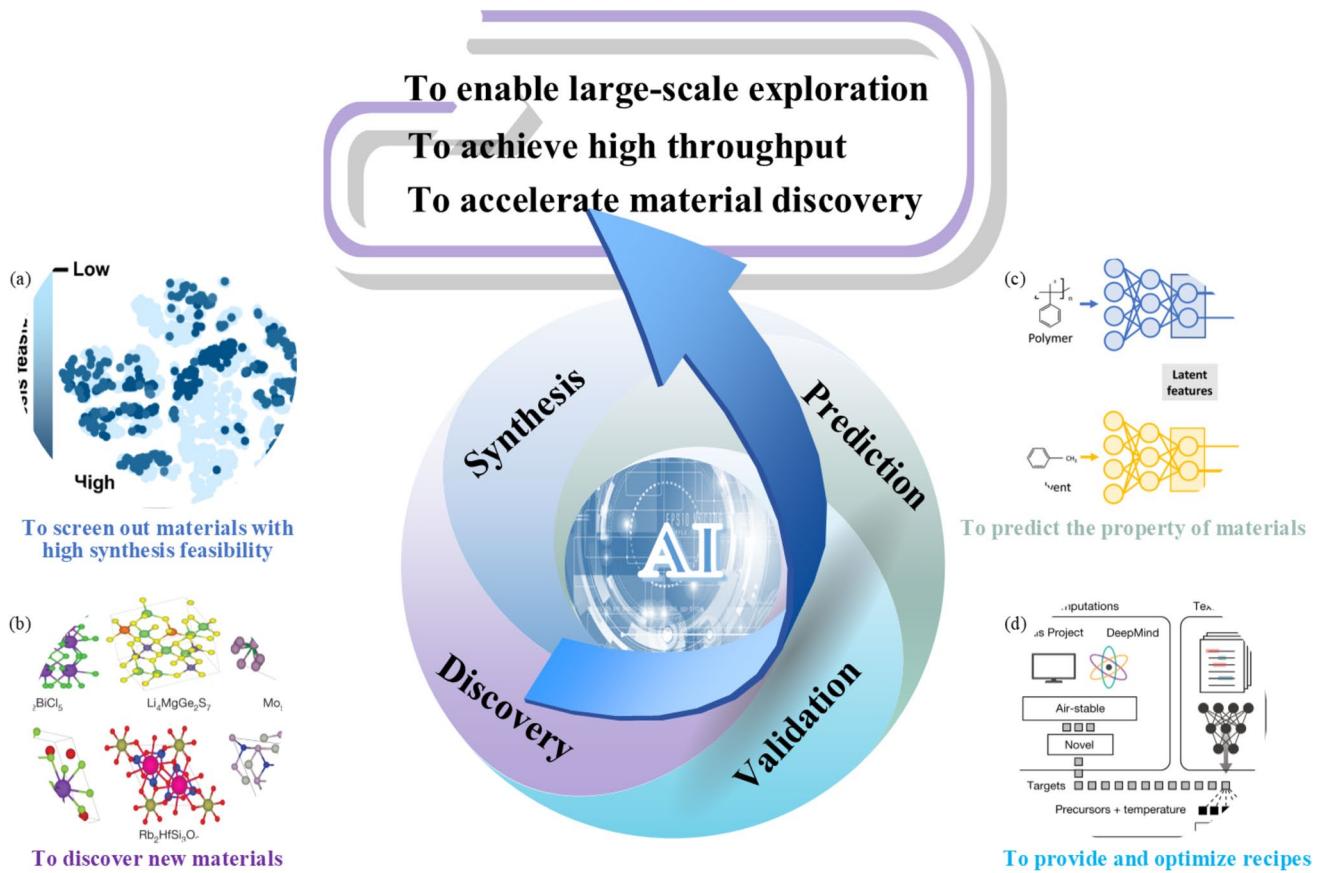
obtain the systems with enhanced performance, the features of the future AI systems for materials, and so on. Last but not least, some ideas with respect to the outlook of AI for materials were proposed.

## 2 Mechanism of AI for Cognizance of Existing Materials and Discovery of Novel Materials

Computing power plays a fundamental role in AI systems for materials, and data and algorithm are also of great importance in these systems [33, 109]. The improved computing power is managed to unlock modeling capabilities, which is beneficial for highly accurate and robust learning [51]. First-principles calculations based on density functional theory (DFT) have been made use of by computational approaches championed by the Materials Project (MP) [110], the Open Quantum Materials Database (OQMD) [111], NOvel MAterials Discovery (NOMAD) [112], and Automatic FLOW for materials discovery (AFLOWLIB) [113]. As shown in Fig. 4, the mechanism related to how AI empowers material research can be

mainly illustrated from the aspects of the existing material cognizance and the novel materials discovery.

AI makes contribution to map molecular structures to their properties in regard to cognition of the existing materials [115], so that the relationships can be got and prediction of the properties for previously uncharacterized materials can be made [53]. It is worthwhile mentioning that efforts have been made to deal with the situation where the structurally similar pair is not the perceptually similar pair, and the predictive modeling in diverse perceptual aspects has been realized by neural networks [116]. As a specific type of graph neural network (GNN), the message passing neural network (MPNN) can be used to map chemical structures to percepts. Each molecule was described as a graph with each atom and bonds represented by a series of characters in details, after which the fragment weights can be optimized. A reference dataset of many molecules described by multiple corresponding property labels is needed to be curated (Fig. 4a). The models are then to be trained with their parameters being optimized to generate the maps (Fig. 4b). The reliability of the model in describing the properties can then be verified by experiments to



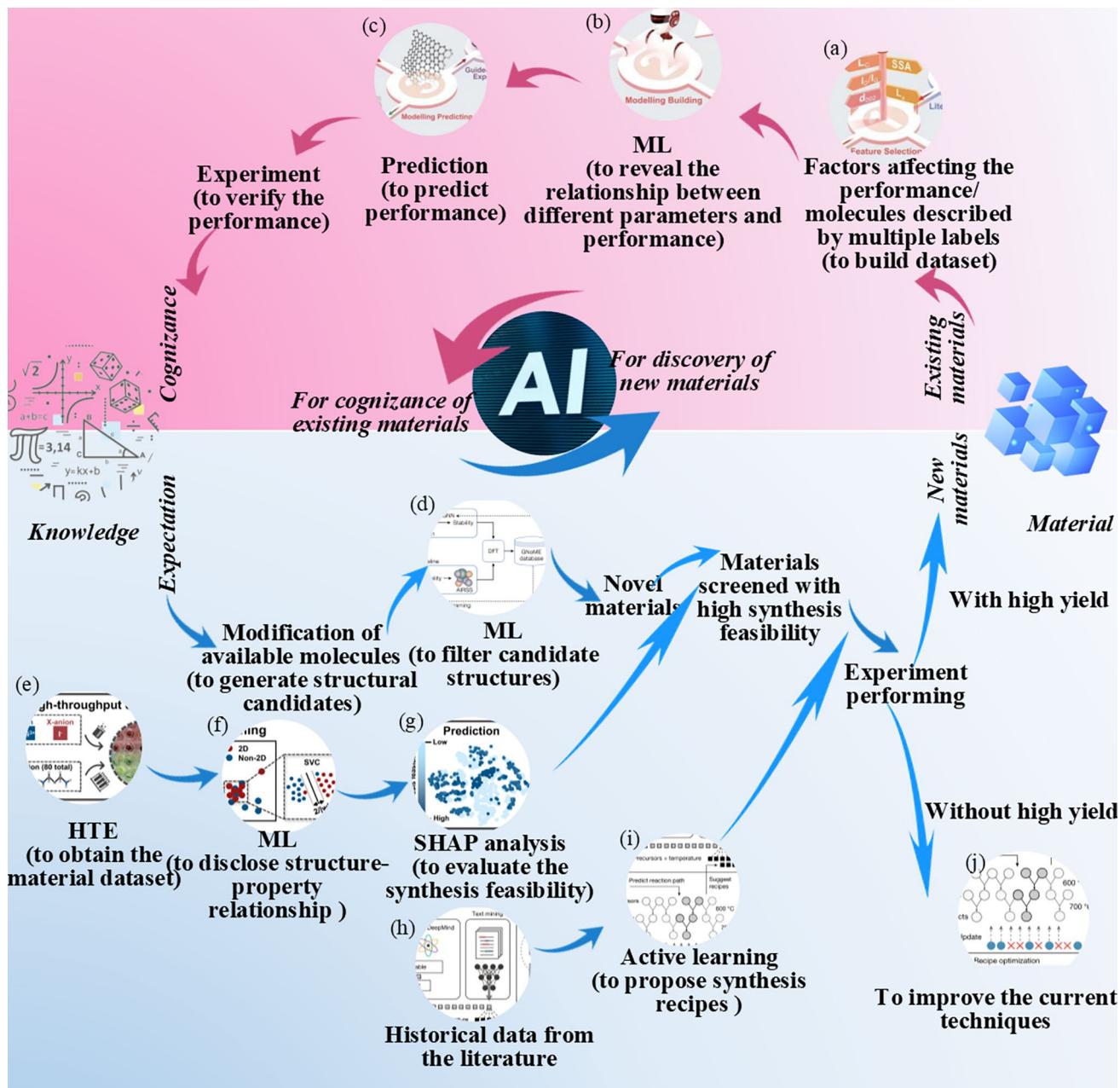
**Fig. 3** Tactics for the AI empowered material synthesis, discovery, prediction, and validation. **a** Reproduced under the terms of the CC-BY license. [87] Copyright 2024, The Authors, published by Nature. **b** Reproduced under the terms of the CC-BY license. [51] Copyright 2023, The Authors, published by Nature. **c** Reproduced under the terms of the CC-BY license. [108] Copyright 2023, The Authors, published by American Chemical Society. **d** Reproduced under the terms of the CC-BY license. [101] Copyright 2023, The Authors, published by Nature

justify whether a generalized description of structure–property relationships can be obtained (Fig. 4c), and the results can also be compared with that of the conventional structure-based maps to verify its efficiency.

In addition to the cognizance of the existing materials, AI also plays a vital role in the discovery of novel material [117–121]. Researchers are capable of conducting searches by substituting similar ions and enumerating prototypes, and endeavors have been made to improve the search efficiency of these approaches. In order to obtain more diverse candidates, neural networks can be applied to guide the searches [51]. It is proposed that a broader exploration can be made by neural networks while maintaining the efficiency. To be specific, structural candidates can be obtained by modifications of available materials, and methods have been built to enable incomplete replacements, so that the set of substitutions can be augmented

largely. DFT, which is an important method for calculating material properties in materials science, plays a fundamental role of bridging between the microscopic electronic structure and the macroscopic properties of materials. As for the structural candidates, the energy and other key properties of materials are calculated through DFT to verify the accuracy of the model's prediction. The new data obtained from DFT calculations can be added to the training set to train more powerful and robust models in the next round of active learning. Large scale of new materials can also be identified by means of high-throughput computation [101]. For instance, large-scale ab initio phase-stability data can be gathered from the MP and Google DeepMind [101].

AI is indispensable for the accelerated realization of new materials and the optimization of their design rules [91, 122–127]. After the identification of new materials, ML can



**Fig. 4** Schematic illustration for the mechanism of AI empowered material discovery and cognizance. **a-c** Reproduced with permission from Ref. [114]. Copyright 2023, The Royal Society of Chemistry. **d** Reproduced under the terms of the CC-BY license. [51] Copyright 2023, The Authors, published by Nature. **e-g** Reproduced under the terms of the CC-BY license. [87] Copyright 2024, The Authors, published by Nature. **h-j** Reproduced under the terms of the CC-BY license. [101] Copyright 2023, The Authors, published by Nature

then be applied to screen the novel materials with excellent performance and high synthesis feasibility (Fig. 4d-g). The critical physicochemical parameters related to the measured performance can be identified as materials genes among many candidate parameters obtained from experiments and initio

simulations. By means of using a ML model, novel materials with high performance are able to be suggested. Various reaction conditions, including precursors, intermediate products, additives, solvents, and temperature, should be taken into considerations for materials synthesis, which usually consumes a

lot of time for the experts [128]. The data-driven techniques are now used for screening out materials with high synthesis feasibilities by means of finding out the structure-property relationships, making it possible for the experimental realization of computational predictions [87].

The autonomous laboratory can be introduced to bridge the computational screening and experimental realization [101]. By the combined usage of computations, historical data, and ML, the plan and interpreting of the experimental outcomes can be made (Fig. 4h). ML model is able to provide the initial synthesis recipes for the proposed compounds, which can promote the material preparation (Fig. 4i). To be specific, the initial synthesis recipes can be obtained by the natural-language models learning to evaluate target ‘similarity’ via natural-language processing of a large database from the literature, which is similar to how a human make an attempt to begin initial synthesis according to known related materials [129]. It is worthwhile mentioning that analysis of the failed syntheses makes sense to offer direct guidance to improve materials screening and synthesis design [101], which is illustrated in Fig. 4j. Experiments will continue by taking advantages of autonomous reaction route optimization and solid-state synthesis, which is an active learning algorithm integrating ab initio computed reaction energies with observed outcomes once the yield does not achieve the expectation [130]. Experiments can be conducted by robotics. It is verified that autonomous workflows based on liquid handling can be demonstrated in organic chemistry [131–134], and recently, it is also possible for A-Lab to handle and characterize solid inorganic powders which used to be a challenge [101].

### 3 Design of the Intelligent Systems for New Materials

#### 3.1 Data Collection

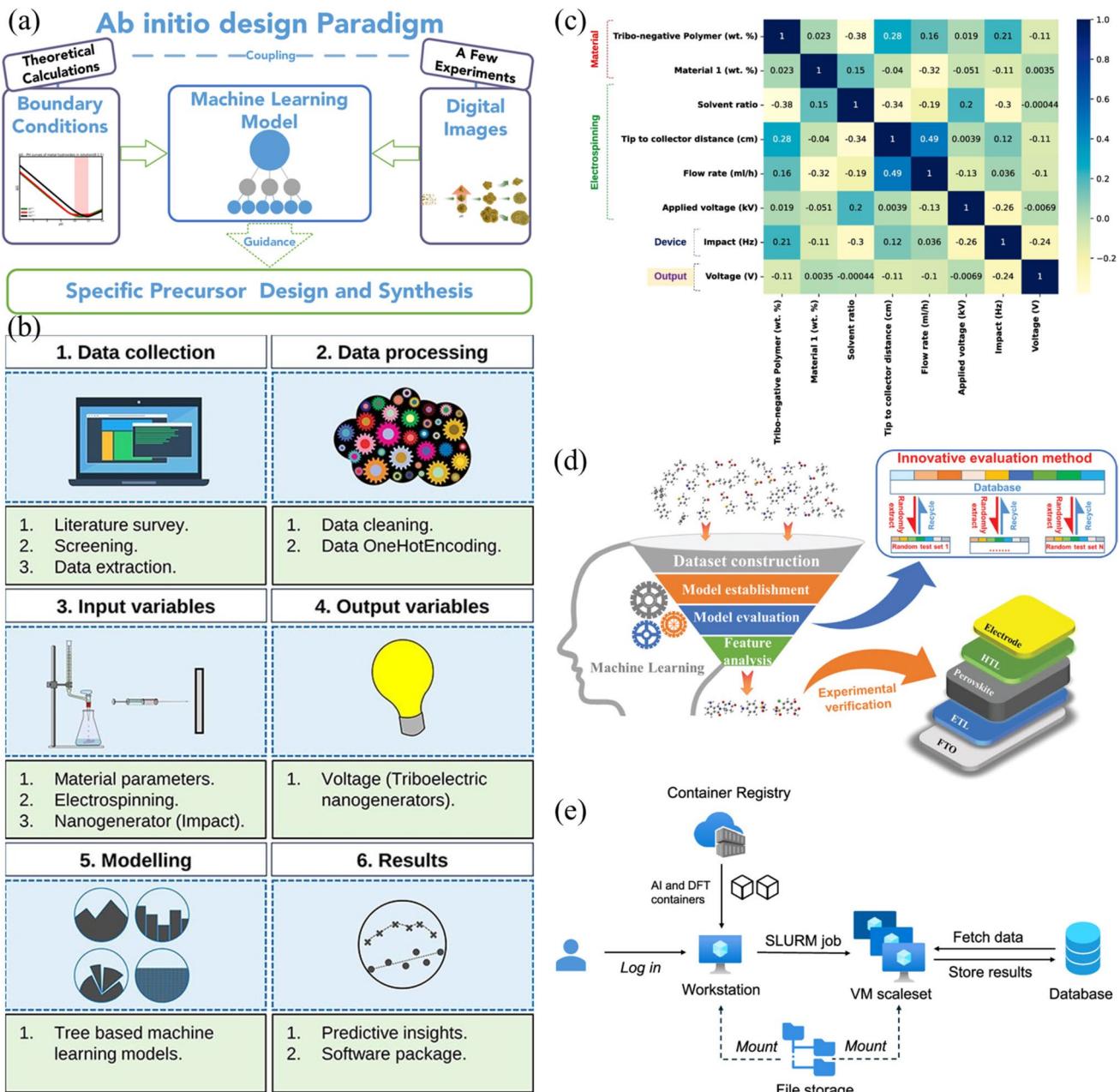
Material data play fundamental and important roles in the intelligent systems [135]. To be specific, the experimental synthesis data provided by studies, the first-principles calculations, and laboratory experience can be served as resources for the database [87]. For example, a design synthesis paradigm incorporated with ML was developed for Ni-rich cathode material, in which the boundary conditions for various reactions of precursors were provided by thermal/kinetic simulations, and a digital image dataset

was constructed by some necessary experiments [18] (Fig. 5a).

Abundant datasets with balanced data to overcome the problems of overfitting, underfitting, and limited extrapolating abilities of ML are expected to be provided [136, 137]. The data processing that includes data cleaning and data transformation can be carried out to make sure that the collected data are integrated [96]. For example, in an attempt to develop the predictive models for real-time voltage output of triboelectric nanogenerator (TENG), data cleaning was conducted to eliminate incomplete or inconsistent data, leading to a refined dataset with 279 reliable data points, which guaranteed the quality and consistency of the dataset (Fig. 5b). Pearson correlation coefficient which revealed the linear relationship between the two variables was demonstrated in Fig. 5c. Particularly, a negative correlation demonstrated that when the values of these parameters increased, there was a high probability for the output voltage to decrease, and then, the specific mechanisms underlying these relationships could be further investigated.

Efforts have been made to meet the challenge of limited available dataset for model evaluation [98]. For instance, a novel evaluation method was developed to screen small molecules served as passivation materials for perovskite solar cell when the available dataset was limited (Fig. 5d). Particularly, 20% of test data were randomly extracted, while the remaining parts were used as the training data, followed by calculating model accuracy which was a critical criterion for the evaluation of classification model. The extracted data were then reintegrated into the established dataset, and another 20% of the data was randomly selected as a test set, the process of which was repeated 100 times. The final model accuracy was then obtained as the averaging of the accuracy values from these 100 calculations.

High-performance computing (HPC) is another strong support for the accelerated and large-scale material discovery and cognizance [99]. Endeavors have been made to put forward the strategies which can offer large-scale computational resources for the screening and experimental validation. It is pointed out that cloud HPC can meet this challenge which has been verified to be managed to train and host large-scale AI models like GPT-4 asking for a massive number of graphical processing units (GPUs), and therefore, it is promising to be applied for material research with an increasing number



**Fig. 5** **a** Schematic illustration of the design synthesis paradigm incorporated with ML, which indicated the resources for the database. Reproduced under the terms of the CC-BY license [97]. Copyright 2024, The Authors, published by Wiley. **b** Schematics of the procedures for modeling the TENG ML models with data cleaning conducted. **c** Pearson correlation coefficient between input and output. Reproduced with permission from Ref. [96] Copyright 2022, Elsevier. **d** Schematic illustration of the ML process with the novel evaluation method to deal with the occasion where the available dataset was limited. Reproduced with permission from Ref. [98] Copyright 2024, Wiley-VCH GmbH. **e** Schematic illustration of the cloud environment for materials discovery workloads. Reproduced with permission from Ref. [99] Copyright 2024, American Chemical Society

of material candidates to be evaluated computationally. One case in point was that ML and cloud HPC were combined, and the schematic illustration of the cloud environment for

materials discovery is shown in Fig. 5e. Particularly, the ML models and DFT code were built into Docker container images. When operated, a workstation virtual machine (VM)

fetched the container images to NetApp Files storage mounted to the workstation and job queues or VM scale sets. The computational jobs were submitted to the VM scale sets via the SLURM job scheduler. Data and metadata were kept in a searchable database. It was proved that the system was managed to quickly navigate through more than 32 million candidates as well as predict around half a million potentially stable materials.

Another issue that cannot be ignored is that the training data used in many studies is often biased toward successful cases reported in the literature or databases, which will lead to the inconsistency between the data distribution and the real-world distribution. This imbalance can leave an impact on the generalization ability and robustness of the models. Some strategies can be taken advantages of to address these issues. For instance, negative sample construction is one of the most fundamental strategies. Besides, active learning enables the model to actively select the samples that most need labeling, prioritizing the supplementation of negative or marginal samples which are most crucial for the model's improvement. In addition, the multi-source data integration is another critical method for solving this problem. By integrating data from different sources and of different types, the sample distribution is enriched, thereby reducing the bias.

### 3.2 Machine Learning Algorithms

The development of data-driven techniques has significantly revolutionized the new material discovery, which is able to provide physical insights from the existing data in depth [87]. ML has developed rapidly to meet the multidimensional challenges in the material field [138–146]. ML can be used to reveal the structure-property relationship hidden behind a large number of experiments. Materials with high synthesis feasibility can be screened out with the assistance from ML, accelerating material synthesis even with limited experimental support [87]. Rapid predictions for structures and properties can be made by ML even for new materials.

The effective transformation of experimental data into model-ready input features plays fundamental and important role in building intelligent models, and some explorations have been made about how to realize the effective transformation. In some cases, the differential features, rather than the original curves or data, are focused. Besides, instead of processing data with a single branch, the integration of

two learning perspectives is carried out, making it possible for the models to learn from two dimensions. Additionally, designs can be conducted by making the input features highly correlated with the prediction target, reducing the learning burden of the model, and therefore, the prediction efficiency can be improved. One case in point was that a deep learning (DL) framework designed for the prediction of battery lifetime was developed by introducing an inter-cell learning mechanism to make prediction of the lifetime differences between two battery cells with the aim to represent the connections between cells cycled under different conditions [147]. In addition, the cycle-level features were fed into two separate branches, which contains the intra-cell difference curves and the inter-cell difference curves. Moreover, the correlations between the constructed features and the prediction targets for both intra-cell and inter-cell learning were investigated, and it was verified that a simple feature computed on inter-cell difference curves was managed to differentiate lifetime differences, even with a reference cell from a different battery chemistry, indicating its direct relationship between the constructed features and the prediction targets.

The appropriate selection and adaptation of models are imperative to develop the AI systems for the material science. It is worthwhile mentioning that the features of the model should be matched with the task requirements and data characteristics. To be specific, the complexity of the data should be evaluated, and then, corresponding models which are suitable for the tasks can be utilized. For instance, some basic linear models relying such as the 'Var.' and 'Dis.' can show commendable performance on the initial datasets, while they are not qualified for some complex datasets. Besides, some models relying on handcrafted features can be suitable for the task scenarios with stable data distribution, while for the tasks with diverse scenarios, models with the features of automatic learning should be given priority. Additionally, diverse and highly challenging test sets can be used to evaluate whether the models are appropriate.

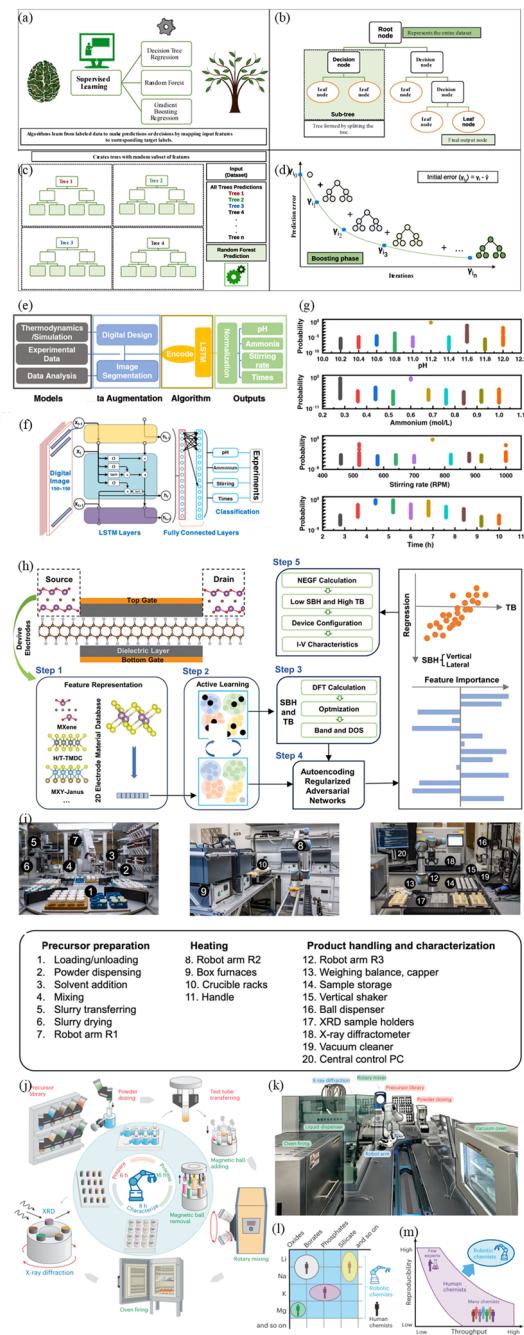
ML algorithms are selected according to classification accuracy, simplicity, computation efficiency, and so on [148–151]. Different ML algorithms have their own characteristics. For modeling with small dataset, support vector machine, linear regression, and gradient boosting are usually suitable [152, 153]. For example, in an attempt to improve the output performance of polyvinylidene fluoride (PVDF) nanogenerators, three decision tree ML models, including decision tree regression (DTR), random forest (RF), and

**Fig. 6** **a** Different algorithms with their own characteristics for the predictive TENG framework. Schematic of **b** DTR, **c** RF, and **d** GBR. Reproduced with permission from Ref. [96] Copyright 2022, Elsevier. **e** Computational flow of the ML for precursor design. **f** Model based on PMD-LSTM algorithm. **g** ML-assisted design for 3  $\mu\text{m}$  precursors. Reproduced under the terms of the CC-BY license [97]. Copyright 2024, The Authors, published by Wiley. **h** Workflow of the efficient screening of 2DEMs. Reproduced with permission from Ref. [157] Copyright 2024, Wiley. **i** Details of the A-Lab. Reproduced under the terms of the CC-BY license. [101] Copyright 2023, The Authors, published by Nature. **j** Schematic illustration of the robotic inorganic materials synthesis, and **k** photograph of the laboratory. **l** Illustration for robotic chemists enabled large-scale exploration compared with human chemists. **m** Illustration for robotic chemists enabled both the high reproducibility and throughput. Reproduced under the terms of the CC-BY license. [88] Copyright 2024, The Authors, published by Nature

gradient boosting regression (GBR), were chosen to develop the predictive models [96], which is illustrated in Fig. 6a. To be specific, DTR was extensively used for regression tasks, for which a binary tree structure was constructed by recursively splitting the TENG data based on the feature values (Fig. 6b). DTR was featured with its interpretability since the tree structure provided a clear visualization of the decision-making process. As shown in Fig. 6c, the RF combined multiple decision trees in order to enhance prediction accuracy, which was able to capture complex relationships between the input and output. As to GBR, decision trees were built sequentially with the subsequent tree reducing the errors made by the previous trees. The predictions of multiple weak models were incorporated to generate a strong model [154–156] (Fig. 6d). Another case in point was that the LSTM algorithm was applied in a design synthesis paradigm assisted with ML for Ni-rich cathode material, since the augmented datasets were still tiny [97] (Fig. 6e). It was proposed that the LSTM unit possessed its own advantages over RNN and CNN in the aspects of dealing with small sample data (Fig. 6f). ML-assisted design for 3  $\mu\text{m}$  precursors is illustrated in Fig. 6g.

Recently, endeavors have also been made to develop atomic-scale models of complex materials [158, 159]. Large datasets and ML have emerged, which are suitable for these more complex systems [160]. One case in point was that active machine learning was applied to offer a unified computational description of the silicon-oxygen systems, which were among the most important materials with complexity [90].

Various approaches have been used to deal with small dataset limitations, such as transfer learning [161], autoencoders [162], and active learning [163, 164]. These methods



can address the issues including noise, data imputation, and some other problems. Active learning strategies have been used more frequently for the classification scenarios compared to rigorous regression predictions. Recently, ML method has been adopted as the core component to screen low-contact electrode when limited data are available [157]. The detailed workflow is shown in Fig. 6h. To be specific, 2D electrode materials (2DEMs) were selected as numerical vectors using

feature descriptors for the reason that it was verified to mitigate the Fermi level pinning (FLP) effect and maintain excellent gate control in low-dimensional devices (Step 1). The feature distribution was deemed as a baseline for active learning, and the representative data points were collected iteratively. The consistency between the feature distribution of the training subset and overall sample features was made use of as the active learning evaluation function (Step 2). An autoencoding regularized adversarial neural network (ARANet) to perform model training in a scenario with a limited contact-property dataset generated through DFT calculations was developed (Step 3 and 4). Moreover, a novel feature-adaptive variational active learning (FAVAL) algorithm was introduced to work with ARANet, obtaining a valuable training subset. It was worthwhile mentioning that the jointly trained FAVAL and ARANet schemes outperformed typical small-data models using the same training datasets. Preliminarily screened materials could be accomplished (Step 5). It was noticeable that this scheme showed exceptional performance when trained with only 15% of the total data points.

The applicability of ML models in dealing with small-sample data and complex material systems and the generalization ability of the models are all important aspects for the AI empowered materials systems. Some novel approaches have been put forward to deal with these problems. For instance, some new mechanisms have been adopted when tailoring the frameworks. It was pointed out that many models for the battery lifetime prediction were developed and validated only across a restricted set of aging conditions, and therefore, efforts should be made to improve their extensive applicability. In contrast to many traditional models that were mainly focused on intra-cell learning by means of capturing early variations of a single cell to implement the prediction of its long-term lifetime, a framework was proposed to integrate inter-cell learning [147]. It was worthwhile mentioning that the stability of lifetime predictions for a target cell under varied aging conditions could be enhanced by combining it with the conventional single-cell learning.

Another factor that should be taken into considerations is the general approaches for interpretability, which can be realized by taking advantages of a series of methods, like the integration of knowledge, algorithm design, and visualization techniques. For instance, material knowledge can be integrated with ML to enhance the model generalization. Besides, design of the algorithm can be conducted. An interpretable ML combining the RF model and the Shapley additive

explanation (SHAP) analysis has been proposed to accelerate the identification of the critical factors that make influence on the formation energy among the complex variables introduced by doping in Ni-rich layered oxide cathodes [165].

The experimental validation of these predictions by quantitative metrics is a critical procedure for these AI systems. It is ideal to conduct the validation by comparing the performance of different models on multiple datasets with a series of quantitative metrics, like root-mean-squared error (r.m.s.e.), mean absolute error (MAE), and mean absolute percentage error (MAPE), for which the smaller deviation between the predicted value and the true value indicates the stronger predictive ability. Besides, it is better to make the comparison between the proposed model and the other models, and different datasets are expected to be used. Moreover, it is also of importance to evaluate the reliability in dealing with critical and complexed tasks in the real-world applications. Additionally, the error range should also be taken into considerations to make a full evaluation of these predictions. For instance, in an attempt to design the DL framework for making a prediction of battery lifetime, the performance comparisons among the proposed models and other models have been made by using a series of indexes, like r.m.s.e. and MAE with the error range being indicated, as well [147]. It was noticeable that error bars when combined with other statistical quantities such as standard deviation could visually represent the extent of variation of the corresponding models across multiple trials. In this case, when comparing the performance of different models, error bars enabled a more comprehensive assessment of the strengths and weaknesses of various models.

### 3.3 Autonomous Laboratory Validation

Material synthesis is featured with complexity with many factors like the kinetics and thermodynamic stability of materials, the synthesis routes, synthetic methods, and precursor species being taken into considerations [87]. The automated synthesis and characterization are important parts of the closed-loop AI systems for materials [166]. Robotic laboratories can be served as excellent platforms for data-driven experimental synthesis science to guide human and robotic chemists [88, 167]. It is worthwhile mentioning that the autonomous workflows based on liquid handling have been successfully realized for organic chemistry [131–134].

In regard to dealing with and characterizing solid inorganic powders, an autonomous laboratory has been built for the accelerated preparation of new materials [101]. The A-Lab performed experiments with three integrated stations for different tasks, including sample preparation, heating and characterization, and robotic arms were responsible for transferring samples and labware, which is illustrated in Fig. 6i. It turned out that the A-Lab was capable of realizing 41 novel compounds from a set of 58 targets with a success rate of 71% after continuous operating over 17 days.

It was pointed out that both the high reproducibility and throughput could be realized by the robotic laboratory simultaneously [88]. For instance, a robotic laboratory was taken advantages of to carry out the large-scale validation of precursor selection. As shown in Fig. 6j, k, a full ceramic synthesis workflow could be accomplished automatedly by a robotic arm, including precursor preparation, ball milling, oven firing, and product characterization. As illustrated in Fig. 6l, exploration of synthesis hypotheses in large-scale could be achieved by robotic laboratory, while it took a lot of human experimentalists many years to finish such intense work. Furthermore, it was difficult to weight the throughput and reproducibility for large-scale human work. In contrast, both the high reproducibility and throughput could be realized simultaneously by the robotic laboratory for the reason that it was possible for a robotic laboratory to produce single-source experimental data with high reproducibility, which is shown in Fig. 6m. It turned out that a comprehensive amount of synthesis hypotheses was managed to be explored rapidly and reproducibly by the robotic laboratory which could be served as a novel platform for the data-driven synthesis science.

Progress has been made to apply autonomy in a diversity of aspects in materials research, with robotic, the optimization of material yield, the improvement of photovoltaic performance, and the enhancement of photocatalysis activity included. However, in contrast to human researchers who have rich background knowledge facilitating their decision-making, some limitations still exist for the A-Lab in these aspects, and therefore, a fusion of encoded domain knowledge, the access to various data sources, and active learning are especially important for the autonomy. In addition to this issue, discrepancies between the current predictions and the experimental outcomes are needed to be further addressed.

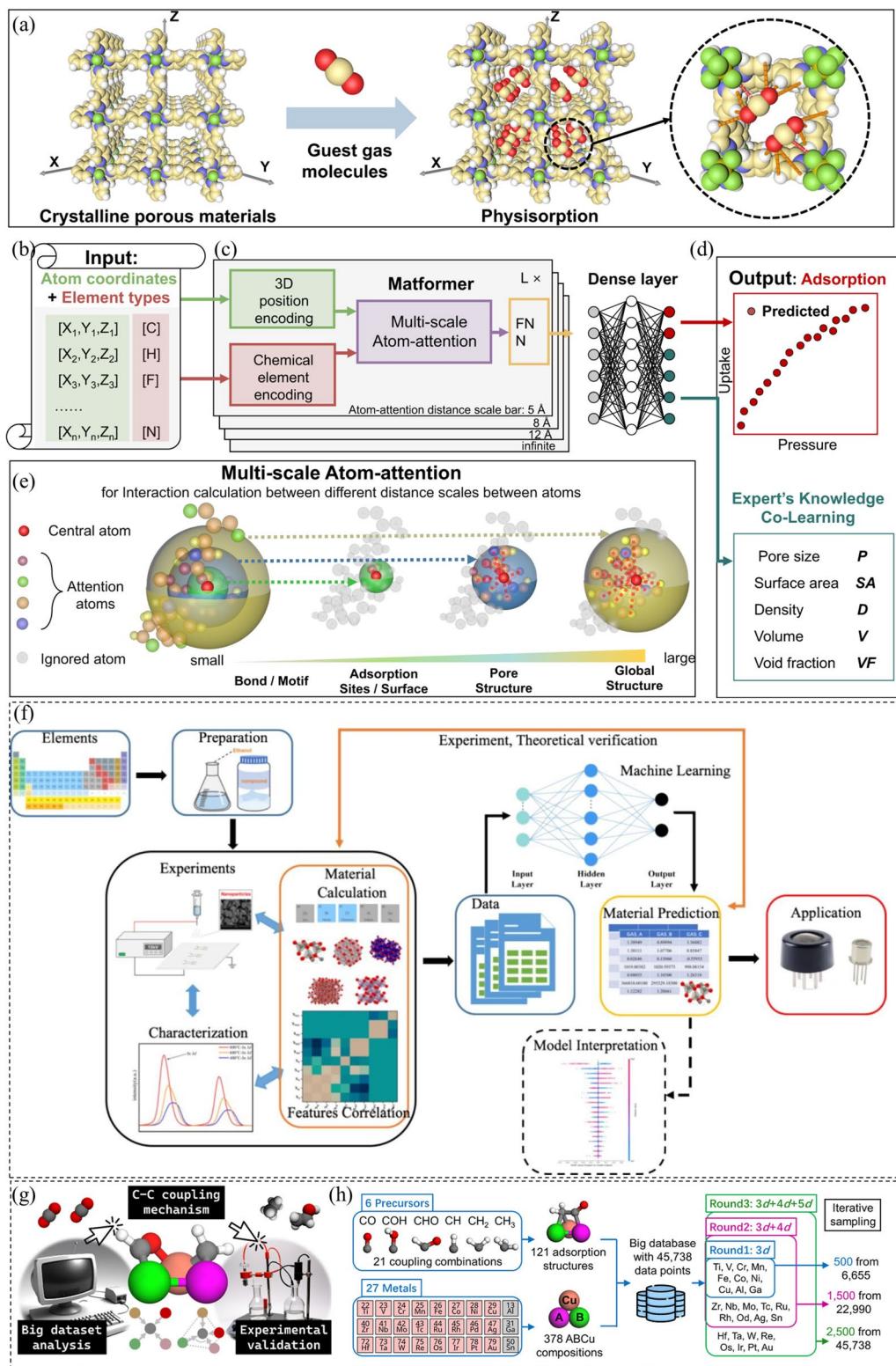
Another challenge that is met for the AI applied in material science is that there is gap between the predicted results and the feasibility of the experiment. Such an issue in result from a series of aspects. For example, in the early stage of new material research and development, the data available is scarce, and there exists the problem of overfitting or underfitting. Besides, the economic imbalance between the verification system and the experimental cost can also lead to the gap between the predicted results and the experiments. These gaps are usually in high relationship with the cognitive gap among data, models, and experiments. Cross-scale data fusion (combining atomic simulation with macroscopic characterization), the human-machine collaborative experimental design (reinforcement learning and domain experts), and other measures can be taken for narrowing the gap between the predictions and practice.

## 4 Strategies to Design AI Systems for Materials with Enhanced Performance

### 4.1 For Cognition of Existing Materials

#### 4.1.1 Existing Data Leverage

The materials used for training can be collected from some datasets [102]. To be specific, the knowledge of porous materials and their physisorption behaviors are beneficial for the ML-enabled rapid discovery of materials with desired adsorption properties [168, 169]. For instance, in order to develop a spatial atom interaction learning network for the prediction of gas adsorption (Fig. 7a), computation-ready, experimental MOF (CoREMOF), hypothetical MOFs (hMOF) and EXPMOF datasets were used. To be specific, CoREMOF dataset includes over 11,000 computation-ready and experimental three-dimensional metal-organic frameworks (MOFs) obtained from Cambridge Structural Dataset and Web of Science, while hMOF dataset includes over 300,000 hypothetical MOFs. Additionally, the EXPMOF dataset is from experiments. In this case, the original data of crystalline materials could be directly used as the input of DeepSorption without information loss, which is illustrated in Fig. 7b, and the outputs including gas adsorption isotherms could then be obtained (Fig. 7d). It is worthwhile mentioning that targeted data processing methods were developed.



**Fig. 7** **a** Schematic of the physisorption process with the materials used for training collected from some datasets. **b** Inputs of DeepSorption. **c** Architecture of MatFormer. **d** Outputs of DeepSorption. **e** Schematic of MSA. Reproduced under the terms of the CC-BY license. [102] Copyright 2023, The Authors, published by Nature. **f** ML-enabled exploitation of gas-sensing descriptors with the computations based on experimental and characterization results. Reproduced with permission from Ref. [170]. Copyright 2024, Elsevier. **g** Schematic illustration of ML big dataset analysis, and **h** the construction and sampling of big dataset. Reproduced with permission from Ref. [171]. Copyright 2024, American Chemical Society

The homemade MatFormer (Fig. 7c) featured with Multi-scale Atom-attention (MSA) was used to process crystalline material data for the reason that it was managed to provide conception of the interactions between different defined atoms, which is shown in Fig. 7e. The judgment of the interatomic interaction at different scales could be promoted by the exchange of information between atom pairs in different distances.

The computations can also be based on experimental and characterization results [170]. Recently, ML has been exploited for exploitation of gas-sensing descriptors, which can predict the gas-sensing performance of oxides (Fig. 7f). To be specific, data were obtained for five pristine oxides that were commonly applied as gas sensors. The input features were based on the characterization, computational results, and physical properties of the materials and gas molecules. The importance of the features was ranked, and six important features were proposed as the descriptors. It was worthwhile mentioning that many properties such as morphology, band structure, and surface composition could exert significant effects on gas-sensing reaction, and therefore, the oxides were characterized by a series of tests.

For some complex cases, it is necessary to construct big dataset to fully reveal the underlying mechanisms and the optimal direction for material design [171]. For example, carbon–carbon (C–C) coupling is of importance in the electrocatalytic reduction of  $\text{CO}_2$  in order to produce green chemical. However, the reaction network is usually complex. To address this problem, big data analysis was introduced into the computational screening of electrocatalysts for complex C–C coupling reaction networks (Fig. 7g). It was worthwhile mentioning that a big dataset with over 45,000 data points was constructed, covering all possible coupling combinations of six precursor species as well as adsorption configurations on the active site. As illustrated in Fig. 7h, 378 adsorption substrates made use of ABCu triatom active sites with 27 metal replacements for A and B, and iterative sampling was taken advantages of to obtain the training set for ML.

In addition to the construction of big dataset, some methods have been proposed for the cases in which the dataset is quantitatively limited and qualitatively biased [108]. For instance, a ML framework was developed for the highly generalizable prediction of temperature-dependent Flory–Huggins  $\chi$  parameters. The experimentally observed

$\chi$  parameters for 1190 samples were used for training the model. However, this dataset was lack of chemical diversity, and the experimental  $\chi$  parameters were biased, which limited the application of the model. Another significant bias existed in the aspect that some observable  $\chi$  parameters would be given only for polymer-solvent molecules in a miscible state due to technical limitations. It could be observed that the majority in the experimental  $\chi$  parameter dataset was consisted of soluble samples. Specifically, it was difficult to realize experimentally determining  $\chi$  parameters for an immiscible polymer-solvent system in which no single phase appeared. In order to address this issue, two auxiliary datasets were constructed, among which one was extracted from PoLyInfo with a list of 29,777 soluble and insoluble polymer–solvent pairs and the other was an in-house dataset obtained by making use of quantum chemistry calculations with COSMO-RS. It was proposed that polymers and solvents in PoLyInfo were distributed over a wider chemical space. It was verified that the applicability domain of the model was managed to be successfully expanded by learning with the two additional large datasets.

#### 4.1.2 Structure and Property Prediction

As to ML for structure or property prediction of the existing materials, it is also essential to make selections of different ML methods. One case in point was that knowledge co-learning (KCL) was chosen when developing a spatial atom interaction learning network [102]. It was proved that the KCL could enhance the convergence of the model in the structure-adsorption space establishment with the assistance from the expert knowledge in the auxiliary tasks by the comparison of the Expert-knowledge-driven learning (Fig. 8a), Data-driven learning (Fig. 8b), and Data-driven knowledge co-learning (Fig. 8c), and therefore, the prediction accuracy could be improved.

Another case in point was that three ML models were developed for the optimal preparation of biochar-based electrodes [172]. As illustrated in Fig. 8d, 14 key parameters from recent articles focused on the preparation of activated biochar-based supercapacitor electrode with urea as the nitrogen source and KOH as the activator were collected. Three classic ML prediction models, with RF, GBR, and extra tree regression (ETR) included, were made used of to make an exploration of the response relationship between

**Fig. 8** Schematic illustration of **a** Expert-knowledge-driven learning, **b** Data-driven learning, and **c** Data-driven knowledge co-learning. Reproduced under the terms of the CC-BY license. [102] Copyright 2023, The Authors, published by Nature. **d** Schematic illustration of ML enabled the optimal preparation method of biochar-based electrodes. Reproduced with permission from Ref. [172]. Copyright 2024, Elsevier. **e** Schematic relationship between different parameters and performances. Reproduced with permission from Ref. [114]. Copyright 2023, The Royal Society of Chemistry. **f** Schematic illustration of frontier digital twin-based battery sustainability platform. Reproduced with permission from Ref. [173]. Copyright 2024, Elsevier

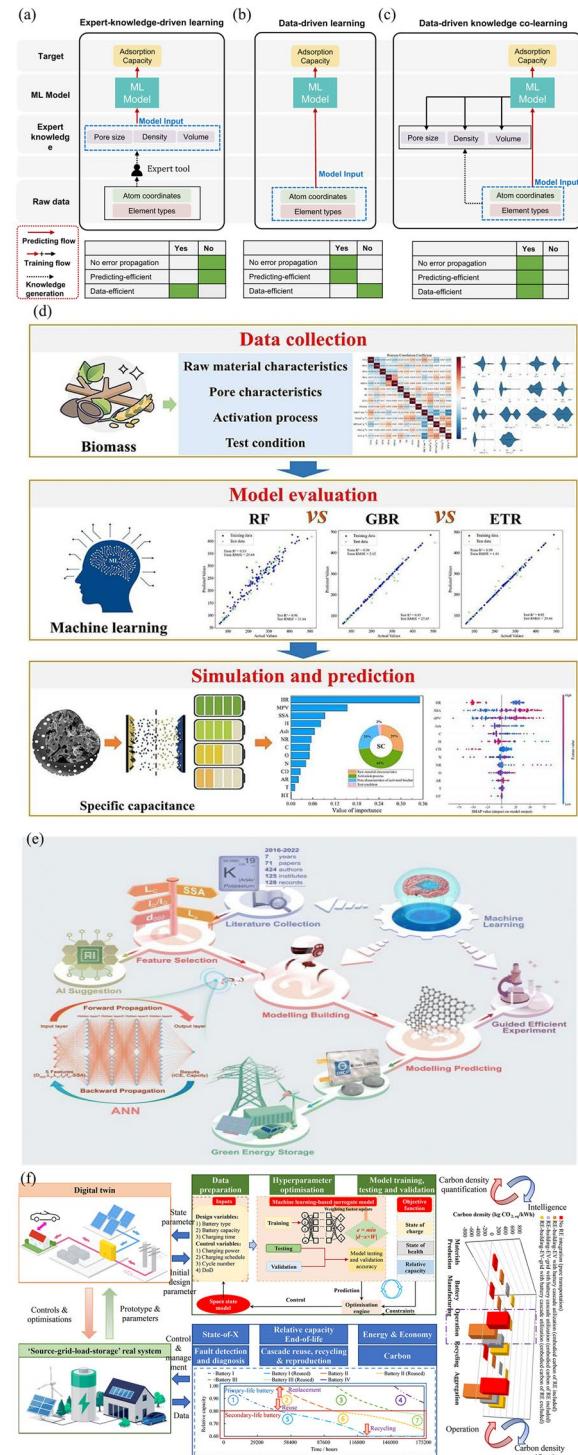
various factors and the energy storage properties. It turned out that the GBR demonstrated the best prediction performance with an  $R^2$  value of 0.93.

Methods have been come up with to handle the issue of limited data supplying in the primary tasks [108]. For instance, in a neural network architecture developed for the prediction of the  $\chi$  parameter with limited data providing for the parameters in the primary task, multitask learning was applied, in which different related tasks with common underlying mechanisms shared were learned simultaneously via a unified model. It was clarified that the multitask learning was able to boost the predictive performance by leveraging and transferring feature representations learned from two auxiliary tasks.

ML can be used to predict the relationship between different parameters and performance with the suitable models [114]. For instance, a strategy to construct hierarchical porous sponge-like carbon was launched for advanced potassium-ion batteries, in which cases ML was taken advantages of to offer further evidence of the excellent performance. Papers focused on layered carbon materials for potassium batteries were made use of to construct the structural parameter performance database. The complete initial coulombic efficiency (ICE) and capacity structural parameters performance database were input into ANN, which is demonstrated in Fig. 8e. It was verified that the predicted capacity and ICE were almost equal to the experimental values.

#### 4.1.3 Experimental Validation

The prediction capacities for structures or properties are usually examined by experiments comprehensively. It is noticeable that the prediction performance could then be evaluated from various aspects and in a diversity of conditions [102].



For example, the spatial atom interaction learning network was employed for prediction of gas adsorption, and it was verified that the predicted gas uptake was consistent with the actual value on CoREMOF-CO<sub>2</sub> and hMOF-CO<sub>2</sub>. In contrast to the other models, there absolute errors were much

smaller and more distributed centralized for DeepSorption. Furthermore, higher coefficient of determination ( $R^2$ ) values could be realized. It turned out that both the highest  $R^2$  value and the lowest MAE could be achieved by DeepSorption compared with the other models.

As a powerful tool for material research, AI has been integrated with other advanced technologies for the formation of more complex platforms [173]. For instance, a cross-scale multi-stage analytic platform featured with inter-disciplinary and trans-disciplinary was developed for the lifecycle carbon intensity investigation of electrochemical batteries, including battery materials, charging/discharging behaviors, recycling, and reproduction (Fig. 8f). ML was applied to address the issues that the collected data from controlled test conditions in the laboratory were not managed to represent various real application scenarios, and the state-of-charge prediction could be made. Besides, ML-assisted computation could promote the sustainability and climate adaption for this framework. Furthermore, by taking advantages of the digital twin, the performance estimation could be cost-saving and time-efficient.

In addition to the theoretical approaches, how these AI systems for material discovery and synthesis make contribution to the real-world examples with experimental implementation and practical validation is another valuable aspect to be explored, and more researches have been carried out focused on how to utilize these systems to address the practical issues. The lithium-ion batteries, which are featured with high energy densities and low production costs, have drawn great attention in many modern industries, serving as renewable energy solutions for many fields, like electric vehicles.

It is worthwhile mentioning that the combination of AI with battery lifetime prediction is also one of the research hotspots, since the capacity of these batteries fades inevitably with cyclic operations, which is attributed to the intrinsic electrochemical mechanisms. Great challenges have been met due to a variety of factors that influence the complex battery capacity degradation, like electrode materials, cycling protocols, ambient temperatures, and so on. Some cutting-edge researches have been conducted with the effective solutions to address these issues. For instance, a DL framework, BatLiNet, which was designed to predict battery lifetime reliably across a variety of aging conditions, was proposed [147]. In contrast to the traditional models which solely focused on individual cells, this framework adopted inter-cell learning which contrasted pairs of battery cells

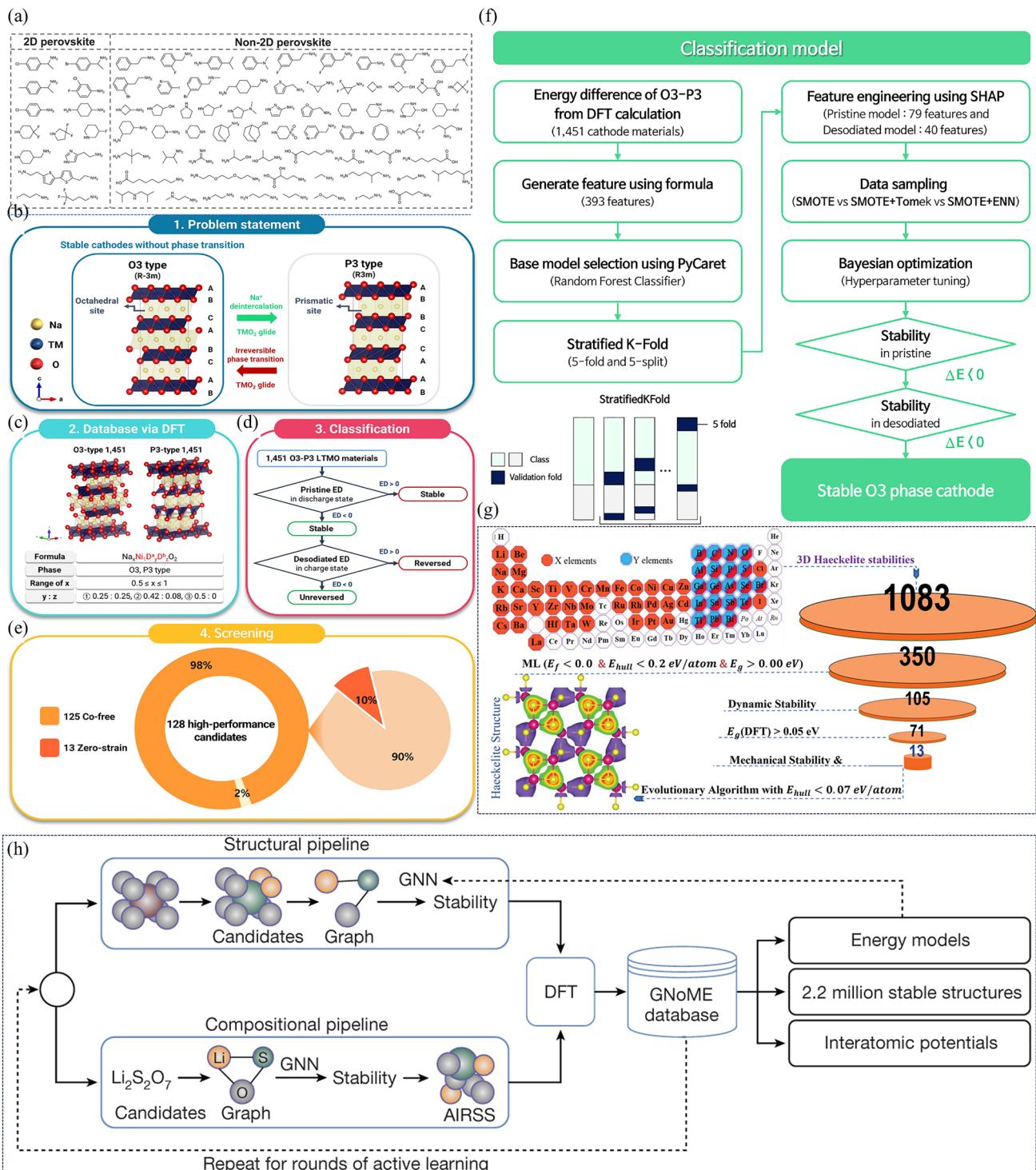
for discerning lifetime differences. It was noticeable that the experimental results, derived from a broad spectrum of aging conditions, verified its superior accuracy and robustness in this research when comparing to other existing models. In addition to the design of the frameworks, efforts have also been made to meet the challenges proposed by the intersection of electrochemical science and ML, and accordingly, an open-source platform with data preprocessing, feature extraction, and the implementation of both conventional and state-of-the art models integrated has been developed, which aims to provide a collaborative platform on which experts from diverse specializations can contribute their own efforts [174].

## 4.2 For Discovery of New Materials

### 4.2.1 Excavating the Existing Data

The dataset used for training is the cornerstone of ML models [175]. The experimental synthesis data provided by studies serve as important resources for the material synthesis. However, only successful cases are usually included in these studies, resulting in the imbalanced distribution of data category. Another important resource is from the first-principles calculations. Besides, previous studies and extensive laboratory experience can offer valuable intuitions for the preparation of new materials. For instance, in an attempt to explore the synthesis feasibility of two-dimensional silver/bismuth (2D AgBi) iodide perovskites, organic spacers from both the previously reported 2D perovskites and the chemical intuitions were exploited [87]. The high-throughput experiments (HTE) were made use of to acquire the material dataset. It was proved that only 13 kinds of organic spacers were able to form 2D AgBi iodide perovskite structures, and the organic spacers were sorted into '2D perovskite' and 'non-2D perovskite' accordingly, which is shown in Fig. 9a.

DFT calculations which are quantum mechanical theory-based tools also play an important role in high-throughput computational material design for the reason that they can characterize material properties and produce data directly [18]. For example, in an attempt to develop Co-free and low strain cathode materials for sodium-ion batteries with the assistance of ML (Fig. 9b), 1451 O3 and P3 layered transition metal oxides (LTMOs) were generated via DFT calculations, which is illustrated in Fig. 9c. The classification



**Fig. 9** **a** Synthesis results of high-throughput experiments. Reproduced under the terms of the CC-BY license. [87] Copyright 2024, The Authors, published by Nature. Schematic illustration of **b** problem statement, **c** database construction, **d** ML classification, and **e** screening result. **f** Schematic of the classification model. Reproduced with permission from Ref. [18]. Copyright 2024, Elsevier. **g** Workflow of ML for new Haeckelite compounds discovery. Reproduced with permission from Ref. [52]. Copyright 2024, Wiley. **h** Schematic illustration of two frameworks to for generation and filtration. Reproduced under the terms of the CC-BY license. [51] Copyright 2023, The Authors, published by Nature

ML models were then constructed to evaluate the structural stability and phase transition (Fig. 9d), leading to the identification of 128 highly reversible high-performance cathode material candidates (Fig. 9e). In this study, endeavors have been made to solve the problem of imbalanced data by a data sampling technique. In particular, a stratified k-fold importing data hierarchically from every class were taken advantages for the construction of a balanced train set, which is shown in Fig. 9f. Given the fact that there were not enough data, it was conducted in fivefold (train set/validation set=8:2), so that the number of validation set could be guaranteed.

Although there are both positive and negative material data in the datasets from HTE, subjective preferences still exist. As a result, it is difficult for ML to obtain reliable conclusions. Efforts have been made to address this issue. For example, in the framework to guide the experimental synthesis of two-dimensional perovskites, data-mining approaches were taken advantages of to identify the applicable subdomains for ML models, and then, models were trained on the identified subdomain, which showed more distinctive descriptors than models training on the whole biased dataset [87]. To be specific, subgroup discovery was used to get the suitable subdomains for ML models. It turned out that the molecular weight and the third ordered kappa index were the two descriptors standing out since they had a high correlation with the synthesis feasibility. Besides, based on the derivation of the rigid sphere model, the width of organic spacers was also of importance for the structural stability. In addition, 2D projections of this 3D data distribution map were generated. The distribution of 2D perovskites and non-2D perovskites was balanced in the determined specific subdomain.

A series of steps are taken for the preparation and preprocessing of data. In an attempt to discover new Haeckelite compounds for optoelectronic devices with the assistant from ML, data preparation and preprocessing were carried out [52], and the workflow of screening and predicting is shown in Fig. 9g. To be specific, the selected chemical space of the ML dataset with X and Y elements was demonstrated in red and blue, and 1083 square-octagon XY form structures were created. After that, the compounds with altered symmetry and duplication were removed. A base dataset for the train structure was selected, which was both quantitatively and qualitatively accurate when compared with other benchmarks, and then,

more compatible optimal features to predict each target with high accuracy and minimum error were constructed. It turned out that 350 materials were got after the investigation of the formation energy, bandgap, and convex hull energy by comparing them with the experimental targets using ML, and 13 semiconducting Haeckelite structures were obtained after the calculations of electronic structures, dynamic stability, and the multistep evolutionary.

It is noticeable that in some cases the space of possible materials is far too large, and it is difficult to sample in an unbiased manner [51]. Under the condition that there is no reliable model available to approximate the energy of candidates cheaply, the substitution of similar ions or enumeration of prototypes has been made use of according to chemical intuition. There still exists limitations in regard to the diversity of candidates even though the search efficiency has been improved, and therefore, it is critical to build new method to make more diverse candidates available. In an attempt to solve this problem, two frameworks were taken advantages of to generate and filtrate these candidates, which is illustrated in Fig. 9h. Particularly, the structural candidates were managed to be generated by modifying available crystals. It was worthwhile mentioning that efforts have been made to augment the set of substitutions by means of adjusting ionic substitution probabilities to give priority to discovery. Moreover, newly proposed symmetry aware partial substitutions (SAPS) were used to enable incomplete replacement efficiently. As to the second framework, compositional models could predict stability free of structural information. The graph networks for materials exploration (GNOME) were trained on available data to filter candidate structures. It was worthwhile mentioning that for the models, the crystal definition, which encoded the lattice, structure, and atom definitions, was served as the input. In particular, each atom was represented as a single node in the graph, and edges were defined on the occasion where the interatomic distance was less than the defined threshold. Nodes were embedded by atom type and edge, and they were embedded on the basis of the interatomic distance. A global feature that was connected in the graph representation to all nodes was also made use of. At every step of the GNN, neighboring nodes and edge features were aggregated, and they were utilized to update the corresponding representations of nodes, edges, or globals individually. After 3-6 layers of message passing, an output layer projected the global vector so as to obtain an estimate of

the energy. It was verified that almost an order of magnitude larger than previous work could be achieved via GNoME.

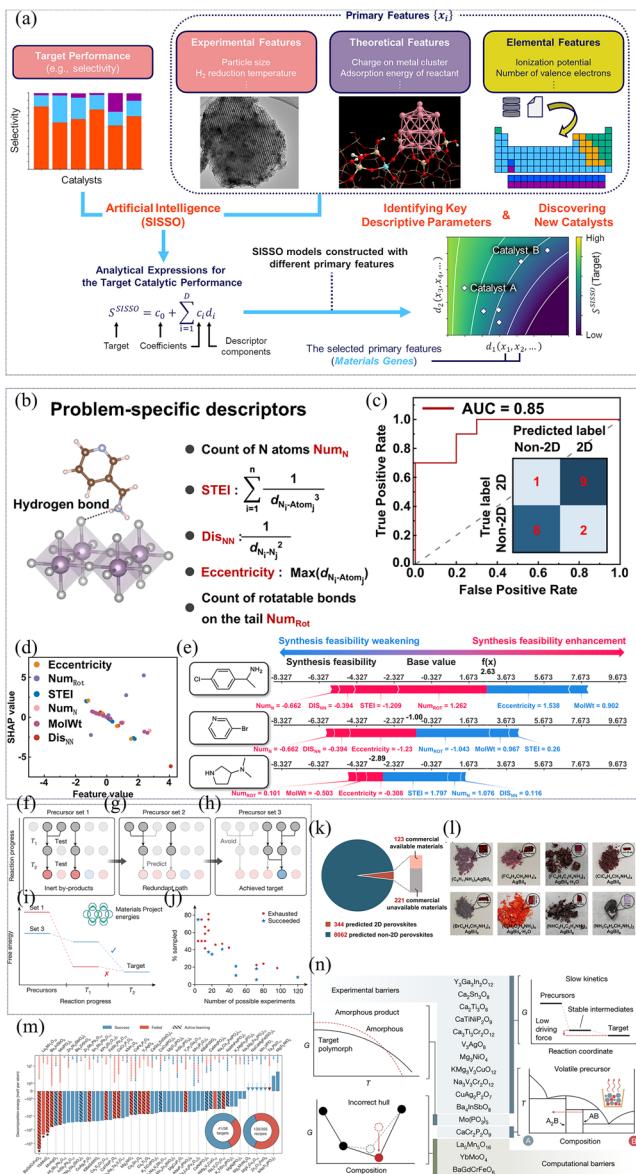
#### 4.2.2 Screening for Excellent Performance and High Synthesis Feasibility

AI can meet the demand for screening the materials with excellent properties and high synthesis feasibilities, which has significantly promoted the discovery and realization of new materials with excellent performance for various applications, such as catalysts, lithium-ion batteries, and perovskite solar cells [91]. Endeavors have been made to identify the key parameters of materials and suggest the new candidates which can be verified experimentally. For instance, in an attempt to address the issue about the catalyst design existing in the aspect that their performance was influenced by an intricate interplay of various multiscale factors, like the chemical reactions on surfaces, the type of support materials, and the material restructuring during the catalytic period, symbolic-regression AI was taken advantages of to extract the key physicochemical parameters related to the performance successfully (Fig. 10a). It was proposed that the Sure-Independence Screening and Sparsifying Operator (SISMO) had been introduced into data-centric methods for heterogeneous catalysis with a series of advantages [176, 177]. To be specific, analytical expressions related to the target catalytic performance could be identified by SISMO with few key parameters out of many offered parameters, which were regarded as materials genes to demonstrate the catalytic function of the materials. The intricate correlations between small datasets could be demonstrated with an immense amount of candidate functions being under consideration during the analysis process. In particular, the theoretical, experimental, and elemental parameters as the primary features were made use of to efficiently model the catalytic performance. The analytical expressions were pointed out. It was noticeable that the key descriptive parameters deemed as materials genes which were in close relationship with the property were obtained. By using the AI model with low costs, new additives could also be suggested, which promoted the discovery of new catalysts with high performance.

When it comes to new material synthesis, a diversity of factors should be taken into consideration, such as the precursors, the by-product, the feasibility of experimental conditions, and the availability of experimental raw materials,

which makes it a more complex and time-consuming task for chemists due to the limitation of experimental instruments and the heavy workload. As a result, only a small subset of the potential conditions can be evaluated with a rather small proportion of theoretically predicted materials being synthesized successfully, and this work heavily relies on the experience of chemists. In this case, the data-driven techniques can be applied for screening out materials with high synthesis feasibilities to provide guidance for the material synthesis [87]. It is worthwhile mentioning that efforts have also been made to extract meaningful physical and chemical insights from trained ML models in order to have a better understand the ML predictions. For instance, ML techniques were made use of to screen two-dimensional hybrid organic–inorganic perovskites (2D HOIPs) with high synthesis feasibility rapidly. A lot of distinctive descriptors relevant to the synthesis feasibility were developed (Fig. 10b). The support vector classification (SVC) algorithm was applied to develop the equation for the synthesis feasibility. The receiver operating characteristic (ROC) curve and confusion matrix which were used to evaluate the accuracy and the error are shown in Fig. 10c, and the area under the ROC curve was as high as 85% where only 1 out of 10 molecules was misclassified. The marginal contribution of individual descriptors was analyzed by SHAP analysis, the result of which demonstrated that the number of rotational bonds in the alkyl tail ( $\text{Num}_{\text{Rot}}$ ) was the most important factor for the synthesis feasibility. The relationships between the feature values and SHAP values are illustrated in Fig. 10d. A positive SHAP value indicated that the feature led to high synthesis feasibility, while a negative one resulted in low synthesis feasibility. The predicted synthesis feasibility is demonstrated in Fig. 10e.

ML also plays an important role in planning and interpreting the outcomes of experiments, which is critical to bridge the gap between computational screening and experimental realization of new materials. For instance, by combining with other tools, including computations and historical data getting from the studies, it was possible for ML model to provide up to five initial synthesis recipes for the proposed compounds [101]. Specifically, target ‘similarity’ was evaluated by means of natural-language processing of a large database extracted from the literature, during which process the behaviors of a human carrying out the initial synthesis referred to known related materials were mimicked [129]. Furthermore, the active learning was made use of to identify



**Fig. 10** **a** Schematic illustration of SISSO AI for integration of materials parameters. Reproduced under the terms of the CC-BY license. [91] Copyright 2024, The Authors, published by American Chemical Society. **b** Schematic of the problem specific descriptors. **c** Receiver operating characteristic (ROC) curve and confusion matrix. **d** SHAP values for the six features. **e** Predicted synthesis feasibility. Reproduced under the terms of the CC-BY license. [87] Copyright 2024, The Authors, published by Nature. **f** Schematic of how to estimate which pairwise reactions occurred from a failed attempt. **g** Schematic of new precursors recommended by substituting a precursor relevant to the unfavorable pairwise reaction and **h** schematic illustration of the successful precursor set. **i** Free energy at each step. **j** Number of experiments for exhausting all unique reaction paths of each target or identifying an optimal path. Reproduced under the terms of the CC-BY license. [101] Copyright 2023, The Authors, published by Nature. **k** Schematic of screening for the prediction set, and **l** the experiment validation. Reproduced under the terms of the CC-BY license. [87] Copyright 2024, The Authors, published by Nature. **m** Experimental outcome. **n** Barriers for the synthesis of the targets. Reproduced under the terms of the CC-BY license. [101] Copyright 2023, The Authors, published by Nature

synthesis routes with improved yield. A database of pairwise reactions was continuously constructed by the autonomous laboratory from its experiments, which made it possible for the products of some recipes to be inferred (Fig. 10f, g). As a result, the search space of possible synthesis recipes could be reduced by up to 80% as many precursor sets reacted to form the same intermediates (Fig. 10j). Besides, knowledge of reaction pathways was used to give priority to the intermediates with large driving force to form the target which could be obtained from the formation energies provided by the Materials Project (Fig. 10h, i).

#### 4.2.3 Experimental Realization

Experiments are usually carried out in order to realize the new materials either by researchers or by autonomous laboratories. Some equations obtained can be used to predict unexplored molecules. For instance, 344 2D perovskites with high synthesis feasibility were screened out [87], which is illustrated in Fig. 10k. Given the fact that organic spacers in the prediction set were gathered from PubChem, some amines were commercial unavailable, leading to only 123 predicted 2D AgBi iodide perovskites to be possible for further experimental synthesis. Experiment validation was then carried out. The synthetic chemical reagents were used as received. In order to eliminate the competing Bi-based phases, an excess amount of  $\text{Ag}_2\text{CO}_3$  was made use of. For instance, an amount of  $\text{Ag}_2\text{CO}_3$  and  $\text{Bi}_2\text{O}_3$  were dissolved in concentrated hydroiodic acid with the heating temperature of 393 K. 1-(4-chlorophenyl) ethan-1- amine was added to  $\text{H}_3\text{PO}_2$  in another beaker, and then, the two solutions were mixed, which was allowed to evaporate at the hot plate with the temperature of 323 K for a day. Finally, brownish red crystals precipitated at the bottom of the beaker could be obtained successfully. As for validation of the ML model, 13 commercially available organic spacers without hydroxyl and ether were unbiased selected, and 8 of 13 predicted 2D AgBi iodide perovskites which showed high synthesis feasibility were successfully synthesized with the success rate of 61.5%, indicating a much higher success rate than chemical intuition (16.4%) (Fig. 10l). It was worthwhile mentioning that the repeatability and stability are critical to the experiment validation, and some measures can be taken to ensure the repeatability. For instance, in this case, ten individual repetitions of the synthesis process for  $(\text{NH}_2\text{C}_5\text{H}_8\text{F}_2)_4\text{AgBiI}_8$

were implemented to make an assess of the experimental reproducibility of the synthesis experiments.

An autonomous laboratory was designed for the accelerated synthesis of novel materials integrated with computations, historical data, ML, and robots to conduct experiments [101]. Recipes for synthesis of the novel materials were tested using a robotic laboratory, which was managed to perform the powder dosing, sample heating, and product characterization. The experimental outcome is demonstrated in Fig. 10m. It is proposed that the robotic experimentation efficiently accelerated the experimental synthesis of materials. The high success rate also verified that it was possible for comprehensive ab initio calculations to discover novel and synthesizable materials effectively.

Some further analysis can be carried out about the barriers for the synthesis of the targets [101]. For example, for the 17 of the 58 targets evaluated by the A-Lab which were not realized even though active learning was taken advantages of, the ‘failure modes’ were classified as experimental barriers which were marked as blue and computational barriers which were marked as green in Fig. 10n. The continued efforts have been made by researchers to create new materials experimentally in order to offer a way to validate the AI findings.

## 5 Design Consideration of the AI Systems for New Materials

### 5.1 To be More Autonomous

The intelligent systems are expected to be more autonomous with the capability to interpret data and make decisions. Autonomy has realized from some aspects in regard to materials science. It is widely considered that a fusion of encoded domain knowledge, access to a variety of data sources, and active learning are critical for the accomplishment of enhanced autonomy [178, 179].

One case in point was that an autonomous laboratory was successfully constructed which was managed to realize 41 novel compounds after more than 17 days of continuous operation [101] (Fig. 11a). To be specific, the targets which were air-stable and unreported were identified via DFT-calculated convex hulls consisting of ground states from the Materials Project and Google DeepMind, after which the synthesis recipes were pointed out by means of ML models

that were trained on synthesis data from the studies. The recipes were then tested via a robotic laboratory through the powder dosing, sample heating, and characterization procedures. Phase purity was evaluated via X-ray diffraction (XRD), which was then analyzed by ML models trained on structures from the Materials Project and the Inorganic Crystal Structure Database (ICSD). In particular, both the phase and weight fractions of the synthesis products were extracted from their XRD patterns by probabilistic ML models that were trained on experimental structures from the ICSD. By inverting the container, the powder was dispensed through the mesh onto an XRD sample holder, after which it was flattened with an acrylic disk. The flattened sample was transferred into the diffractometer for X-ray measurements with 8-min scans that range from  $2\theta = 10^\circ$  to  $100^\circ$ . For a given XRD pattern got from an unknown sample, XRD-Auto Analyzer was utilized to recognize the constituent phases, and their weight fractions were also estimated. In the cases where high ( $> 50\%$ ) target yield was not achieved, new synthesis recipes would then be proposed by means of an active learning algorithm. It was noticeable that the whole sequence was fully automated.

Both the exploration and optimization were able to be realized via a chemical robot for the autonomous synthesis of nanomaterials [89] (Fig. 11b). Particularly, as for the exploration mode, the structural diversity was accomplished via searching for diversity in the behavior space. The fitness was evaluated based on peak prominence and broadness correlated with the yield and mono-dispersity. A new batch of experiments was generated from previous synthetic conditions for the realization of higher-performance samples. As to the optimization cycle, the target spectrum was defined via the extinction spectrum simulation of the nanoparticle with the shape derived from electron micrographs. The similarity to the target spectrum and the sampling density in the synthetic space were taken into consideration by the algorithm in order to come up with multiple optimal conditions. The workflow of the closed loop with the procedures of synthesis, analysis, and design of new experiments is illustrated in Fig. 11c. A chemical reaction module that was able to perform parallel synthesis with up to 24 reactors was served as the core robotic hardware. The rotation of the Geneva wheel was taken advantages of to carry out the synthesis efficiently. The liquid handling, pH control, sample transfer, and spectroscopic analysis could be successfully conducted. CAD design and the experimental setup are shown in

Fig. 11d, e. The detailed setup of the autonomous platform for the exploration and optimization of the nanomaterials is demonstrated in Fig. 11f.

## 5.2 To be More Universally Applicable

For the integration of material science and data-driven techniques, it is always in high demand to provide some practical routes for typical laboratory environment even though limited experimental resources are available [87]. Additionally, for the materials highly dependent on the synthetic conditions, it is expected for the AI systems to be standard and robust, so that the high reproducibility can be realized [89]. When it comes to the guarantee of the synthesis reproducibility, certain design of the platforms should be taken into consideration, and it is also essential to conduct the characterization at each step. For example, in an attempt to construct the autonomous platform for the synthesis of high yield and monodispersed nanomaterials which were very sensitivity to the synthetic conditions with the reagent concentrations, temperature, the order of reagent addition, and many other analogous factors included, workflow of the autonomous multistep synthesis was designed [89] as in Fig. 12a. Three graphs including synthesis, reaction, and hardware were required in this case, which was demonstrated by manes of synthesizing six uniquely shaped gold nanoparticles obtained from the previous exploration (Fig. 12b). It was noted that the synthesis graph represented the multistep synthetic procedure in which each node represented a unique nanoparticle and each directed edge showed the hierarchical relation between these nanoparticles (Fig. 12c). It was worthwhile mentioning that in order to verify the reproducibility, the parallel synthesis of six gold nanoparticles was repeated three times, and the standard deviation in the UV-Vis spectra could then be obtained, which is illustrated in Fig. 12d, e. Moreover, it was proposed that the unique signatures for nanomaterials in accordance with their distinctive synthetic protocols were in high demand, and therefore, the universal chemical description language  $\chi$ DL was taken advantages of to create the unique digital signatures, which is demonstrated in Fig. 12f.

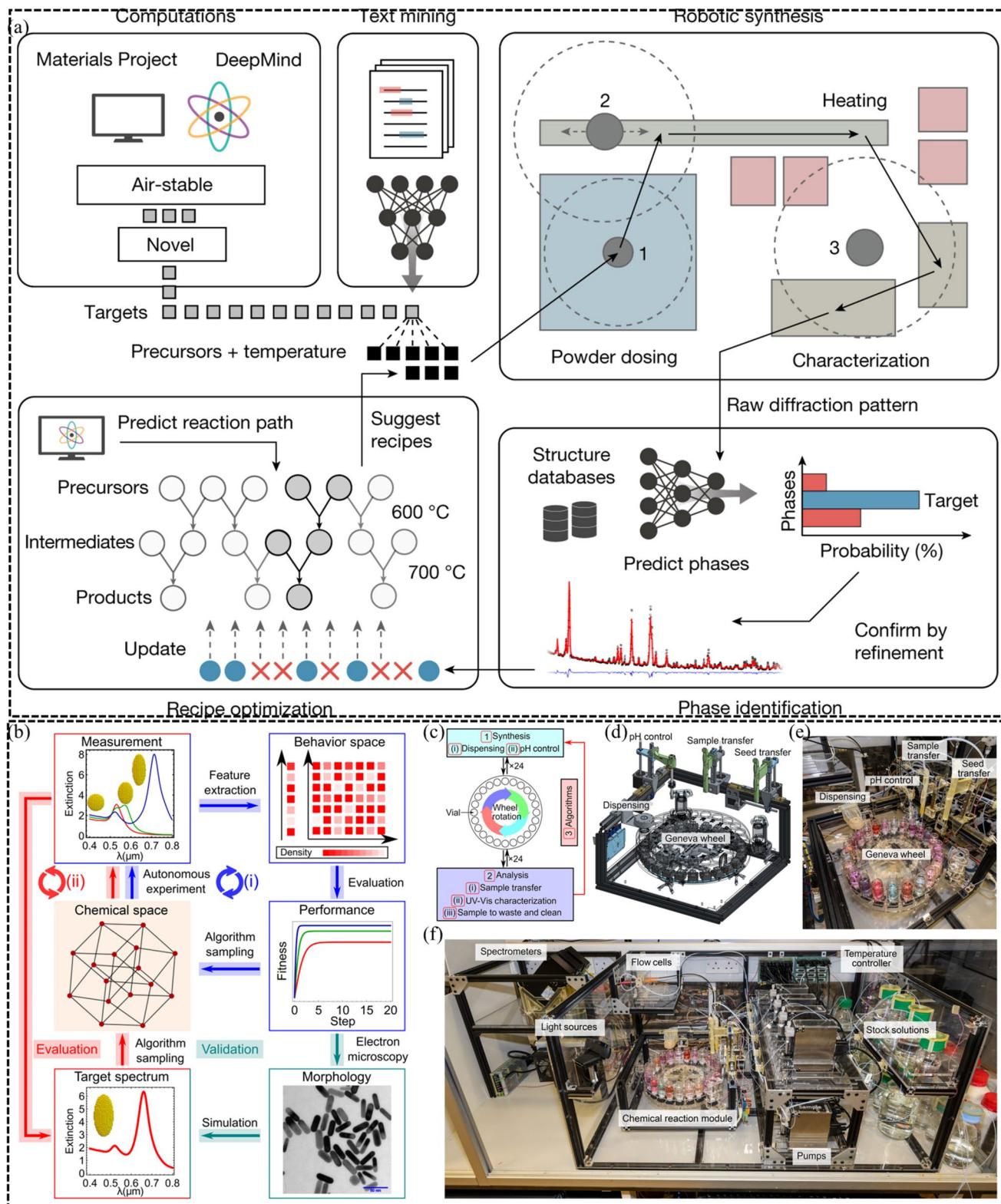
In order to enhance the feasibility, it is ideal for the model trained on one dataset is managed to be used for other occasions operated under different scenarios while using few

training data. One case in point was that the battery charging curve prediction was able to be made by deep neural network (DNN) with 30 points collected in 10 min [180]. Particularly, a piece of the charging curve was applied as the input of a DNN. Key states could then be derived from the estimated entire charging curve. It turned out that a brand new DNN was expensive in regard to cost and time for the reason that extensive battery degradation tests for collecting new training data were required, and data in the real-world were usually incomplete and sparse. DNNs were effective to accomplish this task, which was attributing to its transfer learning feature. The transfer learning was able to resort to the similar knowledge learned from the source dataset so as to improve its performance on the target dataset, reducing the required data amount and saving computational resources. Voltage and capacity sequences gathered from any part of the charging curve could be made use of as the input to the DNN, which facilitated the collection of the input data for the real-world battery management (Fig. 12g, h). The entire constant-current charging curves could be estimated (Fig. 12i). Moreover, the proposed method was able to be quickly adapt to different batteries without much training effort.

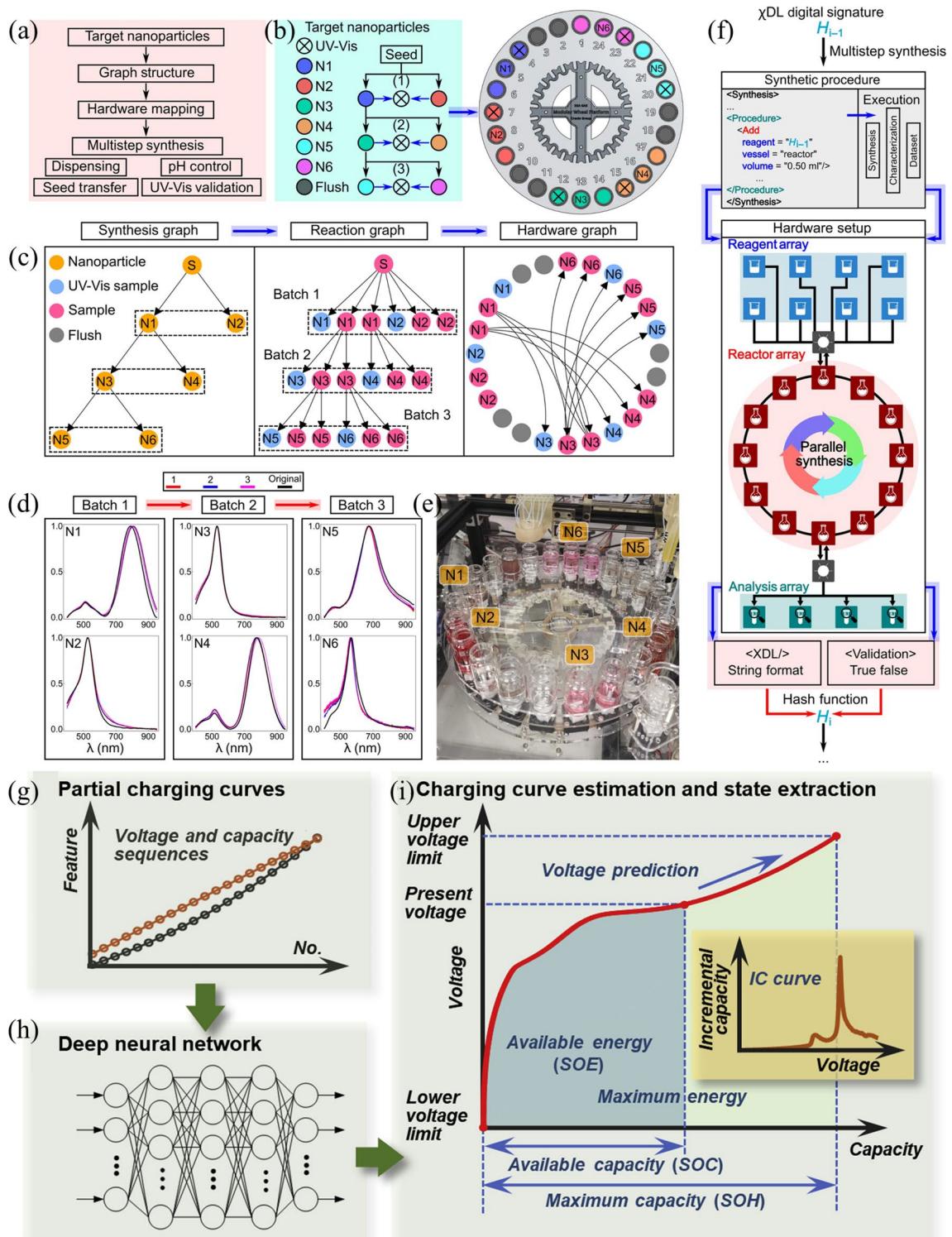
In addition to the methods mentioned above, some other approaches can also be taken advantages of to make the intelligent systems more universally applicable. Particularly, cross-scale data fusion is useful for mapping between microstructure and macroscopic performance by combining atomic simulation with macroscopic characterization, which makes sense for the autonomy and universalization of AI systems for materials. Besides, more comprehensive factors affecting the systems are required to be taken into considerations. For example, the aggregation of most publicly available datasets was utilized for gathering various aging factors when designing the models for making battery lifetime prediction. Moreover, a diversity of dataset is taken advantages of to make an investigation focused on the adaptation of the current models to different cycling protocols.

## 6 Perspectives

Overall, the recent development, including but not limited to AI-assisted cognizance of existing materials and AI empowered new materials discovery, is reviewed in depth. Great progress has been made in the field of AI for material



**Fig. 11** **a** Schematic of A-Lab to discover materials autonomously. Reproduced under the terms of the CC-BY license. [101] Copyright 2023, The Authors, published by Nature. **b** Closed-loop approach of exploration and optimization. **c** Workflow of the closed loop. **d** and **e** CAD design and the experimental setup. **f** Overall setup of the platform. Reproduced under the terms of the CC-BY license. [89] Copyright 2022, The Authors, published by American Association for the Advancement of Science



**Fig. 12** **a** Workflow of the autonomous multistep synthesis platform. **b** Six target nanoparticles. **c** Synthesis, reaction, and hardware graph. **d** UV-Vis spectra of samples. **e** Photographs of the distributed samples. **f** Illustrations of how to generate the unique digital signatures. Reproduced under the terms of the CC-BY license. [89] Copyright 2022, The Authors, published by American Association for the Advancement of Science. **g and h** Voltage and capacity sequences gathered as the input of DNN. **i** Output of the complete charging curve. Reproduced with permission from Ref. [180]. Copyright 2021, Elsevier

science, owing to the enhanced intelligence and high-efficiency (Table 1). The elaborate design lies in every aspect ranging from data exploits to the selection of algorithm (Tables 2 and 3). Progress has been made on the data cleaning, transformation, and processing, as well as proposing the approaches to address the problems when limited data are available. The elaborate selections of models have been made to enable the accuracy, simplicity, and computation efficiency. The diverse experiments have been conducted either by chemists or by robots so that the timely and accurate validation can be guaranteed. AI has boosted the development of various functional materials applied in a diversity of field, including solar cells, nanogenerators, crystals, semiconductors, and so on. There is a growing trend toward the AI systems that are fully autonomous and universally applicable, which lead to the large-scale exploration and more abundant materials to be discovered. This review offers a keen insight into the design ideas for the AI empowered deep cognition of existing materials and fast discovery of novel materials, and some perspectives of the AI systems for materials applied in the future are proposed as follows:

(1) It is ideal for input data more flexible and easier available in the real-world applications. For instance, the intelligent systems for material discovery are expected to be effective even in a standard and simple laboratory. A lot of experimental synthesis data are required by the conventional ML, which proposes big challenge for simple laboratories. Recently, the scheme has been provided for the feasibility of material synthesis in the cases where the available experimental resources are limited, which combined ML techniques with small-scale experiments to accelerate the synthesis of two-dimensional perovskites in a typical laboratory. Additionally, it is also difficult for the investigation of catalysts in which occasion the collection of a large amount of consistent experimental data is always time-consuming. This problem has been successfully solved by the SISSO that can identify potentially nonlinear and intricate correlations between small datasets with a large amount of candidate analytic functions taken into consideration. Besides, the similar issues also exist for the AI-assisted estimation of the maximum battery capacity where a complete charging/discharging curve is needed. However, the complete charge curves are hard to be obtained, since the charging process can start at various states, leading to the record of only pieces of charging/discharging curves. Accordingly, battery charging curve prediction can be made via DNN with

voltage and capacity sequences collected from any part of the charging curve as the input. In the future research, more approaches are expected to be proposed to make the input data much easier to be collected in the real-world applications.

(2) The accurate and comprehensive estimations and predictions about materials assisted by AI are always in high demand. (a) It is expected that the accurate prediction even for some complex issues which are strongly influenced by a series of factors can be realized. Recently, some efforts have been made for the accurate prediction of gas adsorption by the DeepSorption which is a data-driven network with a KCL module, even though every piece of subtle structural information is important for the correct description of adsorption properties. (b) Further to the accurate prediction, the comprehensive reflection of materials is also essential. For the degradation monitoring of battery which calls for the evaluation of battery states over the battery life, method has been developed so that the multiple states can be comprehensively reflected by means of using signals collected from daily battery operation. In the future work, more efforts can be made focused on the fully estimation of the materials with the data obtained from their daily operations when it is required by the practical application.

(3) Endeavors can be made on enhancing the transparency in the predictions of the ML models, which can facilitate the extracting of physical and chemical insights. (a) It is critical to select the models with balanced predictive accuracy and interpretability, which can promote the development of new theories, for the reason that knowledge obtained from the interpretable ML models can accelerate the scientific understanding. For instance, more reliable explanations can be offered by inherently interpretable ML models with functions that can be approximated well via simpler functions concerned with priori knowledge, playing a more important role in extrapolating. Universal ML has been developed for the synthesis of two-dimensional perovskites recently, which can underlie the structure–property relationship in the HTE. (b) Moreover, the unveiling of predictive insights for properties by ML can further facilitate the optimization of the devices with enhanced performance. Efforts have also been made to unveil the complex relationships between piezoelectricity and TENG performance via the principal component analysis (PCA). The in-depth understanding of output offers insights into the energy conversion efficiency, and then the nanogenerator performance can be optimized.

**Table 1** Summary of the state-of-the-art AI systems empowered material science

Material system	Database	Method	Aim	Accuracy	Achievement	Other key features	Refs
Two-dimensional perovskites	by conducting high-throughput experiments	Subgroup discovery and support vector machine	To rapidly screening out materials with high synthesis feasibility	The trained ML model in good performance with an accuracy of 85%	To increase the success rate of the synthesis feasibility by a factor of four relative to traditional approaches	To be effective in a typical laboratory with limited experimental resources	[87]
PVDF-based nanogenerators	Experimental data for various TEENG fabrication parameters	DTR, RF, and GBR	To predict performance	The GBR model exhibiting the highest $R^2$ value of 0.9812 and 0.9370 for train and test	–	With real-time voltage output analysis	[96]
Sodium-ion batteries	DFT	SMOTEENN classifier	To predict the structural stability for materials discovery	With $R^2$ value of 0.962 (Pristine)	To suggest optimal LTMQ candidates with both the high energy density and electrochemical stability	–	[18]
High energy lithium-ion batteries	From experiments	LSTM	To develop high energy Ni-rich cathodes	Prediction accuracy of 80% (except for the stirring rate)	To make a prediction about the reaction conditions for synthesizing cathode precursors with specific morphologies	With the best cathode materials showing a high discharge capacity of $206 \text{ mAh g}^{-1}$ at 0.1C and 83% capacity retention after 200 cycles	[97]
Passivation materials for perovskite solar cells	From a wealth of documented small molecules as passivation materials	SVM, NNM, RF, KNN, and NB	To justify dominant molecular traits and screen excellent passivation materials	SVM's accuracy was the highest, exceeding 80%	To increase absolute efficiency values by over 2% with a champion efficiency of 25.41%	With limited available dataset	[98]
Crystalline porous materials	From the developed MatFormer	KCL	To make prediction of gas adsorption	To realize an 18% increase in $R^2$	To realize a 20–35% decline of the mean absolute error compared to graph neural network CGCNN and ML models based on descriptors	To be universal for predicting the different physicochemical properties of diverse crystalline materials	[102]
Polymer–solvent	From experiments and quantum chemical calculations	DNN	to predict polymer–solvent miscibility	With the solubility classification accuracy of 0.857	To overcome the shortage of the quantitatively limited and qualitatively biased dataset	To offer a highly generalized model for a wide range of polymer solution spaces	[108]

Table 1 (continued)

Material system	Database	Method	Aim	Accuracy	Achievement	Other key features	Refs
Oxides and phosphates	Historical data from the literature	Active learning	For the accelerated synthesis of novel materials	–	To realized 41 novel compounds from a set of 58 targets after continuous operating for over 17 days	To be integrated with robotics	[101]
Multicomponent oxides	Material phases and formation energies got from Materials Project	Thermodynamic strategy	To guide inorganic materials synthesis	–	224 reactions spanning 27 elements with 28 unique precursors could be accomplished by the robot under the operation of 1 human experimenter	To be combined with robotic materials synthesis laboratory	[88]
Nanomaterial	From experiments	Quality-diversity algorithms	To explore and optimize nanomaterials	To realize a yield of up to 95%	To discover five categories of nanoparticles by only performing ca. 1000 experiments in three hierarchically linked chemical spaces	To be integrated with robot, and to transfer materials as seeds between cycles of exploration	[89]
Inorganic material	First-principles calculations	GNNs	For materials discovery	–	To improve the efficiency of materials discovery by an order of magnitude	2.2 million stable crystals have been found by GNoME models with respect to previous work	[51]
Odorants	To curate a reference dataset of ~5000 molecules	MPNN	To map molecular structure to odor perception	–	To outperform chemoinformatic models on several odor prediction tasks	To be contiguous, hierarchical, simple, and parseable	[53]
Battery	The Oxford Battery Degradation Dataset	DNN	To predict battery charging curve	To capture the charging curves accurately with an error of less than 16.9 mAh for 0.74 Ah batteries	To estimate the entire constant-current charging curves	To be adaptable to different batteries without much training effort and to be available with flexible input collected during daily charging	[180]
$\text{CO}_2$ hydrogenation on supported cobalt catalysts	With the experimental, theoretical, and elemental parameters	symbolic-regression AI	To design materials for catalysis	–	To obtain a model well representing the experimental $\text{CH}_3\text{OH}$ selectivity	To identify key parameters as materials genes	[91]



Table 1 (continued)

Material system	Database	Method	Aim	Accuracy	Achievement	Other key features	Refs
Low-contact electrode for 2D semiconductor transistor	From the C2DB database	Autoencoding regularized adversarial neural network and a feature-adaptive variational active learning algorithm	To Screen electrode	With the mean square errors of 0.17 and 0.27 eV for the vertical and lateral Schottky barrier	To be trained with only 15% of the total data points	To be cost-effective and scalable	[157]
Solid-state electrolytes	Structure candidates based on ionic substitution to known crystal structures	With state-of-the-art ML models and traditional physics-based models	To realize screening in large scale and experimental validation	–	To quickly navigate through more than 32 million candidates and predict around half a million potentially stable materials	To combine ML with HPC	[99]
Metal oxides	From experiments	GA-adjusted ANN	To exploit gas-sensing descriptors	With $R^2$ value of 0.92	–	–	[170]
Electrocatalysis for complex C–C coupling reaction networks	Iterative sampling	2D-3D ensemble ML	To reveals C–C electro coupling mechanism	–	To prove the ability of big dataset generated from ML to accelerate quantum chemical computations	With big dataset analysis	[171]
Biochar-based electrodes for supercapacitors	From recent articles	GBR	To clarify relationship between biochar preparation procedures and capacitance characteristics	with an $R^2$ value of 0.93	–	–	[172]
Hierarchical porous carbon	From papers	ANN	To predict advanced potassium-ion batteries	With the MAE values of 8.830 and 2.390 for the capacity and ICE	To predict the relationship between different factors and the performance	–	[114]
Compounds for opto-electronic devices	–	Evolutionary search algorithms	To explore new compounds with structures similar to Haecelites	–	To discover 13 new Haecelite compounds	–	[52]

**Table 2** Summary of some database used in materials science

Name	Contents	Sources and scale	Aim
The Materials Project	Data and associated analysis algorithms	With all known solid-state materials, over thousands more unknown materials, over 30,000 stable materials and over 6,000 materials with computed elastic tensors	To provide open web-based access to computed information on known and predicted materials as well as powerful analysis tools to inspire and design novel materials
QMD	DFT calculated thermodynamic and structural properties	With 1,317,811 materials	To search for material compositions, create phase diagrams, and visualize crystal structures
NOMAD	Data	Combining data from popular sources like the Materials Project, AFLOW, and OQMD	To extract and publish structured data with rich metadata
AFLOW (Automatic AFLOW for Materials Discovery)	Data and calculated properties	Database of 3,929,948 material compounds with over 817,429,184 calculated properties, and growing	–
JARVIS (Joint Automated Repository for Various Integrated Simulations)	Data	Using classical force-field, density functional theory, machine learning, quantum computation calculations, and experiments	To automate materials discovery and optimization
Catalysis-Hub	Data and software	With thousands of reaction energies and barriers from density functional theory (DFT) calculations on surface systems	For computational catalysis research
ICSD	Database for completely determined inorganic crystal structures	Containing an almost exhaustive list of known inorganic crystal structures published since 1913, including their atomic coordinates	To find answers on questions in materials research
Crystallography Open Database (COD)	Data and software	Currently there are 527,197 entries in the COD	To provide open-access collection of crystal structures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers
NIST Materials Data Facility (MDF)	Data	> 650 Datasets, > 80 TB of Materials Data Published, and > 100 Data Sources Indexed	For publish, discover, and access materials datasets
PubChem	Data for chemical information	With 122MCompounds, 338MSubstances, 297MBioactivities, and 44MLiterature	To search chemicals, find chemical and physical properties, biological activities, safety and toxicity information, patents, literature citations
ZINC	Data	With over 230 million purchasable compounds in ready-to-dock, 3D formats. ZINC also contains over 750 million purchasable compounds you can search for analogs in under a minute	For virtual screening



Table 2 (continued)

Name	Contents	Sources and scale	Aim
PolyInfo	Data	The main data source is academic literature on polymers. With 19,227 homopolymers, 8,321 copolymers, 2,788 polymer blends, 3,209 composites, 174,968 polymer sample, 552,427 property points, and 21,793 literature data	To provide various data required for polymeric material design

(4) The explorations on more abundant types and properties of functional materials assisted by AI are expected to be made, so as to make full use of AI. a) More factors and indexes are expected to be investigated by the data-driven techniques in material science. The synthesizability of novel materials with a variety of oxides and phosphates included has been explored by an autonomous laboratory combining computations, historical data, and ML. In the future, more factors with microstructures of materials and performances of various devices can be taken into consideration. b) In addition to inorganic materials and metals, AI can also be applied in the research of organic materials and composites. Multitask ML has been taken advantages of to predict the polymer–solvent miscibility. Active learning has also been used to make exploration of transition metal complexes. In the future work, more efforts can be made to investigate abundant types of materials assisted by AI to achieve the large-scale exploration and accelerated material discovery. c) In regard to the in-depth cognition of existing materials, the mapping of physical stimuli to a variety of perceptual features is in high demand. Detailed maps have led to a better understanding of visual and auditory coding. A principal odor map has also been proposed for olfactory perception. Accordingly, future research can be conducted focused on the revealing of the relationships between the physical stimuli and more diverse perceptual characteristics.

**Table 3** Summary of various algorithms and models for different types of materials

Types of materials	Algorithms/Models	Aim	Experimental performance	Achievement	Refs
Catalysts	Large language models (LLMs) and genetic algorithms (GAs)	For the development of high-entropy alloy (HEA) catalysts	The optimal IrCuNiPdPt/C catalyst exhibits the record-low HER overpotentials at 10 and 100 mA cm <sup>-2</sup> , surpassing commercial Pt/C by 49% and 18%	To slash the discovery time from millennia to hours	[181]
Catalysis	Meta-learning model	For selectivity prediction in asymmetric catalysis	–	To demonstrate significant performance improvement over other popular ML methods, like random forests and graph neural networks	[182]
Polymers/ Hydrogels	Gaussian process (GP) and random forest regression (RFR)	For the design of high-performance adhesive hydrogels	To realize remarkable enhancement in adhesive strength, with a maximum value exceeding 1 MPa	To optimize hydrogel formulations from an initial dataset of 180 bioinspired hydrogels	[183]
Polymers	Feed-forward neural network	To accelerate the discovery of heat-resistant polysulfates	With a polysulfate exhibiting good thermal resilience and ultrahigh discharged energy density with over 90% efficiency at 200 °C	To provide the prediction of key proxy parameters and down selection of polymer candidates from a library of nearly 50,000 polysulfates	[184]
Catalysts	Crystal graph convolutional neural networks	To design a high-entropy intermetallic compound for catalyzing oxygen reduction reaction (ORR)	The catalyst with small particle size is successfully synthesized with ultrahigh mass activity and specific activity	To provide a high prediction accuracy with mean absolute errors of 0.003 for surface strain and 0.011 eV atom 1 for formation energy	[185]
Crystals	Graph networks	For improving the efficiency of materials discovery	Among the stable structures, 736 have already been independently experimentally realized	To enable the discovery of 2.2 million structures below the current convex hull	[51]
Catalysts	Regression models	To accelerate the discovery of catalysis	The screened irregular catalyst shows outstanding sulfur and moisture resistance and long-term stability (> 7000 h, $T_{90} = 345$ °C)	The high prediction accuracy could be realized with a small-size training set	[186]
Catalysts	Extra Trees model	For accelerated screening of single-atom anchored MXenes electrocatalyst	Experimental validation was achieved by 10 synthesized MXene-SACs	With an effective intrinsic descriptor to accelerate the high-throughput screening without additional computations	[187]



**Acknowledgements** This work was supported by the Hong Kong Polytechnic University (Project No. 4-ZZW1, 4-YWER, 97D9, 4-W443).

**Author Contributions** Ying Cao helped in methodology, investigation, and writing—original draft. Jian Lu, Yuejiao Chen, Titao Jing, and Xi Fan contributed to methodology. Hong Fu helped in resources, methodology, and writing—review and editing. Bingang Xu led in conceptualization, funding acquisition, methodology, supervision, and writing—review and editing.

### Declarations

**Conflict of interest** The authors declare no conflict of interest. They have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

**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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### References

1. A.M. Mroz, V. Posligua, A. Tarzia, E.H. Wolpert, K.E. Jelfs, Into the unknown: how computation can help explore uncharted material space. *J. Am. Chem. Soc.* **144**(41), 18730–18743 (2022). <https://doi.org/10.1021/jacs.2c06833>
2. J. Zhou, J. Lin, X. Huang, Y. Zhou, Y. Chen et al., A library of atomically thin metal chalcogenides. *Nature* **556**(7701), 355–359 (2018). <https://doi.org/10.1038/s41586-018-0008-3>
3. S. Jiang, Y. Zhou, W. Xu, Q. Xia, M. Yi et al., Machine learning-driven optimization and application of bimetallic catalysts in peroxyomonosulfate activation for degradation of fluoroquinolone antibiotics. *Chem. Eng. J.* **486**, 150297 (2024). <https://doi.org/10.1016/j.cej.2024.150297>
4. T. Jing, B. Xu, Y. Yang, M. Li, Y. Gao, Organogel electrode enables highly transparent and stretchable triboelectric nanogenerators of high power density for robust and reliable energy harvesting. *Nano Energy* **78**, 105373 (2020). <https://doi.org/10.1016/j.nanoen.2020.105373>
5. Y. Liu, B. Xie, Q. Hu, R. Zhao, Q. Zheng et al., Regulating the Helmholtz plane by trace polarity additive for long-life Zn ion batteries. *Energy Storage Mater.* **66**, 103202 (2024). <https://doi.org/10.1016/j.ensm.2024.103202>
6. J. Wen, B. Xu, J. Zhou, Toward flexible and wearable embroidered supercapacitors from cobalt phosphides-decorated conductive fibers. *Nano-Micro Lett.* **11**(1), 89 (2019). <https://doi.org/10.1007/s40820-019-0321-x>
7. J. Li, B. Xu, Novel highly sensitive and wearable pressure sensors from conductive three-dimensional fabric structures. *Smart Mater. Struct.* **24**(12), 125022 (2015). <https://doi.org/10.1088/0964-1726/24/12/125022>
8. X. Guan, B. Xu, J. Huang, T. Jing, Y. Gao, Fiber-shaped stretchable triboelectric nanogenerator with a novel synergistic structure of opposite Poisson's ratios. *Chem. Eng. J.* **427**, 131698 (2022). <https://doi.org/10.1016/j.cej.2021.131698>
9. B. Niu, T. Hua, B. Xu, Robust deposition of silver nanoparticles on paper assisted by polydopamine for green and flexible electrodes. *ACS Sustainable Chem. Eng.* **8**(34), 12842–12851 (2020). <https://doi.org/10.1021/acssuschemeng.0c03098>
10. K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science. *Nature* **559**(7715), 547–555 (2018). <https://doi.org/10.1038/s41586-018-0337-2>
11. P.M. Attia, A. Grover, N. Jin, K.A. Severson, T.M. Markov et al., Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* **578**(7795), 397–402 (2020). <https://doi.org/10.1038/s41586-020-1994-5>
12. Q. Zhou, S. Lu, Y. Wu, J. Wang, Property-oriented material design based on a data-driven machine learning technique. *J. Phys. Chem. Lett.* **11**(10), 3920–3927 (2020). <https://doi.org/10.1021/acs.jpclett.0c00665>
13. S. Lu, Q. Zhou, X. Chen, Z. Song, J. Wang, Inverse design with deep generative models: next step in materials discovery. *Natl. Sci. Rev.* **9**(8), nwac111 (2022). <https://doi.org/10.1093/nsr/nwac111>
14. H. Shi, Y. Shi, Z. Liang, S. Zhao, B. Qiao et al., Machine learning-enabled discovery of multi-resonance TADF molecules: unraveling PLQY predictions from molecular structures. *Chem. Eng. J.* **494**, 153150 (2024). <https://doi.org/10.1016/j.cej.2024.153150>
15. J. Sun, D. Li, Y. Wang, T. Xie, Y. Zou et al., Accelerating the generation and discovery of high-performance donor materials for organic solar cells by deep learning. *J. Mater. Chem. A* **12**(33), 21813–21823 (2024). <https://doi.org/10.1039/D4TA0394K>
16. Y. Li, X. Zhang, T. Li, Y. Chen, Y. Liu et al., Accelerating materials discovery for electrocatalytic water oxidation *via* center-environment deep learning in spinel oxides. *J. Mater. Chem. A* **12**(30), 19362–19377 (2024). <https://doi.org/10.1039/d4ta02771j>
17. M. Omidvar, H. Zhang, A.A. Ihalage, T.G. Saunders, H. Giddens et al., Accelerated discovery of perovskite solid solutions through automated materials synthesis and characterization. *Nat. Commun.* **15**(1), 6554 (2024). <https://doi.org/10.1038/s41467-024-50884-y>
18. M. Kim, W.-H. Yeo, K. Min, Co-free and low strain cathode materials for sodium-ion batteries: machine learning-based materials discovery. *Energy Storage Mater.* **69**, 103405 (2024). <https://doi.org/10.1016/j.ensm.2024.103405>

19. H. Wang, M. Jiang, G. Xu, C. Wang, X. Xu et al., Machine learning-guided prediction of desalination capacity and rate of porous carbons for capacitive deionization. *Small* **20**(42), e2401214 (2024). <https://doi.org/10.1002/smll.202401214>

20. J. Liu, R. Sun, X. Bao, J. Yang, Y. Chen et al., Machine learning driven atom-thin materials for fragrance sensing. *Small* **21**(28), 2401066 (2025). <https://doi.org/10.1002/smll.202401066>

21. M. Wu, Z. Song, Y. Cui, Z. Fu, K. Hong et al., Machine learning-assisted design of nitrogen-rich covalent triazine frameworks photocatalysts. *Adv. Funct. Mater.* **35**(3), 2413453 (2025). <https://doi.org/10.1002/adfm.202413453>

22. A. Maqbool, A. Khalad, N.Z. Khan, Prediction of corrosion rate for friction stir processed WE43 alloy by combining PSO-based virtual sample generation and machine learning. *J. Magnes. Alloys* **12**(4), 1518–1528 (2024). <https://doi.org/10.1016/j.jma.2024.04.012>

23. C. Li, L. Bao, Y. Ji, Z. Tian, M. Cui et al., Combining machine learning and metal–organic frameworks research: novel modeling, performance prediction, and materials discovery. *Coord. Chem. Rev.* **514**, 215888 (2024). <https://doi.org/10.1016/j.ccr.2024.215888>

24. G. Huang, F. Huang, W. Dong, Machine learning in energy storage material discovery and performance prediction. *Chem. Eng. J.* **492**, 152294 (2024). <https://doi.org/10.1016/j.cej.2024.152294>

25. Y. She, H. Liu, H. Yuan, Y. Li, X. Liu et al., Artificial intelligence-assisted conductive hydrogel dressings for refractory wounds monitoring. *Nano-Micro Lett.* **17**(1), 319 (2025). <https://doi.org/10.1007/s40820-025-01834-w>

26. S. Wang, J. Liu, X. Song, H. Xu, Y. Gu et al., Artificial intelligence empowers solid-state batteries for material screening and performance evaluation. *Nano-Micro Lett.* **17**(1), 287 (2025). <https://doi.org/10.1007/s40820-025-01797-y>

27. K. Zheng, C. Zheng, L. Zhu, B. Yang, X. Jin et al., Machine learning enabled reusable adhesion, entangled network-based hydrogel for long-term, high-fidelity EEG recording and attention assessment. *Nano-Micro Lett.* **17**(1), 281 (2025). <https://doi.org/10.1007/s40820-025-01780-7>

28. F. Zheng, B. Yuan, Y. Cai, H. Xiang, C. Tang et al., Machine learning tailored anodes for efficient hydrogen energy generation in proton-conducting solid oxide electrolysis cells. *Nano-Micro Lett.* **17**(1), 274 (2025). <https://doi.org/10.1007/s40820-025-01764-7>

29. J. Chang, J. Li, J. Ye, B. Zhang, J. Chen et al., AI-enabled piezoelectric wearable for joint torque monitoring. *Nano-Micro Lett.* **17**(1), 247 (2025). <https://doi.org/10.1007/s40820-025-01753-w>

30. Y. Xiao, H. Li, T. Gu, X. Jia, S. Sun et al.,  $Ti_3C_2T_x$  composite aerogels enable pressure sensors for dialect speech recognition assisted by deep learning. *Nano-Micro Lett.* **17**(1), 101 (2024). <https://doi.org/10.1007/s40820-024-01605-z>

31. B.H. Jeong, J. Lee, M. Ku, J. Lee, D. Kim et al., RGB color-discriminable photonic synapse for neuromorphic vision system. *Nano-Micro Lett.* **17**(1), 78 (2024). <https://doi.org/10.1007/s40820-024-01579-y>

32. Z. Zhou, Z. Zhang, P. Li, Z. Guan, Y. Li et al., Low energy consumption photoelectric memristors with multi-level linear conductance modulation in artificial visual systems application. *Nano-Micro Lett.* **17**(1), 317 (2025). <https://doi.org/10.1007/s40820-025-01816-y>

33. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov et al., Highly accurate protein structure prediction with AlphaFold. *Nature* **596**(7873), 583–589 (2021). <https://doi.org/10.1038/s41586-021-03819-2>

34. R. Unni, M. Zhou, P.R. Wiecha, Y. Zheng, Advancing materials science through next-generation machine learning. *Curr. Opin. Solid State Mater. Sci.* **30**, 101157 (2024). <https://doi.org/10.1016/j.cossms.2024.101157>

35. Y. Wang, K. Wang, C. Zhang, Applications of artificial intelligence/machine learning to high-performance composites. *Compos. Part B Eng.* **285**, 111740 (2024). <https://doi.org/10.1016/j.compositesb.2024.111740>

36. R. Du, R. Xin, H. Wang, W. Zhu, R. Li et al., Machine learning: an accelerator for the exploration and application of advanced metal–organic frameworks. *Chem. Eng. J.* **490**, 151828 (2024). <https://doi.org/10.1016/j.cej.2024.151828>

37. N. Ling, Y. Wang, S. Song, C. Liu, F. Yang et al., Experimentally validated screening strategy for alloys as anode in Mg-air battery with multi-target machine learning predictions. *Chem. Eng. J.* **496**, 153824 (2024). <https://doi.org/10.1016/j.cej.2024.153824>

38. S. Gao, Y. Cheng, L. Chen, S. Huang, Rapid discovery of gas response in materials *via* density functional theory and machine learning. *Energy Environ. Mater.* **8**(1), e12816 (2025). <https://doi.org/10.1002/eem2.12816>

39. S. Poudel, R. Thapa, R. Basnet, A. Timofiejczuk, A. Kunwar, PiezoTensorNet: Crystallography informed multi-scale hierarchical machine learning model for rapid piezoelectric performance finetuning. *Appl. Energy* **361**, 122901 (2024). <https://doi.org/10.1016/j.apenergy.2024.122901>

40. L. Wang, H. Chen, L. Yang, J. Li, Y. Li et al., Single-atom catalysts property prediction *via* supervised and self-supervised pre-training models. *Chem. Eng. J.* **487**, 150626 (2024). <https://doi.org/10.1016/j.cej.2024.150626>

41. P. Guo, W. Meng, Y. Bao, Knowledge-guided data-driven design of ultra-high-performance geopolymers (UHPG). *Cem. Concr. Compos.* **153**, 105723 (2024). <https://doi.org/10.1016/j.cemconcomp.2024.105723>

42. L. Chen, M. Ren, J. Zhou, X. Zhou, F. Liu et al., Bioinspired iontronic synapse fibers for ultralow-power multiplexing neuromorphic sensorimotor textiles. *Proc. Natl. Acad. Sci. U.S.A.* **121**(33), e2407971121 (2024). <https://doi.org/10.1073/pnas.2407971121>

43. Y. Li, Z. Qiu, H. Kan, Y. Yang, J. Liu et al., A human-computer interaction strategy for an FPGA platform boosted integrated “perception-memory” system based on electronic tattoos and memristors. *Adv. Sci.* **11**(39), e2402582 (2024). <https://doi.org/10.1002/advs.202402582>

44. H. Zhang, H. Li, Y. Li, Biomimetic electronic skin for robots aiming at superior dynamic-static perception and material cognition based on triboelectric-piezoresistive effects. *Nano*

Lett. **24**(13), 4002–4011 (2024). <https://doi.org/10.1021/acs.nanolett.4c00623>

45. Y. Li, Q. Lin, T. Sun, M. Qin, W. Yue et al., A perceptual and interactive integration strategy toward telemedicine healthcare based on electroluminescent display and triboelectric sensing 3d stacked device. *Adv. Funct. Mater.* **34**(40), 2402356 (2024). <https://doi.org/10.1002/adfm.202402356>

46. W. Yang, H. Kan, G. Shen, Y. Li, A network intrusion detection system with broadband  $\text{WO}_{3-x}/\text{WO}_{3-x}\text{-Ag}/\text{WO}_{3-x}$  optoelectronic memristor. *Adv. Funct. Mater.* **34**(23), 2312885 (2024). <https://doi.org/10.1002/adfm.202312885>

47. S. Deng, B. Xu, J. Zhao, C.W. Kan, X. Liu, Unlocking double redox reaction of metal–organic framework for aqueous zinc-ion battery. *Angew. Chem. Int. Ed.* **63**(17), e202401996 (2024). <https://doi.org/10.1002/anie.202401996>

48. B. Niu, T. Hua, H. Hu, B. Xu, X. Tian et al., A highly durable textile-based sensor as a human-worn material interface for long-term multiple mechanical deformation sensing. *J. Mater. Chem. C* **7**(46), 14651–14663 (2019). <https://doi.org/10.1039/C9TC04006D>

49. J. Xu, K.K. Wang, T. Liu, Y. Peng, B.G. Xu, Br-doped  $\text{Bi}_2\text{O}_2\text{CO}_3$  exposed (001) crystal facets with enhanced photocatalytic activity. *CrystEngComm* **19**(34), 5001–5007 (2017). <https://doi.org/10.1039/c7ce00924k>

50. Y. Wang, Z. Li, H. Fu, B. Xu, Sustainable triboelectric nanogenerators based on recycled materials for biomechanical energy harvesting and self-powered sensing. *Nano Energy* **115**, 108717 (2023). <https://doi.org/10.1016/j.nanoen.2023.108717>

51. A. Merchant, S. Batzner, S.S. Schoenholz, M. Aykol, G. Cheon et al., Scaling deep learning for materials discovery. *Nature* **624**(7990), 80–85 (2023). <https://doi.org/10.1038/s41586-023-06735-9>

52. E. Alibagheri, A. Ranjbar, M. Khazaei, T.D. Kühne, S.M. Vaez Allaei, Remarkable optoelectronic characteristics of synthesizable square-octagon haeckelite structures: machine learning materials discovery. *Adv. Funct. Mater.* **34**(27), 2402390 (2024). <https://doi.org/10.1002/adfm.202402390>

53. B.K. Lee, E.J. Mayhew, B. Sanchez-Lengeling, J.N. Wei, W.W. Qian et al., A principal odor map unifies diverse tasks in olfactory perception. *Science* **381**(6661), 999–1006 (2023). <https://doi.org/10.1126/science.ade4401>

54. V.S. Tuchin, E.A. Stepanidenko, A.A. Vedernikova, S.A. Cherevkov, D. Li et al., Optical properties prediction for red and near-infrared emitting carbon dots using machine learning. *Small* **20**(29), 2310402 (2024). <https://doi.org/10.1002/smll.202310402>

55. S. Lu, A. Jayaraman, Machine learning for analyses and automation of structural characterization of polymer materials. *Prog. Polym. Sci.* **153**, 101828 (2024). <https://doi.org/10.1016/j.progpolymsci.2024.101828>

56. Y. Srivastava, A. Jain, Accelerating thermal conductivity prediction through machine-learning: two orders of magnitude reduction in phonon-phonon scattering rates calculation. *Mater. Today Phys.* **41**, 101345 (2024). <https://doi.org/10.1016/j.mtphys.2024.101345>

57. L. Guo, X. Xu, Q. Wang, J. Park, H. Lei et al., Machine learning-based prediction of heavy metal immobilization rate in the solidification/stabilization of municipal solid waste incineration fly ash (MSWIFA) by geopolymers. *J. Hazard. Mater.* **467**, 133682 (2024). <https://doi.org/10.1016/j.jhazmat.2024.133682>

58. Q. Liu, W. Chen, V. Yakubov, J.J. Kruzic, C.H. Wang et al., Interpretable machine learning approach for exploring process-structure-property relationships in metal additive manufacturing. *Addit. Manuf.* **85**, 104187 (2024). <https://doi.org/10.1016/j.addma.2024.104187>

59. X. Chen, S. Lu, Q. Chen, Q. Zhou, J. Wang, From bulk effective mass to 2D carrier mobility accurate prediction via adversarial transfer learning. *Nat. Commun.* **15**(1), 5391 (2024). <https://doi.org/10.1038/s41467-024-49686-z>

60. D. Kwon, D. Kim, Machine learning interatomic potentials in engineering perspective for developing cathode materials. *J. Mater. Chem. A* **12**(35), 23837–23847 (2024). <https://doi.org/10.1039/D4TA03452J>

61. W.A. Dunlap-Shohl, Y. Meng, P.P. Sunkari, D.A.C. Beck, M. Meilä et al., Physiochemical machine learning models predict operational lifetimes of  $\text{CH}_3\text{NH}_3\text{PbI}_3$  perovskite solar cells. *J. Mater. Chem. A* **12**(16), 9730–9746 (2024). <https://doi.org/10.1039/D3TA06668A>

62. Z. Wang, L. Wang, H. Zhang, H. Xu, X. He, Materials descriptors of machine learning to boost development of lithium-ion batteries. *Nano Convergence* **11**(1), 8 (2024). <https://doi.org/10.1186/s40580-024-00417-6>

63. C. Jin, C. Zhang, P. Yan, M. Jiang, R. Yin et al., A super-human sensing triboelectric nanogenerator with boosted power density and durability via a bio-inspired Janus structure. *Adv. Funct. Mater.* **34**(37), 2402233 (2024). <https://doi.org/10.1002/adfm.202402233>

64. J. Li, M. Zhou, H.-H. Wu, L. Wang, J. Zhang et al., Machine learning-assisted property prediction of solid-state electrolyte. *Adv. Energy Mater.* **14**(20), 2304480 (2024). <https://doi.org/10.1002/aenm.202304480>

65. X. Zhang, B. Ding, Y. Wang, Y. Liu, G. Zhang et al., Machine learning for screening small molecules as passivation materials for enhanced perovskite solar cells. *Adv. Funct. Mater.* **34**(30), 2314529 (2024). <https://doi.org/10.1002/adfm.202314529>

66. Z. Zheng, J. Zhou, Y. Zhu, Computational approach inspired advancements of solid-state electrolytes for lithium secondary batteries: from first-principles to machine learning. *Chem. Soc. Rev.* **53**(6), 3134–3166 (2024). <https://doi.org/10.1039/D3CS00572K>

67. H. Mai, X. Wen, X. Li, N.S.L. Dissanayake, X. Sun et al., Data driven high quantum yield halide perovskite phosphors design and fabrication. *Mater. Today* **74**, 12–21 (2024). <https://doi.org/10.1016/j.mattod.2024.02.002>

68. Y. Liu, X. Tan, P. Xiang, Y. Tu, T. Shao et al., Machine learning as a characterization method for analysis and design of perovskite solar cells. *Mater. Today Phys.* **42**, 101359 (2024). <https://doi.org/10.1016/j.mtphys.2024.101359>

69. Y. Haghshenas, W.P. Wong, V. Sethu, R. Amal, P.V. Kumar et al., Full prediction of band potentials in semiconductor materials. *Mater. Today Phys.* **46**, 101519 (2024). <https://doi.org/10.1016/j.mtphys.2024.101519>

70. H. Ji, Y. Tian, C. Fu, H. Ye, Transfer learning enables prediction of steel corrosion in concrete under natural environments. *Cem. Concr. Compos.* **148**, 105488 (2024). <https://doi.org/10.1016/j.cemconcomp.2024.105488>

71. P. Akbari, M. Zamani, A. Mostafaei, Machine learning prediction of mechanical properties in metal additive manufacturing. *Addit. Manuf.* **91**, 104320 (2024). <https://doi.org/10.1016/j.addma.2024.104320>

72. Y. Fu, A. Howard, C. Zeng, Y. Chen, P. Gao et al., Physics-guided continual learning for predicting emerging aqueous organic redox flow battery material performance. *ACS Energy Lett.* **9**(6), 2767–2774 (2024). <https://doi.org/10.1021/acsenergylett.4c00493>

73. X. Mei, T. Miyoshi, Y. Suzuki, A dynamic model of wrist-worn rotational electret energy harvester: theoretical investigation and experimental validations. *Appl. Energy* **373**, 123888 (2024). <https://doi.org/10.1016/j.apenergy.2024.123888>

74. W. Hamada, M. Hishida, R. Sugiura, H. Tobita, H. Imai et al., Efficient design and synthesis of an amorphous conjugated polymer network for a metal-free electrocatalyst of hydrogen evolution reaction. *J. Mater. Chem. A* **12**(6), 3294–3303 (2024). <https://doi.org/10.1039/D3TA06447F>

75. Y. Gu, S. Tang, X. Liu, X. Liang, Q. Zhu et al., Stability prediction of gold nanoclusters with different ligands and doped metals: deep learning and experimental tests. *J. Mater. Chem. A* **12**(8), 4460–4472 (2024). <https://doi.org/10.1039/d3ta06892g>

76. Z.-P. Lin, Y. Li, S.A. Haque, A.M. Ganose, A. Kafizas, Insights from experiment and machine learning for enhanced  $TiO_2$  coated glazing for photocatalytic  $NO_x$  remediation. *J. Mater. Chem. A* **12**(22), 13281–13298 (2024). <https://doi.org/10.1039/D4TA01319K>

77. W. Li, Y. Wen, K. Wang, Z. Ding, L. Wang et al., Developing a machine learning model for accurate nucleoside hydrogels prediction based on descriptors. *Nat. Commun.* **15**(1), 2603 (2024). <https://doi.org/10.1038/s41467-024-46866-9>

78. X. Sun, L. Yue, L. Yu, C.T. Forte, C.D. Armstrong et al., Machine learning-enabled forward prediction and inverse design of 4D-printed active plates. *Nat. Commun.* **15**(1), 5509 (2024). <https://doi.org/10.1038/s41467-024-49775-z>

79. L. Gao, J. Lin, L. Wang et al., Machine learning-assisted design of advanced polymeric materials. *Acc. Mater. Res.* **5**(5), 571–584 (2024). <https://doi.org/10.1021/accountsmr.3c00288>

80. S. Jiang, W. Xu, Q. Xia, M. Yi, Y. Zhou et al., Application of machine learning in the study of cobalt-based oxide catalysts for antibiotic degradation: an innovative reverse synthesis strategy. *J. Hazard. Mater.* **471**, 134309 (2024). <https://doi.org/10.1016/j.jhazmat.2024.134309>

81. Y. Zhang, C. Wen, P. Dang, T. Lookman, D. Xue et al., Toward ultra-high strength high entropy alloys *via* feature engineering. *J. Mater. Sci. Technol.* **200**, 243–252 (2024). <https://doi.org/10.1016/j.jmst.2024.02.058>

82. D. Park, J. Lee, H. Lee, G.X. Gu, S. Ryu, Deep generative spatiotemporal learning for integrating fracture mechanics in composite materials: inverse design, discovery, and optimization. *Mater. Horiz.* **11**(13), 3048–3065 (2024). <https://doi.org/10.1039/D4MH00337C>

83. M. Seifrid, S. Lo, D.G. Choi, G. Tom, M.L. Le et al., Beyond molecular structure: critically assessing machine learning for designing organic photovoltaic materials and devices. *J. Mater. Chem. A* **12**(24), 14540–14558 (2024). <https://doi.org/10.1039/D4TA01942C>

84. K. Li, K. Choudhary, B. DeCost, M. Greenwood, J. Hattrick-Simpers, Efficient first principles based modeling *via* machine learning: from simple representations to high entropy materials. *J. Mater. Chem. A* **12**(21), 12412–12422 (2024). <https://doi.org/10.1039/D4TA00982G>

85. M. Chen, Z. Yin, Z. Shan, X. Zheng, L. Liu et al., Application of machine learning in perovskite materials and devices: a review. *J. Energy Chem.* **94**, 254–272 (2024). <https://doi.org/10.1016/j.jechem.2024.02.035>

86. R.-L. Liu, J. Wang, Z.-H. Shen, Y. Shen, Ai for dielectric capacitors. *Energy Storage Mater.* **71**, 103612 (2024). <https://doi.org/10.1016/j.ensm.2024.103612>

87. Y. Wu, C.-F. Wang, M.-G. Ju, Q. Jia, Q. Zhou et al., Universal machine learning aided synthesis approach of two-dimensional perovskites in a typical laboratory. *Nat. Commun.* **15**(1), 138 (2024). <https://doi.org/10.1038/s41467-023-44236-5>

88. J. Chen, S.R. Cross, L.J. Miara, J.-J. Cho, Y. Wang et al., Navigating phase diagram complexity to guide robotic inorganic materials synthesis. *Nat. Synth.* **3**(5), 606–614 (2024). <https://doi.org/10.1038/s44160-024-00502-y>

89. Y. Jiang, D. Salley, A. Sharma, G. Keenan, M. Mullin et al., An artificial intelligence enabled chemical synthesis robot for exploration and optimization of nanomaterials. *Sci. Adv.* **8**(40), eabo2626 (2022). <https://doi.org/10.1126/sciadv.abo2626>

90. L.C. Erhard, J. Rohrer, K. Albe, V.L. Deringer, Modelling atomic and nanoscale structure in the silicon–oxygen system through active machine learning. *Nat. Commun.* **15**, 1927 (2024). <https://doi.org/10.1038/s41467-024-45840-9>

91. R. Miyazaki, K.S. Belthle, H. Tüysüz, L. Foppa, M. Scheffler, Materials genes of  $CO_2$  hydrogenation on supported cobalt catalysts: an artificial intelligence approach integrating theoretical and experimental data. *J. Am. Chem. Soc.* **146**(8), 5433–5444 (2024). <https://doi.org/10.1021/jacs.3c12984>

92. Q. Zhai, H. Jiang, N. Long, Q. Kang, X. Meng et al., Machine learning for full lifecycle management of lithium-ion batteries. *Renew. Sustain. Energy Rev.* **202**, 114647 (2024). <https://doi.org/10.1016/j.rser.2024.114647>

93. L. Ma, J. Tian, T. Zhang, Q. Guo, C. Hu, Accurate and efficient remaining useful life prediction of batteries enabled by physics-informed machine learning. *J. Energy Chem.* **91**, 512–521 (2024). <https://doi.org/10.1016/j.jechem.2023.12.043>

94. T. Zheng, Z. Huang, H. Ge, P. Hu, X. Fan et al., Applying machine learning to design delicate amorphous micro-nano materials for rechargeable batteries. *Energy Storage Mater.* **71**, 103614 (2024). <https://doi.org/10.1016/j.ensm.2024.103614>

95. Y. Wang, X. Feng, D. Guo, H. Hsu, J. Hou et al., Temperature excavation to boost machine learning battery thermochemical predictions. *Joule* **8**(9), 2639–2651 (2024). <https://doi.org/10.1016/j.joule.2024.07.002>

96. S. Varun, A.M. Chandran, K.P. Minhaj, V. Shaju, L.A. Varughese et al., Unveiling predictive insights for enhanced performance of PVDF-based nanogenerators *via* machine learning modeling. *Chem. Eng. J.* **484**, 149661 (2024). <https://doi.org/10.1016/j.cej.2024.149661>

97. X. Zhang, D. Mu, S. Lu, Y. Zhang, Y. Zhang et al., Ab initio design of Ni-rich cathode material with assistance of machine learning for high energy lithium-ion batteries. *Energy Environ. Mater.* **7**(6), e12744 (2024). <https://doi.org/10.1002/eem2.12744>

98. X. Zhang, B. Ding, Y. Wang, Y. Liu, G. Zhang et al., Machine learning for screening small molecules as passivation materials for enhanced perovskite solar cells. *Adv. Funct. Mater.* **34**(30), 2314529 (2024). <https://doi.org/10.1002/adfm.202314529>

99. C. Chen, D.T. Nguyen, S.J. Lee, N.A. Baker, A.S. Karakoti et al., Accelerating computational materials discovery with machine learning and cloud high-performance computing: from large-scale screening to experimental validation. *J. Am. Chem. Soc.* **146**(29), 20009–20018 (2024). <https://doi.org/10.1021/jacs.4c03849>

100. J. Liu, M. Liu, J.-P. Liu, Z. Ye, Y. Wang et al., Towards provably efficient quantum algorithms for large-scale machine-learning models. *Nat. Commun.* **15**, 434 (2024). <https://doi.org/10.1038/s41467-023-43957-x>

101. N.J. Szymanski, B. Rendy, Y. Fei, R.E. Kumar, T. He et al., An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**(7990), 86–91 (2023). <https://doi.org/10.1038/s41586-023-06734-w>

102. J. Cui, F. Wu, W. Zhang, L. Yang, J. Hu et al., Direct prediction of gas adsorption *via* spatial atom interaction learning. *Nat. Commun.* **14**(1), 7043 (2023). <https://doi.org/10.1038/s41467-023-42863-6>

103. V.L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti et al., Origins of structural and electronic transitions in disordered silicon. *Nature* **589**(7840), 59–64 (2021). <https://doi.org/10.1038/s41586-020-03072-z>

104. H. Yin, M. Xu, Z. Luo, X. Bi, J. Li et al., Machine learning for membrane design and discovery. *Green Energy Environ.* **9**(1), 54–70 (2024). <https://doi.org/10.1016/j.gee.2022.12.001>

105. N. Han, B.-L. Su, Ai-driven material discovery for energy, catalysis and sustainability. *Natl. Sci. Rev.* **12**(5), nwaf110 (2025). <https://doi.org/10.1093/nsr/nwaf110>

106. A. Jaison, A. Mohan, Y.-C. Lee, Machine learning-enhanced photocatalysis for environmental sustainability: integration and applications. *Mater. Sci. Eng. R. Rep.* **161**, 100880 (2024). <https://doi.org/10.1016/j.mser.2024.100880>

107. X. Jiang, H. Fu, Y. Bai, L. Jiang, H. Zhang et al., Interpretable machine learning applications: a promising prospect of AI for materials. *Adv. Funct. Mater.* **25**(7734) (2025). <https://doi.org/10.1002/adfm.202507734>

108. Y. Aoki, S. Wu, T. Tsurimoto, Y. Hayashi, S. Minami et al., Multitask machine learning to predict polymer–solvent miscibility using flory–huggins interaction parameters. *Macromolecules* **56**(14), 5446–5456 (2023). <https://doi.org/10.1021/acs.macromol.2c02600>

109. V. Kuznetsova, Á. Coogan, D. Botov, Y. Gromova, E.V. Ushakova et al., Expanding the horizons of machine learning in nanomaterials to chiral nanostructures. *Adv. Mater.* **36**(18), 2308912 (2024). <https://doi.org/10.1002/adma.202308912>

110. A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards et al., Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013). <https://doi.org/10.1063/1.4812323>

111. J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **65**(11), 1501–1509 (2013). <https://doi.org/10.1007/s11837-013-0755-4>

112. C. Draxl, M. Scheffler, The NOMAD laboratory: from data sharing to artificial intelligence. *J. Phys. Mater.* **2**(3), 036001 (2019). <https://doi.org/10.1088/2515-7639/ab13bb>

113. S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang et al., AFLOWLIB.ORG: a distributed materials properties repository from high-throughput *ab initio* calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012). <https://doi.org/10.1016/j.commatsci.2012.02.002>

114. K. Bi, Y. Wang, G. Zhou, Hierarchical porous N/S-doped carbon with machine learning to predict advanced potassium-ion batteries. *J. Mater. Chem. A* **11**(22), 11696–11703 (2023). <https://doi.org/10.1039/D3TA00247K>

115. L.E. Vivanco-Benavides, C.L. Martínez-González, C. Mercado-Zúñiga, C. Torres-Torres, Machine learning and materials informatics approaches in the analysis of physical properties of carbon nanotubes: a review. *Comput. Mater. Sci.* **201**, 110939 (2022). <https://doi.org/10.1016/j.commatsci.2021.110939>

116. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**(6), 84–90 (2017). <https://doi.org/10.1145/3065386>

117. P.Z. Moghadam, Y.G. Chung, R.Q. Snurr, Progress toward the computational discovery of new metal–organic framework adsorbents for energy applications. *Nat. Energy* **9**(2), 121–133 (2024). <https://doi.org/10.1038/s41560-023-01417-2>

118. T. Chen, Z. Pang, S. He, Y. Li, S. Shrestha et al., Machine intelligence-accelerated discovery of all-natural plastic substitutes. *Nat. Nanotechnol.* **19**(6), 782–791 (2024). <https://doi.org/10.1038/s41565-024-01635-z>

119. J. Benavides-Hernández, F. Dumeignil, From characterization to discovery: artificial intelligence, machine learning and high-throughput experiments for heterogeneous catalyst design. *ACS Catal.* **14**(15), 11749–11779 (2024). <https://doi.org/10.1021/acscatal.3c06293>

120. L. Xiang, J. Chen, X. Zhao, J. Hu, J. Yu et al., Synergistic machine learning accelerated discovery of nanoporous inorganic crystals as non-absorbable oral drugs. *Adv. Mater.* **36**(32), 2404688 (2024). <https://doi.org/10.1002/adma.202404688>

121. V. Stanev, K. Choudhary, A.G. Kusne, J. Paglione, I. Takeuchi, Artificial intelligence for search and discovery of quantum materials. *Commun. Mater.* **2**, 105 (2021). <https://doi.org/10.1038/s43246-021-00209-z>

122. Y. Wang, Z.-J. Jiang, D.-R. Wang, W. Lu, D. Li, Machine learning-assisted discovery of propane-selective metal-organic frameworks. *J. Am. Chem. Soc.* **146**(10), 6955–6961 (2024). <https://doi.org/10.1021/jacs.3c14610>

123. H.M. Johnson, F. Gusev, J.T. Dull, Y. Seo, R.D. Priestley et al., Discovery of crystallizable organic semiconductors with machine learning. *J. Am. Chem. Soc.* **146**(31), 21583–21590 (2024). <https://doi.org/10.1021/jacs.4c05245>

124. Z. Zhang, F. Pan, S.A. Mohamed, C. Ji, K. Zhang et al., Accelerating discovery of water stable metal–organic frameworks by machine learning. *Small* **20**(42), 2405087 (2024). <https://doi.org/10.1002/smll.202405087>

125. X. Zhou, C. Xu, X. Guo, P. Apostol, A. Vlad et al., Computational and machine learning-assisted discovery and experimental validation of conjugated sulfonamide cathodes for lithium-ion batteries. *Adv. Energy Mater.* **15**(25), 2401658 (2025). <https://doi.org/10.1002/aenm.202401658>

126. X. Liu, K. Fan, X. Huang, J. Ge, Y. Liu et al., Recent advances in artificial intelligence boosting materials design for electrochemical energy storage. *Chem. Eng. J.* **490**, 151625 (2024). <https://doi.org/10.1016/j.cej.2024.151625>

127. G. Xu, M. Jiang, J. Li, X. Xuan, J. Li et al., Machine learning-accelerated discovery and design of electrode materials and electrolytes for lithium ion batteries. *Energy Storage Materials* **72**, 103710 (2024). <https://doi.org/10.1016/j.ensm.2024.103710>

128. B.J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani et al., Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**(7844), 89–96 (2021). <https://doi.org/10.1038/s41586-021-03213-y>

129. T. He, H. Huo, C.J. Bartel, Z. Wang, K. Cruse et al., Precursor recommendation for inorganic synthesis by machine learning materials similarity from scientific literature. *Sci. Adv.* **9**(23), eadg8180 (2023). <https://doi.org/10.1126/sciadv.adg8180>

130. N.J. Szymanski, P. Nevatia, C.J. Bartel, Y. Zeng, G. Ceder, Autonomous and dynamic precursor selection for solid-state materials synthesis. *Nat. Commun.* **14**, 6956 (2023). <https://doi.org/10.1038/s41467-023-42329-9>

131. J. Li, S.G. Ballmer, E.P. Gillis, S. Fujii, M.J. Schmidt et al., Synthesis of many different types of organic small molecules using one automated process. *Science* **347**(6227), 1221–1226 (2015). <https://doi.org/10.1126/science.aaa5414>

132. P.J. Kitson, G. Marie, J.-P. Francoia, S.S. Zalesskiy, R.C. Sigerson et al., Digitization of multistep organic synthesis in reactionware for on-demand pharmaceuticals. *Science* **359**(6373), 314–319 (2018). <https://doi.org/10.1126/science.aoa3466>

133. C.W. Coley, D.A. Thomas 3rd, J.A.M. Lummiss, J.N. Jaworski, C.P. Breen et al., A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**(6453), eaax1566 (2019). <https://doi.org/10.1126/science.aax1566>

134. J.S. Manzano, W. Hou, S.S. Zalesskiy, P. Frei, H. Wang et al., An autonomous portable platform for universal chemical synthesis. *Nat. Chem.* **14**(11), 1311–1318 (2022). <https://doi.org/10.1038/s41557-022-01016-w>

135. A. Chen, Z. Wang, K.L.L. Vidaurre, Y. Han, S. Ye et al., Knowledge-reused transfer learning for molecular and materials science. *J. Energy Chem.* **98**, 149–168 (2024). <https://doi.org/10.1016/j.jec.2024.06.013>

136. X.-Y. Liu, J. Wu, Z.-H. Zhou, Exploratory undersampling for class-imbalance learning. *IEEE Trans. Syst., Man, Cybern. B* **39**(2), 539–550 (2009). <https://doi.org/10.1109/tsmcb.2008.2007853>

137. N. Qu, M. Chen, M. Liao, Y. Cheng, Z. Lai et al., Accelerating density functional calculation of adatom adsorption on graphene via machine learning. *Materials* **16**(7), 2633 (2023). <https://doi.org/10.3390/ma16072633>

138. M. Agarwal, P. Pasupathy, X. Wu, S.S. Recchia, A.A. Pelegri, Multiscale computational and artificial intelligence models of linear and nonlinear composites: a review. *Small Science* **4**(5), 2300185 (2024). <https://doi.org/10.1002/smss.202300185>

139. W.L. Ng, G.L. Goh, G.D. Goh, J.S.J. Ten, W.Y. Yeong, Progress and opportunities for machine learning in materials and processes of additive manufacturing. *Adv. Mater.* **36**(34), 2310006 (2024). <https://doi.org/10.1002/adma.202310006>

140. K. Zhang, X. Gong, Y. Jiang, Machine learning in soft matter: from simulations to experiments. *Adv. Funct. Mater.* **34**(24), 2315177 (2024). <https://doi.org/10.1002/adfm.202315177>

141. M. Delpisheh, B. Ebrahimpour, A. Fattahi, M. Siavashi, H. Mir et al., Leveraging machine learning in porous media. *J. Mater. Chem. A* **12**(32), 20717–20782 (2024). <https://doi.org/10.1039/d4ta00251b>

142. V. Karkaria, A. Goeckner, R. Zha, J. Chen, J. Zhang et al., Towards a digital twin framework in additive manufacturing: machine learning and bayesian optimization for time series process optimization. *J. Manuf. Syst.* **75**, 322–332 (2024). <https://doi.org/10.1016/j.jmsy.2024.04.023>

143. Y. Ren, M. Cui, Y. Zhou, S. Sun, F. Guo et al., Utilizing machine learning for reactive material selection and width design in permeable reactive barrier (PRB). *Water Res.* **251**, 121097 (2024). <https://doi.org/10.1016/j.watres.2023.121097>

144. H. Wang, S.L. Gao, B.T. Wang, Y.T. Ma, Z.J. Guo et al., Recent advances in machine learning-assisted fatigue life prediction of additive manufactured metallic materials: a review. *J. Mater. Sci. Technol.* **198**, 111–136 (2024). <https://doi.org/10.1016/j.jmst.2024.01.086>

145. B. Lu, Y. Xia, Y. Ren, M. Xie, L. Zhou et al., When machine learning meets 2D materials: a review. *Adv. Sci.* **11**(13), 2305277 (2024). <https://doi.org/10.1002/advs.202305277>

146. S.V. Thummalapalli, D. Patil, A. Ramanathan, D. Ravichandran, Y. Zhu et al., Machine learning–enabled direct

ink writing of conductive polymer composites for enhanced performance in thermal management and current protection. *Energy Storage Materials* **71**, 103670 (2024). <https://doi.org/10.1016/j.ensm.2024.103670>

147. H. Zhang, Y. Li, S. Zheng, Z. Lu, X. Gui et al., Battery lifetime prediction across diverse ageing conditions with inter-cell deep learning. *Nat. Mach. Intell.* **7**(2), 270–277 (2025). <https://doi.org/10.1038/s42256-024-00972-x>

148. A. Keddouda, R. Ihaddadene, A. Boukhari, A. Atia, M. Arıcı et al., Photovoltaic module temperature prediction using various machine learning algorithms: performance evaluation. *Appl. Energy* **363**, 123064 (2024). <https://doi.org/10.1016/j.apenergy.2024.123064>

149. M.F. Tahir, M.Z. Yousaf, A. Tzes, M.S. El Moursi, T.H.M. El-Fouly, Enhanced solar photovoltaic power prediction using diverse machine learning algorithms with hyperparameter optimization. *Renew. Sustain. Energy Rev.* **200**, 114581 (2024). <https://doi.org/10.1016/j.rser.2024.114581>

150. B. Bai, L. Wang, F. Guan, Y. Cui, M. Bao et al., Prediction models for bioavailability of Cu and Zn during composting: insights into machine learning. *J. Hazard. Mater.* **471**, 134392 (2024). <https://doi.org/10.1016/j.jhazmat.2024.134392>

151. G. Yin, F.J.I. Alazzawi, S. Mironov, F. Reegu, A.S. El-Shafay et al., Machine learning method for simulation of adsorption separation: comparisons of model's performance in predicting equilibrium concentrations. *Arab. J. Chem.* **15**(3), 103612 (2022). <https://doi.org/10.1016/j.arabjc.2021.103612>

152. R. Lyu, C.E. Moore, T. Liu, Y. Yu, Y. Wu, Predictive design model for low-dimensional organic–inorganic halide perovskites assisted by machine learning. *J. Am. Chem. Soc.* **143**(32), 12766–12776 (2021). <https://doi.org/10.1021/jacs.1c05441>

153. S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li et al., Accelerated discovery of stable lead-free hybrid organic–inorganic perovskites *via* machine learning. *Nat. Commun.* **9**(1), 3405 (2018). <https://doi.org/10.1038/s41467-018-05761-w>

154. A.J. Myles, R.N. Feudale, Y. Liu, N.A. Woody, S.D. Brown, An introduction to decision tree modeling. *J. Chemom.* **18**(6), 275–285 (2004). <https://doi.org/10.1002/cem.873>

155. V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, M. Chica-Rivas, Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* **71**, 804–818 (2015). <https://doi.org/10.1016/j.oregeorev.2015.01.001>

156. A.V. Konstantinov, L.V. Utkin, Interpretable machine learning with an ensemble of gradient boosting machines. *Knowl. Based Syst.* **222**, 106993 (2021). <https://doi.org/10.1016/j.knosys.2021.106993>

157. P. Li, L. Dong, C. Li, Y. Li, J. Zhao et al., Machine learning to promote efficient screening of low-contact electrode for 2D semiconductor transistor under limited data. *Adv. Mater.* **36**(26), 2312887 (2024). <https://doi.org/10.1002/adma.202312887>

158. A.P. Bartók, J. Kermode, N. Bernstein, G. Csányi, Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* **8**(4), 041048 (2018). <https://doi.org/10.1103/physrevx.8.041048>

159. L.C. Erhard, J. Rohrer, K. Albe, V.L. Deringer, A machine-learned interatomic potential for silica and its relation to empirical models. *NPJ Comput. Mater.* **8**, 90 (2022). <https://doi.org/10.1038/s41524-022-00768-w>

160. Z. Liu, M. Jiang, T. Luo, Leveraging low-fidelity data to improve machine learning of sparse high-fidelity thermal conductivity data *via* transfer learning. *Mater. Today Phys.* **28**, 100868 (2022). <https://doi.org/10.1016/j.mtphys.2022.100868>

161. Y. Kang, H. Park, B. Smit, J. Kim, A multi-modal pre-training transformer for universal transfer learning in metal–organic frameworks. *Nat. Mach. Intell.* **5**(3), 309–318 (2023). <https://doi.org/10.1038/s42256-023-00628-2>

162. H. Wang, T. Fu, Y. Du, W. Gao, K. Huang et al., Scientific discovery in the age of artificial intelligence. *Nature* **620**(7972), 47–60 (2023). <https://doi.org/10.1038/s41586-023-06221-2>

163. B. Dou, Z. Zhu, E. Merkurjev, L. Ke, L. Chen et al., Machine learning methods for small data challenges in molecular science. *Chem. Rev.* **123**(13), 8736–8780 (2023). <https://doi.org/10.1021/acs.chemrev.3c00189>

164. D. Azzimonti, C. Rottandi, A. Giusti, M. Tornatore, A. Bianco, Comparison of domain adaptation and active learning techniques for quality of transmission estimation with small-sized training datasets. *J. Opt. Commun. Netw.* **13**(1), A56 (2021). <https://doi.org/10.1364/jocn.401918>

165. Y. Jia, R. Zhang, C. Fang, J. Zheng, Interpretable machine learning to accelerate the analysis of doping effect on Li/Ni exchange in Ni-rich layered oxide cathodes. *J. Phys. Chem. Lett.* **15**(6), 1765–1773 (2024). <https://doi.org/10.1021/acs.jpclett.3c03294>

166. J. Bai, L. Cao, S. Mosbach, J. Akroyd, A.A. Lapkin et al., From platform to knowledge graph: evolution of laboratory automation. *JACS Au* **2**(2), 292–309 (2022). <https://doi.org/10.1021/jacsau.1c00438>

167. O. Bayley, E. Savino, A. Slattery, T. Noël, Autonomous chemistry: navigating self-driving labs in chemical and material sciences. *Matter* **7**(7), 2382–2398 (2024). <https://doi.org/10.1016/j.matt.2024.06.003>

168. P. Chen, R. Jiao, J. Liu, Y. Liu, Y. Lu, Interpretable graph transformer network for predicting adsorption isotherms of metal–organic frameworks. *J. Chem. Inf. Model.* **62**(22), 5446–5456 (2022). <https://doi.org/10.1021/acs.jcim.2c00876>

169. K.M. Jablonka, D. Ongari, S.M. Moosavi, B. Smit, Using collective knowledge to assign oxidation states of metal cations in metal–organic frameworks. *Nat. Chem.* **13**(8), 771–777 (2021). <https://doi.org/10.1038/s41557-021-00717-y>

170. Y. Guo, M. Yang, G. Huang, Y. Zheng, Machine-learning-enabled exploitation of gas-sensing descriptors: a case study of five pristine metal oxides. *Chem. Eng. J.* **492**, 152280 (2024). <https://doi.org/10.1016/j.cej.2024.152280>

171. H. Li, X. Li, P. Wang, Z. Zhang, K. Davey et al., Machine learning big data set analysis reveals C–C electro-coupling mechanism. *J. Am. Chem. Soc.* **146**(32), 22850–22858 (2024). <https://doi.org/10.1021/jacs.4c09079>

172. Y. Sun, P. Sun, J. Jia, Z. Liu, L. Huo et al., Machine learning in clarifying complex relationships: biochar preparation procedures and capacitance characteristics. *Chem. Eng. J.* **485**, 149975 (2024). <https://doi.org/10.1016/j.cej.2024.149975>

173. Y. Zhou, Lifecycle battery carbon footprint analysis for battery sustainability with energy digitalization and artificial intelligence. *Appl. Energy* **371**, 123665 (2024). <https://doi.org/10.1016/j.apenergy.2024.123665>

174. H. Zhang, X. Gui, S. Zheng, Z. Lu, Y. Li, J. Bian. Batteryml: An open-source platform for machine learning on battery degradation. *arXiv preprint arXiv:231014714*. (2023).

175. X. Li, J.M. Hoffman, M.G. Kanatzidis, The 2D halide perovskite rulebook: how the spacer influences everything from the structure to optoelectronic device efficiency. *Chem. Rev.* **121**(4), 2230–2291 (2021). <https://doi.org/10.1021/acs.chemrv.0c01006>

176. R. Ouyang, E. Ahmetcik, C. Carbogno, M. Scheffler, L.M. Ghiringhelli, Simultaneous learning of several materials properties from incomplete databases with multi-task SISSO. *J. Phys. Mater.* **2**(2), 024002 (2019). <https://doi.org/10.1088/2515-7639/ab077b>

177. R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, L.M. Ghiringhelli, SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2**(8), 083802 (2018). <https://doi.org/10.1103/physrevmaterials.2.083802>

178. A. Ludwig, Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods. *NPJ Comput. Mater.* **5**, 70 (2019). <https://doi.org/10.1038/s41524-019-0205-0>

179. Z. Ren, F. Oviedo, M. Thway, S.I.P. Tian, Y. Wang et al., Embedding physics domain knowledge into a Bayesian network enables layer-by-layer process innovation for photovoltaics. *NPJ Comput. Mater.* **6**, 9 (2020). <https://doi.org/10.1038/s41524-020-0277-x>

180. J. Tian, R. Xiong, W. Shen, J. Lu, X.-G. Yang, Deep neural network battery charging curve prediction using 30 points collected in 10 min. *Joule* **5**(6), 1521–1534 (2021). <https://doi.org/10.1016/j.joule.2021.05.012>

181. Z. Fu, P. Huang, X. Wang, W.-D. Liu, L. Kong et al., Artificial intelligence-assisted ultrafast high-throughput screening of high-entropy hydrogen evolution reaction catalysts. *Adv. Energy Mater.* **15**(30), 2500744 (2025). <https://doi.org/10.1002/aenm.202500744>

182. S. Singh, J.M. Hernández-Lobato, A meta-learning approach for selectivity prediction in asymmetric catalysis. *Nat. Commun.* **16**, 3599 (2025). <https://doi.org/10.1038/s41467-025-58854-8>

183. H. Liao, S. Hu, H. Yang, L. Wang, S. Tanaka et al., Data-driven *de novo* design of super-adhesive hydrogels. *Nature* **644**(8075), 89–95 (2025). <https://doi.org/10.1038/s41586-025-09269-4>

184. H. Li, H. Zheng, T. Yue, Z. Xie, S. Yu et al., Machine learning-accelerated discovery of heat-resistant polysulfates for electrostatic energy storage. *Nat. Energy* **10**(1), 90–100 (2025). <https://doi.org/10.1038/s41560-024-01670-z>

185. L. Zhang, X. Zhang, C. Chen, J. Zhang, W. Tan et al., Machine learning-aided discovery of low-Pt high entropy intermetallic compounds for electrochemical oxygen reduction reaction. *Angew. Chem. Int. Ed.* **63**(51), e202411123 (2024). <https://doi.org/10.1002/anie.202411123>

186. X. Duan, Y. Li, J. Zhao, M. Zhang, X. Wang et al., Machine learning accelerated discovery of entropy-stabilized oxide catalysts for catalytic oxidation. *J. Am. Chem. Soc.* **147**(1), 651–661 (2025). <https://doi.org/10.1021/jacs.4c12838>

187. G. Lin, Z. University, T. Guo, Z. University et al., Machine learning accelerated screening advanced single-atom anchored MXenes electrocatalyst for nitrogen fixation. *ACS Catal.* **15**(15), 13534–13548 (2025). <https://doi.org/10.1021/acscatal.4c06914>

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