Toward Autonomous Laboratories: Convergence of Artificial Intelligence and Experimental Automation

Yunchao Xie[1#], Kianoosh Sattari[1#], Chi Zhang[1], and Jian Lin[1*]

[1]Department of Mechanical and Aerospace Engineering
University of Missouri, Columbia, Missouri 65211, United States
[#]Authors contributed equally to this work.
[*]E-mail: LinJian@missouri.edu (J. L.)

Abstract

The ever-increasing demand for novel materials with superior properties inspires retrofitting traditional research paradigms in the era of artificial intelligence and automation. An autonomous experimental platform (AEP) has emerged as an exciting research frontier that achieves full autonomy *via* integrating data-driven algorithms such as machine learning (ML) with experimental automation in the material development loop from synthesis, characterization, and analysis, to decision making. In this review, we started with a primer to describe how to develop data-driven algorithms for solving material problems. Then, we systematically summarized recent progress on automated material synthesis, ML-enabled data analysis, and decision-making. Finally, we discussed challenges and opportunities in an endeavor to develop the next-generation AEPs for ultimately realizing an autonomous or self-driving laboratory. This review will provide insights for researchers aiming to learn the frontier of ML in materials and deploy AEPs in their labs for accelerating material development.

Keywords: artificial intelligence, autonomous experimentation platform, machine learning, materials science

1. Introduction

There is an ever-increasing demand for developing advanced materials with superior properties, which requires extensive investment in research.[1-4] The development process is still largely performed by well-trained and skilled scientists in a structured laboratory set, which is a paradigm that has little evolved over the last several decades.[4-6] Although guided by domain knowledge and explicit physical rules, this process is still a trial-and-error one, which is quite laborious and time-intensive.[5] For example, the filament of incandescent light bulbs was screened from roughly 6000 materials by Thomas Edison and his coworkers.[7] Another example is the discovery of an optimum catalyst for ammonia synthesis, which was conducted by Mittasch and his colleagues in the early 20^{th} century.[8] Moreover, reproducibility and unintentional bias could exist due to unnecessary human interventions. These issues and challenges lead to a development pace greatly falling behind the one demanded by manufacturers and consumers who face a complicated and volatile market. Thus, revolutionizing the current research paradigm into a new one for accelerating material development has become a compelling goal in the field.

Experimental automation with aid of advanced scientific instruments and statistical techniques for automatically screening candidates has attracted enormous interest. It has been adopted in both academy and industry, especially in the field of pharmaceuticals and organic chemistry.[9-20] Automation, which is good at performing consistent tasks, can enormously increase the throughput of materials and chemicals to be studied. Furthermore, it frees the researchers from tedious and repetitive tasks, thus allowing them to investigate more innovative and complex problems than ever before. Nevertheless, there still exist some challenges to taking automation to the next level of autonomy. First, analysis of big characterization data from spectroscopies and microscopies falls behind the data collection rate. Second, heavy dependency

on experts to optimize enormous reaction and chemical spaces undoubtedly lower the exploration efficiency. In each experiment iteration, a new set of reaction and chemical spaces should be decided according to the results obtained from the previous iteration. In a traditional automation setup, such a decision is still made by researchers, thus potentially causing bias and errors. Third, given that the exploration space is enormous and in the high dimension, it is still impossible and impractical for the automated robots to enumerate all combinations, because it would generate too enormous amounts of data to be processed for establishing the synthesis-structure-property relationship. Thus, intelligent data analysis and decision-making algorithms are much needed to drive autonomous experiments, forming the basis for developing an autonomous experimentation platform (AEP).

Recently, machine learning (ML), especially deep learning (DL),[21] has made a giant leap in the fields of computer vision,[22-24] autonomous driving,[25, 26] speech recognition,[27, 28] recommending systems,[29, 30] games,[31-33] protein folding,[34, 35] and biomedical imaging,[36] to name a few. Distinguished from traditional physics-based modeling, ML is usually called "learned from data" due to its ability to learn the hidden knowledge from the data and predict results from unseen data without applying explicit formulas/equations. These recent breakthroughs mainly benefit from three significant advances, i.e., available big data, powerful computational capacity, and advanced algorithms.[37] Over the past few years, the academy has envisioned the implementation of ML in the field of materials science and chemistry for physicochemical/mechanical properties prediction,[38-41] quantifying the processing-structure-property relationship,[42-44] guiding materials synthesis,[45-50] synthesis planning,[51-55] and analyzing characterization data.[56-62]

By integrating ML with automation, the AEP, a newly emerging research paradigm, has shown

great potential in accelerating material development by an order of magnitude.[63-74] The AEP can greatly reduce the total number of experiments needed for discovery *via* actively exploring chemical and reaction spaces.[63, 64] This new ML-enabled research paradigm has greatly shortened the time of material development and fully embraced the vision of autonomy.[69, 74, 75] To determine how an automated apparatus should perform the next-iteration experiment, the ML algorithms, instead of the intuition from the skilled experts, make decisions *via* exploring all collected reaction and characterization data, thus closing the loop by the autonomous iteration thereof. Publications related to topics of automated and autonomous experiments in materials and chemical science have dramatically grown since 2010, and this trend is projected to continue in decades to come.

To reflect this trend, several review papers have been recently published as listed in **Table 1**.[1-3, 5, 6, 14, 60, 76-96] Lapkin and his coworkers have summarized recent advances in techniques and methods that enable closed-loop material development.[76] Buonassisi et al. focused the review on the convergence of high-performance computing, automation, and ML models.[5] Aspuru-Guzik and his coworkers illustrated their perspectives on AI-driven high-throughput virtual screening, automated synthesis planning, automated laboratories, and ML algorithms toward autonomous materials discovery.[79] Jensen's group summarized two main aspects of autonomous discovery.[1, 2] In the first aspect, they defined three broad categories followed by illustration of substantial progress in them. In the second aspect, they proposed a few possible research directions in processing complex data, building empirical models, automating validation and feedback, selecting experiments, and evaluating the performance. Although these reviews have provided invaluable information, there still lacks a comprehensive review that summarizes the recent progress and future trend of autonomous experiments from the perspectives

of how ML algorithms tackle the specific challenges imposed by essential components of an AEP.

Table 1. List of recent reviews on autonomous and automated experiments in materials science, organic chemistry, and drug discovery.

Year	Topic	Ref.
2015	Automatic discovery and optimization	Lapkin et al.[76]
2018	Smart automation	Aspuru-Guzik et al.[3]
	Designing algorithms	Cronin et al.[77]
	Automation, ML, and computing	Buonassisi et al.[5]
	Bioactive molecular discovery and automation	Nelson et al.[14]
2019	Automated and autonomous workflow	Gregoire et al.[78]
	Self-driving laboratories	Aspuru-Guzik et al.[79]
	Autonomous molecular design	Saikin et al.[80]
	Search algorithm and automation	Cronin et al.[6]
2020	Autonomous discovery, machine learning, challenges, and chemical space	Jensen et al.[1, 2]
	The Chemputer	Cronin et al.[81]
	Materials acceleration platforms	Aspuru-Guzik et al.[82]
	Autonomous intelligent agents	Aykol et al.[83]
	Autonomous robotic experimentation, modular microfluidic reactors	Abolhasani et al.[84]
	Automated synthesis and software	Hao et al.[85]
	Microfluidic synthesis, semiconductor materials, and artificial intelligence	Abolhasani et al.[86]
2021	Shape of chemical data	Jacqueline M. Cole[87]
	Automation, ML, and high-throughput experimentation	Jensen et al.[88]
	Automated continuous synthesis and optimization	Jensen et al.[89]
	Synthesis planning, AI, and automation	Engkvist et al.[90]
	Digital transformation, artificial intelligence, and automation	Schubert et al.[91]
	Automated experimentation, data science, and chemistry	Hein et al.[92]
	Automated/autonomous experiment, machine learning, and electron/scanning probe microscopy	Kalinin et al.[60]
	Automation, data-driven approach, and polymer therapeutics	Gormley et al.[93]
	Automated robotic platform, machine learning, and formulations	Lapkin et al.[94]
	Autonomous experimentation, AI	Maruyama et al.[95]
	Automated synthesis, chemical informatics, digital chemistry, and data standards	Cronin et al.[96]

This review is written to fill this gap. The scope of this review is summarized in **Figure 1**. It starts with a section describing an ML primer for beginners to understand the field. This section briefly introduces the concepts, categories, workflow, and evaluation metrics for the ML models. Specially, we discuss the challenge of data scarcity in materials/manufacturing domains and summarize a few methodologies to tackle it. Then the review is followed by showing how the ML

algorithms can promote the AEP development from the three essential components of the AEP. Specifically, to begin with, three types of automated experimentation platforms for material synthesis and characterization are discussed. Then, on-the-fly data analysis with the aid of the ML algorithms is summarized. After that, decision-making enabled by the data-driven algorithms to close the loop is reviewed. In each section, a few representative case studies are discussed to exemplify recent successes in advancing the AEP development. This review highlights the roles of ML in leveraging decades of progress in automation for accelerating material discovery and minimizing human intervention and biases. Finally, ongoing challenges, possible solutions, and future trends to move the research in AEPs forward are discussed and foreseen. We expect that this review will serve as a guideline for beginners to understand the basic principles of the data-driven algorithms and how they can be applied to develop AEPs for applications in materials science and chemistry as well as inspirs experts in the field to explore new frontiers.

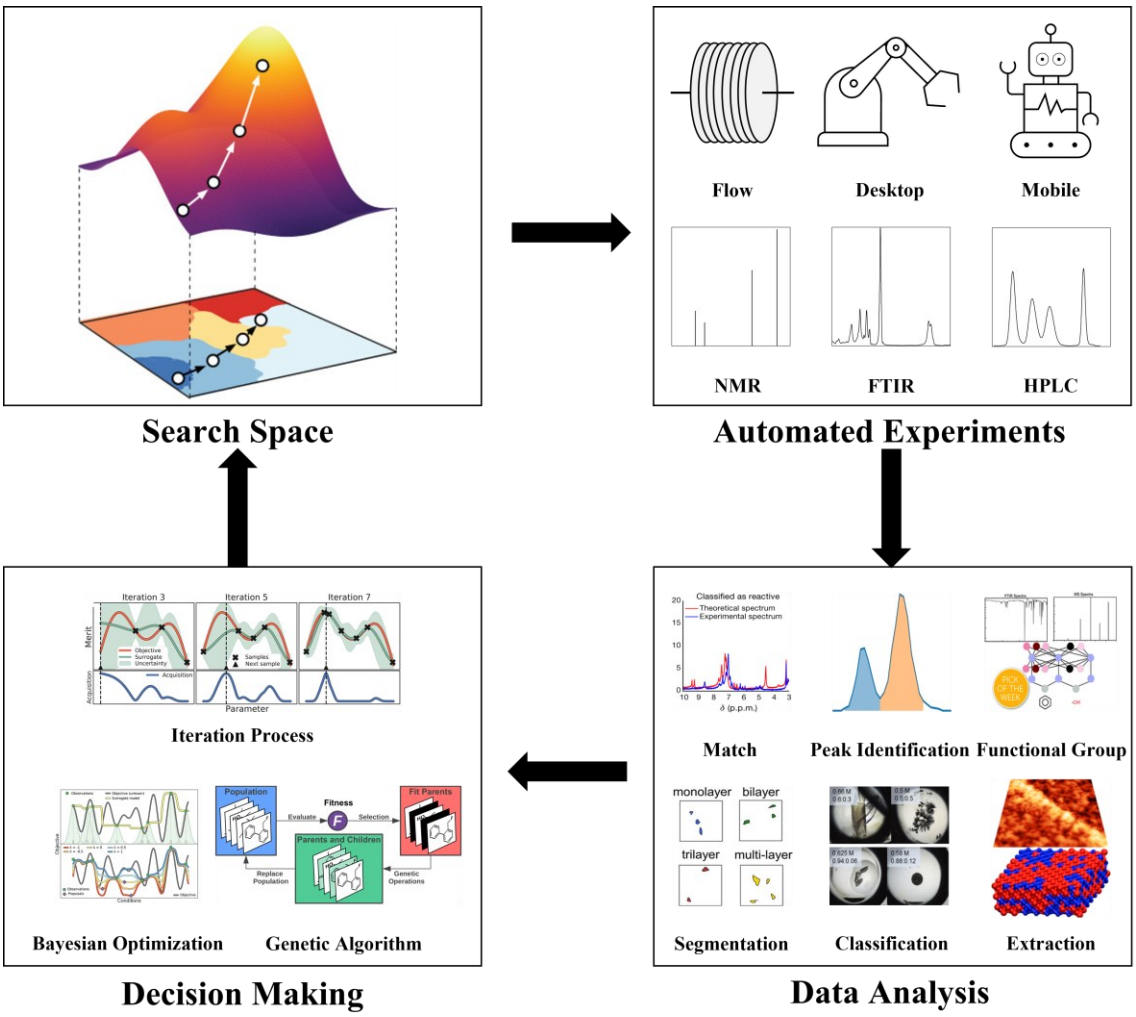


Figure 1. Flowchart showing a fully autonomous experimental platform for the development of novel materials. *Search Space*: Reproduced with permission from Ref.[97], Copyright 2021 American Chemical Society. *Data Analysis*: Reproduced with permission from Ref.[63], Copyright 2018 Springer Nature; Ref.[98], Copyright 2020 American Chemical Society; Ref.[99], Copyright 2020 Royal Society of Chemistry; Ref.[59], Copyright 2020 American Chemical Society; Ref.[100], Copyright 2021 American Chemical Society; Ref.[101], Copyright 2017 American Chemical Society. *Decision Making*: Reproduced with permission from Ref.[102], Copyright 2021 American Chemical Society.

2. A Primer for Developing Data-Driven Algorithms

Data-driven algorithms including ML/DL, the focus of this review, can evaluate or predict the targets/goals from the input features.[37] Particularly, recent breakthroughs in DL have revolutionized the applications in image and speech recognition,[21] which has also created new possibilities for drug discovery,[103] biology,[35, 104, 105] chemistry[52, 54, 106-108] and materials science.[105, 109-111] In this section, we focus on fundamental concepts of ML/DL and discuss how to implement them in the physical and chemical domains. For detailed information, we recommend a few resources,[112-116] some of which have been targeted especially for materials science. Useful textbooks written by professional ML/DL researchers include *Pattern Recognition and Machine Learning*, *Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow*, *Deep Learning*, and *Deep Learning with Python*. In addition, a variety of online tutorials (YouTube, Coursera, Udacity, Udemy, Khan Academy, and Towards Data Science) and lectures are publicly available for beginners to have a basic overview and learn from scratch. A prerequisite for developing the ML/DL models is to master a programming language such as Python (widely used for most ML/DL projects) and useful libraries including Pandas (data manipulation with integrated indexing), NumPy (array type elements and the respective math), Matplotlib or Seaborn (data visualization), scikit-image (image processing), OpenCV (image and video processing), and Scikit-Learn (a versatile and powerful ML library). For those working on DL projects, it is necessary to master at least one DL framework such as Keras, TensorFlow, PyTorch, and MXNET. **Table 2** summarizes some useful textbooks and links for mastering ML and DL.

Table 2. A list of resources for mastering ML and DL.

Content	Textbook	URL
Mathematics	Linear Algebra Done Right	https://link.springer.com/book/10.1007/978-3-319-11080-6

	Linear Algebra and Its Applications	https://www.pearson.com/us/higher-education/product/Lay-Linear-Algebra-and-Its-Applications-5th-Edition/9780321982384.html
	Mathematics for Machine Learning	https://mml-book.github.io/
Python	Python Crash Course	https://nostarch.com/pythoncrashcourse2e
	Learn Python Programming	https://www.packtpub.com/product/learn-python-programming-third-edition/9781801815093
	Think Python	https://www.oreilly.com/library/view/think-python-2nd/9781491939406/
Data science	Python for Data Analysis	https://www.oreilly.com/library/view/python-for-data/9781491957653/
	Data Science from Scratch	https://www.oreilly.com/library/view/data-science-from/9781492041122/
	Python Data Science Handbook	https://www.oreilly.com/library/view/python-data-science/9781491912126/
ML/DL	Introduction to Machine Learning with Python	https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/
	Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow	https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/
	Pattern Recognition and Machine Learning	https://www.springer.com/gp/book/9780387310732
	Machine Learning with Python Cookbook	https://www.oreilly.com/library/view/machine-learning-with/9781491989371/
	Python Machine Learning	https://www.packtpub.com/product/python-machine-learning-third-edition/9781789955750
	Mastering Machine Learning Algorithms	https://www.packtpub.com/product/mastering-machine-learning-algorithms-second-edition/9781838820299
	Deep Learning	https://mitpress.mit.edu/books/deep-learning
	Deep Learning with Python	https://www.manning.com/books/deep-learning-with-python-second-edition
	Deep Learning from Scratch	https://www.oreilly.com/library/view/deep-learning-from/9781492041405/
	Grokking Deep Learning	https://www.manning.com/books/grokking-deep-learning?gclid=Cj0KCQjwtMCKBhDAARIsAG-2Eu8cXUUIAuAs8Ddtkk77u7SV85DtUSufEH84wTvtmr2-KvI2qJ_3b04aAnJKEALw_wcB
	Practical Deep Learning	https://nostarch.com/practical-deep-learning-python
	Deep Learning with TensorFlow 2 and Keras	https://www.packtpub.com/product/deep-learning-with-tensorflow-2-and-keras-second-edition/9781838823412
	Hands-On Deep Learning Algorithms with Python	https://www.packtpub.com/product/hands-on-deep-learning-algorithms-with-python/9781789344158
	Advanced Deep Learning with Python	https://www.packtpub.com/product/advanced-deep-learning-with-python/9781789956177
Libraries	Pandas	https://pandas.pydata.org/
	Numpy	https://numpy.org/

	Matplotlib	https://matplotlib.org/
	Seaborn	https://seaborn.pydata.org/
	Scikit-image	https://scikit-image.org/
	OpenCV	https://opencv.org/
	Scikit-Learn	https://scikit-learn.org/stable/
Frameworks	Keras	https://keras.io/
	Tensorflow	https://www.tensorflow.org/
	PyTorch	https://pytorch.org/
	MXNET	https://mxnet.apache.org/versions/1.8.0/

2.1 Introduction of Machine Learning Algorithms

Based on the amount of supervision in training (**Figure 2**), ML can be categorized into supervised learning,[108, 110, 111, 117, 118] unsupervised learning,[119-121], and semi-supervised learning.[122-125] If the model training fully relies on labeled data, it is called supervised learning (**Figure 2a**), which is commonly used in materials science. According to the types of labels, supervised learning can be used for performing classification and regression tasks. A classification task refers to the situation in which the models are trained with lots of input parameters while their corresponding output classes are represented by discrete values. A regression task is to predict a target numeric value such as conductivity, product yield, and adsorption capacity of the materials when given a set of inputs. In contrast, unsupervised learning is mainly used to seek and deduce potential connections of samples among unlabeled data, which consists of two common methods, i.e., dimensional reduction or data clustering (**Figure 2b**). Dimension reduction involves mapping a high-dimension data matrix to a low-dimension one while preserving information contained in the original data. Main approaches to reducing dimensions include principal component analysis (PCA),[126, 127] singular value decomposition (SVD),[128] Isometric feature mapping (Isomap),[129] Kernel PCA,[130] and t-distributed stochastic neighbor embedding (t-SNE).[131] Clustering is a task of first calculating

the similarities of all samples based on specific metrics, and then assigning them to different groups according to their similarities. K-means[132] and K-Medoids[133] are the two most popular clustering techniques. Semi-supervised learning is the best choice to deal with the situation where there is limited labeled data but plenty of unlabeled data. In semi-supervised learning, an ML model is first trained based on the labeled data, which is then used for predicting the unlabeled data (denoted as pseudo labels). Finally, the ML model is retrained with both the labeled and pseudo data (**Figure 2c**).

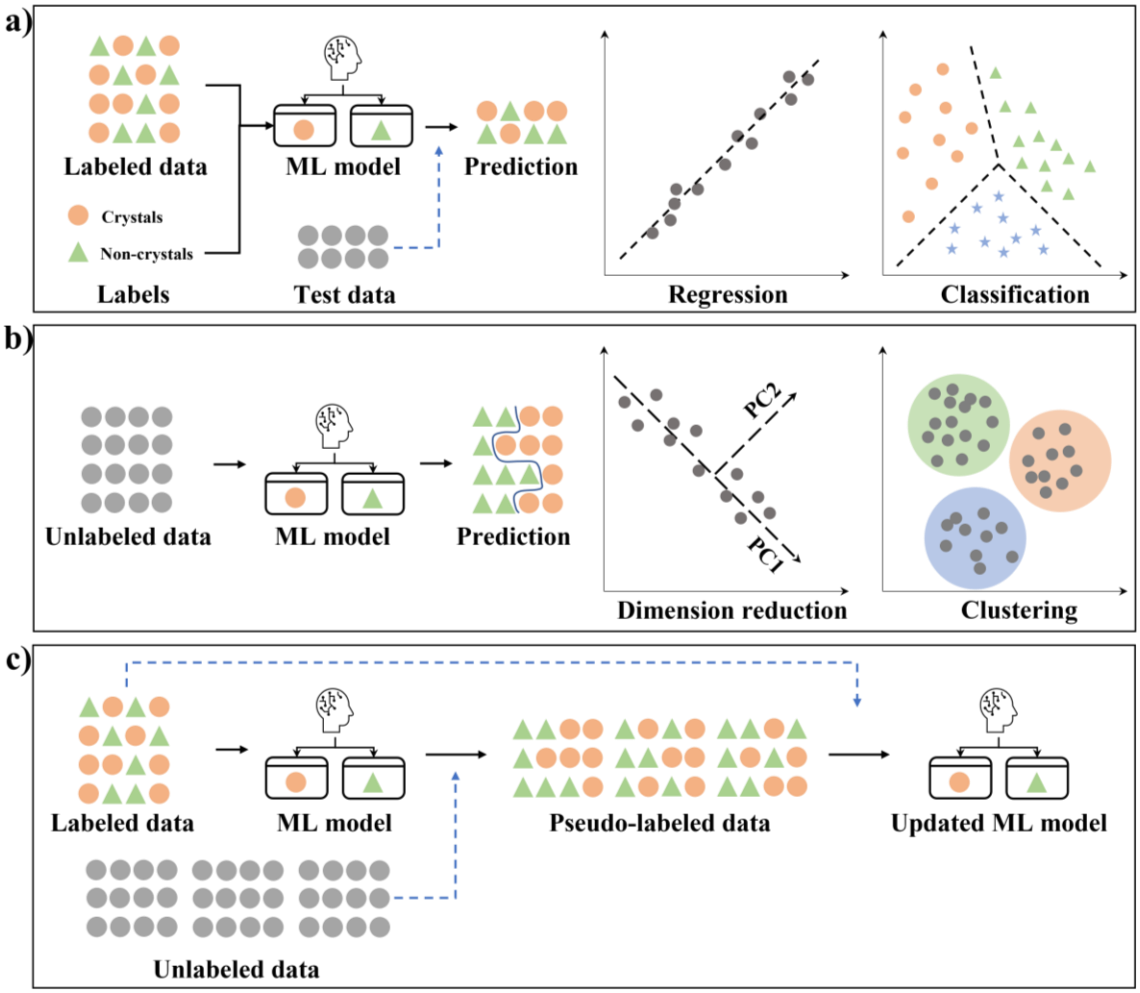


Figure 2. Schematic of three ML categories: (a) supervised learning (regression and classification), (b) unsupervised learning (dimension reduction and clustering), and (c) semi-supervised learning.

Choosing the right ML algorithm is a crucial step toward building an accurate and robust model for solving a material-related problem.[134] Currently, the widely used ML algorithms include *k*-nearest neighbors (KNN, **Figure 3a**),[135] support vector machine (SVM, **Figure 3b**),[136, 137] decision tree (DT, **Figure 3c**),[138] random forest (RF),[139] multilayer perceptron (MLP, **Figure 3c**),[140] naïve Bayes (NB),[141], logistic regression (LR),[142] and gradient boosting-based models including eXtreme Gradient Boosting (XGBoost),[143, 144] adaptive boosting.[145] These traditional ML algorithms are trained on relatively small datasets ($< 10^4$) and can predict the targets with satisfactory performance over many material problems.

However, there still exist some materials-related problems that cannot be well solved through these traditional ML algorithms. These problems come from three main scenarios. The first one is related to the highly data-intensive problems, which contain millions or even billions of training datasets, e.g., computational or characterization data. The second one involves dealing with enormous fingerprints, e.g., de novo drug or molecule design. The third one is related to image segmentation and text mining from materials literature. The DL algorithms[21] including generative adversarial network (GAN, **Figure 3e**),[146] variational autoencoder (VAE, **Figure 3f**),[147] recurrent neural network (RNN, left panel of **Figure 3g**),[148] and long short-term memory (LSTM, the right panel of **Figure 3g**),[149] graph neural network (GNN, **Figure 3h**),[150] and bidirectional encoder representation (BERT)[151] have offered new possible solutions to the aforementioned problems.

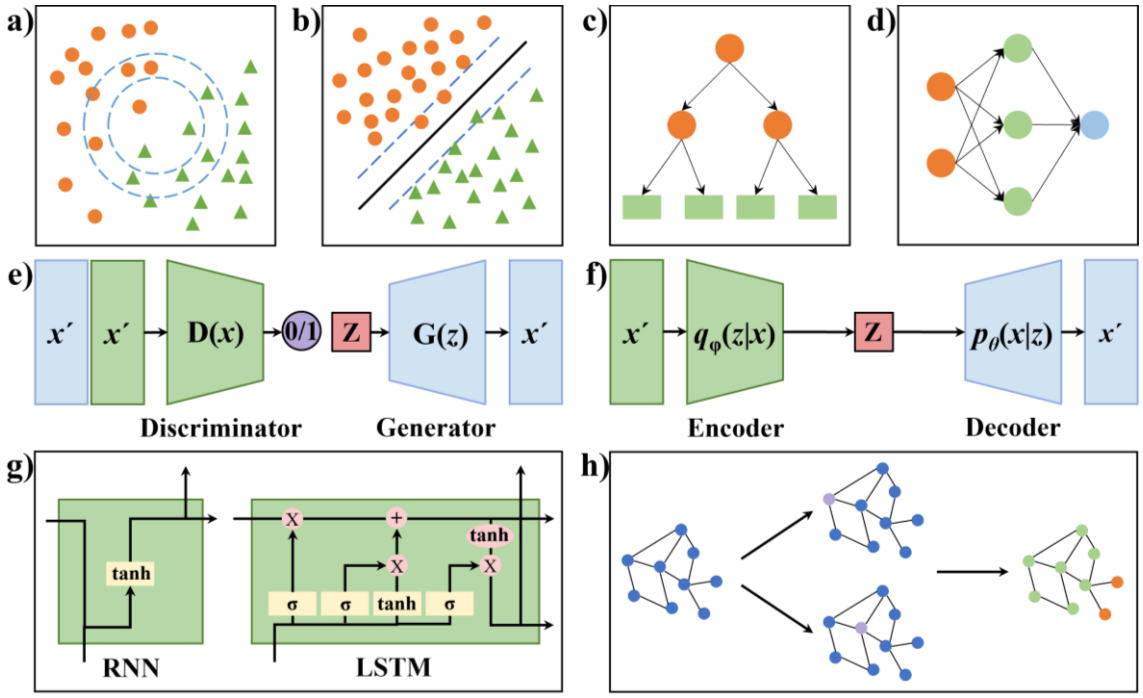


Figure 3. Schematic of widely used ML and DL algorithms: (a) k-nearest neighbors (KNN), (b) support vector machine (SVM), (c) decision tree (DT), (d) multilayer perceptron (MLP), (e) generative adversarial networks (GAN), (f) variational autoencoder (VAE), (g) recurrent neural network (RNN) and long-short term memory (LSTM), and (h) graph neural network (GNN).

Based on the criterion of whether the ML models can learn dynamically from a stream of incoming data, they can be also classified into two main categories.[152] One is static learning, also called batch learning, where all datasets are available before training the models. The other is active learning (AL) or called online/sequential learning, where datasets are fed sequentially to train the models.[153-163] AL trains the models by the streamed experimental data, thus the models can be dynamically updated to reduce the number of needed experiments. AL is well suited for developing AEPs since it can achieve optimal properties with fewer experimental iterations.[63-75, 164-170] When applied to materials simulation, AL is also beneficial in greatly reducing the computational cost.[110, 154, 156, 171]

AL is mainly implemented by Bayesian optimization (BO) and reinforcement learning (RL) algorithms. BO, a global optimization algorithm, is an adaptive approach for optimizing the expensive-to-evaluate objective functions.[172-174] BO utilizes a surrogate model (or belief model) for capturing the relationship between inputs and results, and an acquisition function (or decision policy) for selecting the optimum candidate for the next iteration of operation.[173-175] With the aid of advanced surrogate models and acquisition functions, BO is versatile to tackle various optimization problems like catalytic activity,[157, 176, 177], molecule design,[171] and structure/property prediction.[178-180] Distinguished from BO, RL is a reward-based learning approach that learns how to map situations to actions in an environment for maximizing the reward.[181] In the iteration, an agent acts to change its state simultaneously interacting with the environment. RL has been implemented in several applications such as molecule/drug design,[182-186] reaction synthesis planning,[51, 187-189] and novel material generation.[190]

2.2 Workflow of Constructing ML/DL Models

2.2.1 Data Collection

Figure 4 shows a typical workflow for constructing ML/DL models, including data collection, data preprocessing, model training, and evaluation. The first step is to collect data, from which the hidden knowledge can be extracted or learned by the ML/DL models. Quality, quantity, and diversity of the data largely determine the predictive accuracy, robustness, and generality of the developed ML/DL models.

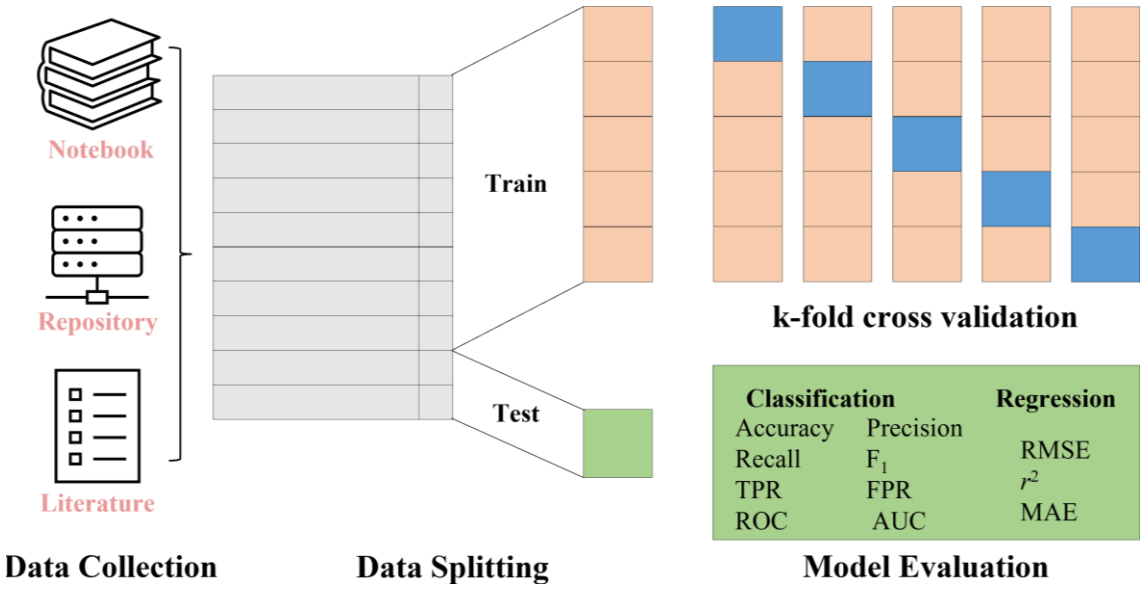


Figure 4. Workflow of building ML/DL models.

Commonly, three primary sources can be used to collect the training data. The first one is the historical data from the lab notebooks[45, 48, 191] or literature.[192, 193] The second one is from open database in the repository websites such as the International Union of Crystallography (IUCr, XRD), International Centre for Diffraction Data (ICDD), National Institute of Standard Technology (IR and MS), Crystallography Open Database (COD)[194] and RRUFF as well as theoretical computational databases such as AFLOW,[195] Materials Project,[196] OQMD,[197] MatNvai,[198] and NOMAD.[199] **Table 3** summarizes these websites. The collected datasets contain data related to physicochemical properties (molecular weight, specific surface area, and melting point), reaction conditions (temperature, pressure, and time), and fingerprints or segments describing the structure information of molecules. These data can be discrete for performing classification tasks or continuous for performing regression tasks. The third data source is from in-situ experimentation where the new data is collected. It should be noted that bias could exist in these data sources. For instance, in the materials science field, data is preferably collected from

the literature that is published in high-impact journals. Or the human researchers may largely determine the diversity and quality of the collected data.

Table 3. List of open data sources for materials.

Name	Full name	Data	URL	Ref.
COD	Crystallography Open Database	XRD	http://www.crystallography.net/cod/	[194]
IUCr	International Union of Crystallography	XRD	https://www.iucr.org/resources/data/databases	
ICDD	International Centre for Diffraction Data	XRD	https://www.icdd.com/	
CSD	Cambridge Structural Database	XRD	https://www.ccdc.cam.ac.uk/solutions/csd-core/components/csd/	
ICSD	Inorganic Crystal Structure Database	XRD	https://www.fiz-karlsruhe.de/en/produkte-und-dienstleistungen/inorganic-crystal-structure-database-icsd	
NIST	National Institute of Standard Technology	Raman, FTIR, MS	https://www.nist.gov/	
RRUFF		Raman, XRD, IR	https://rruff.info/	
AFLOW	Automatic-FLOW for Materials Discovery	Theoretical compound properties	http://aflowlib.org/	[195]
Materials Project		Inorganic compounds properties	https://www.materialsproject.org/	[196]
OQMD	Open Quantum Materials Database	Crystal structures	http://oqmd.org/	[197]
MatNavi	NIMS Materials Database	Polymer, inorganic and metallic materials, computational Electronic Structure	https://mits.nims.go.jp/en/	[198]
NOMAD	Novel Materials Discovery	Computational materials	https://nomad-lab.eu/	[199]
OC20	Open Catalyst 2020	Materials, surfaces, and adsorbates (nitrogen, carbon, and oxygen chemistries)	https://opencatalystproject.org/	[200]
GDB-13		Combinatorially generated library	http://gdb.unibe.ch/downloads/	
ZINC15		Commercially available compounds	http://zinc15.docking.org	
GDB-17		Combinatorially generated library	http://gdb.unibe.ch/downloads/	
QM9		Stable small CHONHF organic molecules taken from GDB-17 with properties calculated	http://quantum-machine.org/datasets/	

		from ab initio density functional theory		

2.2.2 Data Preprocessing

After the data collection, data preprocessing is the next important step to process the information related to the hypothesized problem and convert them into quantifiable data that can be read by a computer. Typical tasks of data preprocessing include (1) representing categorical data, molecules, text, and images; (2) removing duplication and noise; (3) handling missing data points; and (4) scaling features from unstructured raw data.

Representing categorical/text/image data. To represent categorical data, one-hot encoding can be implemented, which creates a binary column for each category.[201] If this category appears in the input, it is coded as 1. Otherwise, it is coded as 0. For example, three alcohols, i.e., methanol, ethanol, and propanol, are used as solvents in a reaction. To represent them using one-hot encoding, methanol, ethanol, and propanol are coded as [1 0 0], [0 1 0], and [0 0 1], respectively. To convert the text features into a set of representative numerical data, the term frequency-inverse document frequency (TF-IDF),[202] a technique of data encoding, is used to take each snippet of text, count the occurrences of each word within it, weight the word counts by a measure of how often they appear in the documents and present the results in a table. Pixel values of images can be directly used to encode the image data.

Removing duplication and noise. Duplication is a row where each value in each column is the same as another row. These duplications may appear in both training and testing datasets after data splitting, resulting in an optimistically biased performance of the model for the unseen data. Thus, they must be removed in the data preprocessing step. Noise is another concern that could make two patterns from two different structures have lower signal-to-noise ratios.[203] This may lead to

poor classification accuracy due to the loss of distinguished peak characteristics. To remove the noise and enhance signal-to-noise ratios, several smoothing strategies can be implemented including the Savitzky-Golay filter,[204] Fourier transformations,[205] and penalized likelihood estimation.[206] VAE has shown promise in automatically reconstructing spectra by removing noise and unwanted spectral artifacts.[207]

Handling missing values. In some situations, features and labels are missing regardless of whether they happen at random or not. Most ML algorithms cannot be successfully trained by datasets with the missing values. A straightforward way is to discard these observations that contain one or more missing values, which can be quickly implemented using NumPy or Pandas. However, directly deleting the observations may introduce bias into the data, thus resulting in some unobserved systematic effects. Another way is to fill these missing values with substitute ones by imputation. Imputation of missing values is implemented using various strategies such as replacing the missing values with the mean, median, or mode of the column and using matrix completion to impute the missing values with the observed elements.[115]

Scaling features. ML algorithms do not perform properly when some features have different scales of their absolute values, which may cause an over-weight of the features with relatively large values. For instance, in a reaction, the volume of solvent ranges from 1 to 1000 mL, while the molar ratios of two chemicals range from 0 to 1. If these values were directly fed into some distance-based algorithms such as KNN, k-means, and SVM, the molar ratio would have a much smaller weight in the prediction outcomes than the solvent would do. Implementation of feature scaling balances the weights of these features, leading to more robust models and better prediction accuracies. For some ML algorithms like LR and MLP which use gradient descent for model training, the feature scaling would make the models converge much faster. The third benefit of

implementing the feature scaling is that the coefficient can be appropriately penalized if regularization is a part of the loss function.

Two common approaches, i.e., *normalization* and *standardization*, can be used to make the features have the same scale.[48] Normalization subtracts the minimum and divides by the difference between maximum and minimum, while standardization first subtracts the mean value. The result is then divided by the standard deviation to make the distribution have unit variance. Unlike normalization, standardization is much less affected by outliers, thus resulting in a more robust and generalizable ML model.

2.2.3 Model Training

Data splitting. To avoid overfitting and increase model robustness and generalization,[48] the dataset is usually split into training and testing ones with a given ratio (typically 0.7, 0.75, 0.8, and 0.85). To train ML models with large training data ($> 10^4$), the splitting ratios among the training, validation, and testing can be 0.90, 0.05, and 0.05. The training and validation datasets are used for training and evaluating the ML models, respectively, while the testing dataset is set aside as never-seen data to evaluate the performance of the ML models.

Cross-validation (CV). A *k*-fold (*k* is usually set to 5 or 10) cross-validation technique is usually implemented to afford the ML models with high robustness and generalizability.[48] In a 5-fold cross-validation, the training data is split into five groups, one of which is used to evaluate the model trained on the remaining four datasets. Evaluation of the trained model is based on an unseen testing dataset.

Hyperparameter tuning. Hyperparameters are the parameters that must be set before training ML/DL models. They can either configure the ML/DL models through hyperparameters like the

number of trees in a decision tree, the number of layers, and the learning rate in a neural network or minimize the loss function by tuning the types of activation functions and optimizer in a neural network, the kernel types in SVM.[208] As a critical and cumbersome task in training the ML/DL models,[209] hyperparameter tuning examines different combinations of hyperparameters to get optimal results.[48] Manual tuning is a traditional way that manually fiddles with the hyperparameters until obtaining satisfactory results. However, it is tedious and is ineffective for many problems arising from the non-convex models, nonlinear hyperparameter interactions, and high dimensionality. Hence, automated hyperparameter optimization (HPO) has become a promising technique that automatically explores the hyperparameters to find the optimal performance. HPO has the advantages such as reducing the required human efforts, improving the performance of ML/DL models, enhancing reproducibility and fairness, lowering the technical threshold, and accelerating the training speed.[208, 209] To perform HPO, the methods like grid search and random search are widely implemented. The grid search evaluates the Cartesian product of hyperparameters,[210] while the random search chooses random combinations of hyperparameters.[211]. To avoid many unnecessary evaluations, BO selects combinations of the hyperparameters based on previous evaluation results.[212] Open source Python libraries including Hyperopt,[213] Talos,[214] Spearmint,[173] Autotune,[215] SMAC,[216] and Vizier[217] have been developed to meet the demand for performing automated HPO. To further automated the end-to-end ML/DL pipelines for freeing experts from the tedious HPS tasks and making ML accessible to non-experts, automated ML (AutoML) frameworks have emerged. These off-the-shelf frameworks include AutoWEKA,[218] Auto-Sklearn 2.0,[219] AutoKeras,[220] Auto-Pytorch,[221] H2O AutoML,[222], and TPOT.[223]

2.2.4 Evaluation Metrics

Classification. Several metrics including accuracy, precision, recall, F_1 score, true positive rate, and false positive rate can be used to evaluate the performance of the ML models. Below are the formulas used to calculate them.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

$$FPR = \frac{FP}{FP + TN} \quad (6)$$

where TP, FP, TN, and FN represent the number of true positives, the number of false positives, the number of true negatives, and the number of false negatives. F_1 score represents the harmonic mean of precision and recall. Precision and recall are used to plot precision-recall curves at different threshold settings, while TPR and FPR are used to plot the receiver operating characteristic (ROC) curves at various decision thresholds. The area under the ROC curve (AUC) can be used to measure how well the ML model distinguishes different classes.

Regression. Root-mean-square error (RMSE), coefficient of determination (r^2) and mean absolute error (MAE) are three main metrics used to evaluate a regression ML model.

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(y_{pred} - y_{true})^2}{n}} \quad (7)$$

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(y_{pred} - y_{true})^2}{\sum_{i=1}^{n}(y_{pred} - y_{mean})^2} \qquad (8)$$

$$\mathrm{MAE} = \sum_{i=1}^{n} \frac{|y_{pred} - y_{true}|}{n} \qquad (9)$$

where y_{pred}, y_{true}, y_{mean}, and n refer to the predicted value, true value, mean value, and the number of samples, respectively.

2.3 Challenge and Solutions of Data Scarcity in Physical Domains

Acquiring sufficient data is always a prerequisite for building robust and generalizable ML/DL models. However, it is time- and cost-intensive, particularly in experimental materials science. Training the ML/DL models with limited data may lead to overfitting of the models. To overcome the issue of data scarcity, some powerful methodologies including data augmentation by wrapping and oversampling techniques, dimension transformation, transfer learning, and data compression can be applied.

Data augmentation by wrapping. The data wrapping technique is usually implemented for generating more training data from the existing data.[56, 58] For example, one-dimension (1D) spectroscopic data can be augmented by random peak elimination, peak scaling, pattern shifting, and noise addition.[56, 224] For peak elimination, a range of specific numbers can be randomly replaced with zero. The peak intensities can be scaled by a factor at the predefined periodic length. To do pattern shifting, the entire spectra are red-shifted (increase in the wavelength) or blue-shifted (decrease in the wavelength) with a given value. To add noise addition to each spectrum, Gaussian noise with a zero mean and variance equal to 0.00001 or white noise can be applied.[224] Through these steps, the spectra can be largely augmented as defined.

The two-dimension (2D) microscopic data can be augmented *via* several random transformations that yield believable images. Trained with these images the DL models can be exposed to more aspects of the data and be more generalizable. Typically, examples of augmentation techniques include geometric transformation, color space transformation, noise injection, random erasing, and kernel filters. The geometric transformation consists of rotation, flipping, coupling, translation, vertical and horizontal shifting, channel shifting, and shearing.[225] The color space transformation, also known as photometric transformation, can be done via splicing individual RGB color matrices, setting certain max/min pixel values, and decreasing or increasing the pixel values by a constant value.[226, 227] Random erasing sets the values of an $n \times m$ patch that are randomly selected from an image to be either 0 s, 255 s, mean pixel values, or other random values.[228] The Kernel filtering sets the pixel values of an $n \times n$ sliding matrix using a Gaussian blur filter or a high contrast vertical/horizontal edge filter.[229] As no new information is produced from data augmentation, eliminating the overfitting is still almost impossible.

Data augmentation by oversampling. Unlike the data augmentation by wrapping that manipulates existing instances, the data augmentation by oversampling generates synthetic instances that are added to the training datasets. There are already several oversampling techniques, e.g., mixing images, augmenting features, transferring neural style, and implementing GAN. The technique of mixing images averages the pixel values of the images after randomly cropping, and flipping, and then assigns the new images in the same way as the originally selected images.[230] The feature augmentation works by first mapping the images into low-dimensional representations, also known as features which are then augmented by methods such as adding random noise.[231] The neural transfer method applies the style of a reference image to a target image via the

sequential representation while preserving the original content of the target image.[232, 233] Augmentation by GAN applies the probability distribution learned from training images to generate new artificial images, which show high diversity but low correlation, thus are statistically indistinguishable from the original ones.[146, 234]

Dimension transformation. Though data augmentation of 1D vector data like XRD, Raman, and FTIR improves the predictive accuracy, it still suffers from several problems. First, the presence of noise and background signals in these spectra leads to a higher false-positive rate and a low detection accuracy of the true peaks. Though a few noise removal and smoothing algorithms are available, the results are not consistent.[235] Second, some temporal correlations across frequencies of the IR and Raman spectra could not be captured using 1D CNNs, resulting in lower prediction accuracy. Encoding 1D spectra to 2D images using data transformation functions has been a practical approach to solving these problems.[236-238]

Several techniques including Gramian Angular Field (GAF),[236] continuous wavelet transformation (CWT)[237] scalogram, spectral short-time Fourier transform (SSTFT),[239, 240] spectra recurrence plot (SRP), and spectral Markov transition field (SMTF)[238] have been successfully implemented to encode 1D vector of spectra to 2D vector of images. These transformed images are used to train transferred models, leading to a higher predictive accuracy than the models trained directly from the spectra. GAF can represent time series in polar coordinates via encoding the intensity as the angular cosine and the time stamp as the radius. Then, various operations can be implemented to transform these angles into symmetric matrices. GAF has several advantages such as preserving the temporal dependency. CWT is another transformation function that can capture characteristic frequencies of the signal.[241] Due to the continuous transformation at every scale, CWT makes the information present in the peak shape

and peak composition more visible and easier to be interpreted.[242] It is demonstrated that the implementation of CWT not only implicitly removes the baseline but also promotes the detection of peaks in the spectra. In addition, with the transformed 2D images, existing CNN models can be transferred for learning less available data, resulting in improved predictive accuracy.[237] SSTFT can convert Raman spectra into 2D spectrograms via Fourier transformation, which confers much discriminatory information and removes redundant features.[239] SRP can convert the spectra based on the internal structure of the wavenumber, while SMTF does the conversion based on the position information of the wavenumber.[238]

Transfer Learning. Transferring learning (TL) of a pre-trained network is another common and highly effective approach to solving the issue of data scarcity. Recently, a variety of pre-trained networks such as VGG19,[243] ResNet152 V2,[244] Inception V3,[245] Inception-ResNet,[246] Xception,[247] DenseNet201,[248] and EfficientNetB7[249] are publicly available. They are typically trained by very large image databases. For instance, ImageNet is trained by 14 million labeled images from 1000 object classes.[250] Thus, they can be transferred as effective generic models for applications in the physical domains.

There are two common approaches to performing TL in the pre-trained networks: *feature extraction* and *fine-tuning*. Feature extraction effectively extracts new features from data using representations learned from the model. Training is done *via* freezing convolutional layers of the pre-trained network while unfreezing the fully connected layers or called the classifier. The reasons why feature extraction can improve the training accuracy are as follows. First, the representations learned from the convolutional layers are generic and reusable, while the representations learned from the classifier are related to the set of classes. Second, local information of an object that appears in the data is lost in the classifier while such information remains in the convolutional

layers. Fine-tuning is a training procedure that unfreezes the first few convolutional layers while jointly training the newly added fully connected layers. The last few convolutional layers encode more generic, reusable, and specialized features.

Data Compression. Material characterization is crucial to understanding the structures, properties, and performances of the target materials. The collected data is usually in high-resolution and contains enriched information. However, the large data size delays data acquisition and increases the storage and communication burden, thus increasing the cost of data collection and analysis. While most data contains redundant information, they can be discarded with almost no perceptual information loss.[251] A method of directly sampling the compressed representations rather than the complete raw data may promote the efficacy of AEPs.

Compressed sensing (CS), a new data technique, has been demonstrated as a practical solution for directly collecting data in a form of compressed representation.[252] Instead of acquiring data and then post-eliminating redundancy using various compression schemes, CS allows the researchers to collect useful data from real-time experiments. CS can improve data sampling and acquisition rates and reduce the communication burden. Moreover, it allows the usage of low-quality and low-resolution data for training models with high prediction accuracy. To significantly improve the sparsity level of the recovered vectors and compression ratios, physics-based compressive sensing (PCBS), which uses domain knowledge and physical models to define the transformation and sparse vectors in CS, was proposed to monitor the temperature and melting of metals in additive manufacturing.[253]

3. Essential Components of an AEP

In this section, we will introduce the research progress and efforts made in AEPs with a focus

on how data-driven algorithms, especially ML/DL, can be implemented and/or integrated with the systems. Accordingly, we consider the three essential components that consist of an AEP for realizing a closed-loop self-driving lab. They include (1) automated experimentations, (2) on-the-fly data analysis, and (3) decision making.[68, 69, 74, 75]

First, an automated synthesis/processing platform is made of easily controllable and programmable equipment for automatically synthesizing materials of interest in a high-throughput manner. Despite various types of platforms, continuous flow reactors, desktop robots, and mobile robots have been mostly deployed in AEPs. In-situ and on-site characterization equipment—including spectroscopies such as UV-Vis/Infrared (IR)/photoluminescence (PL), gas chromatography-mass spectroscopy (GC-MS), liquid chromatography-mass spectroscopy (LC-MS), high-performance liquid chromatography (HPLC), nuclear magnetic resonance (NMR), electron paramagnetic resonance (EPR), as well as microscopies like atom force microscopy (AFM), scanning electron microscopy (SEM), transmission electron microscopy (TEM), scanning probe microscopy (SPM), scanning tunnel microscopy (STM), scanning transmission electron microscopy (STEM), piezo-response force microscopy (PFM), optical microscopy (OM), and digital imaging—can be integrated within the synthesis/processing platforms to characterize structures and properties of materials or classify the reaction results. Second, the on-the-fly data analysis with the aid of ML/DL can automatically process raw data to qualify or quantify the outcomes such as microstructures, product yield, catalytic activity, and reaction kinetics. It can undoubtedly accelerate the development pace and provide real-time feedback for the following decision-making step. In this section, we focus on two types of data, i.e., spectroscopic data collected from LC-MS,[62, 98, 254, 255] GC-MS,[256-258] NMR,[57, 63, 259, 260] IR,[99, 261, 262] XRD,[56, 58, 203, 263-265] and microscopic data collected from AFM,[266-268] SEM,[117,

269-271] TEM,[272-274] SPM,[275-279] STM,[118, 280, 281] STEM,[282-287] PFM,[288-290] OM,[59, 61, 291] and digital imaging.[100] Last but not the least, intelligent decision-making algorithms can actively learn from previous experimental outcomes to suggest a new set of reaction parameters for the next iteration of the experiments. Herein, four widely investigated decision-making algorithms, namely BO, RL, Evolutionary Algorithm (EA), and Random Goal Exploration Algorithm (RGEA), are discussed. Distinguished from brute-force search of the exploration space adopted by high-throughput workflows,[292] the intelligent decision-making algorithms are devoted to finding a shortened path to a global optimal. They not only ensure the accelerated discovery of novel materials with superior properties but also minimize human effort and lower the cost.

3.1 Automated Experimentations

In a traditional chemical reaction procedure, reactors such as flask, beaker, vial, and autoclave are commonly used under the guidance of the design of experiment (DoE). The products are collected *via* washing, centrifuging, and drying, followed by physicochemical characterization and performance evaluation. Though it has promoted scientific progress for centuries, this type of research paradigm has its intrinsic disadvantages. For instance, it is time- and cost-intensive, and imprecise, thus limiting the ability for rapid materials discovery. To remain competitive and deliver the expected benefit, efficient algorithms must be deployed to optimize the processes. With the advances in hardware and software, the focus of the academy and industry has gradually switched to automation or so-called high-throughput experimentation (HTE). HTE can conduct parallel experiments (from a hundred to tens of thousands) that combine reaction variables such as solvents, reactants, molar ratios of compositions, and temperatures. In addition to HTE for reactions, many

high-throughput characterization tools have been developed for online monitoring. Ever since the first demonstration of complete automation for molecule synthesis in the 1960s-1970s,[293-295] a few research groups such as Lauterbach,[296-299] and M. Ahamdi[300, 301] have made remarkable success in the field of pharmaceutical, organic chemical and DNA-sequencing industries in the past half-century. Nowadays, automation is expanding to the field of materials science.

A combinatorial approach, a watershed for accelerating materials discovery, development, and optimization, refers to parallelly synthesizing and characterizing many compounds in a matrix form for rapidly investigating the large compositional and structural landscapes. This field is pioneered by X. Xiang,[302, 303] I. Takeuchi,[304-308] J. Gregoire,[309-314] H. Christen,[315, 316] P. Rack[317-319], T. Unold,[320-322] to name a few. Meanwhile, the discovery of novel materials has changed radically with the introduction of HTE enabled by liquid handling robots (Chemspeed, Tecan, Hamilton, Hudson, Sartorius, Gilson, ThermoFisher, INTEGRA, Opentrons, and Andrew Alliance).

Herein, we review three main types of automated experimental platforms for THE: continuous flow reactors (CFRs),[11, 63, 67, 68, 70, 71, 73, 170, 323-343] desktop robots,[17, 66, 69, 74, 75, 164-169, 300, 344-351] and mobile robots.[64, 352-355] Table 4 shows the detailed comparison of these three types of platforms in terms of their application scenarios, deployment sizes, throughput, cost, and efforts to deploy.

Table 4. Comparison among three types of automated experimental platforms

No	Types	Scenario	Size	Throughput	Cost	Effort
1	Continuous Flow Reactors	Molecules, nanoparticles, drugs, polymers	Small	Medium	Low ~10K	Low
2	Desktop Robots	Molecules, nanoparticles, drugs, polymers, thin films, single crystals, solids	Medium	High	Medium 10K~100K	Medium
3	Mobile Robots	All	Large	Low	High >100K	High

3.1.1 Continuous Flow Reactors

Restrained by limited heat and mass transfer, traditional reactors such as beakers, flasks, and autoclaves usually suffer from significant batch-to-batch variability and generate unexpected byproducts. Hence, it is not appropriate for accurate investigation of reaction kinetics when dealing with process-dependent organic/inorganic materials synthesis, which requires fast, reproducible, and controlled reactions enabled by the rapid heat and mass transport.[84, 356]

In recent years, CFRs, where reactants are continuously fed, have achieved enormous progress in producing fine chemicals and specialty materials (**Figure 5a**).[356-361] They have the following advantages. First, they enable fast mass transport for efficient reactant mixing as channel miniaturization leads to large and well-defined interfacial areas in CFRs. This prevents the formation of byproducts generated from localized concentration gradients. Second, efficient heat transfer is realized due to the small channel diameters in CFRs. This advantage can largely avoid locally hot spots and byproduct formation. Third, the usage of both solvents and reactants can be reduced. Reactions in traditional reactors require from milliliters to liters of reactants and solvents, while they can be largely minimized to a few microliters or even nanoliters for the same reactions. Fourth, CFRs have shown the capacity to substantially increase reaction selectivity and reproducibility. Fifth, CFRs can be modularized for sequential reaction, filtering, and extraction all in one continuously streamlined flow process. Last but not the least, CFRs can increase operating safety. Due to the automated operation and less usage of reactants, it makes the research procedure much safer when handling hazardous, toxic, or even radioactive chemicals.

A typical CFR has the following essential components: 1) precursor formulation modules including a precursor stock and micromixer; 2) tubes (plastic or stainless steel) equipped with heating units such as an oil bath and a heating coil if necessary; 3) separators for purification and

collection. They are easy and favorable to be integrated with many in-situ or in-line analytical instruments. The integration enables real-time monitoring and analysis of reaction outcomes. It allows the researchers to efficiently explore chemical space and easily extend the systems with increased complexity.

Case studies. Cronin and his coworkers designed and built a synthesis robot based on CFRs for fully autonomous organic compound synthesis (**Figure 5b**).[63] All the reactors were connected *via* syringe pumps and switch valves. To determine the reaction results, real-time analytic tools including NMR, MS, and ATR-IR systems were integrated. The collected data was real-time analyzed and then fed into an optimization algorithm for decision making. Four new reactions were discovered through the chemical robot. Abolhasani et al. developed a fully autonomous CFR named Artificial Chemist for raid synthesis of perovskite quantum dots (QDs) (**Figure 5c**).[68] The Artificial Chemist consists of three main modules, i.e., a precursor formulation module, a flow reaction module, and an in-situ QD characterization module. To enable in situ monitoring, a flow cell with reduced path length was designed and integrated into the CFR for recoding PL and UV-Vis spectra. By a multivariate process optimization algorithm, this Artificial Chemist can synthesize QDs with target optoelectronic properties even without prior knowledge about the QD synthesis. Just recently, Reis and his colleagues built a CFR robot capable of polymerizing multiple samples simultaneously (**Figure 5d**).[325] A droplet-based flow system was employed for high-throughput polymer synthesis.[362] The polymer properties can be optimized over their compositions, molar masses, and dispersity. In this CFR robot, a few ^{19}F MRI copolymer agents with high imaging sensitivities were discovered.

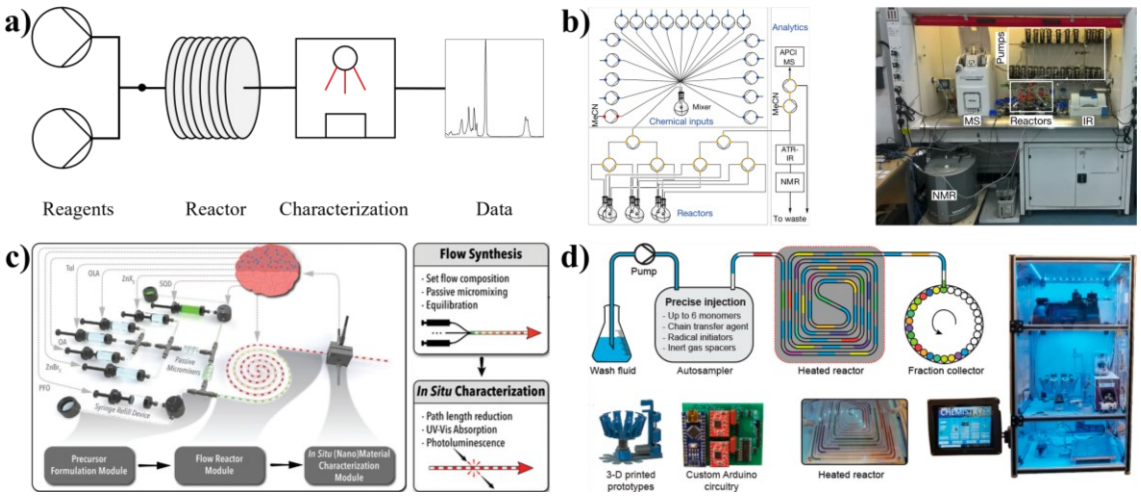


Figure 5. (A) Scheme showing reaction workflow in a continuous-flow reactor (CRF) based AEP. (B) Schematics (Left) and Photograph (Right) of a chemical robot based on a CFR. Reproduced with permission from Ref.[63], Copyright 2018 Springer Nature. (C) Schematics of an Artificial Chemist for autonomous synthetic path discovery and optimization of colloidal QDs. Reproduced with permission from Ref.[68], Copyright 2020 Wiley. (D) Schematic of a CFR for autonomous polymer synthesis. Reproduced with permission from Ref.[325], Copyright 2021 America Chemical Society.

CFRs are usually designed for the synthesis of specific materials. To increase their agility, modularizing them for new projects is an efficient way. Several groups have made a great contribution to modulating CFRs for conducting new reactions without redesigning the system. Jensen and Jamison et al. developed a compact, fully integrated, and easily reconfigurable platform for automatically optimizing a wide range of chemical reactions (**Figure 6a** and **b**).[341] Six available modules were designed for plug-and-play operations. Meanwhile, characterization instruments including HPLC, MS, and vibrational spectroscopies were integrated for real-time monitoring. Cronin and his coworkers designed an automated modular synthesis platform, called

Chemputer, to synthesize the organic compounds with minimal human intervention (**Figure 6c**,**d**).[67] The platform has a backbone structure that enables facile switching of modules for routine synthesis tasks such as heating or phase separation. The backbone has a six-port valve that connects pumps to the modules so reagents or reaction mixtures can flow to the appropriate module. Three drugs including diphenhydramine hydrochloride, rufinamide, and sildenafil were synthesized in 38-100 hours with yields comparable to the reported ones by traditional batch synthesis ways.

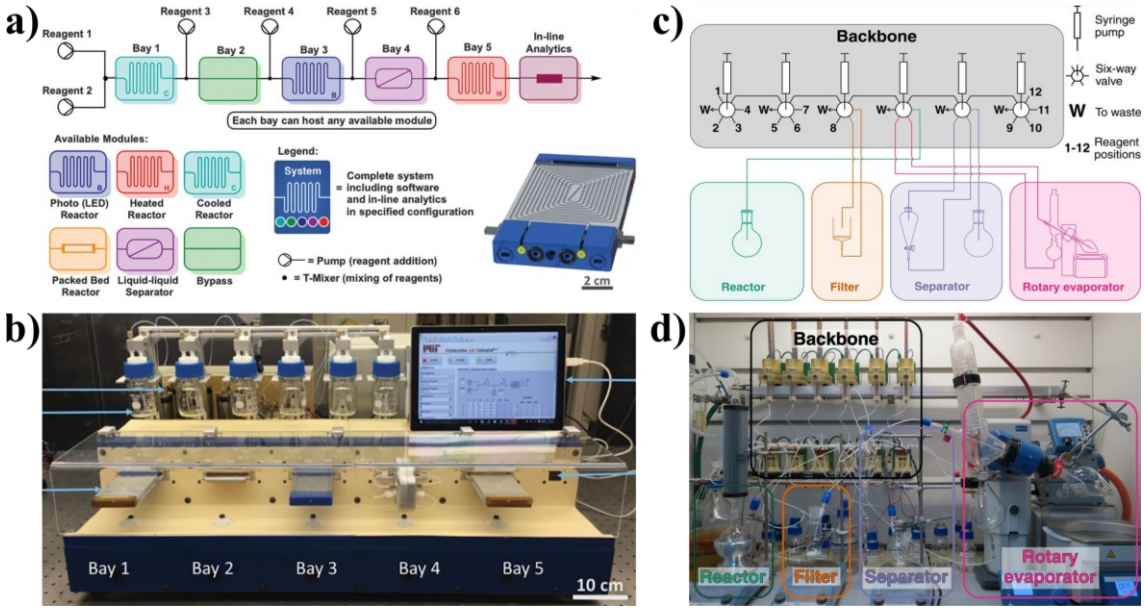


Figure 6. (a) Schematics and (b) photograph of a plug-and-play, reconfigurable, continuous-flow chemical synthesis system. Reproduced with permission from Ref.[341], Copyright 2018 AAAS. (c) Schematics and (d) photograph of the Chemputer setup. Reproduced with permission from Ref. [67], Copyright 2019 AAAS.

Although CFRs have shown numerous benefits in material synthesis, they face challenges that deter their widespread applications in large-scale materials, pharmaceutical, fine, and specialty chemical production. The first challenge is how to prevent the formation of solid precipitates in

these cases that use solid and/or high-viscosity liquid.[363, 364] The solid formation accumulates fouling in channels, leading to blockage, which not only causes fluctuation of the flow velocity but also induces a major setback when optimizing the reaction parameters. To effectively circumvent the fouling, feasible approaches such as the use of dilute liquid,[365] tubes with larger diameters,[366] a micro-flow focusing technique,[367] multiphase systems[368, 369], and ultrasound acoustic irradiation can be implemented.[370, 371] In addition, the use of solid in reactions can cause clogging. To alleviate this problem, several strategies can be implemented, i.e., adding magnetic/mechanical stirring in the feeding system,[372] introducing continuously stirred tank reactors,[373] and coating packed column reactors with heterogeneous catalysts.[374] The second challenge is how to integrate in-line purification for obtaining the desired high-purity products. Purification not only helps to determine the yield but also mitigates the side effect of the by-products on the subsequent reactions. The viability of in-line purification techniques such as liquid-liquid extraction,[375, 376] micro-distillation,[377] micro-crystallization,[378] and free-flow electrophoresis[379] have been investigated. The third challenge is the relatively higher cost needed to build CFRs compared with the traditional batch reactors.[380] The commercially available syringe pumps, rotary valves, various reactors, and multichannel connectors are the components that cost the most in building the CFRs. A few research groups are devoted to developing affordable devices via assembling necessitate components with 3D printed parts,[381-386] greatly reducing the deployment cost.

3.1.2 Desktop Robots

Though CFRs have provided a promising way for experimental automation, they still lack enough agility when participating in various types of chemical reactions. Because agile chemical

synthesis calls for a redesign of a synthetic route, interchange of different chemicals, selection of appropriate reactors with suitable sizes, reassembly of different hardware, and optimization of the whole process. For example, lots of enzymes used in pharmaceutical products should be well stored and prepared just before the reactions. Desktop robots with ingenious arms possess great advantages for mixing, processing emulsions, and handling chemical solid and liquid (**Figure 7a**). The application of desktop robots in the field of life science and drug discovery has been a tremendous success. An early demonstration of the desktop robots for biomedical research came from a group of scientists at Aberystwyth University.[387-392] They built two prototype robots: Adam and Eve. Adam was designed to test genes and enzyme functions, while Eve was devoted to screening and designing drugs of interest. Implementation of desktop robots in developing novel materials has also been demonstrated by different research groups.

Case studies. Cronin et al. designed and constructed a droplet-generating desktop robot named "Dropfactory", a robust platform with easy maintenance, for investigating droplet behaviors (**Figure 7b**).[75] Dropfactory has three main mechanisms: an XYZ CNC frame that provides both the structural support and the motion, working stations that perform only one task at each running, and two Geneva wheels that move containers from one station to the others. Dynamics of the oil-in-water droplets including movement, division, fusion, and chemotaxis were recorded using a commercial camera to construct promised protocell models. All operations including mixing, droplet placing, recording, cleaning, and drying were parallelly performed for 300 experiments per day in full autonomy, showing 6 times increase in throughput compared with their previously developed non-autonomous platform.[393, 394] Based on their previous work,[341] Jensen and his coworkers integrated a six-axis robotic arm with modularized CFRs to develop an automated and scalable synthesis of organic compounds (**Figure 7c**).[66] This robotic arm allowed automatic

and on-demand selection of modules from storage locations and arranged them in a required reaction sequence suggested by AI planning algorithms for investigating the amide coupling and reduction reactions. This reconfigurable platform yielded the target compounds at a high rate of 100 mg/h. Besides successful demonstration of synthesizing 15 drug or drug-like molecules, it can back-to-back synthesize complex molecules. A collaborative team led by Aspuru-Guzik, Hein, and Berlinguette designed and built an "Ada" desktop robot that was capable of autonomously synthesizing, processing, and characterizing organic thin films (**Figure 7d**).[74] This robot consists of 1) a robotic arm for handling vials and slides, 2) a weigh scale for preparing precursors, 3) a spin coater for thin-film coating, and 4) a furnace for thin-film annealing, 5) a camera for dark-field imaging, 6) a four-point probe unit for electrical conductance measurement, and 7) an ultraviolet-visible-near-infrared (UV-Vis-NIR) spectrometer for recording spectra. Thin films were automatically prepared with the recommended chemical compositions and processing conditions by the AI optimization algorithms. From the measured absorbance and electrical conductance, the pseudomobility, which is proportional to the hole mobility, of the thin film materials was derived as the optimization target of the AI algorithms. Finally, maximum pseudomobility of thin-film materials with value of 750 s was successfully screened out within 30 hours.

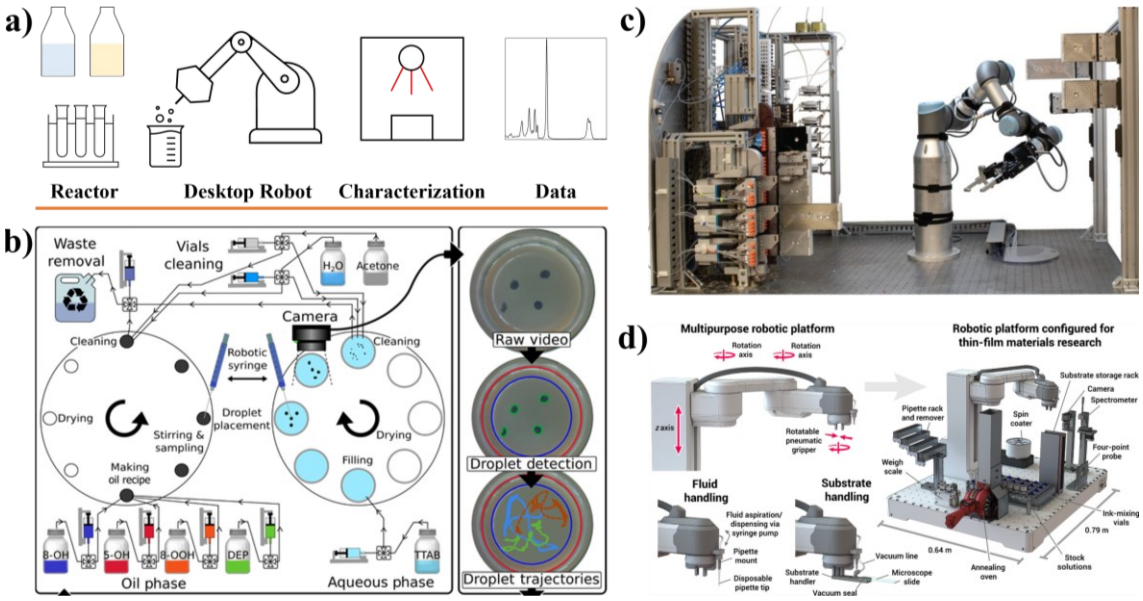


Figure 7. (a) Schematic showing a role of a desktop robot in an AEP. (b) Schematic of a high-throughput droplet-generation robot named "Dropfactory". Reproduced under a Creative Commons Attribution License 4.0 (CC BY) from Ref.[75], Copyright 2020 AAAS. (c) Photograph of a reconfigurable flow chemistry platform enabled by a desktop robot for performing multistep chemical synthesis. Reproduced with permission from Ref.[66], Copyright 2019 AAAS. (d) Schematic of an "Ada" self-driving laboratory for fabrication and characterization of thin-film materials. Reproduced under a Creative Commons Attribution License 4.0 (CC BY) from Ref.[74], Copyright 2020 AAAS.

3.1.3 Mobile Robots

Lots of synthesis and characterization instruments are spatially big and functionally complex. In some cases, they require special working environments, e.g., isolation of UV, noise, and vibration. Thus, they are often distributed in different locations, making the direct integration into a single platform often impractical. In this case, realizing full autonomy would demand a mobile robot that can serve as an operator like a human researcher. Recent advances in industrial robots

produced by KUKA, Fanuc, ABB, and Yaskawa have inspired much interest in exploring their applications in autonomous laboratories for material and chemical development.[352-354] A mobile robot with a built-in scheduling software can physically move among different components of a synthesis lab to perform tasks like handling chemicals (solid and solvent) from cabinets to synthesis stations (dissolution, distillation, and centrifuge) and characterization stations (HPLC, LC-MS, UV-Vis and GC) without human invention (**Figure 8a**).[64, 355] This mobile robot can realize unmanned intelligent labs, showing superiority to humans in consistence, efficiency, flexibility, and dealing with toxic and explosive chemicals/gases.

Case studies. In 2018, Li et al. proposed an authentic intelligent robot for use in a chemistry laboratory (AIR-Chem), which automatically executed the synthesis of inorganic perovskite quantum dots (IPQDs) (**Figure 8b**).[355] AIR-Chem consists of an automated guided vehicle (AGV) and a real-time computer vision (CV) system. The AGV can navigate the chemical cabinets and conduct IPQD synthesis experiments with the aid of CV. The embedded CV integrated with a PL device can monitor the IPQD growth in real time. In another recent work, a commercially available mobile robot was used to replace a human researcher in conducting experiments (**Figure 8c**).[64] Using laser scanning and touch as the feedback medium, the robot chemist can move freely and accurately in a standard laboratory under a dark environment, which is required for handling light-sensitive chemicals or photochemical reactions. In addition, it can work continuously except for charging, which takes ~2.4 h per day. Compared with other automated platforms that only handle liquid, it can accurately and reliably dispense both solid and liquid. As a demonstration, it was used to search for efficient photocatalysts for hydrogen production from water. Without any instruction from human researchers and prior knowledge, it synthesized and tested the catalysts, and then obtained an optimized recipe from a ten-variable reaction space in

just 8 days. It is worth mentioning that unlike the previously reported flow synthesis-based robotic platforms upon which many modules are customized, in this work, all the stations except the capping and photolysis stations are commercially available. Thus, no hardware modification is needed.

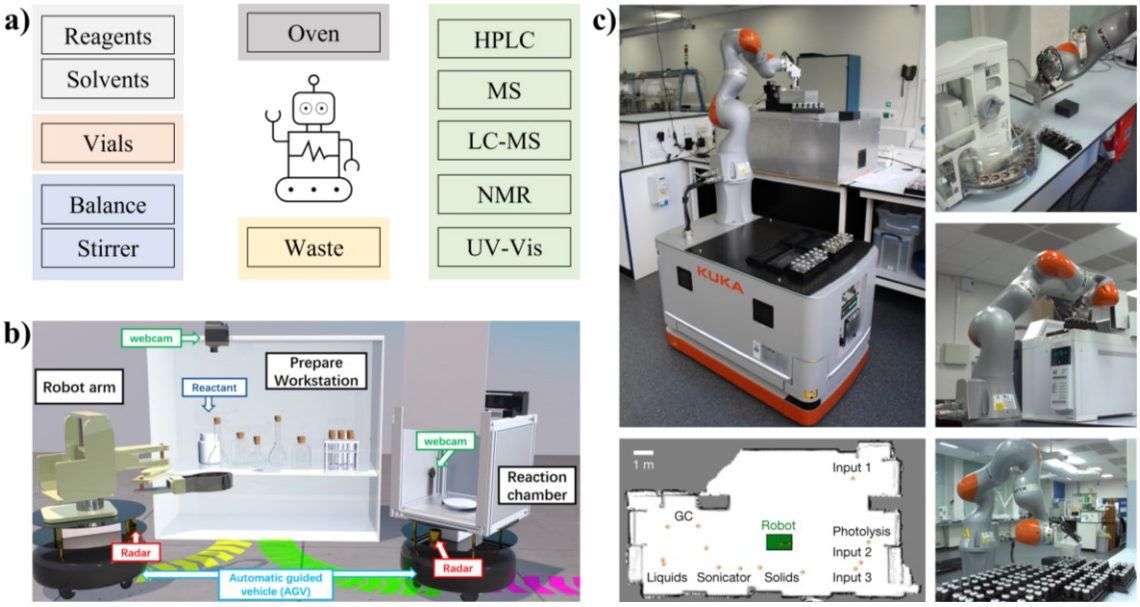


Figure 8. (a) Schematics showing a role of a mobile robot in performing experiments. (b) Schematics showing components of AIR-Chem and their functionalities. Reproduced with permission from Ref.[355], Copyright 2018 American Chemical Society. (c) Photographs showing a Kuka mobile robot handling samples for synthesis and characterization. Reproduced with permission from Ref.[64], Copyright 2020 Springer Nature.

Similarly, the deployment of desktop and mobile robots in labs also faces grand challenges. First, high cost and long investment time remain the biggest constraints.[395-398] The cost associated with a liquid handling robot, robotic arm, and mobile platform is still an important factor for large-scale applications in the lab. It is also a time-consuming process to develop the software to execute commands and communicate among different hardware. To mitigate these

issues, hardware with affordable prices and open-source software is highly desired. The second challenge is associated with precise and repetitive positioning and fine manipulation of both desktop and mobile robots after a long-time operation.[64] Unlike the CFRs that liquids flow through the tubes, robots should allow 1) the fine manipulations, such as placement of vials, substrates, and tips, measurement of solid and liquid, and on/off of specific instruments, and 2) precise and repetitive positioning among various experimental stations in the labs. To alleviate these concerns, touch-sensitive multiple points calibration, and advanced sensing techniques of using laser and radar for robotic navigation can be implemented. The third challenge arises from the capability of the currently available robots to work only in a structured environment, where a spatial arrangement is organized and determined, making the robots less adaptive to emergent situations. In the future, advances in collaborative robots, and human-robot interactions would make these commercial robots more applicable in AEPs.

3.2 On-the-fly Data Analysis

After the collection of characterization data, a subsequent step is to analyze them followed by presenting and visualizing the results for decision making. The ever-increasing acquisition rate from real-time experiments by modern instruments leads to an exponential increase in data size.[5] However, data analysis typically requires domain knowledge, costing an expert much time and effort to process, interpret, and convert the data. To make the best use of the fast acquisition rate, it is necessary to boost the speed and efficiency of data analysis. Recently, ML has been deeply integrated into the characterization instruments for achieving the on-the-fly data analysis. This aspect of the analytic workflow focuses on spectroscopic data from LC-MS,[62, 98, 254, 255] GC-MS,[256-258] NMR,[57, 63, 259, 260] IR,[99, 261, 262] XRD,[56, 58, 263-265, 399] and

microscopic data from AFM,[266-268] SEM,[117, 269-271] TEM,[272-274] SPM,[275-279] STM,[118, 280, 281] STEM,[282-287] PFM,[288-290] OM,[59, 61, 291] and digital imaging.[100] A few reviews summarize recent progress in the application of DL for microscopic data analysis.[60, 400, 401] Herein, we focus our review on a general procedure of on-the-fly data analysis of both spectroscopic and microscopic data, and discuss how to extract insightful information for new knowledge generation, which can be used to achieve predefined targets, such as establishing a processing-structure-property relationship.[402]

3.2.1 Spectroscopic Data Analysis

Data preprocessing. Usually, spectra collected from different equipment and/or by different operators contain different amounts of data points. For most of the ML/DL models, the input vectors should possess the same length. To convert the raw spectrum data to a vector of a specific length within a given range, the interpolation is usually first implemented.[56, 58, 224] In addition, the data is usually normalized to the range between 0 and 1 to make the data on the same scale for better comparison. In the case of experimental spectra containing much background noise, Savitzky-Golay filtering and polynomial fitting techniques can be applied to smooth the spectra and correct the baseline, respectively.[403] More information about the spectra data preprocessing can be referred to in the previous section.

Case studies. Grand et al. trained an SVM model to classify NMR and IR spectra of reactive and non-reactive mixtures (**Figure 9a**).[63] These data were manually labeled by domain experts. The model trained on 72 datasets afforded an accuracy of 86%. This well-trained SVM model was then used for real-time distinguishing the spectra of the starting materials from those of the final products. Finally, the difference between the two types of spectra was registered as reactivity hits to classify reactive and non-reactive outcomes. In another work, an algorithm named *peakonly*,

consisting of two CNN models, was developed to detect true positive peaks of raw LC-MS spectra (**Figure 9b**).[98] The first one was used to classify the regions of interest (ROI) into three categories (noise, peaks, and uncertain peaks), while the second one was used to output the area of the detected peaks. *Peakonly* shows superior performance in labeling the true positive peaks with a precision of 97 %. To capture temporal correlations, Zinchik and his colleagues used GAF to encode the 1D mid-infrared (MIR) spectra to 2D matrices for training a CNN model (**Figure 9c**).[236] To reduce the dimension of the input GAF matrices, a piecewise aggregate approximation (PAA) technique was used. The results showed that the CNN model reached an overall classification accuracy of ~100% at a much faster prediction rate than the model trained directly with 1D data.

Tremendous progress has been recently made in the application of ML/DL models in analyzing XRD, Raman, and FTIR spectra. Buonassisi et al. proposed a CNN model to predict crystallographic space groups of XRD patterns of perovskite thin films.[56] Data was augmented from the theoretic spectra to overcome the issue of scarcity in experimental data. To validate its effectiveness, they integrated the CNN model with high-throughput synthesis to accelerate the development of perovskite-inspired materials.[263] Such an integrated approach achieved a classification accuracy of 90% and a classification speed of > 10 times faster than manual analysis. Recently, our group trained a CNN model from theoretical XRD patterns combined with very limited experimental spectra.[58] Rather than classifying the materials into crystallographic space groups, this CNN model enables rapid identification of individual metal-organic framework (MOF). It affords a prediction accuracy of 96.7% for the top-5 ranking among > 1000 MOFs. Fan and his colleagues developed a novel approach of DL-based component identification (DeepCID) to identify the presence of species in mixtures from Raman spectra.[404] The well-trained

DeepCID exhibits a prediction accuracy of 98.8% for 167 compounds and 99.5% for 160 compounds with significantly lower false-positive rates.

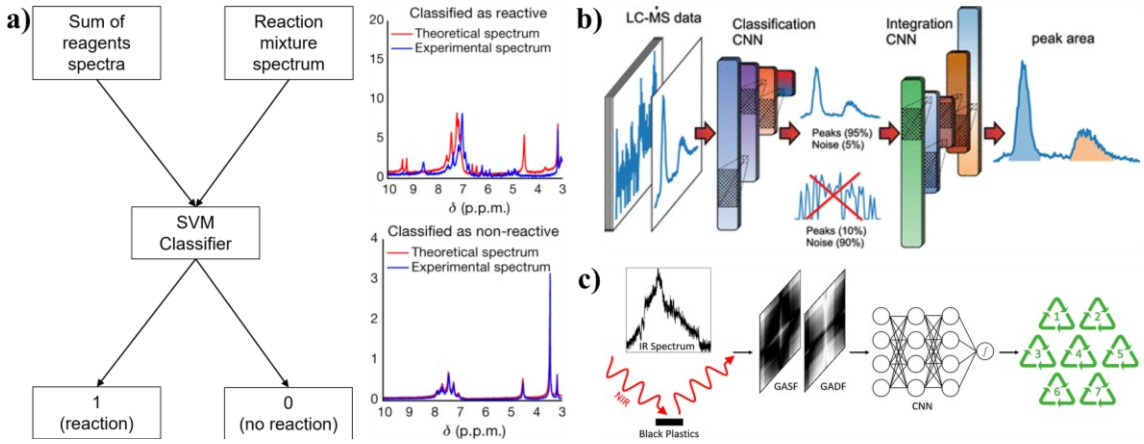


Figure 9. (a) Schematic showing workflow and results of an SVM classifier for reaction outcome detection from NMR spectra. Reproduced with permission from Ref.[63], Copyright 2018 Springer Nature. (b) The architecture of the *peakonly* model for peak classification and integration. Reproduced with permission from Ref.[98], Copyright 2020 American Chemical Society. (c) Workflow of a CNN framework for classifying different types of plastic using a GAF transformation method. Reproduced with permission from Ref.[236], Copyright 2021 American Chemical Society.

3.2.2 Microscopic Data Analysis

Microscopic images taken on advanced instruments such as SEM, TEM, and STM have been widely used to retrieve the relationship between microstructures and properties.[405, 406] Analyzing these microscopy images includes tasks of segmenting areas of interest such as defects and phases and determining the thickness and number of layers. However, such a manual workflow is tedious and time-consuming. In addition, some important information hidden in the image data may be missing due to intrinsic limitations of the equipment or unintentional ignorance of human

researchers. In contrast, the DL models such as DenseNet,[407] ShuffleNet,[408, 409], and Mask R-CNN[410] were proved to perform better tasks like image recognition, segmentation, reconstruction of missing information, and retrofitting new information from the images.

Data preprocessing. Data preprocessing is usually needed before training DL models for analyzing microscopic data. Preprocessing includes fixing constant aspect ratios, scaling, and normalizing values, reducing dimensions, sharpening white-black contrast, and filtering image noise. Fixing the aspect ratios ensures that the input images are square and cropped properly as presumed by the CNNs.[411] Normalization is essential to afford the same data structure for each image.[412, 413] It includes rescaling, standardization, and stretching. Rescaling is to rescale the images to smaller ones. For example, an image with 256 × 256 pixels was rescaled to the one with 128 × 128 pixels.[414] Rescaling also increases training speed and inference. The CNNs converge faster with the aid of normalization. Dimension reduction is to collapse multiple channels of an RGB image into a single grayscale channel when the CNNs are dimensionally invariant.[415] White-black sharpening can avoid gradient vanishing by enhancing the features, while filtering helps to remove the noise.[416]

Case studies. Segmentation of microscopy images helps to analyze the objects or features of the images.[417] For example, the shapes and size distribution of nanoparticles would be outlined and calculated from the segmented SEM images.[409] There are two main categories of image segmentation, i.e., semantic segmentation and instance segmentation. The former assigns each part of an image a label after the image is partitioned into semantically meaningful parts, while the latter would exhaustively identify each instance of a class in the image. Dong et al. developed a multimodal multiclass segmentation model (DALM) for determining the number of layers in 2D materials, i.e., MoS_2 flakes (**Figure 10a**).[59] To increase the prediction accuracy, the RGB images

were merged with the hyperspectral images to train the DALM, which showed a higher prediction accuracy than those trained with only RGB images. It also exhibited satisfactory robustness even if the images showed high illumination and contrast variations. Besides the image segmentation, the DL models can also predict the reaction outcomes from the collected in-situ images. Sargent and coworkers developed a CNN model based on VGGNet for classifying the images taken from reactors into two categories: (1) bad crystals or no crystals, and (2) good crystals (**Figure 10b**).[100]

Investigation of adatom-adatom and adatom-substrate interactions is beneficial for understanding the physical and chemical reactivity of novel materials. STM and AFM can be used to visualize structures of the surface atoms, which makes the correlation between the structures and the surface properties easier. Kalinin et al. proposed an ML-based algorithm to seamlessly transform STM images to atomic coordinates of surface and adatoms (**Figure 10c**).[282] They used $Co_3Sn_2S_2$ as a model material to demonstrate a family member of Shandite $A_3M_2X_2$ crystal. They have a rhombohedral structure, which shows a CoSn Kagome lattice sandwiched by the S and Sn layers. A series of STEM images were analyzed using Laplacian of the Gaussian filter in the scikit-image library[418] to reconstruct the coordinates of the surface atoms. To match the experimental observations with the ones derived from a lattice Hamiltonian model, BO was further implemented to minimize the statistical errors in distances.

Though providing high-resolution imaging with enriched information, high-does electron beams of the electron microscopy may cause devastating damage to samples such as nano-catalysts[419] and biological samples.[420] Reducing the beam dose may mitigate this issue while scarifying the image quality. The CS technique has been an alternative strategy for collecting the images with reduced doses, acquisition time, and data volume. Browning and coworkers developed CS via the Bayesian dictionary learning as a low-dose acquisition method to obtain high-resolution

STEM images (**Figure 10d**).[421] With only 20 % of the pixels sampled from the real image of the ZSM-5 zeolite catalyst, the reconstructed images maintained high image quality. This approach not only automatically reduces the beam doses and variances caused by noise, but also increases the acquisition rate. In his follow-up work, Browning proposed an efficient sparse sampling strategy that randomly samples only a few rows of the pixels as the electron beam moves along the scanning direction.[422] This approach accelerates the acquisition rate and lowers the electron beam by a factor of > 5 times. In 2018, Browning developed a deliberately sub-sampling method that showed a much higher acquisition rate of the STEM images than conventional low-dose methods do.[423] This method acquires STEM images of ZnSe at least an order of magnitude faster and reduces data storage and communication. When integrated with an adaptive sampling strategy, this method shows a significant increase in the rate, speed, and sensitivity of images.

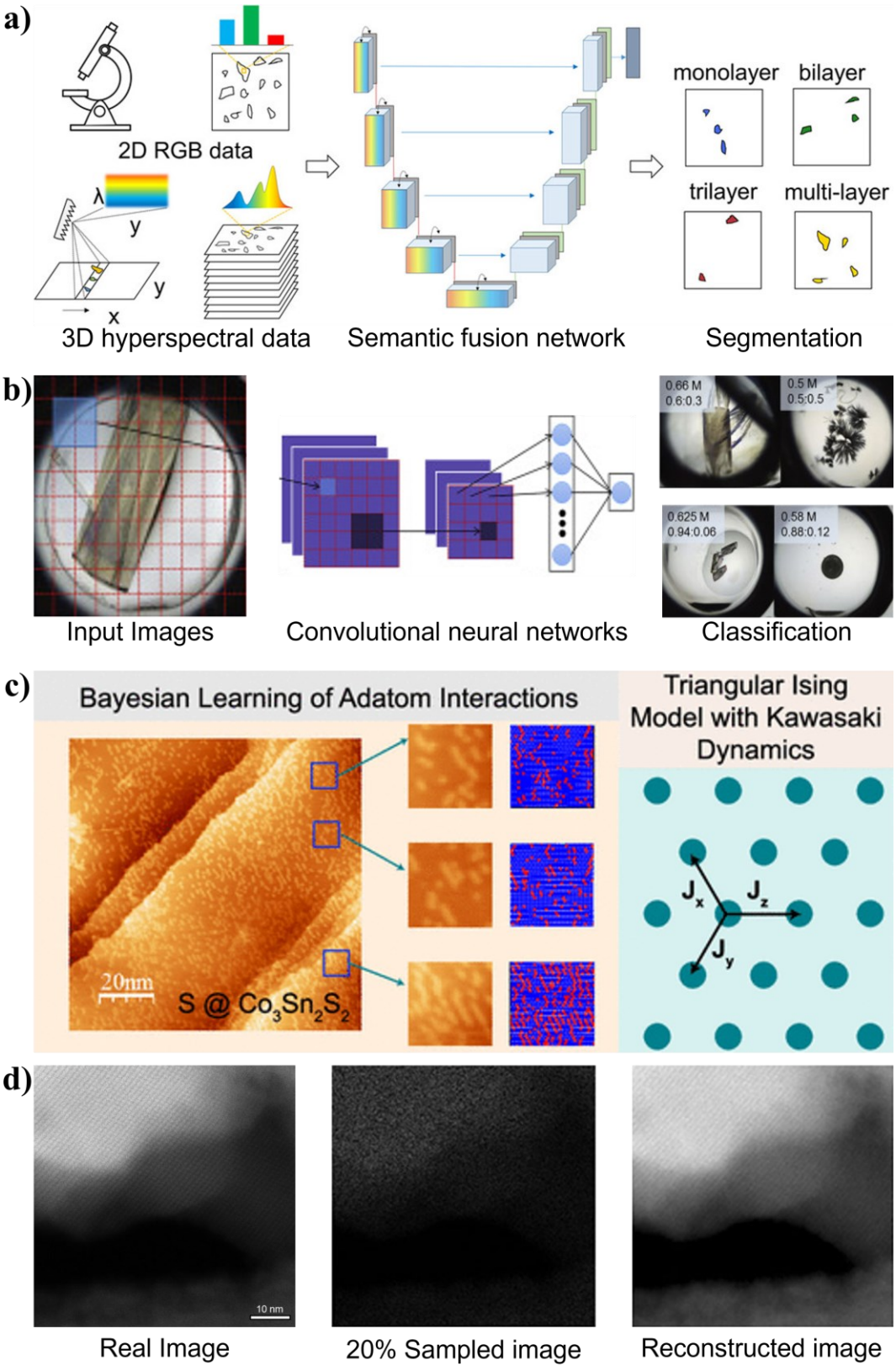

a)
2D RGB data
λ
y
y
x
3D hyperspectral data
Semantic fusion network
monolayer
bilayer
trilayer
multi-layer
Segmentation
b)
Input Images
Convolutional neural networks
0.66 M
0.6:0.3
0.5 M
0.5:0.5
0.625 M
0.94:0.06
0.58 M
0.88:0.12
Classification
c)
Bayesian Learning of Adatom Interactions
20nm
S @ Co3Sn2S2
Triangular Ising Model with Kawasaki Dynamics
Jx
Jz
Jy
d)
10 nm
Real Image
20% Sampled image
Reconstructed image

Figure 10. (a) Schematic showing a workflow of using DALM for mapping the atomic layer of 2D materials. Reproduced with permission from Ref.[59], Copyright 2020 American Chemical Society. (b) A workflow showing a CNN model for crystal formation prediction. Reproduced with permission from Ref.[100], Copyright 2020 Elsevier. (c) Bayesian learning of adatom interactions from STM images. Reproduced with permission from Ref.[282], Copyright 2021 American Chemical Society. (d) CS reconstructed a STEM image of the ZSM-5 zeolite catalyst. Real image (Left), 20 % sampled image (middle) and reconstructed image (right). Reproduced with permission from Ref.[421], Copyright 2014 Oxford University Press.

3.3 Decision-making Algorithms

Materials discovery can be considered as an optimization process in which input parameters must be tuned to reach a global optimal. In many applications of autonomous experiments, the objective functions are "black boxes", meaning that there are no specific functions that can define the objectives. Exhaustive or brute-force search is a general problem-solving method, while it is only suitable for problems with inexpensive and easily parallelized experiments within a relatively small chemical space. It is powerless when the chemical space is enormous.[424] Hence, intelligent decision-making algorithms are needed to efficiently explore the chemical space and save cost and time.[1, 95, 102] They can suggest optimum candidates based on the previous observations, thus avoiding redundant or biased evaluations. Also, they can help to maximize the yield of the product by adjusting a synthetic procedure or tuning the structure of a material to realize desired properties. Here, we focus on widely investigated decision-making algorithms, i.e., BO, RL, evolutionary algorithm (EA), Stable Noisy Optimization and Brach and FIT (SNOBFIT), and curiosity algorithm (CA), as summarized in **Table 5**.

Table 5. Examples of decision-making algorithms for autonomous experiments.

Algorithms		Model	Components		Task	Ref.
	# of objectives		Surrogate Model	Acquisition Function		
BO	SOO	Classical BO	GP	EI	Optimize the toughness of additively manufactured structures	[69]
			GP	UCB	Optimize the growth rates of carbon nanotubes	[425]
		Dragonfly	GP	EI/PI/ UCB/TS	Optimize the battery electrolyte	[73]
		Phoenics	BNN	ES	Optimize the pseudomobility of thin-film materials	[74]
					Improve the efficiency of quaternary organic photovoltaic blends	[165]
					Screen hydrogen evolution photocatalyst	[64]
		Gryffin	BNN	ES	Optimize stereoselective Suzuki-Miyaura coupling reactions	[167]
					Optimize the synthesis of o-xylenyl adducts of Buckminsterfullerene	[426]
					Design the redox-active materials for non-aqueous flow batteries	
	MOO	Chimera	BNN	ES	Optimize parameters for real-time reaction monitoring of High-Performance Liquid Chromatography (HPLC)	[427]
					Inverse design of efficient excitation energy transport (EET)	
		BO	NNE	UCB/EI/ EPLT	Optimize the synthesis of perovskite quantum dots	[68]
		TSEMO	GP	TS-EHVI	Optimize S_NAr and N-benzylation reactions	[428]
					Optimize Sogogashira reaction and multiple-step Claisen-Schmidt condensation reaction	[429]
					Screen the formulated products	[347]
		GP-qEHVI	GP	qEHVI	Optimize the electrical conductivity of metallic films	[169]
		GP-TS	GP	TS	Optimize the mechanical performance of polymers	[430]
RL		Model	Policy function		Task	
		DRO	RNN		Optimize Pomeranz-Fritsch, Friedländer, Ribose synthesis, and reaction between DCIP and AA	[188]
					Optimize the synthesis of silver nanoparticles	
		RL	CNN		Optimize the synthesis of MoS_2	[187]
		SNOBFIT-RL	CNN		Optimize the circular dichroism signal of perovskites	[71]
EA		Model			Task	
		GA/RF			Optimize the growth rates of CNTs	[431]
		GA/RF			Optimize the crystallinity of MOFs	[46]
		GA			Guide the synthesis of gold nanoparticles with different shapes	[432]
Others		SNOBFIT			Optimize the synthesis of EGFR kinase inhibitor AZD9291	[433]

		Optimize the yields of organic products	[341]
	CA	Explore the droplet behaviors	[75]

Note: BO: Bayesian optimization, SOO: single-objective optimization, MOO: multi-objective optimization, EI: expected improvement, PI: probability of improvement, EHVI: expected hypervolume improvement, TS: Thompson sampling, UCB: upper confidence bound, BNN: Bayesian neural network, EPLT: pure exploration, TSEMO: Thompson Sampling Efficient Multi-objective Optimization, RL: reinforcement learning, DRO: deep reaction optimizer, RNN: recurrent neural network, SNOBFIT: Stable Noisy Optimization by Branch and FIT.EA: evolutionary algorithm, RF: random forest, EA: evolutionary algorithm, GA: genetic algorithm, CA: curiosity algorithm.

3.3.1 Bayesian Optimization (BO)

BO is well suitable for solving *black-box* optimization problems because it has no prior assumption of any functional form (**Figure 11a**).[102, 173, 434] To implement this task, BO needs a surrogate model and an acquisition function. The surrogate model approximates the expensive objective function, while the acquisition function calculates a criterion that indicates how desirable it is to sample the next candidate. There are several choices of surrogate models such as probabilistic ones, e.g., Gaussian process (GP),[347, 430, 435-437] sparse pseudoinput GP (SPGP),[438, 439] and sparse spectrum GP (SSGP),[440] or non-probabilistic ones, e.g., Bayesian neural networks (BNNs)[441, 442] and RF.[216, 443] Choices of the acquisition functions include pure exploration (EPLT), expected improvement (EI),[444] probability of improvement (PI),[445] maximum variance (MV), upper confidence bound (UCB),[446] Thompson sampling (TS),[447] entropy search (ES),[448, 449] and knowledge gradient (KG).[450]

BO has the following advantages. First, Bo enables to search for the candidates actively and efficiently with an optimal property given a predefined task. It can greatly reduce the number of experiments to be evaluated. Second, BO is noise-tolerant since it can introduce noise when calculating the covariance. Third, BO can balance the trade-off between the exploitation of the best local optima and the exploration of high uncertainty to allow for the determination of the

global optima. These advantages make BO well suitable for AEPs.

Case studies. Gongora et al. proposed a Bayesian Experimental Autonomous Researcher (BEAR) to identify the optimal toughness of additively manufactured structures (**Figure 11b**).[69] BEAR utilizes GP as the surrogate model and EI to select the optimal design parameters for the next experiment. BEAR enables to identify high-performing structures within 100 experimental iterations. In comparison with the grid search strategy, BEAR reduces the number of experiments by 60-fold. Maruyama and his colleagues incorporated BO in an Autonomous REsearch System (ARES)[431, 451] to optimize the growth rate of carbon nanotubes (CNTs) (**Figure 11c**).[425] BO successfully improves the growth rate by a factor of 8 in comparison to that of a seed dataset within ~ 100 experiments. Dave and coworkers integrated Dragonfly, a BO software package, with a robotic platform to autonomously optimize battery electrolytes (**Figure 11d**).[73] Dragonfly implements an adaptive sampling strategy that actively learns which one of the four acquisition functions (EI, PI, UCB, and TS), performs the best during each optimization cycle. In only 40 hours, Dragonfly screens out a mixed anion sodium electrolyte that has a potential window of ~ 3.0 V. Takeuchi et al. developed a Closed-loop, Autonomous system for Materials Exploration and Optimization (CAMEO) to synthesize the Ge-Sb-Te alloy with a maximum bandgap (**Figure 11e**).[65] To minimize the experimental iterations, they first used the raw ellipsometry spectra of Fe-Ga-Pd to train the CAMEO model. CAMEO can make use of the phase distribution information learned from the Fe-Ga-Pd alloy to identify the optimal candidate for Ge-Sb-Te within 35 experimental iterations, which is superior to the classical Gaussian process-upper confidence bound (GP-UCB) algorithm.

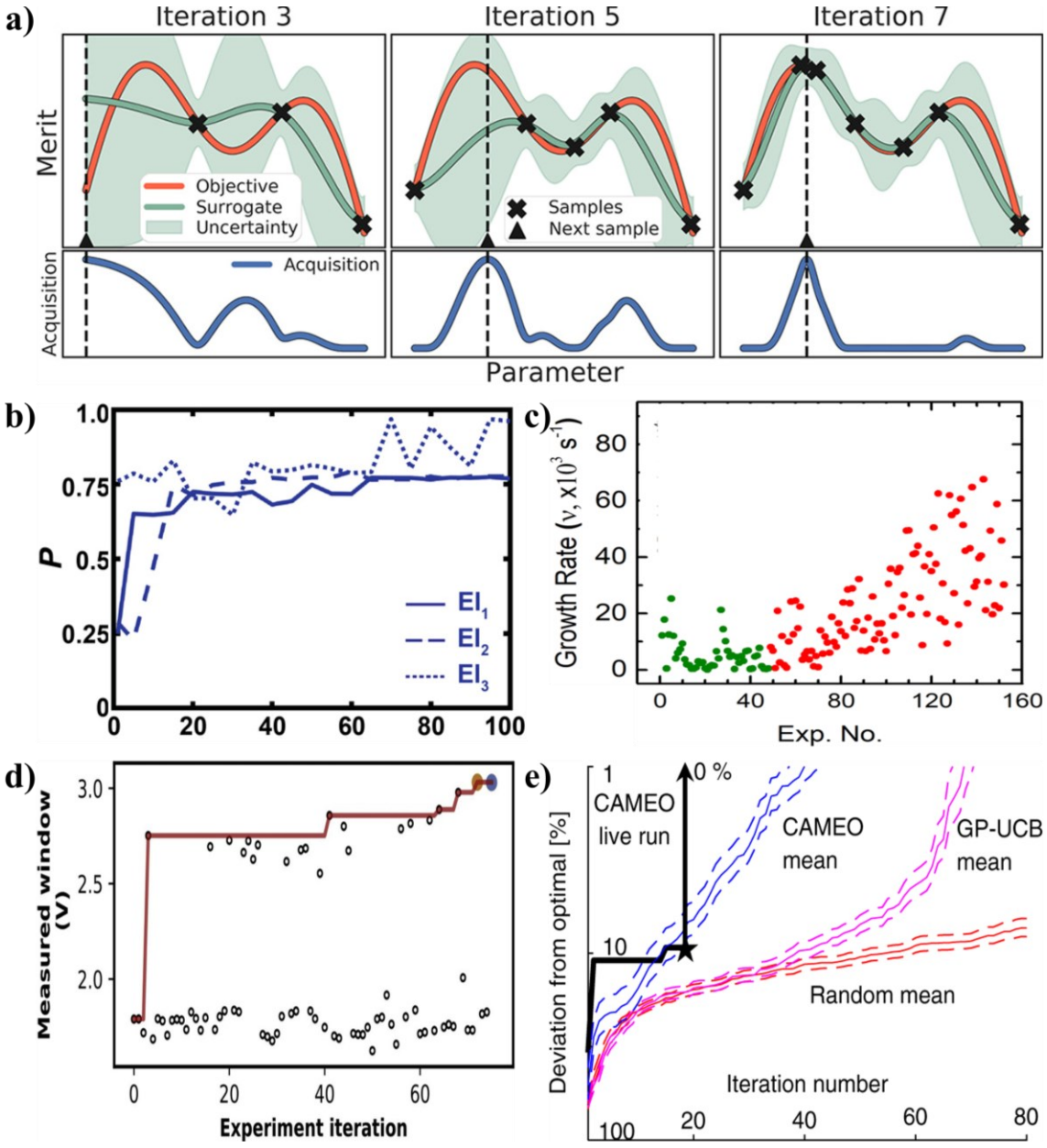


Figure 11. (a) Schematics of BO. Reproduced with permission from Ref.[102], Copyright 2021 American Chemical Society. (b) Evolution of the performance of the mechanical structures obtained from BEAR. Reproduced under a Creative Commons Attribution License 4.0 (CC BY) from Ref.[69], Copyright 2020 AAAS. (c) Evolution of the growth rates of CNTs obtained from BO. Reproduced under a Creative Commons Attribution 4.0 International License from Ref. [425], Copyright 2020 Springer Nature. (d) Evolution of the potential window of sodium ions electrolytes

obtained from Dragonfly. Reproduced with permission from Ref.[73], Copyright 2020 Elsevier. (e) Evolution of the phases of solid-state materials obtained from CAMEO. Reproduced with permission from Ref.[65], Copyright 2020 AAAS.

Due to their inherent sequential characteristics and heavy computational load, typical BO approaches can be costly for applications in AEPs. To tackle these problems, Aspuru-Guzik and coworkers developed a Probabilistic Harvard Optimizer Exploring Non-Intuitive Complex Surfaces (Phoenics) algorithm (**Figure 12a**).[452] Phoenics employs BNN to estimate the objective function, resulting in reduced training time. Phoenics formulates an inexpensive acquisition function that allows the batch evaluations to be run in parallel. Aspuru-Guzik et al. also developed Gryffin to optimize the problems that involve categorical inputs, which were relaxed into the continuous ones using categorical kernel density (**Figure 12b**).[453] Fruitful achievements have been realized via deploying Phoenics and Gryffin in ChemOS[454, 455] for performing autonomous experiments in optimizing the pseudomobility of thin-film materials (**Figure 12c**),[74] improving quaternary organic photovoltaic (OPV) blends,[165] screening optimal photocatalysts for hydrogen evolution (**Figure 12d**),[64] optimizing stereoselective Suzuki-Miyaura coupling reactions (**Figure 12e**),[167] optimizing the synthesis of o-xylenyl adducts of Buckminsterfullerene, and designing the redox active materials for non-aqueous flow batteries.[426]

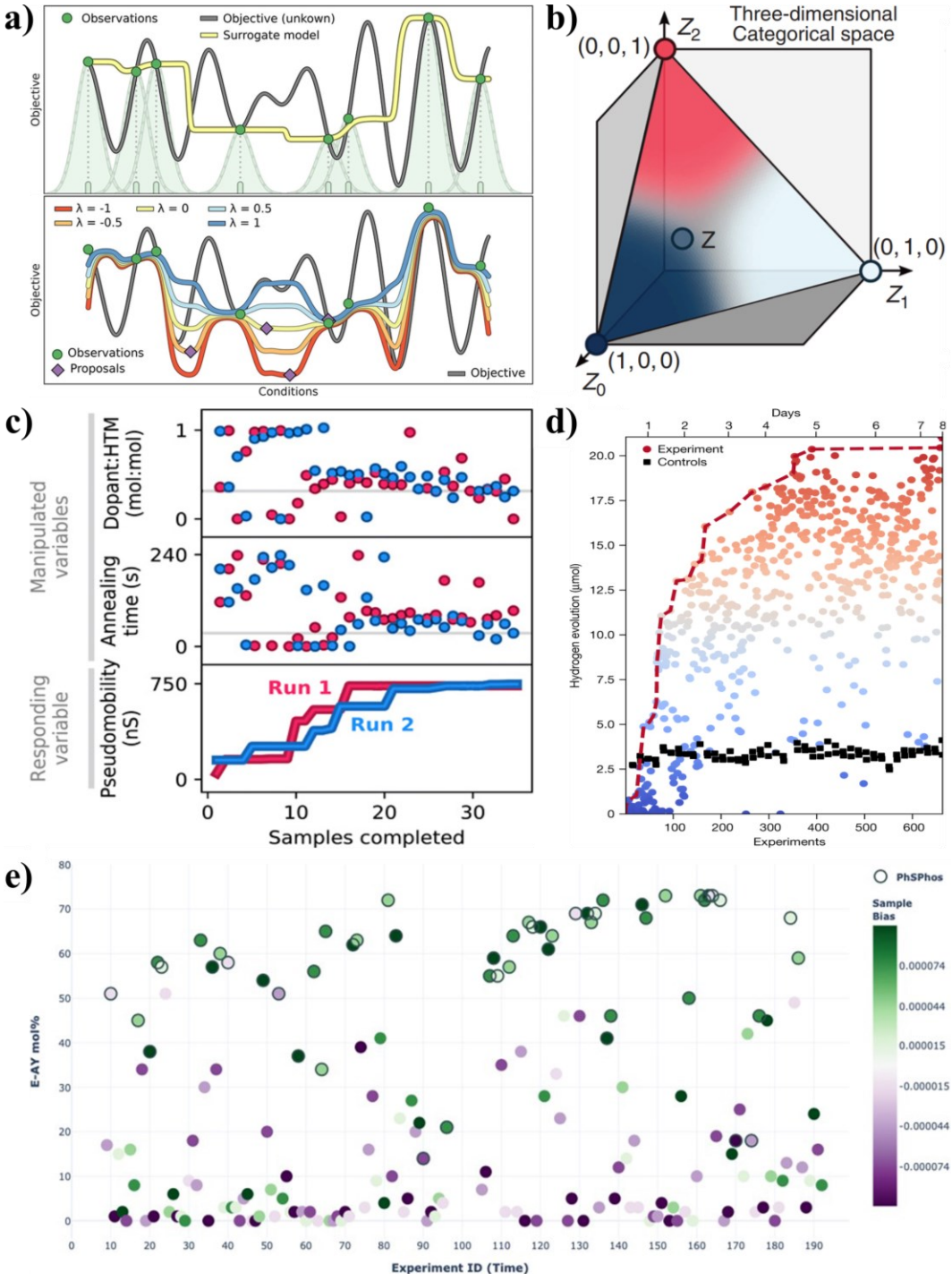


Figure 12. (a) Schematics of Phoenics for optimizing continuous parameters. Reproduced with permission from Ref.[452], Copyright 2018 American Chemical Society. (b) Schematics of Gryffin for optimizing categorical inputs. Reproduced with permission from Ref.[453], Copyright 2021 AIP Publishing. (c) Evolution of the pseudomobilty of thin-film materials obtained from Phoenics.

Reproduced under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC) from Ref.[74], Copyright 2020 AAAS. (d) Evolution of hydrogen evolution performance obtained from Phoenics. Reproduced with permission from Ref.[64], Copyright 2020 Springer Nature. (c) Process optimization of Suzuki-Miyaura coupling reactions by Phoenics and Gryffin. Reproduced with permission from Ref.[167], Copyright 2021 Springer Nature.

The abovementioned studies are mainly focused on single-objective optimization (SOO), e.g., yield, growth rate, and yield strength, while many cases involve multi-objective optimization (MOO), where optimizing one objective usually results in penalizing the others.[435] Unlike SOO, the solution of MOO is not a single point in the design space, but rather consists of a set of points, named the Pareto set. The optimal points derived from the corresponding objective function are named the Pareto front.[456] Two main approaches can be implemented to solve the MOO problems: transforming multiple objectives into one objective and identifying a Pareto front that trades off among the multiple objectives.

Aspuru-Guzik proposed Chimera as a generalized approach for MOO, where multi-objectives were converted to a single one using a concept of a *priori* scalarized with the lexicographic approaches (**Figure 13a**).[427] To avoid degradation of the objectives, Chimera strictly follows the predefined hierarchy. The hierarchy can construct a single objective function, which shapes a response surface that can be optimized by SOO algorithms. Chimera was successfully demonstrated in two different cases. The first case was to optimize parameters for realizing three objectives: maximizing the response of High-Performance Liquid Chromatography (HPLC), reducing the sampling volume, and minimizing the overall running time. The second case was about the inverse design of efficient excitation energy transport (EET) for realizing three objectives:

maximizing the transferring efficiency and total distance and minimizing the energy gradient. The results showed that Chimera achieved the goal by following the defined hierarchy. To simultaneously optimize the objectives of the improved photoluminescence quantum yield (PLQY), desired peak emission energy (E_P), and emission linewidth (E_{FWHM}) of halides produced by the Artificial Chemist, Abolhasani and coworkers developed an ensemble neural network (NNE)-based BO algorithm (**Figure 13b**).[68] NNE, the surrogate model, was trained to map the five input reaction conditions to the three objectives, which were further converted into a single quality metric using an objective function *Z* for the subsequent optimization. Then, combined with three different acquisition functions, i.e., UCB, EI, and EPLT, respectively, NNE was used to build three corresponding BO algorithms: NNE-UCB, NNE-EI, and NNE-EPLT. Among them, NNE-UCB showed the fast convergence of Z value as the increase of the experimental iterations. Although NNE-EPLT performed the worst, if it was pre-trained with the collected data, it also identified the optimal synthesis conditions.

Lapkin et al. proposed a Thompson Sampling Efficient Multi-objective Optimization (TSEMO) algorithm to simultaneously optimize the multiple objectives.[435] TSEMO builds an independent GP surrogate model for each objective and identifies a set of new evaluation points from the Pareto set with the maximum hypervolume at each iteration. TSEMO has advantages such as no requirement of prior knowledge, reduced hypervolume calculations, the capability of handling noise, and batch-sequential design, making it have performance comparable to Pareto Efficient Global Optimization (ParEGO), Expected Hypervolume Improvement (EHI), and Non-dominated Sorting Genetic Algorithm II (NSGA-II). Thus, it has been widely applied to optimize materials synthesis.[327, 347, 428, 429, 457-459] Lapkin and coworkers incorporated TSEMO with a continuous flow reactor (CFR) for the automated optimization of four exemplar chemical

reactions.[428, 429] Several conflicting objectives, i.e., maximizing spacetime yield (STY) vs minimizing E-factor, impurity, and ingredients, were simultaneously optimized. TSEMO fast converged to output the Pareto front within a minimized number of experiments. In his latest work, Lapkin et al. proposed a pipeline consisting of TSEMO and a Bayesian classifier in conjunction with a robotic experimental platform for screening the formulated products (**Figure 13b**).[347] They targeted four optimization objectives of high stability, low turbidity, honey-like viscosity, and low-cost precursors. A naïve Bayes algorithm was developed to classify the new evaluation points chosen by TSEMO based on the objective of stability while saving time and cost of the precursors. With the aid of two desktop robots, nine formulated recipes that are easily implemented and cost-effective were successfully screened out within 15 working days.

MacLeod et al. implemented *a posteriori* MOO algorithm to identify the Pareto front between the electrical conductivity of palladium films and their processing temperatures in a self-driving laboratory (**Figure 13c**).[169] This MOO algorithm used GP and *q*-Expected Hypervolume Improvement (*q*EHVI) as the surrogate model and the acquisition function, respectively. *q*EHVI can identify the Pareto front in a few experimental iterations.[460] In comparison with the other acquisition functions, qEHVI is superior in many ways such as parallelization, a constrained evaluation that excludes impossible or impractical data points, and efficient and effective optimization via auto-differentiation.[460] With the aid of qEHVI, the self-driving laboratory discovered new synthesis conditions that yielded uniform palladium films with moderate conductivity but processed at a lower temperature. Erps and his colleagues coupled a MOO algorithm with a 3D printer to optimize the mechanical performance of polymers that were produced from inks consisting of six primary photocurable monomers (**Figure 13d** and **13e**).[430] Using the data collected by the Thompson sampling strategy, the algorithm simultaneously

optimized three conflicting objectives of toughness, compression modulus, and strength. The MOO algorithm uncovered 12 optimal formulations after 30 experimental iterations, where the hypervolume indicator was increased by a factor of 1.65.

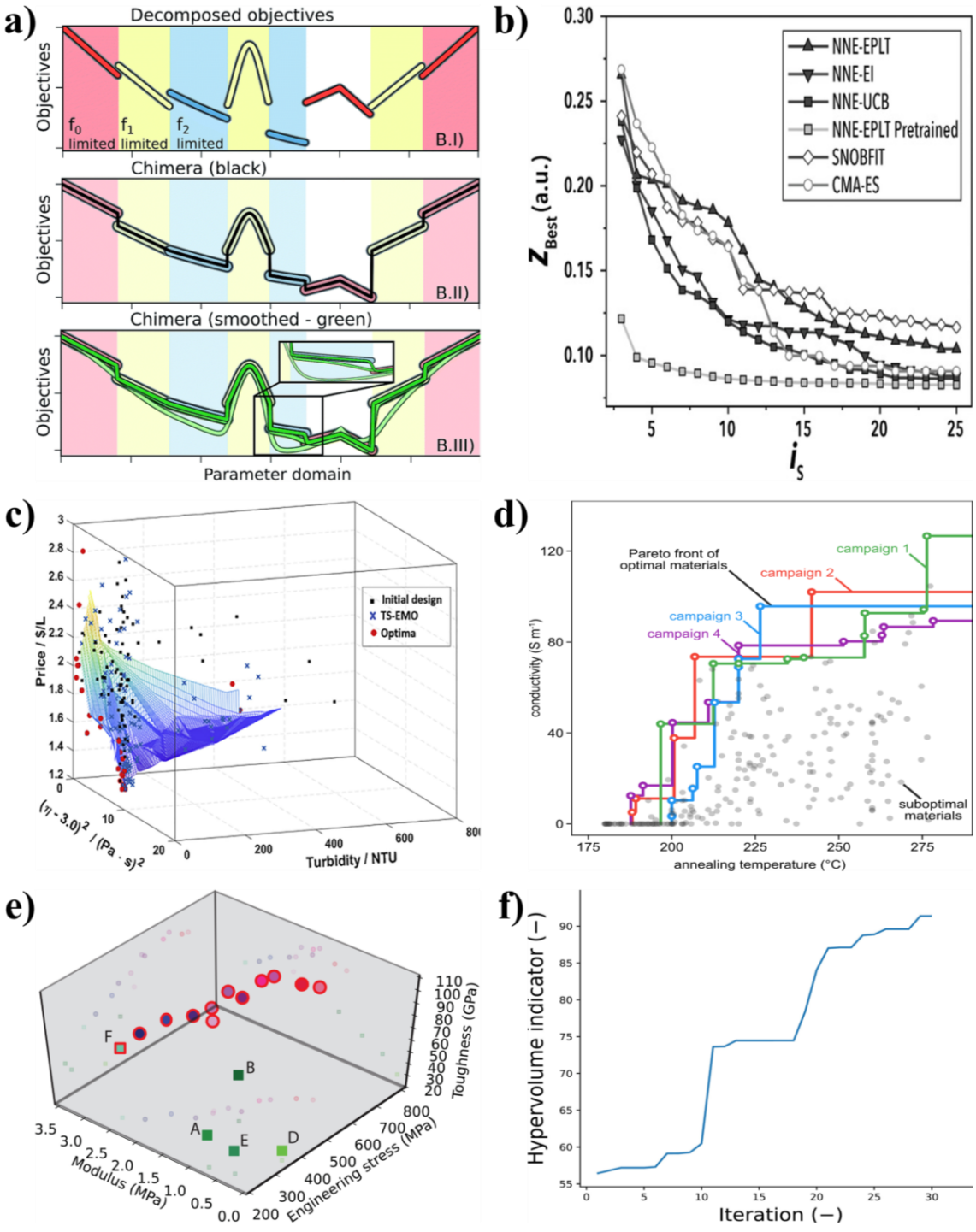


Figure 13. (a) Schematics of Chimera for MOO. Reproduced with permission from Ref.[427],

Copyright 2018 Royal Chemical Society. (b) The evolution of the Z value obtained from a few optimization algorithms. Reproduced with permission from Ref.[68], Copyright 2020 Wiley. (c) TSEMO for optimizing formulated products. Reproduced with permission from Ref.[347], Copyright 2022 Springer Nature. (d) qEHVI for optimizing the synthesis conditions of metallic films with quadruplicate. Reproduced with permission from Ref.[169], Copyright 2022 Springer Nature. (e) GP-TS for optimizing the mechanical performance of the 3D printed polymer. (f) A hypervolume evolution plot showing improvement of the Pareto front over iterations. Reproduced with permission from Ref.[430], Copyright 2021 AAAS.

3.3.2 Reinforcement Learning (RL)

RL is widely used to solve dynamic decision problems. It makes the sequential actions possible in a prescribed environment and estimates the statistical relationship between the actions and their possible outcomes to maximize the cumulative reward.[182] Mathematically, RL uses a Markov decision process defined by a set of states (S), a set of actions (A), a probability of transition from the state (s) to (s') under action a, $P_a(s, s')$, and a reward function (R). In the case of reaction optimization, S is the set of all possible experimental conditions, A is a set of all possible changes made to the experimental conditions, and r is the desired reaction outcome. Reward function (R) is applied to map a certain experimental condition (s) to a reaction outcome (r). Also, P_a defines the probability of transition of the experimental conditions by applying change a, given the inaccuracy in operating equipment. The purpose of RL is to learn an optimal policy that maximizes the reward function. The policy function maps the current and previous experimental conditions to the next ones. Due to the new advances in the DL algorithms and the availability of big data, deep reinforcement learning (DRL) is developed for solving many materials-related

dynamic decision-making problems.

Case studies. Zhou et al. developed a Deep Reaction Optimizer (DRO) to optimize a series of chemical reactions (**Figure 14a**).[188] DRO uses an RNN as the policy function to deicide the next reaction (action) that would realize an improved yield (reward) for the chemical reaction (environment). DRO reduces the number of reaction steps by 71 % and finds the optimal conditions for four real microdroplet reactions within 30 min. DRO also shows superior performance to other *black-box* optimization algorithms in optimizing the synthesis of silver nanoparticles (**Figure 14b**). Moreover, it can learn hidden information from both similar and dissimilar reactions for understanding the microdroplet reaction mechanism. Rajak et al. implemented a RL algorithm to identify the optimal synthesis of MoS_2 via a simulated Neural Autoregressive Density Estimator-Chemical Vapor Deposition (NADE-CVD) platform (**Figure 14c**).[187] The RL agent learns the policy for designing the optimal synthesis conditions via a policy gradient algorithm as informed by the NADE-CVD simulation results. The proposed RL algorithm prefers to synthesize MoS_2 with more 2H phase than that of MoS_2 generated by the random search method (**Figure 14d**). Li et al. employed a SNOBFIT-based RL algorithm to optimize the circular dichroism (CD) signal from the $CsPbBr_3$ nanocrystals produced by a Materials Acceleration Operating System In Cloud (MAOSIC) platform (**Figure 14e**).[71] SNOBFIT is well suitable for screening and optimizing chemical reactions because it can search randomly in the global region while applying the gradient descent method in a local region.[461] RL maps a set of actions and receives a reward based on the differences between the experimental and the targeted outcomes. Their policy function directs the local optimization toward the optimal conditions while searching for unexplored regions to obtain the global optima. Significant improvement in the CD signal was successfully achieved within 250 experimental iterations.

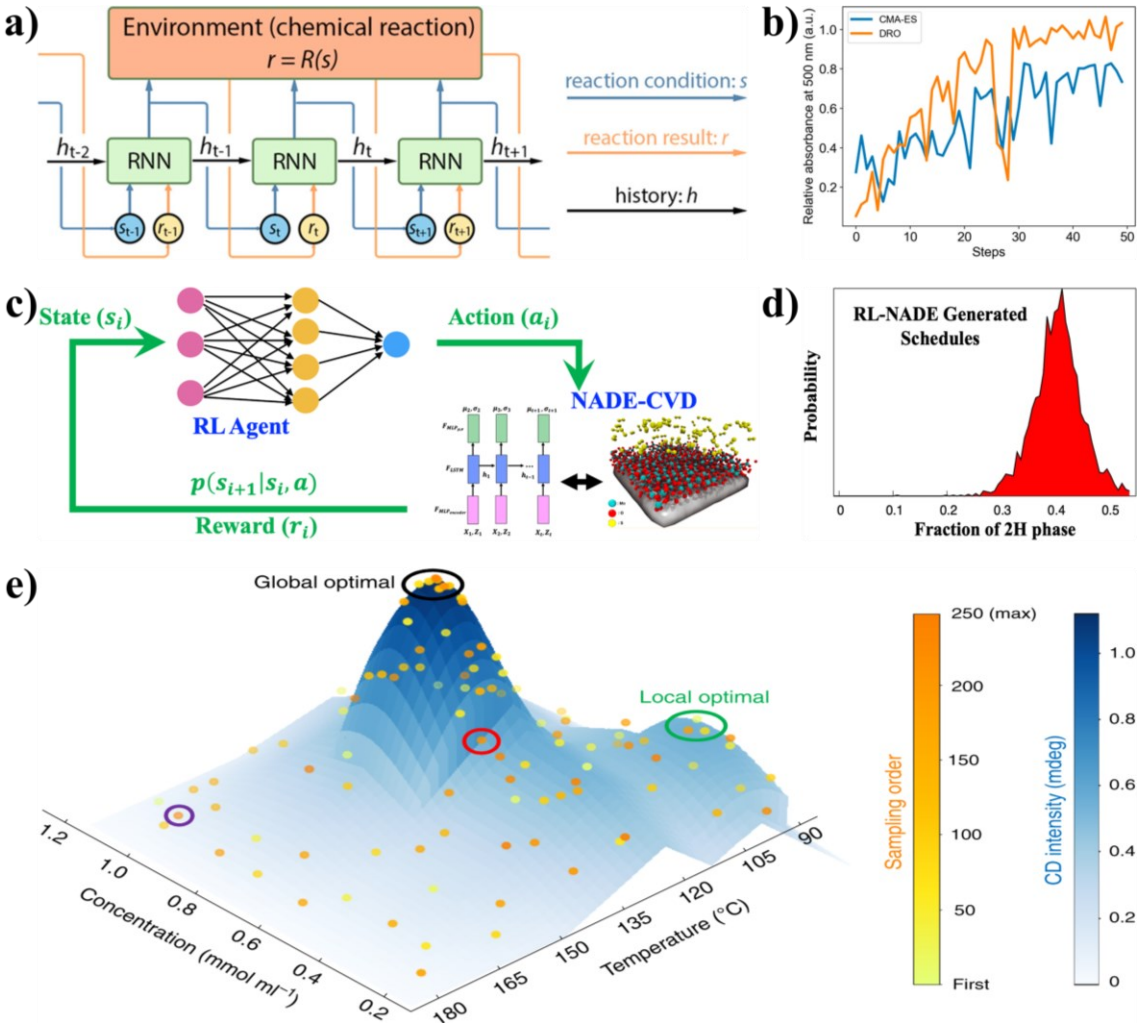


Figure 14. (a) Schematics of the deep reaction optimizer (DRO) that consists of an RNN architecture for optimizing chemical reactions. (b) Evolution of the relative absorbance of silver nanoparticles obtained from DRO. Reproduced with permission from Ref.[188], Copyright 2017 American Chemical Society. (c) Schematics of the RL-NADE algorithm for predicting optimal synthesis of MoS_2. (d) The synthesis schedules of MoS_2 obtained from the RL-NADE model. Reproduced with permission from Ref.[187], Copyright 2021 Springer Nature. (e) Evolution of circular dichroism (CD) signal obtained from SNOBFIT-RL. Reproduced with permission from Ref.[71], Copyright 2020 Springer Nature.

3.3.3 Evolutionary Algorithm (EA)

An evolutionary algorithm (EA), a population-based metaheuristic optimization algorithm, iteratively selects the optimal candidates with the highest-ranking scores of defined properties.[462, 463] EA does not make any assumptions about the nature of the fitness landscape, thus making it generic in solving optimization problems across many areas such as drug discovery, molecule design, and materials science. There are various EAs including genetic algorithm (GA),[464-466] particle swarm optimization (PSO),[467-469] ant colony optimization (ACO),[470, 471] and evolutionary programming (EP).[472-474]

Among these EAs, GA is the dominant one in materials science and has achieved enormous progress in exploring large chemical spaces for materials development (**Figure 15a**). GA first creates the initial population of individuals from chromosomes by a random process or by incorporating prior knowledge. Chromosomes are a set of genes represented by a string/sequence, and genes are the input variables represented in a binary format (0 or 1). Then, a fitness function ranks the fitness of individual candidates among a population. Then the top-ranked ones are selected as the parents for subsequent crossover and mutation operations to create a new generation. The crossover changes the subsequence between two parents at a random locus, while the mutation randomly flips some bits of individual parents based on the probability. The operation of GA terminates when either the properties of individuals exceed the threshold or the iteration cycles reach the set number of generations.[102, 463, 475] GA has shown applications in designing polymers with desired glass transition temperatures,[476] semiconducting polymers for OPVs,[477] polymer dielectrics,[478] and MOFs for carbon capture.[479] It is also applied in AEPs.[46, 431, 432]

Case studies. Maruyama and coworkers combined GA and RF as an AI experimental planner

to optimize the growth rates of CNTs (**Figure 15b**).[431] GA guided exploration of the search space, while RF recommended the reaction conditions that resulted in high growth rates. The growth rates of CNTs gradually converged after > 600 experiments were performed. To optimize the crystallinity of MOFs, Moosavi and his colleagues adopted the same strategy to explore the search space by a robotic platform (**Figure 15c**).[46] A gradual converge in the crystallinity was observed after three generations of 90 experiments. Salley et al. used GA to guide the synthesis of gold nanoparticles with different shapes in an autonomous robotic platform (**Figure 15d**).[432] GA recommended experimental parameters for the next generation after analyzing previous results of UV-Vis spectra. The fitness factor for three different shapes (spheres, rods, and octahedrons) finally converged to higher values with the evolution of the generations. In addition to the well-known sphere and rod shapes, GA also discovered a complex octahedron shape.

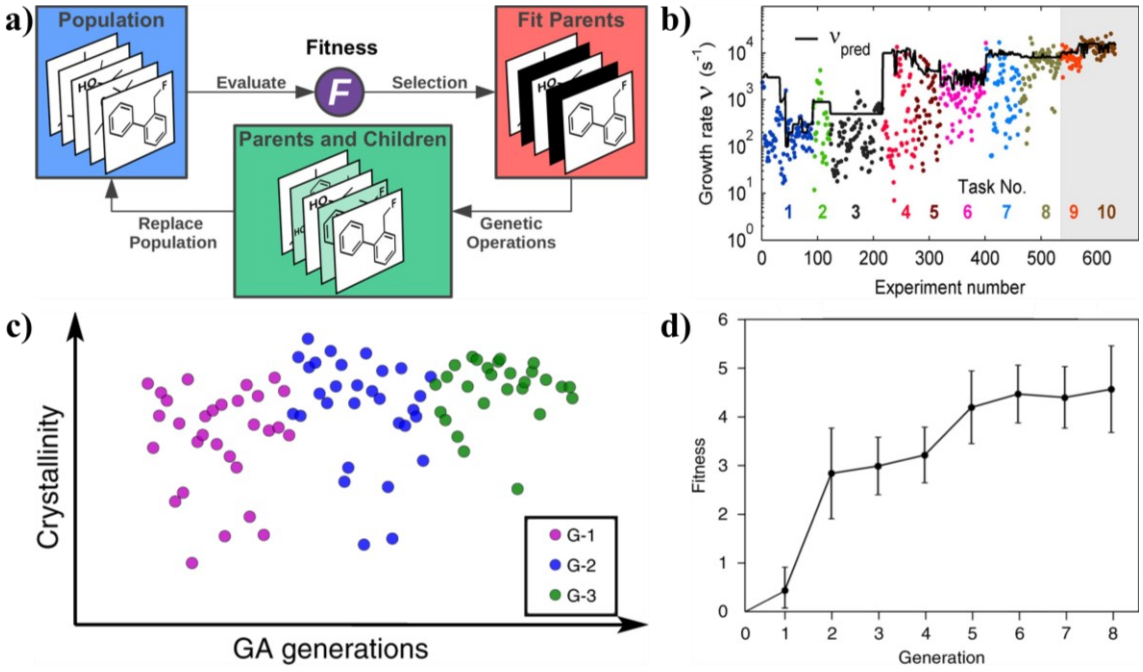


Figure 15. (a) Schematic of a typical GA. Reproduced with permission from Ref.[102], Copyright 2021 American Chemical Society. (b) Evolution of the growth rates of CNTs by GA/RF. Reproduced with permission from Ref.[431], Copyright 2016 Springer Nature. (c) Evolution of

the crystallinity of MOFs by GA/RF. Reproduced with permission from Ref.[46], Copyright 2019 Springer Nature. (d) Evolution of the fitness function of gold nanoparticles with octahedron shape by GA. Reproduced with permission from Ref.[432], Copyright 2020 Springer Nature.

3.3.4 Other Decision-Making Algorithms

SNOBFIT. SNOBFIT, a global optimization algorithm, includes a constraint and a fit function that fits polynomials to the obtained experimental data for identifying the multiple optima.[480] SNOBFIT has a higher chance of finding the global optima than that of finding the local optima since it generates a set of experimental variables widely distributed across the search space. Moreover, SNOBFIT can avoid false optimization directions due to its ability to take the experimental noise into account. Hence, SNOBFIT has achieved much progress in optimizing chemical reactions. Bourne et al. incorporated SNOBFIT with CFR to automatically optimize the synthesis of EGFR kinase inhibitor AZD9291 (**Figure 16a**).[433] SNOBFIT successfully synthesized AZD9291 at a yield of 89% within 42 experimental iterations. Jensen and coworkers applied SNOBFIT to optimize the yield of organic products from 3-5 manipulated variables (temperatures, flowrates, and catalyst mass) in a reconfigurable CFR (**Figure 16b**).[341] The yield successfully converged to the optimal values within 30-45 experimental iterations.

Curiosity Algorithm (CA). CA, the simplest random goal exploration algorithm,[481] can actively and autonomously choose the candidates that maximize the number of new and reproducible observations (**Figure 16c**). Rather than optimizing the target properties chosen by the user with prior knowledge, CA focuses on exploration with goals randomly chosen from observation space. CA starts with a random observation with defined experimental parameters. Then it uses the data collected from the previous experiments to build a regression model, which

recommends the parameters from the search space to reach the target. Grizou and coworkers applied CA in a robotic platform to autonomously explore droplet behaviors.[75] The search space consisted of the mixture ratios of four oils, while the observation space was defined as the speed of droplets and the number of divisions. Under the same conditions, CA explored 3.3 times more search space and identified more extreme cases of the droplet behaviors—showing obvious response to a slight change in temperature—than the random search method did (**Figure 16b**).

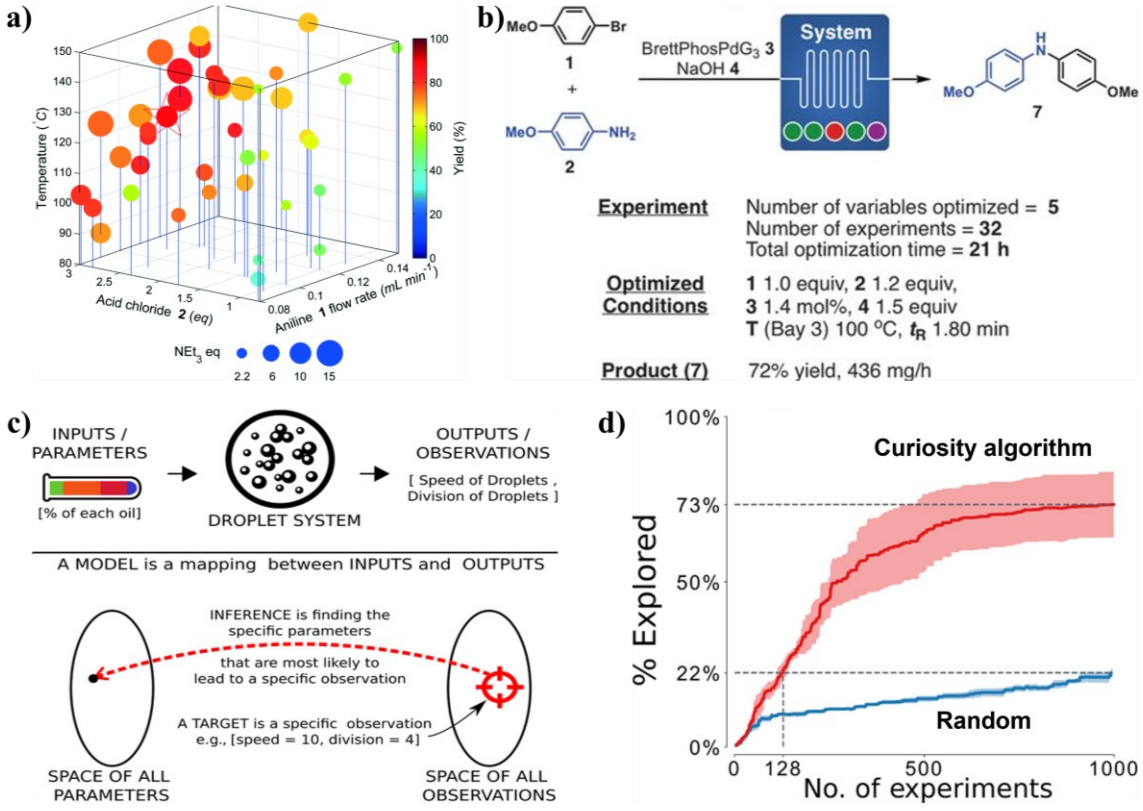


Figure 16. (a) SNOBFIT for optimizing AZD9291. Reproduced with permission from Ref.[433], Copyright 2016 Royal Society of Chemistry. (b) SNOBFIT for optimizing the reaction conditions of Buchwald-Hartwig amination. Reproduced with permission from Ref.[341], Copyright 2018 AAAS. (c) Schematic of a CA algorithm. (d) Comparison of the explored space over CA against the random search. Reproduced under a Creative Commons Attribution License 4.0 (CC BY) from Ref.[75], Copyright 2020 AAAS.

3.3.5 New Advances in Decision-Making Algorithms

These state-of-the-art decision-making algorithms have shown promising applications in AEPs. However, due to the inherently agnostic treatment of the objectives, they tend to become intractable when (a) the high-dimensional search space leads to too many iterations to reach the optimal properties; (b) the trivial or sluggish progress persists after many iterations or even no convergence occurs at end of the experiment; (c) further analysis is required to illustrate the experimental trends. To tackle these challenges, several reach groups including Buonassisi,[482-484] Kalinin,[485, 486] Kusne,[65, 487-490] Ghiringhelli,[491-494] and Lipson[495] have made new advances in embedding prior knowledge and implementing feature selection in the algorithm development.

The first approach of embedding the prior knowledge into the optimization procedure presents several benefits compared to these algorithms that solely decide the next iteration experiment via mapping the reaction variables to the results. First, it can greatly increase the efficiency in optimizing the explored search space. Second, it can enhance the researchers' confidence in identification of the global optimum. Third, it helps researchers understand the mechanisms. The prior knowledge can be obtained from several sources such as explicit physicochemical equations, theoretical simulations, archived experimental results, and expert intuitions and insights and then be applied to the decision-making algorithms.[496] Buonassisi and coworkers designed a two-step BO framework that embeds domain knowledge to optimize the growth conditions of photovoltaics (**Figure 17a-b**).[482] The first Bayesian interference network maps the process conditions (e.g. growth temperature) to the materials descriptors, while the second neural network as a surrogate model links the materials descriptors with the device

performance parameters (**Figure 17a**). This BO framework improves the efficiency of the solar cells by 6.5% compared to a gird search method (**Figure 17b**). Kusne and his colleagues designed a physics-informed Bayesian AL framework in Autonomous Neutron Diffraction Explorer (ANDiE) to autonomously control neuron diffraction experiments with a goal of reducing the operation time[487] Three physics based models, the Weiss equation, the first-order model, and the Ising model, were incorporated into a Markov Chain Monte Carlo (MCMC) framework to capture uncertainty and restrict the analysis results. ANDiE reduces the number of the neuron diffraction measurements by a factor of ~ 5 for identifying T_N of both MnO and $Fe_{1.09}Te$. Ziatdinov and Kalinin et al. designed structured GP (sGP) in an AL framework for exploring phase transitions of Sm-doped $BiFeO_3$ using piezoresponse force microscopy (**Figure 17c**).[497] They included both the statistical descriptors such as latent variables and the physics-informed descriptors like conductivity and polarization of the materials in the framework, which showed improved performance in determining the models of the hysteresis loop behavior. Buonassisi et al. incorporated a probabilistic constraint that was derived from the experimental results into the acquisition function of BO (**Figure 17d**).[484] Such an experimental constraint excluded the perovskites whose compositions are susceptible to phase segregation. In another work, Buonassisi et al. reported a data fusion approach that integrated high-throughput degradation tests with theoretical simulations for identifying the most stable perovskites (**Figure 17e**).[483] Aspuru-Guzik updated Gryffin by including the known experimental and design constraints.[426] These constraints were formulated in the acquisition function by the gradient-based and the hill-climbing-based approaches or the GA algorithm. This updated Gryffin showed superior performance on two practical and simulated chemical reactions to those of the random search method and GA.

The second approach is to implement feature selection, which chooses the most informative

variables to reduce the dimensions of the search space. The CS technique has been one of the feature selection approaches. Ghiringhelli and coworkers proposed a CS-based Sure Independence Screening and Sparsifying Operators (SISSO) method to identify low-dimensional variables that predicted material properties.[493] The sure independence screening technique selectes a subset of variables, which is further reduced by a sparsifying operator. SISSO showed superior performance to those of Least Absolute Shrinkage and Selection Operator (LASSO),[498] orthogonal matching pursuit (OMP),[499] and EUREQA,[495] and was also benchmarked in predicting the ground-state enthalpies of theoretically octet binary materials and classifying metals and insulators.

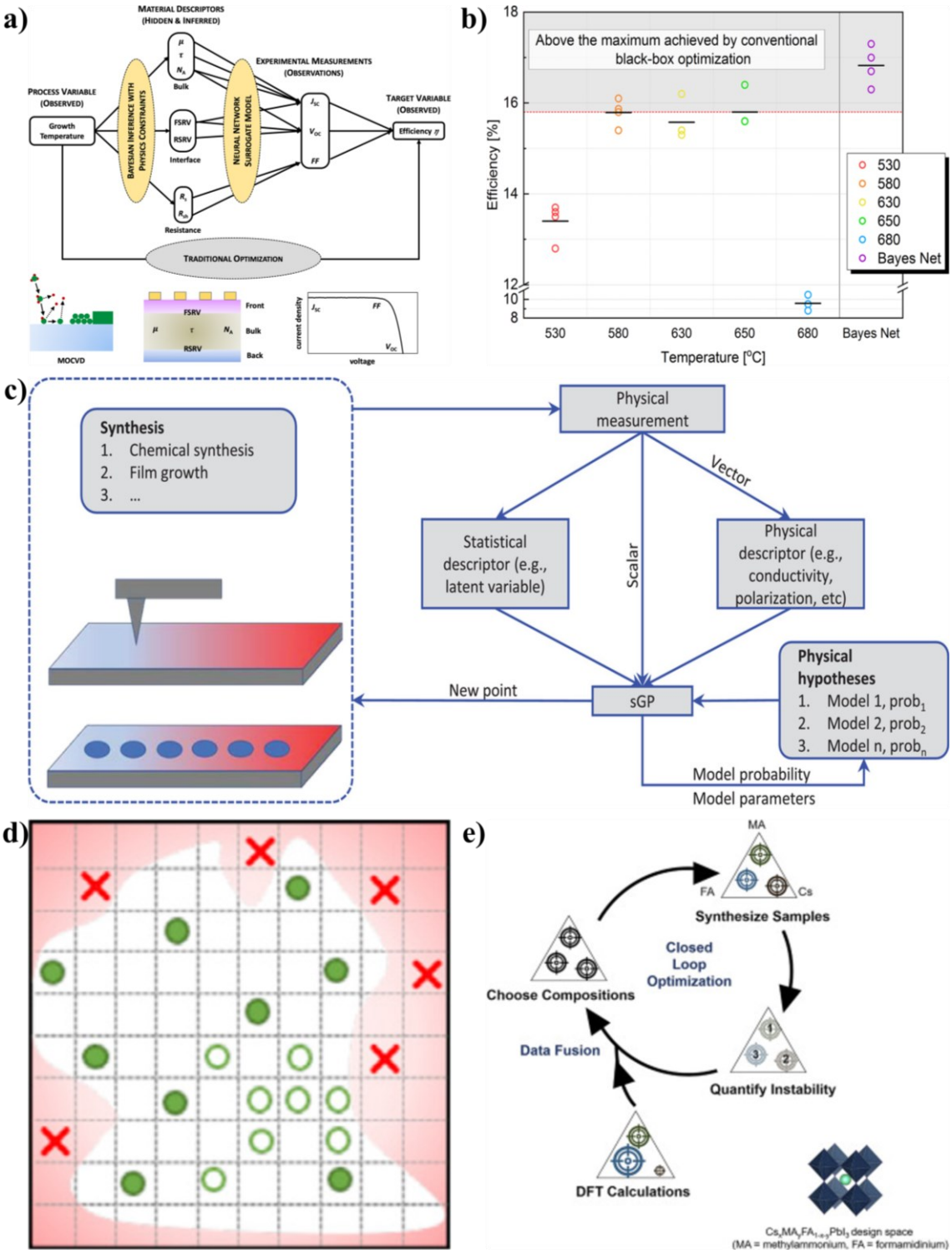


Figure 17. (a) Workflow of a two-step BO network. (b) Comparison of photovoltaic efficiency obtained from the BO net and the random search method. Reproduced with permission from Ref.[482], Copyright 2020 Springer Nature. (c) Schematic of exploring phase transition of

materials via incorporating physics-informed descriptors. Reproduced with permission from Ref.[497], Copyright 2022 Wiley. (d) Schematic of probabilistic constraints for the acquisition function of BO from experimental data. Reproduced with permission from Ref.[484], Copyright 2022 Elsevier. (e) Closed-loop optimization of halide perovskite stability with data fusion from DFT calculations. Reproduced with permission from Ref. [483], Copyright 2022 Elsevier.

4. Challenges and Future Directions

Despite much progress, the research in AEPs for materials development is still in its fancy stage. Much effort is required to solve some key challenges. Herein, we put forward some of them as well as potential solutions, from which the future trend is envisioned. As summarized in Figure 18, we envision a future self-driving laboratory that can perform de novo material design enabled by physics informed ML/DL models trained by computation and prior knowledge extracted from literature and open databases, and then synthesize and characterize target materials by the robots and the data-driven optimization algorithms under the guidance of researchers and be digitalized by a digital twin (DT) for scalable manufacturing.

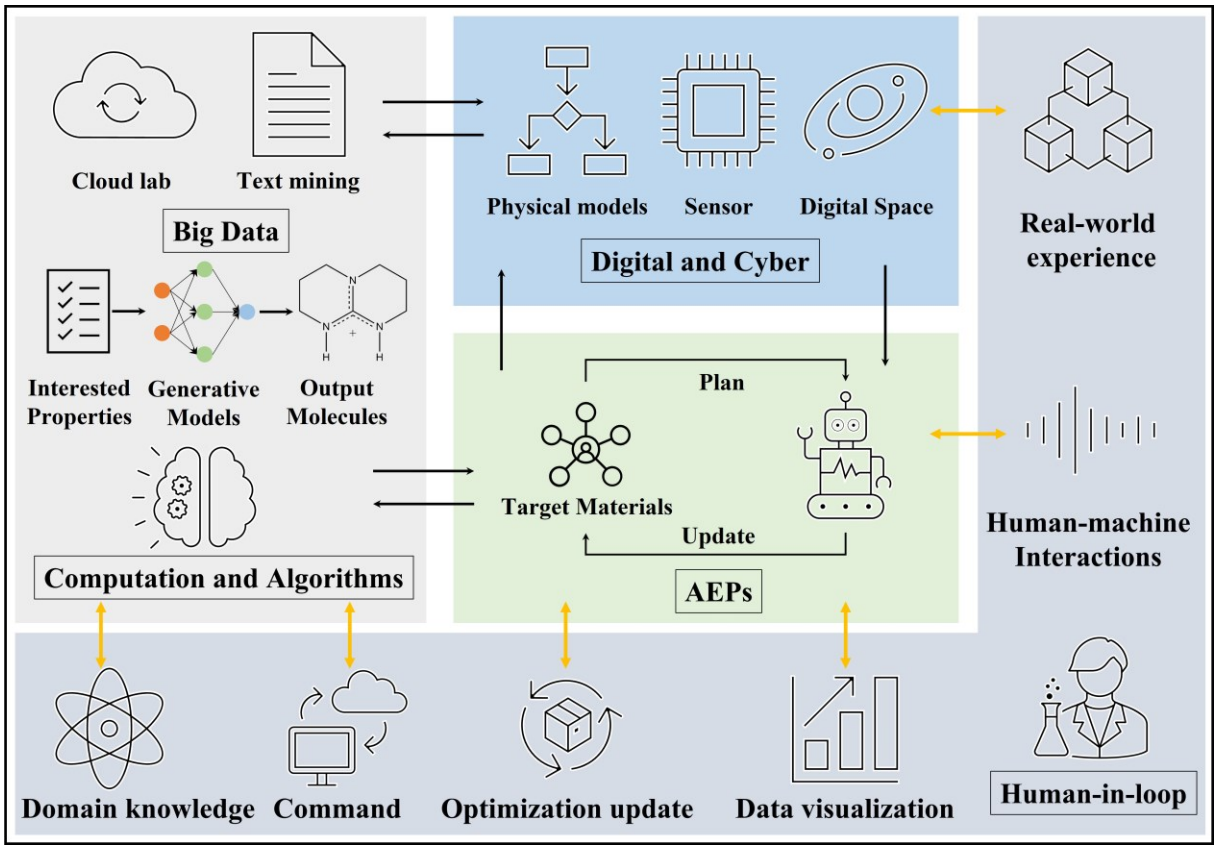


Figure 18. A future trend of AEPs enabled by big data, physics-based computations and AI algorithms, human-machine interactions, and digital/cyber manufacturing methodologies.

4.1 Data Standardization and Sharing

Big data and AI have been referred to as the basis of both the "fourth paradigm of science" and the "fourth industrial revolution". They have greatly passivated various domains at an astounding pace.[500, 501] Over the past two decades, we have seen the availability of several public databases containing millions of biological assay results, such as ChEMBL[502] and PubChem.[503] They have provided data for training the ML/DL models to predict a variety of biological activities or physical properties of molecules. Nevertheless, proper scientific data stewardship and management are much needed to make data searchable, accessible, interoperable, and reusable to the public.[504] Data should also follow the simple ALCOA (Attributable, Legible,

Contemporaneous, Original, and Accurate) principles by US FDA guidance. Such guidelines should be updated appropriately.[505] In the next decades, there will be a trend of standardizing the experimental and computational data for widespread sharing. Data standardization and sharing can also improve research reproducibility.[506, 507] For instance, recording lab details in an electronic lab notebook (ELN) is a promising way in materials science, chemistry, and biology[508] since it improves data acquisition, archiving, accessibility, sharing, and real-time data presentation. Currently, PerkinElmer E-Notebook, Evernote,[509] Microsoft OneNote,[510] and Google Docs[511] have been explored. With a predefined framework by researchers, text, images, audio and video can be fused for recording at the point of creation and then be used for future data mining. Moreover, the wide availability of the ELN in smartphones, tablets, and smartwatches makes data sharing easy among different stakeholders and through repository websites such as GitHub.

4.2 Text Mining for Knowledge Extraction

A major obstacle to implementing ML/DL for materials discovery is the lack of large publicly available and structured data. Although scientific data is available in literature, patents, and handbooks, manually mining them for extracting hidden information is very challenging.[512] Recent advances in natural language processing (NLP) enable the automatic mining of text, tables, and images from various data sources.[513-515] Several research groups led by Cole,[516-522] Olivetti,[523-526] Ceder,[123, 527-531] Schwaller,[120] and Jain[119, 532] have conducted NLP-driven automatic text mining and achieved great progress.

Integration of text mining in AEPs for guiding chemical synthesis starts to emerge, which may become a promising research field. A computer-aided synthesis planning (CASP) algorithm takes the target molecules as input and recommends chemically feasible reaction steps to

synthesize the molecules. Jensen's group proposed an automated system for knowledge-based continuous organic synthesis (ASKCOS) that extracts knowledge from millions of reactions stored in the U.S. Patent and Trademark Office (USPTO) or tabulated in Reaxys.[66] ASKCOS is an open-source software and can do retrosynthetic planning, reaction condition recommendation, and pathway evaluation. After that, experienced chemists can refine the recipes for the automatic synthesis platforms to perform the experiments. Cronin's group devoted their efforts to digitizing published reaction protocols with NLP, named as SynthReader. With the Chemical Description Language (XDL), they standardized description of the synthesis procedures for easy execution by the robots.[323, 331, 533, 534] We expect that the new transformers in NLP, e.g., BERT,[151] Transformer-XL,[535] XLNet,[536] RoBERTa,[537] and generative pretrained transformer (GPT)[538] would promote the development of a future AEP that can automatically mine the literature and execute experiments.

4.3 Incorporation of Inverse Materials Design with AEPs

Due to an enormous chemical space, even high-throughput screening is often powerless.[82, 539] Inverse design starts with target properties and proceeds toward desired structures, which is also called the de novo materials discovery. VAE,[540, 541] GAN,[182, 542, 543], and hybrid models[544] have been widely used for this purpose. Combined with RL[182, 545] and BO,[97] these generative models can generate candidates with target properties. For instance, Yao et al. demonstrated a VAE-based autonomous materials discovery platform for inverse design of reticular materials.[546]

Both BO and GA algorithms show poor scalability when the search space exceeds the limit.[547] To tackle this problem, Monte Carlo tree search (MCTS), a powerful global

optimization method, has found widespread applications in materials science such as screening Si-Ge alloy with high thermal conductivity, planning the synthesis of organic molecules, predicting the partition coefficient of organic molecules.[31, 51, 548, 549] For instance, Patra and coworkers developed an inverse design framework by combining MCTS with MD simulations as to identify sequence-specific copolymers that lead to interfacial energy between two immiscible homopolymers. Though the search space varies from 2^{10} (1024) to 2^{30} (~ 1 billion), the MCTS-MD framework showed excellent performance in identifying target sequences within a few hundred evaluations.[550] We expect that incoperation of the decision-making algorithms with the generative algorithms in an AEP can greatly shrink the initial search space for the AEP to design, plan, execute, and analyze the hypothesized experiments.

4.4 Interpretability of ML/DL Models

Distinguished from chemical/physical simulations that relied on explicit formulas, the ML/DL models provide impressive prediction power by learning knowledge from data. However, the *black-box* nature of ML/DL models makes them difficult to be explained or interpreted, which may impose an obstacle for widely deploying AEPs.[551] Developing strategies to demystify the inner working mechanism of these ML/DL models has become a compelling research task. The interpretable ML/DL models have three major advantages including troubleshooting, novel insights, and trust.[552] First, the interpretability improves the understanding of the prediction mechanism of the ML/DL models. Second, it can help the researchers quickly identify the errors or biases happening in the training process. Third, the interpretability improves the trustworthiness of the ML/DL models.

Interpretation of the ML/DL models can be empowered by the intrinsic characteristics of the

models or performed by the post-hoc interpretability techniques.[552-554] The intrinsic or model-based interpretability is inherited in the structures of ML/DL models. There are two ways of building intrinsically interpretable models.[555] The first is to add interpretability constraints by enforcing sparsity[556] and imposing semantic monotonicity.[557] The second is to use interpretable models such as a decision tree, rule-based model, or a linear model.[558, 559]

The post-hoc interpretability refers to illuminating the parameters or representation in an intuitive way that can be understood by researchers, which can be realized by three main strategies. The first strategy is to permutate feature importance for doing model-agnostic explanation, which calculates how the accuracy varies as a permutation of the values of a specific feature.[48, 560] The second strategy is to calculate the accuracy gain or feature coverage in tree-based ensemble models such as RF and XGBoost. The accuracy gain removes a new split to a branch of a feature, resulting in poor predictive accuracy. The feature coverage calculates the relative quantity of observations related to a feature. The third strategy is to visualize the intermediate or last layers of CNN models, which helps researchers understand the representation captured by the neurons.[56, 58] In the future, new advances in data visualization techniques or models/architectures, e.g., the physics-informed ML/DL models, will push the research in improving the model interpretability to a higher level.

4.5 Human-in-Loop AEPs

Currently, the involvement of researchers in AEPs is desired while remaining a challenge. Communication between researchers and machines is crucial to complementing the capabilities of data-driven algorithms with human expertise to realize a human-in-loop AEP. They usually lack generalizability when applied to real-world problems even after being trained by large-scaled data.

In contrast, human researchers have intuition and are better at learning unexpected events and knowledge from small data than a machine. Thus, involving human intelligence in the loop of AEPs can maximize the chance of obtaining the global optimum via intuitively understanding the most promising regions of the design space. Such a human-in-loop AEP is particularly desired due to the higher interpretability, better detection of failures/errors, and easier bug-fixing, improved generalizability and robustness of AEPs in practical applications.

The involvement of humans in AEPs can be done through data visualization, real-time updating optimization algorithms, and executing new commands remotely. The data visualization reduces the dimensions of the data for better visualization by humans. There are several data visualization techniques such as principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), uniform manifold approximation and projection (UMAP), and Isomap. They transform data into visual contexts such as graphs and maps, thus helping researchers to visualize the search space topology and partial dependencies of performance over the reaction parameters. Researchers can better localize the regions that show a higher chance of finding promising performance or target properties. Researchers can infer the robots to explore promising regions. The data visualization can also leverage the trust in ML/DL models and correct the optimization direction in time. To interact with the algorithms and facilitate the communication between researchers and platforms, Aspuru-Guzik and his coworker developed a software package named ChemOS.[454, 455] As a key component, the communication module was realized using common social media platforms such as Twitter, Gmail, and Slack. According to the optimization trend, the human researchers can advise the machines to adjust initial conditions or change search domains so that the machines can achieve the global optimum with a reduced number of iterations.

4.6 Digitalization of AEPs

Virtual lab. Virtual reality (VR), a classical immersive technology, has achieved much progress over the past decade.[561] Previously, VR mainly focuses on video game entertainment *via* generating 3D hologram-like objects in artificial environments to allow players to interact with the virtual objects. Recently, VR is gaining increasing attention from materials scientists. Zhu *et al*. built a Materials Acceleration Operation System (MAOS) for realizing the "on-demand" synthesis of quantum dots.[562] The MAOS has a customed interface, UI-VR (user interface and virtual reality, an isomorphic reflection of the real lab) to interact with MAOS. Communicating with reality *via a* 5G network through the TCP protocol, the UI-VR interface allows the researchers to control a virtual robot in the lab.

Cloud Lab. To enable the collaboration of researchers across the world through remote control, a cloud lab has been proposed.[71, 335, 345, 563, 564] A cloud lab integrates robotic platforms, cloud servers, sensing devices, communication tools, and managing software. Ley and his coworkers demonstrated the ability to automate the optimization and synthesis of pharmaceutical agents through operations in could.[564] In their work, the servers were operated in Japan, while devices and chemicals were located in Cambridge, UK, and the operation commands were delivered from Los Angeles, USA. Such a Cloud lab not only demonstrates the possibility of remote operation of experiments but also avoids machine redundancy since the system can be rapidly modified for new experiments.

Digital Twin. Digital Twin (DT), initially introduced in 2003,[565] has been one of the most promising technologies for realizing smart manufacturing and Industry 4.0.[566] Through seamless data transmission between the physical and virtual world, DT allows researchers to monitor, understand, and optimize the functions of all involved physical entities.[567]

Combination of DT and AEPs would enable the integration of major physical components for evolving the properties in a tractable numerical framework. Such a combination has several advantages including (i) further minimizing optimization iterations, (ii) shortening the search path to global optimization, and (iii) better understanding the optimization mechanism.

5. Conclusions

AEPs are poised to develop new materials with target properties or to search for parameters that realize improved efficiency under constraints of budget and time. In this review, we systematically summarize the recent progress on AEPs toward autonomous laboratories. This review first describes the fundamentals, concepts, and workflow of AI algorithms, then, how these AI algorithms advance three essential components of an AEP. Each section starts with a brief introduction of background and summary of methodologies followed by non-exhaustive examples.

We also outline future research directions that may lead to scientific and technological breakthroughs in AEPs. They include advances in data sharing, text mining, explainable ML/DL models, de novo materials discovery, human-in-loop AEPs, and digitalization of AEPs. We expect that these new advances could push the research in AEPs to a new height and catalyze novel materials discovery at a record development pace. We believe that this review would meet the needs of both beginners in the field and experts who aim to pursue new research frontiers.

Notes

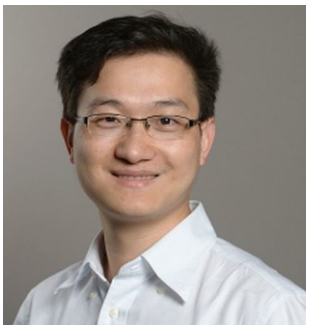
The authors declare no competing financial interests.

Biographies

Yunchao Xie is a postdoctoral researcher in the Department of Mechanical and Aerospace Engineering at the University of Missouri (MU). He earned his Ph.D. from MU under the supervision of Dr. Jian Lin in 2020. Currently, his research focuses on the integration of ML/DL with automated experimentation for accelerating the development of novel materials.

Kianoosh Sattari is a Ph.D. student in the Department of Mechanical and Aerospace Engineering at MU under the supervision of Dr. Jian Lin. He completed his master's degree in Mechanical Engineering from Saint Louis University in 2019. His research focuses on developing the physics-informed data-driven algorithm for materials discovery.

Chi Zhang is an Associate Professor at Hangzhou Institute for Advanced Study, UCAS. He has been a postdoc at the University of Missouri (MU) until 2021. In 2019, he completed his Ph.D. and Master degrees in Mechanical Engineering under the supervision of Dr. Jian Lin. His current research interests include laser materials processing and the development of data-driven models for process optimization.



Jian Lin is the William R. Kimel Associate Professor of Mechanical and Aerospace Engineering at MU. Before joining MU in 2014, he was a postdoctoral fellow with Professor James M. Tour at Rice University. He received his Ph.D. in Mechanical Engineering in 2011 from the University of California at Riverside. His current research interests include artificial intelligence and robotics for materials development and 3D/4D printing of smart polymeric materials.

Acknowledgement

This work has been financially supported by National Science Foundation (award numbers:

1825352, 1933861, and 2154428), U.S. Army Corps of Engineers, ERDC (grant number: W912HZ-21-2-0050), and DOE National Energy Technology Laboratory (award number: DE-FE0031988).

References

[1] Coley CW, Eyke NS, Jensen KF. Autonomous Discovery in the Chemical Sciences Part I: Progress. Angew Chem Int Ed 2020;59:22858-93.
[2] Coley CW, Eyke NS, Jensen KF. Autonomous Discovery in the Chemical Sciences Part II: Outlook. Angew Chem Int Ed 2020;59:23414-36.
[3] Tabor DP, Roch LM, Saikin SK, Kreisbeck C, Sheberla D, Montoya JH, et al. Accelerating the discovery of materials for clean energy in the era of smart automation. Nat Rev Mater 2018;3:5-20.
[4] Werber JR, Osuji CO, Elimelech M. Materials for next-generation desalination and water purification membranes. Nat Rev Mater 2016;1:16018.
[5] Correa-Baena J-P, Hippalgaonkar K, van Duren J, Jaffer S, Chandrasekhar VR, Stevanovic V, et al. Accelerating Materials Development via Automation, Machine Learning, and High-Performance Computing. Joule 2018;2:1410-20.
[6] Gromski PS, Henson AB, Granda JM, Cronin L. How to explore chemical space using algorithms and automation. Nat Rev Chem 2019;3:119-28.
[7] Potyrailo RA, Amis EJ. High-throughput analysis: a tool for combinatorial materials science: Springer Science & Business Media; 2012.
[8] Mittasch A, Frankenburg W. Early Studies of Multicomponent Catalysts. In: Frankenburg WG, Komarewsky VI, Rideal EK. Advances in Catalysis: Academic Press; 1950. p. 81-104.
[9] Torrance CJ, Agrawal V, Vogelstein B, Kinzler KW. Use of isogenic human cancer cells for high-throughput screening and drug discovery. Nat Biotechnol 2001;19:940-5.
[10] Chapman T. Lab automation and robotics: Automation on the move. Nature 2003;421:661-3.
[11] Perera D, Tucker JW, Brahmbhatt S, Helal CJ, Chong A, Farrell W, et al. A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. Science 2018;359:429-34.
[12] Broecker J, Morizumi T, Ou W-L, Klingel V, Kuo A, Kissick DJ, et al. High-throughput in situ X-ray screening of and data collection from protein crystals at room temperature and under cryogenic conditions. Nat Protoc 2018;13:260-92.
[13] Goud NR, Zhang X, Brédas J-L, Coropceanu V, Matzger AJ. Discovery of Non-linear Optical Materials by Function-Based Screening of Multi-component Solids. Chem 2018;4:150-61.
[14] Chow S, Liver S, Nelson A. Streamlining bioactive molecular discovery through integration and automation. Nat Rev Chem 2018;2:174-83.
[15] Laramy CR, Brown KA, O'Brien MN, Mirkin CA. High-Throughput, Algorithmic Determination of Nanoparticle Structure from Electron Microscopy Images. ACS Nano 2015;9:12488-95.
[16] Kelty ML, Morris W, Gallagher AT, Anderson JS, Brown KA, Mirkin CA, et al. High-throughput synthesis and characterization of nanocrystalline porphyrinic zirconium metal–organic

frameworks. Chem Commun 2016;52:7854-7.
[17] Li Z, Najeeb MA, Alves L, Sherman AZ, Shekar V, Cruz Parrilla P, et al. Robot-Accelerated Perovskite Investigation and Discovery. Chem Mater 2020;32:5650-63.
[18] Nega PW, Li Z, Ghosh V, Thapa J, Sun S, Hartono NTP, et al. Using automated serendipity to discover how trace water promotes and inhibits lead halide perovskite crystal formation. Appl Phys Lett 2021;119:041903.
[19] Lee XY, Saha SK, Sarkar S, Giera B. Automated detection of part quality during two-photon lithography via deep learning. Addit Manuf 2020;36:101444.
[20] Chu A, Nguyen D, Talathi SS, Wilson AC, Ye C, Smith WL, et al. Automated detection and sorting of microencapsulation via machine learning. Lab Chip 2019;19:1808-17.
[21] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
[22] Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 2117-25.
[23] Erhan D, Szegedy C, Toshev A, Anguelov D. Scalable object detection using deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2014. p. 2147-54.
[24] Szegedy C, Toshev A, Erhan D. Deep Neural Networks for Object Detection. In: Advances in Neural Information Processing Systems; 2013. p. 2553-61.
[25] Chen X, Ma H, Wan J, Li B, Xia T. Multi-view 3d object detection network for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 1907-15.
[26] Chen C, Seff A, Kornhauser A, Xiao J. Deepdriving: Learning affordance for direct perception in autonomous driving. In: Proceedings of the IEEE International Conference on Computer Vision; 2015. p. 2722-30.
[27] Hannun A, Case C, Casper J, Catanzaro B, Diamos G, Elsen E, et al. Deep Speech: Scaling up end-to-end speech recognition. 2014. p. arXiv:1412.5567.
[28] Amodei D, Ananthanarayanan S, Anubhai R, Bai J, Battenberg E, Case C, et al. Deep speech 2: end-to-end speech recognition in English and mandarin. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning, New York, NY, USA; 2016. p. 173–82.
[29] Wang X, Wang Y. Improving Content-based and Hybrid Music Recommendation using Deep Learning. In: Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, Florida, USA; 2014. p. 627–36.
[30] Elkahky AM, Song Y, He X. A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems. In: Proceedings of the 24th International Conference on World Wide Web, Florence, Italy; 2015. p. 278–88.
[31] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016;529:484-9.
[32] Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of Go without human knowledge. Nature 2017;550:354-9.
[33] Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science 2018;362:1140-4.
[34] Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature 2020;577:706-10.

[35] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583-9.
[36] Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. npj Digit Med 2021;4:65.
[37] Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science 2015;349:255-60.
[38] Ward L, Agrawal A, Choudhary A, Wolverton C. A general-purpose machine learning framework for predicting properties of inorganic materials. npj Comput Mater 2016;2:16028.
[39] Rosen AS, Iyer SM, Ray D, Yao Z, Aspuru-Guzik A, Gagliardi L, et al. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. Matter 2021;4:1578-97.
[40] Dong Y, Wu C, Zhang C, Liu Y, Cheng J, Lin J. Bandgap prediction by deep learning in configurationally hybridized graphene and boron nitride. npj Comput Mater 2019;5:26.
[41] Lu Z, Chen X, Liu X, Lin D, Wu Y, Zhang Y, et al. Interpretable machine-learning strategy for soft-magnetic property and thermal stability in Fe-based metallic glasses. npj Comput Mater 2020;6:187.
[42] Esterhuizen JA, Goldsmith BR, Linic S. Theory-Guided Machine Learning Finds Geometric Structure-Property Relationships for Chemisorption on Subsurface Alloys. Chem 2020;6:3100-17.
[43] Torrisi SB, Carbone MR, Rohr BA, Montoya JH, Ha Y, Yano J, et al. Random forest machine learning models for interpretable X-ray absorption near-edge structure spectrum-property relationships. npj Comput Mater 2020;6:109.
[44] Barnard AS, Opletal G. Predicting structure/property relationships in multi-dimensional nanoparticle data using t-distributed stochastic neighbour embedding and machine learning. Nanoscale 2019;11:23165-72.
[45] Raccuglia P, Elbert KC, Adler PDF, Falk C, Wenny MB, Mollo A, et al. Machine-learning-assisted materials discovery using failed experiments. Nature 2016;533:73-6.
[46] Moosavi SM, Chidambaram A, Talirz L, Haranczyk M, Stylianou KC, Smit B. Capturing chemical intuition in synthesis of metal-organic frameworks. Nat Commun 2019;10:539.
[47] Voznyy O, Levina L, Fan JZ, Askerka M, Jain A, Choi M-J, et al. Machine Learning Accelerates Discovery of Optimal Colloidal Quantum Dot Synthesis. ACS Nano 2019;13:11122-8.
[48] Xie Y, Zhang C, Hu X, Zhang C, Kelley SP, Atwood JL, et al. Machine Learning Assisted Synthesis of Metal–Organic Nanocapsules. J Am Chem Soc 2020;142:1475-81.
[49] Xu M, Tang B, Lu Y, Zhu C, Lu Q, Zhu C, et al. Machine Learning Driven Synthesis of Few-Layered WTe_2 with Geometrical Control. J Am Chem Soc 2021;143:18103-13.
[50] Beckham JL, Wyss KM, Xie Y, McHugh EA, Li JT, Advincula PA, et al. Machine Learning Guided Synthesis of Flash Graphene. Adv Mater 2022;34:2106506.
[51] Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. Nature 2018;555:604-10.
[52] Mikulak-Klucznik B, Gołębiowska P, Bayly AA, Popik O, Klucznik T, Szymkuć S, et al. Computational planning of the synthesis of complex natural products. Nature 2020;588:83-8.
[53] Badowski T, Gajewska EP, Molga K, Grzybowski BA. Synergy Between Expert and Machine-Learning Approaches Allows for Improved Retrosynthetic Planning. Angew Chem Int Ed 2020;59:725-30.
[54] Coley CW, Green WH, Jensen KF. Machine Learning in Computer-Aided Synthesis Planning.

Acc Chem Res 2018;51:1281-9.
[55] Finnigan W, Hepworth LJ, Flitsch SL, Turner NJ. RetroBioCat as a computer-aided synthesis planning tool for biocatalytic reactions and cascades. Nat Catal 2021;4:98-104.
[56] Oviedo F, Ren Z, Sun S, Settens C, Liu Z, Hartono NTP, et al. Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks. npj Comput Mater 2019;5:60.
[57] Qu X, Huang Y, Lu H, Qiu T, Guo D, Agback T, et al. Accelerated Nuclear Magnetic Resonance Spectroscopy with Deep Learning. Angew Chem Int Ed 2020;59:10297-300.
[58] Wang H, Xie Y, Li D, Deng H, Zhao Y, Xin M, et al. Rapid Identification of X-Ray Diffraction Patterns Based on Very Limited Data by Interpretable Convolutional Neural Networks. J Chem Inf Model 2020;60:2004-11.
[59] Dong X, Li H, Jiang Z, Grünleitner T, Güler İ, Dong J, et al. 3D Deep Learning Enables Accurate Layer Mapping of 2D Materials. ACS Nano 2021;15:3139-51.
[60] Kalinin SV, Ziatdinov M, Hinkle J, Jesse S, Ghosh A, Kelley KP, et al. Automated and Autonomous Experiments in Electron and Scanning Probe Microscopy. ACS Nano 2021;15:12604-27.
[61] Han B, Lin Y, Yang Y, Mao N, Li W, Wang H, et al. Deep-Learning-Enabled Fast Optical Identification and Characterization of 2D Materials. Adv Mater 2020;32:2000953.
[62] Kantz ED, Tiwari S, Watrous JD, Cheng S, Jain M. Deep Neural Networks for Classification of LC-MS Spectral Peaks. Anal Chem 2019;91:12407-13.
[63] Granda JM, Donina L, Dragone V, Long D-L, Cronin L. Controlling an organic synthesis robot with machine learning to search for new reactivity. Nature 2018;559:377-81.
[64] Burger B, Maffettone PM, Gusev VV, Aitchison CM, Bai Y, Wang X, et al. A mobile robotic chemist. Nature 2020;583:237-41.
[65] Kusne AG, Yu H, Wu C, Zhang H, Hattrick-Simpers J, DeCost B, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. Nat Commun 2020;11:5966.
[66] Coley CW, Thomas DA, Lummiss JAM, Jaworski JN, Breen CP, Schultz V, et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. Science 2019;365:eaax1566.
[67] Steiner S, Wolf J, Glatzel S, Andreou A, Granda JM, Keenan G, et al. Organic synthesis in a modular robotic system driven by a chemical programming language. Science 2019;363:eaav2211.
[68] Epps RW, Bowen MS, Volk AA, Abdel-Latif K, Han S, Reyes KG, et al. Artificial Chemist: An Autonomous Quantum Dot Synthesis Bot. Adv Mater 2020;32:2001626.
[69] Gongora AE, Xu B, Perry W, Okoye C, Riley P, Reyes KG, et al. A Bayesian experimental autonomous researcher for mechanical design. Sci Adv 2020;6:eaaz1708.
[70] Rizkin BA, Shkolnik AS, Ferraro NJ, Hartman RL. Combining automated microfluidic experimentation with machine learning for efficient polymerization design. Nat Mach Intell 2020;2:200-9.
[71] Li J, Li J, Liu R, Tu Y, Li Y, Cheng J, et al. Autonomous discovery of optically active chiral inorganic perovskite nanocrystals through an intelligent cloud lab. Nat Commun 2020;11:2046.
[72] Lee EC, Parrilla-Gutierrez JM, Henson A, Brechin EK, Cronin L. A Crystallization Robot for Generating True Random Numbers Based on Stochastic Chemical Processes. Matter 2020;2:649-57.
[73] Dave A, Mitchell J, Kandasamy K, Wang H, Burke S, Paria B, et al. Autonomous Discovery of Battery Electrolytes with Robotic Experimentation and Machine Learning. Cell Rep Phys Sci 2020;1:100264.

[74] MacLeod BP, Parlane FGL, Morrissey TD, Häse F, Roch LM, Dettelbach KE, et al. Self-driving laboratory for accelerated discovery of thin-film materials. Sci Adv 2020;6:eaaz8867.
[75] Grizou J, Points LJ, Sharma A, Cronin L. A curious formulation robot enables the discovery of a novel protocell behavior. Sci Adv 2020;6:eaay4237.
[76] Houben C, Lapkin AA. Automatic discovery and optimization of chemical processes. Curr Opin Chem Eng 2015;9:1-7.
[77] Henson AB, Gromski PS, Cronin L. Designing Algorithms To Aid Discovery by Chemical Robots. ACS Cent Sci 2018;4:793-804.
[78] Stein HS, Gregoire JM. Progress and prospects for accelerating materials science with automated and autonomous workflows. Chem Sci 2019;10:9640-9.
[79] Häse F, Roch LM, Aspuru-Guzik A. Next-generation experimentation with self-driving laboratories. Trends Chem 2019;1:282-91.
[80] Dimitrov T, Kreisbeck C, Becker JS, Aspuru-Guzik A, Saikin SK. Autonomous Molecular Design: Then and Now. ACS Appl Mater Interfaces 2019;11:24825-36.
[81] Gromski PS, Granda JM, Cronin L. Universal chemical synthesis and discovery with 'the chemputer'. Trends Chem 2020;2:4-12.
[82] Flores-Leonar MM, Mejía-Mendoza LM, Aguilar-Granda A, Sanchez-Lengeling B, Tribukait H, Amador-Bedolla C, et al. Materials Acceleration Platforms: On the way to autonomous experimentation. Curr Opin Green Sustain Chem 2020;25:100370.
[83] Montoya JH, Winther KT, Flores RA, Bligaard T, Hummelshøj JS, Aykol M. Autonomous intelligent agents for accelerated materials discovery. Chem Sci 2020;11:8517-32.
[84] Volk AA, Epps RW, Abolhasani M. Accelerated Development of Colloidal Nanomaterials Enabled by Modular Microfluidic Reactors: Toward Autonomous Robotic Experimentation. Adv Mater 2021;33:2004495.
[85] Wang Z, Zhao W, Hao G-F, Song B-A. Automated synthesis: current platforms and further needs. Drug Discov Today 2020;25:2006-11.
[86] Campbell ZS, Bateni F, Volk AA, Abdel-Latif K, Abolhasani M. Microfluidic Synthesis of Semiconductor Materials: Toward Accelerated Materials Development in Flow. Part Part Syst Char 2020;37:2000256.
[87] Cole JM. How the Shape of Chemical Data Can Enable Data-Driven Materials Discovery. Trends Chem 2021;3:111-9.
[88] Eyke NS, Koscher BA, Jensen KF. Toward Machine Learning-Enhanced High-Throughput Experimentation. Trends Chem 2021;3:120-32.
[89] Breen CP, Nambiar AMK, Jamison TF, Jensen KF. Ready, Set, Flow! Automated Continuous Synthesis and Optimization. Trends Chem 2021;3:373-86.
[90] Thakkar A, Johansson S, Jorner K, Buttar D, Reymond J-L, Engkvist O. Artificial intelligence and automation in computer aided synthesis planning. React Chem Eng 2021;6:27-51.
[91] Kimmig J, Zechel S, Schubert US. Digital Transformation in Materials Science: A Paradigm Change in Material's Development. Adv Mater 2021;33:2004940.
[92] Shi Y, Prieto PL, Zepel T, Grunert S, Hein JE. Automated Experimentation Powers Data Science in Chemistry. Acc Chem Res 2021;54:546-55.
[93] Upadhya R, Kosuri S, Tamasi M, Meyer TA, Atta S, Webb MA, et al. Automation and data-driven design of polymer therapeutics. Adv Drug Delivery Rev 2021;171:1-28.
[94] Cao L, Russo D, Lapkin AA. Automated robotic platforms in design and development of formulations. AlChE J 2021;67:e17248.
[95] Stach E, DeCost B, Kusne AG, Hattrick-Simpers J, Brown KA, Reyes KG, et al. Autonomous

experimentation systems for materials development: A community perspective. Matter 2021;4:2702-26.
[96] Hammer AJS, Leonov AI, Bell NL, Cronin L. Chemputation and the Standardization of Chemical Informatics. JACS Au 2021;1:1572-87.
[97] Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. ACS Cent Sci 2018;4:268-76.
[98] Melnikov AD, Tsentalovich YP, Yanshole VV. Deep Learning for the Precise Peak Detection in High-Resolution LC–MS Data. Anal Chem 2020;92:588-92.
[99] Fine JA, Rajasekar AA, Jethava KP, Chopra G. Spectral deep learning for prediction and prospective validation of functional groups. Chem Sci 2020;11:4618-30.
[100] Kirman J, Johnston A, Kuntz DA, Askerka M, Gao Y, Todorović P, et al. Machine-Learning-Accelerated Perovskite Crystallization. Matter 2020;2:938-47.
[101] Vlcek L, Maksov A, Pan M, Vasudevan RK, Kalinin SV. Knowledge Extraction from Atomically Resolved Images. ACS Nano 2017;11:10313-20.
[102] Pollice R, dos Passos Gomes G, Aldeghi M, Hickman RJ, Krenn M, Lavigne C, et al. Data-Driven Strategies for Accelerated Materials Design. Acc Chem Res 2021;54:849-60.
[103] !!! INVALID CITATION !!! [103-105].
[104] Weeks KM. Piercing the fog of the RNA structure-ome. Science 2021;373:964-5.
[105] Bar N, Korem T, Weissbrod O, Zeevi D, Rothschild D, Leviatan S, et al. A reference map of potential determinants for the human serum metabolome. Nature 2020;588:135-40.
[106] Gormley AJ, Webb MA. Machine learning in combinatorial polymer chemistry. Nat Rev Mater 2021;6:642-4.
[107] Meuwly M. Machine Learning for Chemical Reactions. Chem Rev 2021;121:10218-39.
[108] Ahneman DT, Estrada JG, Lin S, Dreher SD, Doyle AG. Predicting reaction performance in C–N cross-coupling using machine learning. Science 2018;360:186-90.
[109] Hart GLW, Mueller T, Toher C, Curtarolo S. Machine learning for alloys. Nat Rev Mater 2021;6:730-55.
[110] Zhong M, Tran K, Min Y, Wang C, Wang Z, Dinh C-T, et al. Accelerated discovery of CO_2 electrocatalysts using active machine learning. Nature 2020;581:178-83.
[111] Zahrt AF, Henle JJ, Rose BT, Wang Y, Darrow WT, Denmark SE. Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. Science 2019;363:eaau5631.
[112] Bzdok D, Krzywinski M, Altman N. Machine learning: a primer. Nat Methods 2017;14:1119-20.
[113] Artrith N, Butler KT, Coudert F-X, Han S, Isayev O, Jain A, et al. Best practices in machine learning for chemistry. Nat Chem 2021;13:505-8.
[114] Palkovits S. A Primer about Machine Learning in Catalysis – A Tutorial with Code. ChemCatChem 2020;12:3995-4008.
[115] Wang AY-T, Murdock RJ, Kauwe SK, Oliynyk AO, Gurlo A, Brgoch J, et al. Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices. Chem Mater 2020;32:4954-65.
[116] Jablonka KM, Ongari D, Moosavi SM, Smit B. Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. Chem Rev 2020;120:8066-129.
[117] Kaufmann K, Zhu C, Rosengarten AS, Maryanovsky D, Harrington TJ, Marin E, et al. Crystal symmetry determination in electron diffraction using machine learning. Science

2020;367:564-8.
[118] Zhang Y, Mesaros A, Fujita K, Edkins SD, Hamidian MH, Ch'ng K, et al. Machine learning in electronic-quantum-matter imaging experiments. Nature 2019;570:484-90.
[119] Tshitoyan V, Dagdelen J, Weston L, Dunn A, Rong Z, Kononova O, et al. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature 2019;571:95-8.
[120] Schwaller P, Hoover B, Reymond J-L, Strobelt H, Laino T. Extraction of organic chemistry grammar from unsupervised learning of chemical reactions. Sci Adv 2021;7:eabe4166.
[121] Zhang Y, He X, Chen Z, Bai Q, Nolan AM, Roberts CA, et al. Unsupervised discovery of solid-state lithium ion conductors. Nat Commun 2019;10:5260.
[122] Ma W, Cheng F, Xu Y, Wen Q, Liu Y. Probabilistic Representation and Inverse Design of Metamaterials Based on a Deep Generative Model with Semi-Supervised Learning Strategy. Adv Mater 2019;31:1901111.
[123] Huo H, Rong Z, Kononova O, Sun W, Botari T, He T, et al. Semi-supervised machine-learning classification of materials synthesis procedures. npj Comput Mater 2019;5:62.
[124] Okaro IA, Jayasinghe S, Sutcliffe C, Black K, Paoletti P, Green PL. Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning. Addit Manuf 2019;27:42-53.
[125] Glielmo A, Husic BE, Rodriguez A, Clementi C, Noé F, Laio A. Unsupervised Learning Methods for Molecular Simulation Data. Chem Rev 2021;121:9722-58.
[126] Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 2016;374:20150202.
[127] Mukherjee SP, Sinha BK, Chattopadhyay AK. Principal Component Analysis. Statistical Methods in Social Science Research. Singapore: Springer Singapore; 2018. p. 95-102.
[128] Koren Y, Bell R, Volinsky C. Matrix Factorization Techniques for Recommender Systems. Computer 2009;42:30-7.
[129] Tenenbaum JB, Silva Vd, Langford JC. A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 2000;290:2319-23.
[130] Schölkopf B, Smola A, Müller K-R. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Comput 1998;10:1299-319.
[131] Van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res 2008;9:2579-605.
[132] Lloyd S. Least squares quantization in PCM. IEEE Trans Inf Theory 1982;28:129-37.
[133] Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis: John Wiley & Sons; 2009.
[134] Zhang R, Li X, Zhang X, Qin H, Xiao W. Machine learning approaches for elucidating the biological effects of natural products. Nat Prod Rep 2021;38:346-61.
[135] Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inf Theory 1967;13:21-7.
[136] Burges CJC. A Tutorial on Support Vector Machines for Pattern Recognition. Data Min Knowl Discov 1998;2:121-67.
[137] Rao R, Carpena-Núñez J, Nikolaev P, Susner MA, Reyes KG, Maruyama B. Advanced machine learning decision policies for diameter control of carbon nanotubes. npj Comput Mater 2021;7:157.
[138] Ture M, Tokatli F, Kurt I. Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients. Expert Syst Appl 2009;36:2017-26.

[139] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825-30.
[140] Jaeger S, Fulle S, Turk S. Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. J Chem Inf Model 2018;58:27-35.
[141] Friedman N, Geiger D, Goldszmidt M. Bayesian Network Classifiers. Mach Learn 1997;29:131-63.
[142] Schultz C, Alegría AC, Cornelis J, Sahli H. Comparison of spatial and aspatial logistic regression models for landmine risk mapping. Appl Geogr 2016;66:52-63.
[143] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA; 2016. p. 785-94.
[144] Su J-W, Li D, Xie Y, Zhou T, Gao W, Deng H, et al. A machine learning workflow for 4D printing: understand and predict morphing behaviors of printed active structures. Smart Mater Struct 2020;30:015028.
[145] Freund Y, Schapire R, Abe N. A short introduction to boosting. Jpn Soc Artif Intell 1999;14:1612.
[146] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. Advances in Neural Information Processing Systems 2014;27.
[147] Kingma DP, Welling M. Auto-Encoding Variational Bayes. 2013. p. arXiv:1312.6114.
[148] Lukoševičius M, Jaeger H. Reservoir computing approaches to recurrent neural network training. Comput Sci Rev 2009;3:127-49.
[149] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput 1997;9:1735-80.
[150] Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The Graph Neural Network Model. IEEE Trans Neural Netw 2009;20:61-80.
[151] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2018. p. arXiv:1810.04805.
[152] Géron A. Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 2nd ed: O'Reilly Media, Inc.; 2019.
[153] Yuan R, Liu Z, Balachandran PV, Xue D, Zhou Y, Ding X, et al. Accelerated Discovery of Large Electrostrains in $BaTiO_3$-Based Piezoelectrics Using Active Learning. Adv Mater 2018;30:1702884.
[154] Tran K, Ulissi ZW. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. Nat Catal 2018;1:696-703.
[155] Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Comput Mater 2019;5:21.
[156] Wang W, Yang T, Harris WH, Gómez-Bombarelli R. Active learning and neural network potentials accelerate molecular screening of ether-based solvate ionic liquids. Chem Commun 2020;56:8920-3.
[157] Rohr B, Stein HS, Guevarra D, Wang Y, Haber JA, Aykol M, et al. Benchmarking the acceleration of materials discovery by sequential learning. Chem Sci 2020;11:2696-706.
[158] Min K, Cho E. Accelerated discovery of potential ferroelectric perovskite via active learning. J Mater Chem C 2020;8:7866-72.
[159] Tian Y, Yuan R, Xue D, Zhou Y, Wang Y, Ding X, et al. Determining Multi-Component Phase Diagrams with Desired Characteristics Using Active Learning. Adv Sci 2021;8:2003165.
[160] Zhao S, Cai T, Zhang L, Li W, Lin J. Autonomous Construction of Phase Diagrams of Block

Copolymers by Theory-Assisted Active Machine Learning. ACS Macro Lett 2021;10:598-602.
[161] Ueno T, Ishibashi H, Hino H, Ono K. Automated stopping criterion for spectral measurements with active learning. npj Comput Mater 2021;7:139.
[162] Kim Y, Kim Y, Yang C, Park K, Gu GX, Ryu S. Deep learning framework for material design space exploration using active transfer learning and data augmentation. npj Comput Mater 2021;7:140.
[163] Kim M, Ha MY, Jung W-B, Yoon J, Shin E, Kim I-d, et al. Searching for an Optimal Multi-Metallic Alloy Catalyst by Active Learning Combined with Experiments. Adv Mater 2022;n/a:2108900.
[164] Shimizu R, Kobayashi S, Watanabe Y, Ando Y, Hitosugi T. Autonomous materials synthesis by machine learning and robotics. APL Mater 2020;8:111110.
[165] Langner S, Häse F, Perea JD, Stubhan T, Hauch J, Roch LM, et al. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems. Adv Mater 2020;32:1907801.
[166] Ament S, Amsler M, Sutherland DR, Chang M-C, Guevarra D, Connolly AB, et al. Autonomous synthesis of metastable materials. 2021. p. arXiv:2101.07385.
[167] Christensen M, Yunker LPE, Adedeji F, Häse F, Roch LM, Gensch T, et al. Data-science driven autonomous process optimization. Commun Chem 2021;4:112.
[168] Gongora AE, Snapp KL, Whiting E, Riley P, Reyes KG, Morgan EF, et al. Using simulation to accelerate autonomous experimentation: A case study using mechanics. iScience 2021;24:102262.
[169] MacLeod BP, Parlane FGL, Rupnow CC, Dettelbach KE, Elliott MS, Morrissey TD, et al. A self-driving laboratory advances the Pareto front for material properties. Nat Commun 2022;13:995.
[170] Porwol L, Kowalski DJ, Henson A, Long D-L, Bell NL, Cronin L. An Autonomous Chemical Robot Discovers the Rules of Inorganic Coordination Chemistry without Prior Knowledge. Angew Chem Int Ed 2020;59:11256-61.
[171] Graff DE, Shakhnovich EI, Coley CW. Accelerating high-throughput virtual screening through molecular pool-based active learning. Chem Sci 2021;12:7866-81.
[172] Brochu E, Cora VM, de Freitas N. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. 2010. p. arXiv:1012.2599.
[173] Snoek J, Larochelle H, Adams RP. Practical Bayesian Optimization of Machine Learning Algorithms. Advances in Neural Information Processing Systems 2012;25.
[174] Shahriari B, Swersky K, Wang Z, Adams RP, Freitas Nd. Taking the Human Out of the Loop: A Review of Bayesian Optimization. Proc IEEE 2016;104:148-75.
[175] Lei B, Kirk TQ, Bhattacharya A, Pati D, Qian X, Arroyave R, et al. Bayesian optimization with adaptive surrogate models for automated experimental design. npj Comput Mater 2021;7:194.
[176] Pedersen JK, Clausen CM, Krysiak OA, Xiao B, Batchelor TAA, Löffler T, et al. Bayesian Optimization of High-Entropy Alloy Compositions for Electrocatalytic Oxygen Reduction. Angew Chem Int Ed 2021;60:24144-52.
[177] Nugraha AS, Lambard G, Na J, Hossain MSA, Asahi T, Chaikittisilp W, et al. Mesoporous trimetallic PtPdAu alloy films toward enhanced electrocatalytic activity in methanol oxidation: unexpected chemical compositions discovered by Bayesian optimization. J Mater Chem A 2020;8:13532-40.
[178] Zuo Y, Qin M, Chen C, Ye W, Li X, Luo J, et al. Accelerating materials discovery with

Bayesian optimization and graph deep learning. Mater Today 2021;51:126-35.
[179] Deshwal A, Simon CM, Doppa JR. Bayesian optimization of nanoporous materials. Mol Syst Des Eng 2021;6:1066-86.
[180] Yamashita T, Sato N, Kino H, Miyake T, Tsuda K, Oguchi T. Crystal structure prediction accelerated by Bayesian optimization. Phys Rev Mater 2018;2:013803.
[181] Sutton RS, Barto AG. Reinforcement learning: An introduction: MIT press; 2018.
[182] Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci Adv 2018;4:eaap7885.
[183] Yang X, Zhang J, Yoshizoe K, Terayama K, Tsuda K. ChemTS: an efficient python library for de novo molecular generation. Sci Technol Adv Mater 2017;18:972-6.
[184] Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de-novo design through deep reinforcement learning. J Cheminformatics 2017;9:48.
[185] Thiede LA, Krenn M, Nigam A, Aspuru-Guzik A. Curiosity in exploring chemical space: Intrinsic rewards for deep molecular reinforcement learning. 2020. p. arXiv:2012.11293.
[186] Li H, Collins CR, Ribelli TG, Matyjaszewski K, Gordon GJ, Kowalewski T, et al. Tuning the molecular weight distribution from atom transfer radical polymerization using deep reinforcement learning. Mol Syst Des Eng 2018;3:496-508.
[187] Rajak P, Krishnamoorthy A, Mishra A, Kalia R, Nakano A, Vashishta P. Autonomous reinforcement learning agent for chemical vapor deposition synthesis of quantum materials. npj Comput Mater 2021;7:108.
[188] Zhou Z, Li X, Zare RN. Optimizing Chemical Reactions with Deep Reinforcement Learning. ACS Cent Sci 2017;3:1337-44.
[189] Wang X, Qian Y, Gao H, Coley Connor W, Mo Y, Barzilay R, et al. Towards efficient discovery of green synthetic pathways with Monte Carlo tree search and reinforcement learning. Chem Sci 2020;11:10959-72.
[190] Rajak P, Wang B, Nomura K-i, Luo Y, Nakano A, Kalia R, et al. Autonomous reinforcement learning agent for stretchable kirigami design of 2D materials. npj Comput Mater 2021;7:102.
[191] Jia X, Lynch A, Huang Y, Danielson M, Lang'at I, Milder A, et al. Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis. Nature 2019;573:251-5.
[192] Zhang Z, Schott JA, Liu M, Chen H, Lu X, Sumpter BG, et al. Prediction of Carbon Dioxide Adsorption via Deep Learning. Angew Chem Int Ed 2019;58:259-63.
[193] Zhang C, Li D, Xie Y, Stalla D, Hua P, Nguyen DT, et al. Machine learning assisted rediscovery of methane storage and separation in porous carbon from material literature. Fuel 2021;290:120080.
[194] Grazulis S, Chateigner D, Downs RT, Yokochi AFT, Quiros M, Lutterotti L, et al. Crystallography Open Database–an open-access collection of crystal structures. J Appl Crystallogr 2009;42:726-9.
[195] Curtarolo S, Setyawan W, Hart GLW, Jahnatek M, Chepulskii RV, Taylor RH, et al. AFLOW: An automatic framework for high-throughput materials discovery. Comput Mater Sci 2012;58:218-26.
[196] Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Mater 2013;1:011002.
[197] Saal JE, Kirklin S, Aykol M, Meredig B, Wolverton C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). JOM 2013;65:1501-9.

[198] Yamazaki M, Xu Y, Murata M, Tanaka H, Kamihira K, Kimura K. NIMS structural materials databases and cross search engine-MatNavi. In, Finland; 2007. p. 193-207.
[199] Draxl C, Scheffler M. The NOMAD laboratory: from data sharing to artificial intelligence. J Phys Mater 2019;2:036001.
[200] Chanussot L, Das A, Goyal S, Lavril T, Shuaibi M, Riviere M, et al. Open Catalyst 2020 (OC20) Dataset and Community Challenges. ACS Catal 2021;11:6059-72.
[201] Lucas A. Ising formulations of many NP problems. Front Phys 2014;2.
[202] Aizawa A. An information-theoretic perspective of tf–idf measures. Inf Process Manag 2003;39:45-65.
[203] Iwasaki Y, Kusne AG, Takeuchi I. Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries. npj Comput Mater 2017;3:4.
[204] Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal Chem 1964;36:1627-39.
[205] Artursson T, Hagman A, Björk S, Trygg J, Wold S, Jacobsson SP. Study of Preprocessing Methods for the Determination of Crystalline Phases in Binary Mixtures of Drug Substances by X-ray Powder Diffraction and Multivariate Calibration. Appl Spectrosc 2000;54:1222-30.
[206] de Rooi JJ, van der Pers NM, Hendrikx RWA, Delhez R, Bottger AJ, Eilers PHC. Smoothing of X-ray diffraction data and $K\alpha_2$ elimination using penalized likelihood and the composite link model. J Appl Crystallogr 2014;47:852-60.
[207] Brandt J, Mattsson K, Hassellöv M. Deep Learning for Reconstructing Low-Quality FTIR and Raman Spectra─A Case Study in Microplastic Analyses. Anal Chem 2021;93:16360-8.
[208] Hutter F, Kotthoff L, Vanschoren J. Automated Machine Learning: Methods, Systems, Challenges: Springer Nature; 2019.
[209] Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 2020;415:295-316.
[210] Montgomery DC. Design and Analysis of Experiments: John wiley & sons; 2017.
[211] Bergstra J, Bengio Y. Random Search for Hyper-Parameter Optimization. J Mach Learn Res 2012;13:281–305.
[212] Eggensperger K, Feurer M, Hutter F, Bergstra J, Snoek J, Hoos H, et al. Towards an Empirical Foundation for Assessing Bayesian Optimization of Hyperparameters. In: NIPS workshop on Bayesian Optimization in Theory and Practice; 2013
[213] Bergstra J, Komer B, Eliasmith C, Yamins D, Cox DD. Hyperopt: a Python library for model selection and hyperparameter optimization. Comput Sci Discov 2015;8:014008.
[214] Autonomio Talos. 2020.
[215] Koch P, Golovidov O, Gardner S, Wujek B, Griffin J, Xu Y. Autotune: A Derivative-free Optimization Framework for Hyperparameter Tuning. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2018. p. 443-52.
[216] Hutter F, Hoos HH, Leyton-Brown K. Sequential Model-based Optimization for General Algorithm Configuration. In: International Conference on Learning and Intelligent Optimization; 2011. p. 507-23.
[217] Golovin D, Solnik B, Moitra S, Kochanski G, Karro J, Sculley D. Google Vizier: A Service for Black-Box Optimization. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada; 2017. p. 1487–95.
[218] Thornton C, Hutter F, Hoos HH, Leyton-Brown K. Auto-WEKA: Combined Selection and Hyperparameter Optimization of Classification Algorithms. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, Chicago, Illinois,

USA; 2013. p. 847–55.
[219] Feurer M, Eggensperger K, Falkner S, Lindauer M, Hutter F. Auto-Sklearn 2.0: Hands-free AutoML via Meta-Learning. 2020. p. arXiv:2007.04074.
[220] Jin H, Song Q, Hu X. Auto-Keras: An Efficient Neural Architecture Search System. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA; 2019. p. 1946–56.
[221] Zimmer L, Lindauer M, Hutter F. Auto-PyTorch Tabular: Multi-Fidelity MetaLearning for Efficient and Robust AutoDL. 2020. p. arXiv:2006.13799.
[222] LeDell E, Poirier S. H2O AutoML: Scalable Automatic Machine Learning. In: Proceedings of the AutoML Workshop at ICML; 2020
[223] Olson RS, Urbanowicz RJ, Andrews PC, Lavender NA, Kidd LC, Moore JH. Automating Biomedical Data Science Through Tree-Based Pipeline Optimization. In: Applications of Evolutionary Computation, Cham; 2016. p. 123-37.
[224] Zhang R, Xie H, Cai S, Hu Y, Liu G-k, Hong W, et al. Transfer-learning-based Raman spectra identification. J Raman Spectrosc 2020;51:176-86.
[225] Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. J Big Data 2019;6:60.
[226] Wu R, Yan S, Shan Y, Dang Q, Sun G. Deep Image: Scaling up Image Recognition. 2015. p. arXiv:1501.02876.
[227] Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW); 2018. p. 117-22.
[228] Zhong Z, Zheng L, Kang G, Li S, Yang Y. Random Erasing Data Augmentation. Proceedings of the AAAI Conference on Artificial Intelligence 2020;34:13001-8.
[229] Kang G, Dong X, Zheng L, Yang Y. PatchShuffle Regularization. 2017. p. arXiv:1707.07103.
[230] Summers C, Dinneen MJ. Improved Mixed-Example Data Augmentation. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV); 2019. p. 1262-70.
[231] DeVries T, Taylor GW. Dataset Augmentation in Feature Space. 2017. p. arXiv:1702.05538.
[232] Gatys LA, Ecker AS, Bethge M. Image Style Transfer Using Convolutional Neural Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 2414-23.
[233] Ma B, Wei X, Liu C, Ban X, Huang H, Wang H, et al. Data augmentation in microscopic images for material data mining. npj Comput Mater 2020;6:125.
[234] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks. Commun ACM 2020;63:139–44.
[235] Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics 2004;20:777-85.
[236] Zinchik S, Jiang S, Friis S, Long F, Høgstedt L, Zavala VM, et al. Accurate Characterization of Mixed Plastic Waste Using Machine Learning and Fast Infrared Spectroscopy. ACS Sustain Chem Eng 2021;9:14143-51.
[237] Huang T-Y, Yu JCC. Development of Crime Scene Intelligence Using a Hand-Held Raman Spectrometer and Transfer Learning. Anal Chem 2021;93:8889-96.
[238] Qi Y, Zhang G, Yang L, Liu B, Zeng H, Xue Q, et al. High-Precision Intelligent Cancer Diagnosis Method: 2D Raman Figures Combined with Deep Learning. Anal Chem 2022.
[239] Qi Y, Yang L, Liu B, Liu L, Liu Y, Zheng Q, et al. Accurate diagnosis of lung tissues for 2D Raman spectrogram by deep learning based on short-time Fourier transform. Anal Chim Acta

2021;1179:338821.
[240] Qi Y, Yang L, Liu B, Liu L, Liu Y, Zheng Q, et al. Highly accurate diagnosis of lung adenocarcinoma and squamous cell carcinoma tissues by deep learning. Spectrochim Acta A Mol Biomol Spectrosc 2022;265:120400.
[241] Ramos PM, Ruisánchez I. Noise and background removal in Raman spectra of ancient pigments using wavelet transform. J Raman Spectrosc 2005;36:848-56.
[242] Du P, Kibbe WA, Lin SM. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. Bioinformatics 2006;22:2059-65.
[243] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014. p. arXiv:1409.556.
[244] He K, Zhang X, Ren S, Sun J. Identity Mappings in Deep Residual Networks. In: Computer Vision – ECCV 2016, Cham; 2016. p. 630-45.
[245] Xiaoling X, Cui X, Bing N. Inception-v3 for flower classification. In: 2017 2nd International Conference on Image, Vision and Computing (ICIVC); 2017. p. 783-7.
[246] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In: Thirty-First AAAI Conference on Artificial Intelligence 2017
[247] Chollet F. Xception: Deep Learning With Depthwise Separable Convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 1251-8.
[248] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 4700-8.
[249] Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research; 2019. p. 6105--14.
[250] Deng J, Dong W, Socher R, Li L, Kai L, Li F-F. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009. p. 248-55.
[251] Tsaig Y, Donoho DL. Extensions of compressed sensing. Signal Process 2006;86:549-71.
[252] Donoho DL. Compressed sensing. IEEE Trans Inf Theory 2006;52:1289-306.
[253] Lu Y, Wang Y. Physics based compressive sensing to monitor temperature and melt flow in laser powder bed fusion. Addit Manuf 2021;47:102304.
[254] Woldegebriel M, Derks E. Artificial Neural Network for Probabilistic Feature Recognition in Liquid Chromatography Coupled to High-Resolution Mass Spectrometry. Anal Chem 2017;89:1212-21.
[255] Zohora FT, Rahman MZ, Tran NH, Xin L, Shan B, Li M. DeepIso: A Deep Learning Model for Peptide Feature Detection from LC-MS map. Sci Rep 2019;9:17168.
[256] Borgsmüller N, Gloaguen Y, Opialla T, Blanc E, Sicard E, Royer A-L, et al. WiPP: Workflow for Improved Peak Picking for Gas Chromatography-Mass Spectrometry (GC-MS) Data. Metabolites 2019;9:171.
[257] Matyushin DD, Sholokhova AY, Buryak AK. Deep Learning Driven GC-MS Library Search and Its Application for Metabolomics. Anal Chem 2020;92:11818-25.
[258] Risum AB, Bro R. Using deep learning to evaluate peaks in chromatographic data. Talanta 2019;204:255-60.
[259] Wu K, Luo J, Zeng Q, Dong X, Chen J, Zhan C, et al. Improvement in Signal-to-Noise Ratio

of Liquid-State NMR Spectroscopy via a Deep Neural Network DN-Unet. Anal Chem 2021;93:1377-82.
[260] Klukowski P, Augoff M, Zięba M, Drwal M, Gonczarek A, Walczak MJ. NMRNet: a deep learning approach to automated peak picking of protein NMR spectra. Bioinformatics 2018;34:2590-7.
[261] Enders AA, North NM, Fensore CM, Velez-Alvarez J, Allen HC. Functional Group Identification for FTIR Spectra Using Image-Based Machine Learning Models. Anal Chem 2021;93:9711-8.
[262] Lansford JL, Vlachos DG. Infrared spectroscopy data- and physics-driven machine learning for characterizing surface microstructure of complex materials. Nat Commun 2020;11:1513.
[263] Sun S, Hartono NTP, Ren ZD, Oviedo F, Buscemi AM, Layurova M, et al. Accelerated Development of Perovskite-Inspired Materials via High-Throughput Synthesis and Machine-Learning Diagnosis. Joule 2019;3:1437-51.
[264] Stanev V, Vesselinov VV, Kusne AG, Antoszewski G, Takeuchi I, Alexandrov BS. Unsupervised phase mapping of X-ray diffraction data by nonnegative matrix factorization integrated with custom clustering. npj Comput Mater 2018;4:43.
[265] Suzuki Y, Hino H, Kotsugi M, Ono K. Automated estimation of materials parameter from X-ray absorption and electron energy-loss spectra with similarity measures. npj Comput Mater 2019;5:39.
[266] Gordon OM, Hodgkinson JEA, Farley SM, Hunsicker EL, Moriarty PJ. Automated Searching and Identification of Self-Organized Nanostructures. Nano Lett 2020;20:7688-93.
[267] Alldritt B, Hapala P, Oinonen N, Urtev F, Krejci O, Canova FF, et al. Automated structure discovery in atomic force microscopy. Sci Adv 2020;6:eaay6913.
[268] Ziatdinov M, Zhang S, Dollar O, Pfaendtner J, Mundy CJ, Li X, et al. Quantifying the Dynamics of Protein Self-Organization Using Deep Learning Analysis of Atomic Force Microscopy Data. Nano Lett 2021;21:158-65.
[269] Kim H, Han J, Han TY-J. Machine vision-driven automatic recognition of particle size and morphology in SEM images. Nanoscale 2020;12:19461-9.
[270] Yang W, Wang Z, Yang T, He L, Song X, Liu Y, et al. Exploration of the Underlying Space in Microscopic Images via Deep Learning for Additively Manufactured Piezoceramics. ACS Appl Mater Interfaces 2021;13:53439-53.
[271] Sheng Y, Deng T, Qiu P, Shi X, Xi J, Han Y, et al. Accelerating the Discovery of Cu–Sn–S Thermoelectric Compounds via High-Throughput Synthesis, Characterization, and Machine Learning-Assisted Image Analysis. Chem Mater 2021;33:6918-24.
[272] Lee B, Yoon S, Lee JW, Kim Y, Chang J, Yun J, et al. Statistical Characterization of the Morphologies of Nanoparticles through Machine Learning Based Electron Microscopy Image Analysis. ACS Nano 2020;14:17125-33.
[273] Horwath JP, Zakharov DN, Mégret R, Stach EA. Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images. npj Comput Mater 2020;6:108.
[274] Yao L, Ou Z, Luo B, Xu C, Chen Q. Machine Learning to Reveal Nanoparticle Dynamics from Liquid-Phase TEM Videos. ACS Cent Sci 2020;6:1421-30.
[275] Li J, Telychko M, Yin J, Zhu Y, Li G, Song S, et al. Machine Vision Automated Chiral Molecule Detection and Classification in Molecular Imaging. J Am Chem Soc 2021;143:10177-88.
[276] Vasudevan RK, Kelley KP, Hinkle J, Funakubo H, Jesse S, Kalinin SV, et al. Autonomous

Experiments in Scanning Probe Microscopy and Spectroscopy: Choosing Where to Explore Polarization Dynamics in Ferroelectrics. ACS Nano 2021;15:11253-62.
[277] Kelley KP, Ziatdinov M, Collins L, Susner MA, Vasudevan RK, Balke N, et al. Fast Scanning Probe Microscopy via Machine Learning: Non-Rectangular Scans with Compressed Sensing and Gaussian Process Optimization. Small 2020;16:2002878.
[278] Farley S, Hodgkinson JEA, Gordon OM, Turner J, Soltoggio A, Moriarty PJ, et al. Improving the segmentation of scanning probe microscope images using convolutional neural networks. Mach Learn Sci Technol 2020;2:015015.
[279] Liu Y, Fields SS, Mimura T, Kelley KP, Ihlefeld JF, Kalinin SV. Exploring leakage in dielectric films via automated experiment in scanning probe microscopy. 2021. p. arXiv:2111.09918.
[280] Roccapriore KM, Zou Q, Zhang L, Xue R, Yan J, Ziatdinov M, et al. Revealing the Chemical Bonding in Adatom Arrays via Machine Learning of Hyperspectral Scanning Tunneling Spectroscopy Data. ACS Nano 2021;15:11806-16.
[281] Usman M, Wong YZ, Hill CD, Hollenberg LCL. Framework for atomic-level characterisation of quantum computer arrays by machine learning. npj Comput Mater 2020;6:19.
[282] Valleti SMP, Zou Q, Xue R, Vlcek L, Ziatdinov M, Vasudevan R, et al. Bayesian Learning of Adatom Interactions from Atomically Resolved Imaging Data. ACS Nano 2021;15:9649-57.
[283] Ghosh A, Sumpter BG, Dyck O, Kalinin SV, Ziatdinov M. Ensemble learning-iterative training machine learning for uncertainty quantification and automated experiment in atom-resolved microscopy. npj Comput Mater 2021;7:100.
[284] Nelson CT, Ghosh A, Oxley M, Zhang X, Ziatdinov M, Takeuchi I, et al. Deep learning ferroelectric polarization distributions from STEM data via with and without atom finding. npj Comput Mater 2021;7:149.
[285] Guo Y, Kalinin SV, Cai H, Xiao K, Krylyuk S, Davydov AV, et al. Defect detection in atomic-resolution images via unsupervised learning with translational invariance. npj Comput Mater 2021;7:180.
[286] Kalinin SV, Oxley MP, Valleti M, Zhang J, Hermann RP, Zheng H, et al. Deep Bayesian local crystallography. npj Comput Mater 2021;7:181.
[287] Creange N, Dyck O, Vasudevan RK, Ziatdinov M, Kalinin SV. Towards Automating Structural Discovery in Scanning Transmission Electron Microscopy. Mach Learn Sci Technol 2021.
[288] Kalinin SV, Steffes JJ, Liu Y, Huey BD, Ziatdinov M. Disentangling ferroelectric domain wall geometries and pathways in dynamic piezoresponse force microscopy via unsupervised machine learning. Nanotechnol 2021;33:055707.
[289] Kelley KP, Ren Y, Dasgupta A, Kavle P, Jesse S, Vasudevan RK, et al. Probing Metastable Domain Dynamics via Automated Experimentation in Piezoresponse Force Microscopy. ACS Nano 2021;15:15096-103.
[290] Liu Y, Kelley KP, Vasudevan RK, Funakubo H, Ziatdinov MA, Kalinin SV. Experimental discovery of structure-property relationships in ferroelectric materials via active learning. 2021. p. arXiv:2108.06037.
[291] Saito Y, Shin K, Terayama K, Desai S, Onga M, Nakagawa Y, et al. Deep-learning-based quality filtering of mechanically exfoliated 2D crystals. npj Comput Mater 2019;5:124.
[292] Cheng L, Assary RS, Qu X, Jain A, Ong SP, Rajput NN, et al. Accelerating Electrolyte Discovery for Energy Storage with High-Throughput Screening. J Phys Chem Lett 2015;6:283-91.
[293] Merrifield RB, Stewart JM, Jernberg N. Instrument for automated synthesis of peptides. Anal

Chem 1966;38:1905-14.
[294] Deming SN, Pardue HL. Automated instrumental system for fundamental characterization of chemical reactions. Anal Chem 1971;43:192-200.
[295] Winicov H, Schainbaum J, Buckley J, Longino G, Hill J, Berkoff CE. Chemical process optimization by computer — a self-directed chemical synthesis system. Anal Chim Acta 1978;103:469-76.
[296] Snively CM, Oskarsdottir G, Lauterbach J. Chemically Sensitive High Throughput Parallel Analysis of Solid Phase Supported Library Members. J Comb Chem 2000;2:243-5.
[297] McCullough K, Chiang P-H, Jimenez JD, Lauterbach JA. Material Discovery and High Throughput Exploration of Ru Based Catalysts for Low Temperature Ammonia Decomposition. Materials 2020;13:1869.
[298] McCullough K, Williams T, Mingle K, Jamshidi P, Lauterbach J. High-throughput experimentation meets artificial intelligence: a new pathway to catalyst discovery. Phys Chem Chem Phys 2020;22:11174-96.
[299] Sasmaz E, Mingle K, Lauterbach J. High-Throughput Screening Using Fourier-Transform Infrared Imaging. Engineering 2015;1:234-42.
[300] Higgins K, Valleti SM, Ziatdinov M, Kalinin SV, Ahmadi M. Chemical Robotics Enabled Exploration of Stability in Multicomponent Lead Halide Perovskites via Machine Learning. ACS Energy Lett 2020;5:3426-36.
[301] Higgins K, Ziatdinov M, Kalinin SV, Ahmadi M. High-Throughput Study of Antisolvents on the Stability of Multicomponent Metal Halide Perovskites through Robotics-Based Synthesis and Machine Learning Approaches. J Am Chem Soc 2021;143:19945-55.
[302] Xiang X-D, Sun X, Briceño G, Lou Y, Wang K-A, Chang H, et al. A Combinatorial Approach to Materials Discovery. Science 1995;268:1738-40.
[303] Schultz PG, Xiang X-D. Combinatorial approaches to materials science. Curr Opin Solid State Mater Sci 1998;3:153-8.
[304] Wang J, Yoo Y, Gao C, Takeuchi I, Sun X, Chang H, et al. Identification of a Blue Photoluminescent Composite Material from a Combinatorial Library. Science 1998;279:1712-4.
[305] Koinuma H, Takeuchi I. Combinatorial solid-state chemistry of inorganic materials. Nat Mater 2004;3:429-38.
[306] Takeuchi I, Lauterbach J, Fasolka MJ. Combinatorial materials synthesis. Mater Today 2005;8:18-26.
[307] Cui J, Chu YS, Famodu OO, Furuya Y, Hattrick-Simpers J, James RD, et al. Combinatorial search of thermoelastic shape-memory alloys with extremely small hysteresis width. Nat Mater 2006;5:286-90.
[308] Takeuchi I, Famodu OO, Read JC, Aronova MA, Chang KS, Craciunescu C, et al. Identification of novel compositions of ferromagnetic shape-memory alloys using composition spreads. Nat Mater 2003;2:180-4.
[309] Haber JA, Cai Y, Jung S, Xiang C, Mitrovic S, Jin J, et al. Discovering Ce-rich oxygen evolution catalysts, from high throughput screening to water electrolysis. Energy Environ Sci 2014;7:682-8.
[310] Yan Q, Yu J, Suram SK, Zhou L, Shinde A, Newhouse PF, et al. Solar fuels photoanode materials discovery by integrating high-throughput theory and experiment. PNAS 2017;114:3040-3.
[311] Green ML, Choi CL, Hattrick-Simpers JR, Joshi AM, Takeuchi I, Barron SC, et al. Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies.

Appl Phys Rev 2017;4:011105.
[312] Zhou L, Yan Q, Shinde A, Guevarra D, Newhouse PF, Becerra-Stasiewicz N, et al. High Throughput Discovery of Solar Fuels Photoanodes in the $CuO–V_2O_5$ System. Adv Energy Mater 2015;5:1500968.
[313] Haber JA, Xiang C, Guevarra D, Jung S, Jin J, Gregoire JM. High-Throughput Mapping of the Electrochemical Properties of $(Ni-Fe-Co-Ce)O_x$ Oxygen-Evolution Catalysts. ChemElectroChem 2014;1:524-8.
[314] Gregoire JM, Van Campen DG, Miller CE, Jones RJR, Suram SK, Mehta A. High-throughput synchrotron X-ray diffraction for combinatorial phase mapping. J Synchrotron Radiat 2014;21:1262-8.
[315] Christen HM, Silliman SD, Harshavardhan KS. Continuous compositional-spread technique based on pulsed-laser deposition and applied to the growth of epitaxial films. Rev Sci Instrum 2001;72:2673-8.
[316] Christen HM, Eres G. Recent advances in pulsed-laser deposition of complex oxides. J Phys: Condens Matter 2008;20:264005.
[317] Fowlkes JD, Fitz-Gerald JM, Rack PD. Ultraviolet emitting $(Y_{1-x}Gd_x)_2O_{3-\delta}$ thin films deposited by radio frequency magnetron sputtering; structure-property-thin film processing relationships. Thin Solid Films 2007;515:3488-98.
[318] Deng Y, Guan YF, Rack PD. Combinatorial synthesis and sputter parameter optimization of chromium-doped yttrium aluminum garnet photoluminescent thin films. Thin Solid Films 2006;515:1721-6.
[319] Shi Y, Yang B, Rack PD, Guo S, Liaw PK, Zhao Y. High-throughput synthesis and corrosion behavior of sputter-deposited nanocrystalline $Al_x(CoCrFeNi)_{100-x}$ combinatorial high-entropy alloys. Mater Des 2020;195:109018.
[320] Näsström H, Becker P, Márquez JA, Shargaieva O, Mainz R, Unger E, et al. Dependence of phase transitions on halide ratio in inorganic $CsPb(Br_xI_{1-x})_3$ perovskite thin films obtained from high-throughput experimentation. J Mater Chem A 2020;8:22626-31.
[321] Becker P, Márquez JA, Just J, Al-Ashouri A, Hages C, Hempel H, et al. Low Temperature Synthesis of Stable γ-$CsPbI_3$ Perovskite Layers for Solar Cells Obtained by High Throughput Experimentation. Adv Energy Mater 2019;9:1900555.
[322] Näsström H, Shargaieva O, Becker P, Mathies F, Zizak I, Schröder VRF, et al. Combinatorial inkjet printing for compositional tuning of metal-halide perovskite thin films. J Mater Chem A 2022;10:4906-14.
[323] Angelone D, Hammer AJS, Rohrbach S, Krambeck S, Granda JM, Wolf J, et al. Convergence of multiple synthetic paradigms in a universally programmable chemical synthesis machine. Nat Chem 2021;13:63-9.
[324] Vikram A, Brudnak K, Zahid A, Shim M, Kenis PJA. Accelerated screening of colloidal nanocrystals using artificial neural network-assisted autonomous flow reactor technology. Nanoscale 2021;13:17028-39.
[325] Reis M, Gusev F, Taylor NG, Chung SH, Verber MD, Lee YZ, et al. Machine-Learning-Guided Discovery of ^{19}F MRI Agents Enabled by Automated Copolymer Synthesis. J Am Chem Soc 2021;143:17677-89.
[326] Epps RW, Volk AA, Reyes KG, Abolhasani M. Accelerated AI development for autonomous materials synthesis in flow. Chem Sci 2021;12:6025-36.
[327] Jeraal MI, Sung S, Lapkin AA. A Machine Learning-Enabled Autonomous Flow Chemistry Platform for Process Optimization of Multiple Reaction Metrics. Chemistry–Methods 2021;1:71-

7.
[328] Hall BL, Taylor CJ, Labes R, Massey AF, Menzel R, Bourne RA, et al. Autonomous optimisation of a nanoparticle catalysed reduction reaction in continuous flow. Chem Commun 2021;57:4926-9.
[329] Tao H, Wu T, Kheiri S, Aldeghi M, Aspuru-Guzik A, Kumacheva E. Self-Driving Platform for Metal Nanoparticle Synthesis: Combining Microfluidics and Machine Learning. Adv Funct Mater 2021:2106725.
[330] Abdel-Latif K, Epps RW, Bateni F, Han S, Reyes KG, Abolhasani M. Self-Driven Multistep Quantum Dot Synthesis Enabled by Autonomous Robotic Experimentation in Flow. Adv Intell Syst 2021;3:2000245.
[331] Bornemann-Pfeiffer M, Wolf J, Meyer K, Kern S, Angelone D, Leonov A, et al. Standardization and Control of Grignard Reactions in a Universal Chemical Synthesis Machine using online NMR. Angew Chem Int Ed 2021;60:23202-6.
[332] Chatterjee S, Guidi M, Seeberger PH, Gilmore K. Automated radial synthesis of organic molecules. Nature 2020;579:379-84.
[333] Hartrampf N, Saebi A, Poskus M, Gates ZP, Callahan AJ, Cowfer AE, et al. Synthesis of proteins by automated flow chemistry. Science 2020;368:980-7.
[334] Vasudevan N, Wimmer E, Barré E, Cortés-Borda D, Rodriguez-Zubiri M, Felpin F-X. Direct C−H Arylation of Indole-3-Acetic Acid Derivatives Enabled by an Autonomous Self-Optimizing Flow Reactor. Adv Synth Catal 2021;363:791-9.
[335] Li J, Tu Y, Liu R, Lu Y, Zhu X. Toward "On-Demand" Materials Synthesis and Scientific Discovery through Intelligent Robots. Adv Sci 2020;7:1901957.
[336] Waldron C, Pankajakshan A, Quaglio M, Cao E, Galvanin F, Gavriilidis A. An autonomous microreactor platform for the rapid identification of kinetic models. React Chem Eng 2019;4:1623-36.
[337] Wimmer E, Cortés-Borda D, Brochard S, Barré E, Truchet C, Felpin F-X. An autonomous self-optimizing flow machine for the synthesis of pyridine–oxazoline (PyOX) ligands. React Chem Eng 2019;4:1608-15.
[338] Aka EC, Wimmer E, Barré E, Vasudevan N, Cortés-Borda D, Ekou T, et al. Reconfigurable Flow Platform for Automated Reagent Screening and Autonomous Optimization for Bioinspired Lignans Synthesis. J Org Chem 2019;84:14101-12.
[339] Whitacre JF, Mitchell J, Dave A, Wu W, Burke S, Viswanathan V. An Autonomous Electrochemical Test Stand for Machine Learning Informed Electrolyte Optimization. J Electrochem Soc 2019;166:A4181-A7.
[340] Rubens M, Vrijsen JH, Laun J, Junkers T. Precise Polymer Synthesis by Autonomous Self-Optimizing Flow Reactors. Angew Chem Int Ed 2019;58:3183-7.
[341] Bédard A-C, Adamo A, Aroh KC, Russell MG, Bedermann AA, Torosian J, et al. Reconfigurable system for automated optimization of diverse chemical reactions. Science 2018;361:1220-5.
[342] Cortés-Borda D, Wimmer E, Gouilleux B, Barré E, Oger N, Goulamaly L, et al. An Autonomous Self-Optimizing Flow Reactor for the Synthesis of Natural Product Carpanone. J Org Chem 2018;83:14286-99.
[343] Dragone V, Sans V, Henson AB, Granda JM, Cronin L. An autonomous organic reaction search engine for chemical reactivity. Nat Commun 2017;8:15733.
[344] Deneault JR, Chang J, Myung J, Hooper D, Armstrong A, Pitt M, et al. Toward autonomous additive manufacturing: Bayesian optimization on a 3D printer. MRS Bull 2021;46:566-75.

[345] Liang J, Xu S, Hu L, Zhao Y, Zhu X. Machine-learning-assisted low dielectric constant polymer discovery. Mater Chem Front 2021;5:3823-9.
[346] Shiri P, Lai V, Zepel T, Griffin D, Reifman J, Clark S, et al. Automated solubility screening platform using computer vision. iScience 2021;24:102176.
[347] Cao L, Russo D, Felton K, Salley D, Sharma A, Keenan G, et al. Optimization of Formulations Using Robotic Experiments Driven by Machine Learning DoE. Cell Rep Phys Sci 2021;2:100295.
[348] Li X, Maffettone PM, Che Y, Liu T, Chen L, Cooper AI. Combining machine learning and high-throughput experimentation to discover photocatalytically active organic molecules. Chem Sci 2021;12:10742-54.
[349] Boyce BL, Uchic MD. Progress toward autonomous experimental systems for alloy development. MRS Bull 2019;44:273-80.
[350] Noack MM, Yager KG, Fukuto M, Doerk GS, Li R, Sethian JA. A Kriging-Based Approach to Autonomous Experimentation with Applications to X-Ray Scattering. Sci Rep 2019;9:11809.
[351] Masubuchi S, Morimoto M, Morikawa S, Onodera M, Asakawa Y, Watanabe K, et al. Autonomous robotic searching and assembly of two-dimensional crystals to build van der Waals superlattices. Nat Commun 2018;9:1413.
[352] Liu H, Stoll N, Junginger S, Thurow K. Mobile Robot for Life Science Automation. Int J Adv Robot Syst 2013;10:288.
[353] Liu H, Stoll N, Junginger S, Thurow K. A Fast Approach to Arm Blind Grasping and Placing for Mobile Robot Transportation in Laboratories. Int J Adv Robot Syst 2014;11:43.
[354] Abdulla AA, Liu H, Stoll N, Thurow K. A New Robust Method for Mobile Robot Multifloor Navigation in Distributed Life Science Laboratories. J Control Sci Eng 2016;2016:3589395.
[355] Li J, Lu Y, Xu Y, Liu C, Tu Y, Ye S, et al. AIR-Chem: Authentic Intelligent Robotics for Chemistry. J Phys Chem A 2018;122:9142-8.
[356] Sambiagio C, Noël T. Flow Photochemistry: Shine Some Light on Those Tubes! Trends Chem 2020;2:92-106.
[357] Cambié D, Bottecchia C, Straathof NJW, Hessel V, Noël T. Applications of Continuous-Flow Photochemistry in Organic Synthesis, Material Science, and Water Treatment. Chem Rev 2016;116:10276-341.
[358] Volk AA, Abolhasani M. Autonomous flow reactors for discovery and invention. Trends Chem 2021;3:519-22.
[359] Fu WC, MacQueen PM, Jamison TF. Continuous flow strategies for using fluorinated greenhouse gases in fluoroalkylations. Chem Soc Rev 2021;50:7378-94.
[360] Neyt NC, Riley DL. Application of reactor engineering concepts in continuous flow chemistry: a review. React Chem Eng 2021;6:1295-326.
[361] Mateos C, Nieves-Remacha MJ, Rincón JA. Automated platforms for reaction self-optimization in flow. React Chem Eng 2019;4:1536-44.
[362] Reis MH, Varner TP, Leibfarth FA. The Influence of Residence Time Distribution on Continuous-Flow Polymerization. Macromolecules 2019;52:3551-7.
[363] Laue S, Haverkamp V, Mleczko L. Experience with Scale-Up of Low-Temperature Organometallic Reactions in Continuous Flow. Org Process Res Dev 2016;20:480-6.
[364] Hartman RL, Naber JR, Zaborenko N, Buchwald SL, Jensen KF. Overcoming the Challenges of Solid Bridging and Constriction during Pd-Catalyzed C−N Bond Formation in Microreactors. Org Process Res Dev 2010;14:1347-57.
[365] White TD, Alt CA, Cole KP, Groh JM, Johnson MD, Miller RD. How to Convert a Walk-in

Hood into a Manufacturing Facility: Demonstration of a Continuous, High-Temperature Cyclization to Process Solids in Flow. Org Process Res Dev 2014;18:1482-91.
[366] Tsaoulidis D, Angeli P. Effect of channel size on mass transfer during liquid–liquid plug flow in small scale extractors. Chem Eng J 2015;262:785-93.
[367] Horie T, Sumino M, Tanaka T, Matsushita Y, Ichimura T, Yoshida J-i. Photodimerization of Maleic Anhydride in a Microreactor Without Clogging. Org Process Res Dev 2010;14:405-10.
[368] Olivon K, Sarrazin F. Heterogeneous reaction with solid catalyst in droplet-flow millifluidic device. Chem Eng J 2013;227:97-102.
[369] Nightingale AM, Phillips TW, Bannock JH, de Mello JC. Controlled multistep synthesis in a three-phase droplet reactor. Nat Commun 2014;5:3777.
[370] Dong Z, Yao C, Zhang X, Xu J, Chen G, Zhao Y, et al. A high-power ultrasonic microreactor and its application in gas–liquid mass transfer intensification. Lab Chip 2015;15:1145-52.
[371] Zhang L, Geng M, Teng P, Zhao D, Lu X, Li J-X. Ultrasound-promoted intramolecular direct arylation in a capillary flow microreactor. Ultrason Sonochem 2012;19:250-6.
[372] Scheiff F, Agar DW. Solid Particle Handling in Microreaction Technology: Practical Challenges and Application of Microfluid Segments for Particle-Based Processes. In: Köhler JM, Cahill BP. Micro-Segmented Flow: Applications in Chemistry and Biology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 103-48.
[373] Chapman MR, Kwan MHT, King G, Jolley KE, Hussain M, Hussain S, et al. Simple and Versatile Laboratory Scale CSTR for Multiphasic Continuous-Flow Chemistry and Long Residence Times. Org Process Res Dev 2017;21:1294-301.
[374] Baran T, Sargin I, Menteş A, Kaya M. Exceptionally high turnover frequencies recorded for a new chitosan-based palladium(II) catalyst. Appl Catal A Gen 2016;523:12-20.
[375] Vural Gürsel I, Kockmann N, Hessel V. Fluidic separation in microstructured devices – Concepts and their Integration into process flow networks. Chem Eng Sci 2017;169:3-17.
[376] Imbrogno J, Rogers L, Thomas DA, Jensen KF. Continuous purification of active pharmaceutical ingredients utilizing polymer membrane surface wettability. Chem Commun 2018;54:70-3.
[377] Yang R-J, Liu C-C, Wang Y-N, Hou H-H, Fu L-M. A comprehensive review of micro-distillation methods. Chem Eng J 2017;313:1509-20.
[378] Escribà-Gelonch M, Hessel V, Maier MC, Noël T, Neira d'Angelo MF, Gruber-Woelfler H. Continuous-Flow In-Line Solvent-Swap Crystallization of Vitamin D_3. Org Process Res Dev 2018;22:178-89.
[379] Agostino FJ, Krylov SN. Advances in steady-state continuous-flow purification by small-scale free-flow electrophoresis. TrAC 2015;72:68-79.
[380] Britton J, Jamison TF. The assembly and use of continuous flow systems for chemical synthesis. Nat Protoc 2017;12:2423-46.
[381] Sweet EC, Mehta RR, Lin R, Lin L. Finger-powred, 3D printed microfluidic pumps. In: 2017 19th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS); 2017. p. 1766-9.
[382] Capel AJ, Rimington RP, Lewis MP, Christie SDR. 3D printing for chemical, pharmaceutical and biological applications. Nat Rev Chem 2018;2:422-36.
[383] Hou W, Bubliauskas A, Kitson PJ, Francoia J-P, Powell-Davies H, Gutierrez JMP, et al. Automatic Generation of 3D-Printed Reactionware for Chemical Synthesis Digitization using ChemSCAD. ACS Cent Sci 2021;7:212-8.
[384] Kitson PJ, Symes MD, Dragone V, Cronin L. Combining 3D printing and liquid handling to

produce user-friendly reactionware for chemical synthesis and purification. Chem Sci 2013;4:3099-103.
[385] Gutmann B, Köckinger M, Glotz G, Ciaglia T, Slama E, Zadravec M, et al. Design and 3D printing of a stainless steel reactor for continuous difluoromethylations using fluoroform. React Chem Eng 2017;2:919-27.
[386] Kitson PJ, Marie G, Francoia J-P, Zalesskiy SS, Sigerson RC, Mathieson JS, et al. Digitization of multistep organic synthesis in reactionware for on-demand pharmaceuticals. Science 2018;359:314-9.
[387] King RD, Rowland J, Aubrey W, Liakata M, Markham M, Soldatova LN, et al. The Robot Scientist Adam. Computer 2009;42:46-54.
[388] Sparkes A, Aubrey W, Byrne E, Clare A, Khan MN, Liakata M, et al. Towards Robot Scientists for autonomous scientific discovery. Autom Exp 2010;2:1.
[389] King RD. Rise of the Robo Scientists. Sci Am 2011;304:72-7.
[390] Williams K, Bilsland E, Sparkes A, Aubrey W, Young M, Soldatova LN, et al. Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases. J R Soc Inteface 2015;12:20141289.
[391] King RD, Costa VS, Mellingwood C, Soldatova LN. Automating Sciences: Philosophical and Social Dimensions. IEEE Technol Soc Mag 2018;37:40-6.
[392] Coutant A, Roper K, Trejo-Banos D, Bouthinon D, Carpenter M, Grzebyta J, et al. Closed-loop cycles of experiment design, execution, and learning accelerate systems biology model development in yeast. PNAS 2019;116:18142-7.
[393] Points LJ, Taylor JW, Grizou J, Donkers K, Cronin L. Artificial intelligence exploration of unstable protocells leads to predictable properties and discovery of collective behavior. PNAS 2018;115:885-90.
[394] Gutierrez JMP, Hinkley T, Taylor JW, Yanev K, Cronin L. Evolution of oil droplets in a chemorobotic platform. Nat Commun 2014;5:5571.
[395] Check Hayden E. The automated lab. Nature 2014;516:131-2.
[396] Peplow M. Organic synthesis: The robo-chemist. Nature 2014;512:20-2.
[397] Sanderson K. Automation: Chemistry shoots for the Moon. Nature 2019;568:577-80.
[398] May M. Automated Science on a Shoestring. NATURE PUBLISHING GROUP MACMILLAN BUILDING, 4 CRINAN ST, LONDON N1 9XW, ENGLAND; 2019. p. 587-8.
[399] Iwasaki Y, Kusne AG, Takeuchi I. Comparison of dissimilarity measures for cluster analysis of X-ray diffraction data from combinatorial libraries. npj Comput Mater 2017;3:1-9.
[400] Ge M, Su F, Zhao Z, Su D. Deep learning analysis on microscopic imaging in materials science. Mater Today Nano 2020;11:100087.
[401] Hong S, Liow CH, Yuk JM, Byon HR, Yang Y, Cho E, et al. Reducing Time to Discovery: Materials and Molecular Modeling, Imaging, Informatics, and Integration. ACS Nano 2021;15:3971-95.
[402] Sans V, Cronin L. Towards dial-a-molecule by integrating continuous flow, analytics and self-optimisation. Chem Soc Rev 2016;45:2032-43.
[403] Pozo C, Rodríguez-Llamazares S, Bouza R, Barral L, Castaño J, Müller N, et al. Study of the structural order of native starch granules using combined FTIR and XRD analysis. J Polym Res 2018;25:266.
[404] Fan X, Ming W, Zeng H, Zhang Z, Lu H. Deep learning-based component identification for the Raman spectra of mixtures. Analyst 2019;144:1789-98.
[405] Lin J, Peng Z, Liu Y, Ruiz-Zepeda F, Ye R, Samuel ELG, et al. Laser-induced porous

graphene films from commercial polymers. Nat Commun 2014;5:5714.
[406] Xie Y, Zhang C, Su J-W, Deng H, Zhang C, Lin J. Rapid Synthesis of Zeolitic Imidazole Frameworks in Laser-Induced Graphene Microreactors. ChemSusChem 2019;12:473-9.
[407] Iandola F, Moskewicz M, Karayev S, Girshick R, Darrell T, Keutzer K. DenseNet: Implementing Efficient ConvNet Descriptor Pyramids. 2014. p. arXiv:1404.869.
[408] Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018. p. 6848-56.
[409] Ma N, Zhang X, Zheng H-T, Sun J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In: Computer Vision – ECCV 2018, Cham; 2018. p. 122-38.
[410] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision (ICCV); 2017. p. 2980-8.
[411] Chatterjee S, Hore S, Dey N, Chakraborty S, Ashour AS. Dengue Fever Classification Using Gene Expression Data: A PSO Based Artificial Neural Network Approach. In: Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications, Singapore; 2017. p. 331-41.
[412] Santosh KC, Nattee C. Template-based Nepali Natural Handwritten Alphanumeric Character Recognition. Thammasat Int J Sci Tech 2007;12:20-30.
[413] Hore S, Chatterjee S, Sarkar S, Dey N, Ashour AS, Balas-Timar D, et al. Neural-based prediction of structural failure of multistoried RC buildings. Struct Eng Mech 2016;58:459-73.
[414] Talebi H, Milanfar P. Learning to Resize Images for Computer Vision Tasks. 2021. p. arXiv:2103.09950.
[415] Maji P, Chatterjee S, Chakraborty S, Kausar N, Samanta S, Dey N. Effect of Euler number as a feature in gender recognition system from offline handwritten signature using neural networks. In: 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom); 2015. p. 1869-73.
[416] Mishra SP, Rahul MR. A comparative study and development of a novel deep learning architecture for accelerated identification of microstructure in materials science. Comput Mater Sci 2021;200:110815.
[417] Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. Deep learning for cellular image analysis. Nat Methods 2019;16:1233-46.
[418] Van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, et al. scikit-image: image processing in Python. PeerJ 2014;2:e453.
[419] Pan M, Crozier PA. Low-dose high-resolution electron microscopy of zeolite materials with a slow-scan CCD camera. Ultramicroscopy 1993;48:332-40.
[420] Fujiyoshi Y. The structural study of membrane proteins by electron crystallography. Adv Biophys 1998;35:25-80.
[421] Stevens A, Yang H, Carin L, Arslan I, Browning ND. The potential for Bayesian compressive sensing to significantly reduce electron dose in high-resolution STEM images. Microscopy 2014;63:41-51.
[422] Kovarik L, Stevens A, Liyu A, Browning ND. Implementing an accurate and rapid sparse sampling approach for low-dose atomic resolution STEM imaging. Appl Phys Lett 2016;109:164102.
[423] Stevens A, Luzi L, Yang H, Kovarik L, Mehdi BL, Liyu A, et al. A sub-sampled approach to extremely low-dose STEM. Appl Phys Lett 2018;112:043104.
[424] Szymanski NJ, Zeng Y, Huo H, Bartel CJ, Kim H, Ceder G. Toward autonomous design and

synthesis of novel inorganic materials. Mater Horizon 2021;8:2169-98.
[425] Chang J, Nikolaev P, Carpena-Núñez J, Rao R, Decker K, Islam AE, et al. Efficient Closed-loop Maximization of Carbon Nanotube Growth Rate using Bayesian Optimization. Sci Rep 2020;10:9040.
[426] Hickman RJ, Aldeghi M, Häse F, Aspuru-Guzik A. Bayesian optimization with known experimental and design constraints for chemistry applications. 2022. p. arXiv:2203.17241.
[427] Häse F, Roch LM, Aspuru-Guzik A. Chimera: enabling hierarchy based multi-objective optimization for self-driving laboratories. Chem Sci 2018;9:7642-55.
[428] Schweidtmann AM, Clayton AD, Holmes N, Bradford E, Bourne RA, Lapkin AA. Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives. Chem Eng J 2018;352:277-82.
[429] Clayton AD, Schweidtmann AM, Clemens G, Manson JA, Taylor CJ, Niño CG, et al. Automated self-optimisation of multi-step reaction and separation processes using machine learning. Chem Eng J 2020;384:123340.
[430] Erps T, Foshey M, Luković MK, Shou W, Goetzke HH, Dietsch H, et al. Accelerated discovery of 3D printing materials using data-driven multiobjective optimization. Sci Adv 2021;7:eabf7435.
[431] Nikolaev P, Hooper D, Webber F, Rao R, Decker K, Krein M, et al. Autonomy in materials research: a case study in carbon nanotube growth. npj Comput Mater 2016;2:16031.
[432] Salley D, Keenan G, Grizou J, Sharma A, Martín S, Cronin L. A nanomaterials discovery robot for the Darwinian evolution of shape programmable gold nanoparticles. Nat Commun 2020;11:2771.
[433] Holmes N, Akien GR, Blacker AJ, Woodward RL, Meadows RE, Bourne RA. Self-optimisation of the final stage in the synthesis of EGFR kinase inhibitor AZD9291 using an automated flow reactor. React Chem Eng 2016;1:366-71.
[434] Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N. Taking the human out of the loop: A review of Bayesian optimization. Proc IEEE 2015;104:148-75.
[435] Bradford E, Schweidtmann AM, Lapkin A. Efficient multiobjective optimization employing Gaussian processes, spectral sampling and a genetic algorithm. J Glob Optim 2018;71:407-38.
[436] Galuzio PP, de Vasconcelos Segundo EH, Coelho LdS, Mariani VC. MOBOpt — multi-objective Bayesian optimization. SoftwareX 2020;12:100520.
[437] Osborne MA, Garnett R, Roberts SJ. Gaussian processes for global optimization. In: 3rd international conference on learning and intelligent optimization (LION3); 2009. p. 1-15.
[438] Seeger MW, Williams CKI, Lawrence ND. Fast Forward Selection to Speed Up Sparse Gaussian Process Regression. In: Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research; 2003. p. 254--61.
[439] Snelson E, Ghahramani Z. Sparse Gaussian Processes using Pseudo-inputs. In: Advances in Neural Information Processing Systems; 2005
[440] Lázaro-Gredilla M, Quinonero-Candela J, Rasmussen CE, Figueiras-Vidal AR. Sparse Spectrum Gaussian Process Regression. J Mach Learn Res 2010;11:1865-81.
[441] Springenberg JT, Klein A, Falkner S, Hutter F. Bayesian optimization with robust Bayesian neural networks. Advances in Neural Information Processing Systems 2016;29:4134-42.
[442] Snoek J, Rippel O, Swersky K, Kiros R, Satish N, Sundaram N, et al. Scalable Bayesian Optimization Using Deep Neural Networks. In: Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research; 2015. p. 2171-80.
[443] Hutter F, Hoos HH, Leyton-Brown K. Parallel Algorithm Configuration. In: Learning and

Intelligent Optimization, Berlin, Heidelberg; 2012. p. 55-70.
[444] Močkus J. On Bayesian Methods for Seeking the Extremum. In: Marchuk GI. Optimization Techniques IFIP Technical Conference: Novosibirsk, July 1–7, 1974. Berlin, Heidelberg: Springer Berlin Heidelberg; 1975. p. 400-4.
[445] Kushner HJ. A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise. J Basic Eng 1964;86:97-106.
[446] Srinivas N, Krause A, Kakade SM, Seeger M. Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. 2009. p. arXiv:0912.3995.
[447] Thompson WR. On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. Biometrika 1933;25:285-94.
[448] Villemonteix J, Vazquez E, Walter E. An informational approach to the global optimization of expensive-to-evaluate functions. J Glob Optim 2008;44:509.
[449] Hennig P, Schuler CJ. Entropy Search for Information-Efficient Global Optimization. 2011. p. arXiv:1112.217.
[450] Wu J, Poloczek M, Wilson AG, Frazier PI. Bayesian Optimization with Gradients. 2017. p. arXiv:1703.04389.
[451] Nikolaev P, Hooper D, Perea-López N, Terrones M, Maruyama B. Discovery of Wall-Selective Carbon Nanotube Growth Conditions via Automated Experimentation. ACS Nano 2014;8:10214-22.
[452] Häse F, Roch LM, Kreisbeck C, Aspuru-Guzik A. Phoenics: A Bayesian Optimizer for Chemistry. ACS Cent Sci 2018;4:1134-45.
[453] Häse F, Aldeghi M, Hickman RJ, Roch LM, Aspuru-Guzik A. Gryffin: An algorithm for Bayesian optimization of categorical variables informed by expert knowledge. Appl Phys Rev 2021;8:031406.
[454] Roch LM, Häse F, Kreisbeck C, Tamayo-Mendoza T, Yunker LPE, Hein JE, et al. ChemOS: Orchestrating autonomous experimentation. Sci Robot 2018;3:eaat5559.
[455] Roch LM, Häse F, Kreisbeck C, Tamayo-Mendoza T, Yunker LP, Hein JE, et al. ChemOS: An orchestration software to democratize autonomous discovery. PLoS One 2020;15:e0229862.
[456] Naranjani Y, Hernández C, Xiong F-R, Schütze O, Sun J-Q. A Hybrid Algorithm for the Simple Cell Mapping Method in Multi-objective Optimization. In: EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation IV, Heidelberg; 2013. p. 207-23.
[457] Helmdach D, Yaseneva P, Heer PK, Schweidtmann AM, Lapkin AA. A Multiobjective Optimization Including Results of Life Cycle Assessment in Developing Biorenewables-Based Processes. ChemSusChem 2017;10:3632-43.
[458] Amar Y, Schweidtmann Artur M, Deutsch P, Cao L, Lapkin A. Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis. Chem Sci 2019;10:6697-706.
[459] Knox ST, Parkinson SJ, Wilding CYP, Bourne RA, Warren NJ. Autonomous polymer synthesis delivered by multi-objective closed-loop optimisation. Polymer Chemistry 2022;13:1576-85.
[460] Daulton S, Balandat M, Bakshy E. Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization. Advances in Neural Information Processing Systems 2020;33:9851-64.
[461] Huyer W, Neumaier A. SNOBFIT -- Stable Noisy Optimization by Branch and Fit. ACM Trans Math Softw 2008;35:1-25.

[462] Whitley D. An overview of evolutionary algorithms: practical issues and common pitfalls. Inf Softw Technol 2001;43:817-31.
[463] Le TC, Winkler DA. Discovery and Optimization of Materials Using Evolutionary Approaches. Chem Rev 2016;116:6107-32.
[464] Whitley D. A genetic algorithm tutorial. Statistics and Computing 1994;4:65-85.
[465] Mirjalili S. Genetic Algorithm. Evolutionary Algorithms and Neural Networks: Theory and Applications. Cham: Springer International Publishing; 2019. p. 43-55.
[466] Katoch S, Chauhan SS, Kumar V. A review on genetic algorithm: past, present, and future. Multimedia Tools and Applications 2021;80:8091-126.
[467] Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks; 1995. p. 1942-8.
[468] Venter G, Sobieszczanski-Sobieski J. Particle Swarm Optimization. AIAA Journal 2003;41:1583-9.
[469] Poli R, Kennedy J, Blackwell T. Particle swarm optimization. Swarm Intell 2007;1:33-57.
[470] Blum C. Ant colony optimization: Introduction and recent trends. Phys Life Rev 2005;2:353-73.
[471] Dorigo M, Birattari M, Stutzle T. Ant colony optimization. IEEE Computational Intelligence Magazine 2006;1:28-39.
[472] Cao YJ, Wu QH. Evolutionary programming. In: Proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC '97); 1997. p. 443-6.
[473] Fogel DB, Fogel LJ. An introduction to evolutionary programming. In: Artificial Evolution, Berlin, Heidelberg; 1996. p. 21-33.
[474] Sinha N, Chakrabarti R, Chattopadhyay PK. Evolutionary programming techniques for economic load dispatch. IEEE Trans Evol Comput 2003;7:83-94.
[475] Sattari K, Xie Y, Lin J. Data-driven algorithms for inverse design of polymers. Soft Matter 2021;17:7607-22.
[476] Kim C, Batra R, Chen L, Tran H, Ramprasad R. Polymer design using genetic algorithm and machine learning. Comput Mater Sci 2021;186:110067.
[477] O'Boyle NM, Campbell CM, Hutchison GR. Computational design and selection of optimal organic photovoltaic materials. J Phys Chem C 2011;115:16200-10.
[478] Mannodi-Kanakkithodi A, Pilania G, Huan TD, Lookman T, Ramprasad R. Machine Learning Strategy for Accelerated Design of Polymer Dielectrics. Sci Rep 2016;6:20952.
[479] Chung YG, Gómez-Gualdrón DA, Li P, Leperi KT, Deria P, Zhang H, et al. In silico discovery of metal-organic frameworks for precombustion CO_2 capture using a genetic algorithm. Sci Adv 2016;2:e1600909.
[480] Huyer W, Neumaier A. SNOBFIT -- Stable Noisy Optimization by Branch and Fit. ACM Trans Math Softw 2008;35:9.
[481] Baranes A, Oudeyer P-Y. Active learning of inverse models with intrinsically motivated goal exploration in robots. Robot Auton Syst 2013;61:49-73.
[482] Ren Z, Oviedo F, Thway M, Tian SIP, Wang Y, Xue H, et al. Embedding physics domain knowledge into a Bayesian network enables layer-by-layer process innovation for photovoltaics. npj Comput Mater 2020;6:9.
[483] Sun S, Tiihonen A, Oviedo F, Liu Z, Thapa J, Zhao Y, et al. A data fusion approach to optimize compositional stability of halide perovskites. Matter 2021;4:1305-22.
[484] Liu Z, Rolston N, Flick AC, Colburn TW, Ren Z, Dauskardt RH, et al. Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell

manufacturing. Joule 2022;6:834-49.
[485] Ziatdinov MA, Liu Y, Morozovska AN, Eliseev EA, Zhang X, Takeuchi I, et al. Hypothesis Learning in Automated Experiment: Application to Combinatorial Materials Libraries. Adv Mater;n/a:2201345.
[486] Ziatdinov MA, Ghosh A, Kalinin SV. Physics makes the difference: Bayesian optimization and active learning via augmented Gaussian process. Mach Learn Sci Technol 2022;3:015003.
[487] McDannald A, Frontzek M, Savici AT, Doucet M, Rodriguez EE, Meuse K, et al. On-the-fly autonomous control of neutron diffraction via physics-informed Bayesian active learning. Appl Phys Rev 2022;9:021408.
[488] Saar L, Liang H, Wang A, McDannald A, Rodriguez E, Takeuchi I, et al. A Low-Cost Robot Science Kit for Education with Symbolic Regression for Hypothesis Discovery and Validation. 2022. p. arXiv:2204.04187.
[489] Wang A, Liang H, McDannald A, Takeuchi I, Gilad Kusne A. Benchmarking Active Learning Strategies for Materials Optimization and Discovery. 2022. p. arXiv:2204.05838.
[490] Gilad Kusne A, McDannald A, DeCost B, Oses C, Toher C, Curtarolo S, et al. Physics in the Machine: Integrating Physical Knowledge in Autonomous Phase-Mapping. 2021. p. arXiv:2111.07478.
[491] Ghiringhelli LM, Vybiral J, Levchenko SV, Draxl C, Scheffler M. Big Data of Materials Science: Critical Role of the Descriptor. Phys Rev Lett 2015;114:105503.
[492] Ghiringhelli LM, Vybiral J, Ahmetcik E, Ouyang R, Levchenko SV, Draxl C, et al. Learning physical descriptors for materials science by compressed sensing. New Journal of Physics 2017;19:023017.
[493] Ouyang R, Curtarolo S, Ahmetcik E, Scheffler M, Ghiringhelli LM. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. Phys Rev Mater 2018;2:083802.
[494] Ouyang R, Ahmetcik E, Carbogno C, Scheffler M, Ghiringhelli LM. Simultaneous learning of several materials properties from incomplete databases with multi-task SISSO. J Phys Mater 2019;2:024002.
[495] Schmidt M, Lipson H. Distilling Free-Form Natural Laws from Experimental Data. Science 2009;324:81-5.
[496] von Rueden L, Mayer S, Beckh K, Georgiev B, Giesselbach S, Heese R, et al. Informed Machine Learning -- A Taxonomy and Survey of Integrating Knowledge into Learning Systems. 2019. p. arXiv:1903.12394.
[497] Ziatdinov MA, Liu Y, Morozovska AN, Eliseev EA, Zhang X, Takeuchi I, et al. Hypothesis Learning in Automated Experiment: Application to Combinatorial Materials Libraries. Adv Mater 2022;n/a:2201345.
[498] Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological) 1996;58:267-88.
[499] Tropp JA, Gilbert AC. Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. IEEE Trans Inf Theory 2007;53:4655-66.
[500] Agrawal A, Choudhary A. Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science. APL Mater 2016;4:053208.
[501] Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature 2018;559:547-55.
[502] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 2011;40:D1100-D7.

[503] Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res 2009;37:W623-W33.
[504] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018.
[505] Computerized systems used in clinical investigations. 2007. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computerized-systems-used-clinical-investigations.
[506] Baker M. 1,500 scientists lift the lid on reproducibility. Nature 2016;533:452-4.
[507] Li J, Liu L, Le TD, Liu J. Accurate data-driven prediction does not mean high reproducibility. Nat Mach Intell 2020;2:13-5.
[508] Kwok R. How to pick an electronic laboratory notebook. Nature 2018;560:269-71.
[509] Van Dyke AR, Smith-Carpenter J. Bring Your Own Device: A Digital Notebook for Undergraduate Biochemistry Laboratory Using a Free, Cross-Platform Application. J Chem Educ 2017;94:656-61.
[510] Guerrero S, López-Cortés A, García-Cárdenas JM, Saa P, Indacochea A, Armendáriz-Castillo I, et al. A quick guide for using Microsoft OneNote as an electronic laboratory notebook. PLoS Computational Biology 2019;15:e1006918.
[511] Bromfield Lee D. Implementation and Student Perceptions on Google Docs as an Electronic Laboratory Notebook in Organic Chemistry. J Chem Educ 2018;95:1102-11.
[512] Landhuis E. Scientific literature: Information overload. Nature 2016;535:457-8.
[513] Hong Z, Ward L, Chard K, Blaiszik B, Foster I. Challenges and Advances in Information Extraction from Scientific Literature: a Review. JOM 2021;73:3383-400.
[514] Olivetti EA, Cole JM, Kim E, Kononova O, Ceder G, Han TY-J, et al. Data-driven materials research enabled by natural language processing and information extraction. Appl Phys Rev 2020;7:041317.
[515] Kononova O, He T, Huo H, Trewartha A, Olivetti EA, Ceder G. Opportunities and challenges of text mining in materials research. iScience 2021;24:102155.
[516] Swain MC, Cole JM. ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature. J Chem Inf Model 2016;56:1894-904.
[517] Court CJ, Cole JM. Auto-generated materials database of Curie and Néel temperatures via semi-supervised relationship extraction. Sci Data 2018;5:180111.
[518] Huang S, Cole JM. A database of battery materials auto-generated using ChemDataExtractor. Sci Data 2020;7:260.
[519] Court CJ, Cole JM. Magnetic and superconducting phase diagrams and transition temperatures predicted using text mining and machine learning. npj Comput Mater 2020;6:18.
[520] Beard EJ, Cole JM. ChemSchematicResolver: A Toolkit to Decode 2D Chemical Diagrams with Labels and R-Groups into Annotated Chemical Named Entities. J Chem Inf Model 2020;60:2059-72.
[521] Mavračić J, Court CJ, Isazawa T, Elliott SR, Cole JM. ChemDataExtractor 2.0: Autopopulated Ontologies for Materials Science. J Chem Inf Model 2021;61:4280-9.
[522] Court CJ, Jain A, Cole JM. Inverse Design of Materials That Exhibit the Magnetocaloric Effect by Text-Mining of the Scientific Literature and Generative Deep Learning. Chem Mater 2021;33:7217-31.
[523] Kim E, Huang K, Saunders A, McCallum A, Ceder G, Olivetti E. Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. Chem Mater 2017;29:9436-44.
[524] Kim E, Huang K, Jegelka S, Olivetti E. Virtual screening of inorganic materials synthesis

parameters with deep learning. npj Comput Mater 2017;3:53.
[525] Jensen Z, Kim E, Kwon S, Gani TZH, Román-Leshkov Y, Moliner M, et al. A Machine Learning Approach to Zeolite Synthesis Enabled by Automatic Literature Data Extraction. ACS Cent Sci 2019;5:892-9.
[526] Kim E, Jensen Z, van Grootel A, Huang K, Staib M, Mysore S, et al. Inorganic Materials Synthesis Planning with Literature-Trained Neural Networks. J Chem Inf Model 2020;60:1194-201.
[527] He T, Sun W, Huo H, Kononova O, Rong Z, Tshitoyan V, et al. Similarity of Precursors in Solid-State Synthesis as Text-Mined from Scientific Literature. Chem Mater 2020;32:7861-73.
[528] Wang Z, Kononova O, Cruse K, He T, Huo H, Fei Y, et al. Dataset of Solution-based Inorganic Materials Synthesis Recipes Extracted from the Scientific Literature. 2021. p. arXiv:2111.10874.
[529] Wang Z, Cruse K, Fei Y, Chia A, Zeng Y, Huo H, et al. ULSA: unified language of synthesis actions for the representation of inorganic synthesis protocols. Digital Discovery 2022.
[530] Cruse K, Trewartha A, Lee S, Wang Z, Huo H, He T, et al. Text-mined dataset of gold nanoparticle synthesis procedures, morphologies, and size entities. 2022. p. arXiv:2204.10379.
[531] Subramanian A, Cruse K, Trewartha A, Wang X, Alivisatos AP, Ceder G. Dataset of gold nanoparticle sizes and morphologies extracted from literature-mined microscopy images. 2021. p. arXiv:2112.01689.
[532] Weston L, Tshitoyan V, Dagdelen J, Kononova O, Trewartha A, Persson KA, et al. Named Entity Recognition and Normalization Applied to Large-Scale Information Extraction from the Materials Science Literature. J Chem Inf Model 2019;59:3692-702.
[533] Mehr SHM, Craven M, Leonov AI, Keenan G, Cronin L. A universal system for digitization and automatic execution of the chemical synthesis literature. Science 2020;370:101-8.
[534] Wilbraham L, Mehr SHM, Cronin L. Digitizing Chemistry Using the Chemical Processing Unit: From Synthesis to Discovery. Acc Chem Res 2021;54:253-62.
[535] Dai Z, Yang Z, Yang Y, Carbonell J, Le QV, Salakhutdinov R. Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context. 2019. p. arXiv:1901.02860.
[536] Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In; 2019
[537] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. 2019. p. arXiv:1907.11692.
[538] Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. 2018.
[539] Drew KL, Baiman H, Khwaounjoo P, Yu B, Reynisson J. Size estimation of chemical space: how big is it? J Pharm Pharmacol 2012;64:490-5.
[540] Kusner MJ, Paige B, Hernández-Lobato JM. Grammar variational autoencoder. In: Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia; 2017. p. 1945-54.
[541] Lim J, Ryu S, Kim JW, Kim WY. Molecular generative model based on conditional variational autoencoder for de novo molecular design. J Cheminformatics 2018;10:1-9.
[542] Dong Y, Li D, Zhang C, Wu C, Wang H, Xin M, et al. Inverse design of two-dimensional graphene/h-BN hybrids by a regressional and conditional GAN. Carbon 2020;169:9-16.
[543] Nouira A, Sokolovska N, Crivello J-C. CrystalGAN: Learning to Discover Crystallographic Structures with Generative Adversarial Networks. 2018. p. arXiv:1810.11203.
[544] You J, Liu B, Ying Z, Pande V, Leskovec J. Graph Convolutional Policy Network for Goal-

Directed Molecular Graph Generation. In: Advances in Neural Information Processing Systems; 2018
[545] Lima Guimaraes G, Sanchez-Lengeling B, Outeiral C, Cunha Farias PL, Aspuru-Guzik A. Objective-Reinforced Generative Adversarial Networks (ORGAN) for Sequence Generation Models. 2017. p. arXiv:1705.10843.
[546] Yao Z, Sánchez-Lengeling B, Bobbitt NS, Bucior BJ, Kumar SGH, Collins SP, et al. Inverse design of nanoporous crystalline reticular materials with deep generative models. Nat Mach Intell 2021;3:76-86.
[547] M. Dieb T, Ju S, Yoshizoe K, Hou Z, Shiomi J, Tsuda K. MDTS: automatic complex materials design using Monte Carlo tree search. Sci Technol Adv Mater 2017;18:498-503.
[548] Dieb TM, Ju S, Shiomi J, Tsuda K. Monte Carlo tree search for materials design and discovery. MRS Commun 2019;9:532-6.
[549] Shin K, Tran DP, Takemura K, Kitao A, Terayama K, Tsuda K. Enhancing Biomolecular Sampling with Reinforcement Learning: A Tree Search Molecular Dynamics Simulation Method. ACS Omega 2019;4:13853-62.
[550] Patra TK, Loeffler TD, Sankaranarayanan SKRS. Accelerating copolymer inverse design using monte carlo tree search. Nanoscale 2020;12:23653-62.
[551] Lipton ZC. The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue 2018;16:31-57.
[552] Oviedo F, Lavista Ferres J, Buonassisi T, Butler K. Interpretable and Explainable Machine Learning for Materials Science and Chemistry. 2021. p. arXiv:2111.01037.
[553] Molnar C. Interpretable Machine Learning: Lulu.com; 2020.
[554] Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, methods, and applications in interpretable machine learning. PNAS 2019;116:22071-80.
[555] Freitas AA. Comprehensible classification models: a position paper. ACM SIGKDD explorations newsletter 2014;15:1-10.
[556] Quinlan JR. Simplifying decision trees. Int J Man Mach Stud 1987;27:221-34.
[557] Zhang Q, Wu YN, Zhu S-C. Interpretable Convolutional Neural Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018. p. 8827-36.
[558] Vandewiele G, Janssens O, Ongenae F, De Turck F, Van Hoecke S. GENESIM: genetic extraction of a single, interpretable model. 2016. p. arXiv:1611.05722.
[559] Bastani O, Kim C, Bastani H. Interpretability via Model Extraction. 2017. p. arXiv:1706.09773.
[560] Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: A corrected feature importance measure. Bioinformatics 2010;26:1340-7.
[561] Qin T, Cook M, Courtney M. Exploring Chemistry with Wireless, PC-Less Portable Virtual Reality Laboratories. J Chem Educ 2021;98:521-9.
[562] Li J, Tu Y, Liu R, Lu Y, Zhu X. Toward “On-Demand” Materials Synthesis and Scientific Discovery through Intelligent Robots. Adv Sci 2020;7:1901957.
[563] Fitzpatrick DE, Battilocchio C, Ley SV. A Novel Internet-Based Reaction Monitoring, Control and Autonomous Self-Optimization Platform for Chemical Synthesis. Org Process Res Dev 2016;20:386-94.
[564] Fitzpatrick DE, Maujean T, Evans AC, Ley SV. Across-the-World Automated Optimization and Continuous-Flow Synthesis of Pharmaceutical Agents Operating Through a Cloud-Based Server. Angew Chem Int Ed 2018;57:15128-32.
[565] Grieves M. Digital twin: manufacturing excellence through virtual factory replication. White

paper 2014;1:1-7.
[566] Tao F, Zhang H, Liu A, Nee AYC. Digital Twin in Industry: State-of-the-Art. IEEE Trans Industr Inform 2019;15:2405-15.
[567] Saddik AE. Digital Twins: The Convergence of Multimedia Technologies. IEEE MultiMedia 2018;25:87-92.