Autonomous chemical science and engineering enabled by self-driving laboratories

Jeffrey A Bennett and Milad Abolhasani*

Recent advances in machine learning (ML) and artificial intelligence have provided an exciting opportunity to computerize the fundamental and applied studies of complex reaction systems *via* self-driving laboratories. Autonomous robotic experimentation can enable time-, material-, and resource-efficient exploration and/or optimization of high-dimensional space reaction systems. Furthermore, interpretation of the ML models trained on the experimental data can unveil the underlying reaction mechanisms. In this article, we discuss different elements of a self-driving lab, and present recent efforts in autonomous reaction modeling and optimization. Further development and adoption of ML-guided closed-loop experimentation strategies can realize the full potential of autonomous chemical science and engineering to accelerate the discovery and development of advanced materials and molecules.

Address
Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695, USA

Corresponding author: Milad Abolhasani (abolhasani@ncsu.edu)
* Webpage: www.abolhasanilab.com

Introduction

With the recent advancements of data science and rapidly growing power of machine learning (ML) and artificial intelligence (AI), there has been an increasing interest to create autonomous experimentation platforms by combining ML and AI modeling with automated experimentation and reaction-control systems as a time-efficient and material-efficient approach to augment reaction engineering and process optimization campaigns. Automated experimentation strictly follows a sequence of physical/chemical processes identified by the scientist; however, autonomous experimentation leverages autonomy (defined by the scientist) to make decisions about the sequence of physical/chemical processes in an iterative manner. Many strategies in chemical engineering utilize the ML component as a strict n-dimensional black-box optimization problem [1–5], with n corresponding to the number of controlled process parameters. While the technique can be powerful, the black-box approach is generally highly system specific, while neglecting mechanistic information that could be more widely applied. Black-box optimization methods normally have underlying assumptions about the behavior of the reaction surface being smooth, continuous, not chaotic, and not stiff on the order of the variation in input parameters.

While ML can be applied to a manually sampled system [6,7], it becomes much more powerful when combined with a fully automated system in a closed-loop fashion to autonomously perform the experiments selected based on the ML model and decision-making algorithm. Coupled with *in situ* or online reaction-characterization techniques, it becomes feasible to have a truly autonomous reaction system, where the ML algorithm can explore the accessible reaction universe, only limited by the physical parameters of the reaction system, collect data, and utilize the newly performed experiment(s) to inform future experiments. As long as the experimental conditions can be automatically varied and the data collection and processing can be automated, the closed-loop autonomous experimentation strategy can be realized for batch [6,8–12], flow [1–4,13–16], and hybrid systems [17].

ML strategies in chemical science and engineering can be divided into two categories: (i) data mining from literature and reaction databases and (ii) closed-loop intelligent experimentation. If a given reaction or class of reactions have abundant data present in the literature (e.g. organic synthesis), ML can be used to form predictions of reaction outcomes or perform retrosynthesis of target compounds. However, when targeting a specific class of reactions with a lack of experimental data with high batch-to-batch, reactor-to-reactor, and lab-to-lab variation, ML prediction becomes challenging. As access to data from literature becomes scarce, autonomous experimentation strategies become important to rapidly generate the minimum set of experimental data required to accurately model the synthesis-process-property relationship.

Apart from black-box optimization campaigns, there has been significant effort in using ML and physics-based approaches [18] to form a surrogate model (i.e. digital twin) of the reaction system; essentially, the ML model [19,20] (e.g. neural networks, NNs) or physics-based model [21–24] simulates the complete system behavior and then can be subsequently used for performing rapid mechanistic studies, without the need to run additional experiments. Once an accurate surrogate model is developed, the model is able to be probed for system behavior more rapidly and more cost effectively than physical experimentation. Then, any unique or interesting behaviors identified in the ML model can be validated with automated experimentation. Building the surrogate model of a reaction system involves a learning step where the ML model familiarizes itself with the reaction space (known as active learning) followed by cross-validation (dividing the data into training, testing, and validation sets). The active learning step searches the reaction space through techniques such as maximum variance or an initial experimental space-filling set (e.g. Latin hypercube). The ML model can additionally identify the Pareto front for a multiobjective optimization during closed-loop reaction-space exploration [25,26].

Other than individual reaction optimization and modeling, ML techniques have been utilized to perform predictions of reaction conditions for novel organic synthesis reactions, for example, by training with data from the Reaxys database [27]. The more generalized the ML predictions of reactions and process conditions are desired, the more experimental data for ML model training it requires, as the scope of possible parameters (i.e. design space of intrinsic and extrinsic parameters) becomes much larger when introducing discrete variables such as chemical species. Various data mining and *in silico* techniques (physics-based modeling and data augmentation [28]) can be used to supply the increased demand for data. It should be noted that in contrast to organic synthesis, the process-sensitive nature of nanomaterial syntheses limits the generalizability of the ML predictions beyond a specific class of materials.

Here, we present an overview of recent autonomous experimentation efforts in chemical science and engineering enabled by self-driving laboratories, and discuss different models and algorithms that have been implemented in the design of such systems. A self-driving laboratory in chemical science and engineering consists of four main components, which can be individually adjusted to suit the specific goals of the reaction campaign, either process exploitation or probing the reaction system for more fundamental understanding (Figure 1). The first component is reagent preparation and chemical reaction execution: the physical equipment for performing a given experimental condition and

extracting the information of interest from the result. This would commonly be an automated system being robust enough to perform experiments without any human supervision after being given a set of initial parameters. The second piece of the system is interpretation or data processing, the ability to extract and distill relevant information from the raw data supplied by the execution step. Accurate and precise data acquisition is essential to minimize system sampling variance to reduce the uncertainty in the model prediction, resulting in reducing time and cost for optimization campaigns. The third element of the system is building or updating the ML model (i.e. current belief of the reaction model). The specific model and architecture chosen can have a dramatic impact on the performance and training behavior of the ML model, depending on the types of input parameters. Commonly used ML models are Gaussian process regression (GPR) as well as deep NNs for relatively small dimensional spaces such as process conditions, and convolutional NNs for larger and more complex dimensional spaces such as chemical structure information [29–31]. Additionally, the NNs can be grouped to form an ensemble model to enable statistical and probabilistic evaluation of the reaction outputs [32•,33••]. The last component required for a complete autonomous experimentation system is the intelligent experiment-selection algorithm, or how the system will choose the next experiment in a closed-loop campaign. Selection can be exploitative, such as a strict parameter optimization, exploratory, as is the case for minimizing maximum model variance, or some combination of the two. The first two aspects of an autonomous experimentation system are highly dependent on the individual reaction system being studied, and as such, this review will be primarily focused on the modeling and decision policies.

Autonomous chemical science and engineering enabled by self-driving laboratories

Fully realized self-driving laboratories enable several different use cases ranging from reaction optimization to process-space mapping and investigation into more fundamental understanding of system mechanisms and kinetics. While systems with rapid sampling rates — fast reaction kinetics O (0.1–100 s) and rapid characterization (e.g. in-line analysis) — can move through the experimental selection process to more quickly find an optimum or explore reaction space, systems with slower reaction timescales, ~O(1 h), or characterization (offline analysis), can still benefit in acceleration from both the automation and the intelligent experimentation (ML-guided experimental planning). Relatively slow chemical processes can additionally be conducted in a single-droplet oscillatory flow format [34–36] to decouple reaction time from reactor volume to minimize

Figure 1
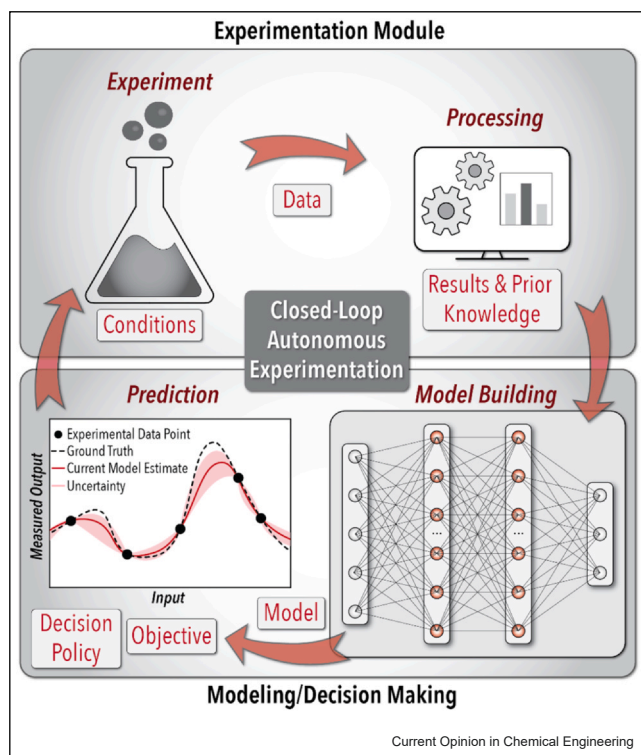
Process flow of a self-driving lab in chemical science and engineering.

reactor footprint for screening conditions. One way to help accelerate the automated reaction campaigns is to use *in silico* surrogate models (ML or physics-based) to simulate and supplement the system for more rapidly selecting optimization hyperparameters for systems with slow or costly sampling [37]. Bateni et al. [20] harnessed the power of this directed intelligent experimentation to construct a digital twin of a complex multistage material synthesis (Mn doping of $CsPbCl_3$ nanocrystals). A maximum variance decision policy for experiment selection returned a model with prediction accuracy > 80% after only 60 autonomously conducted experiments.

Closed-loop autonomous reaction optimization

Closed-loop autonomous reaction optimization is intended to effectively search a high-dimensional reaction parameter space to provide a prediction that gets as close to a user-specified objective as possible. Common objective parameters include maximum reaction yield, minimum experimental cost, desired product selectivity, and relevant material properties in the case of materials development. ML and AI strategies for closed-loop optimization generally involve predicting the response of the system and selecting an improved set of reaction conditions sequentially, until reaching a predefined target. Reaction optimization can also be performed with no model of the system as is the case for a strictly

evolutionary method (i.e. a population-based meta-heuristic optimization algorithm); however, forming an ML model of the system (e.g. response surface based or NN), can help to reduce the number of experiments (cost) that must be performed in closed loop to identify an optimum. As the dimensionality of the input space increases, linear models do not work as well and NN approaches begin to shine. In this section, we present a few recent examples of closed-loop autonomous reaction optimization.

In one example of an autonomous robotic experimentation technique, Li et al. [38] identified and optimized for chirality in inorganic perovskite nanocrystals using an automated microfluidic reactor and robotic sampling system coupled with a cloud-based stable noisy optimization by branch-and-fit (SNOBFIT) [39] strategy, combining random global search and steepest-descent gradient methods. Circular dichroism of inorganic perovskite nanocrystals was optimized by varying reaction temperature and precursor concentration, and perovskites with screw dislocations were obtained. The optimum in the two-dimensional input space was identified after ~120 experiments and was not improved upon after 250 experiments. The SNOBFIT algorithm allows the optimization of an arbitrary function through local gradient quadratic fitting, without a need to have access to the full underlying model.

Evolutionary approaches can also be utilized with little to no information about the underlying system at the cost of potentially requiring additional experiments to identify an optimum. The evolutionary algorithms use the parameters from the best candidates of the previous generation of experimental conditions before applying crossover between candidates and random mutation to create the conditions for the next set of experiments to be conducted autonomously. In one example, Salley et al. [8] utilized a sequential evolutionary algorithm campaign to synthesize gold (Au) nanoparticles on demand with different morphologies (spheres, rods, and octahedral), as shown in Figure 2a. The initial population generation was produced by random sampling of the initial parameter space. Once optimized conditions of a morphology were identified, those conditions were used as seed material for the subsequent material optimization campaigns. Each optimization campaign required 10 generations of 15 experiments (90 min of sampling time) for 150 total experiments per optimization. Evolutionary approaches are primarily feasible in systems with a high degree of parallelization or rapid sampling due to the potentially large number of required experiments.

Reinforcement learning (RL) has not been yet widely adapted in autonomous chemical science and engineering. RL aims to maximize a user-supplied reward function using the state of the system in addition to the

Figure 2
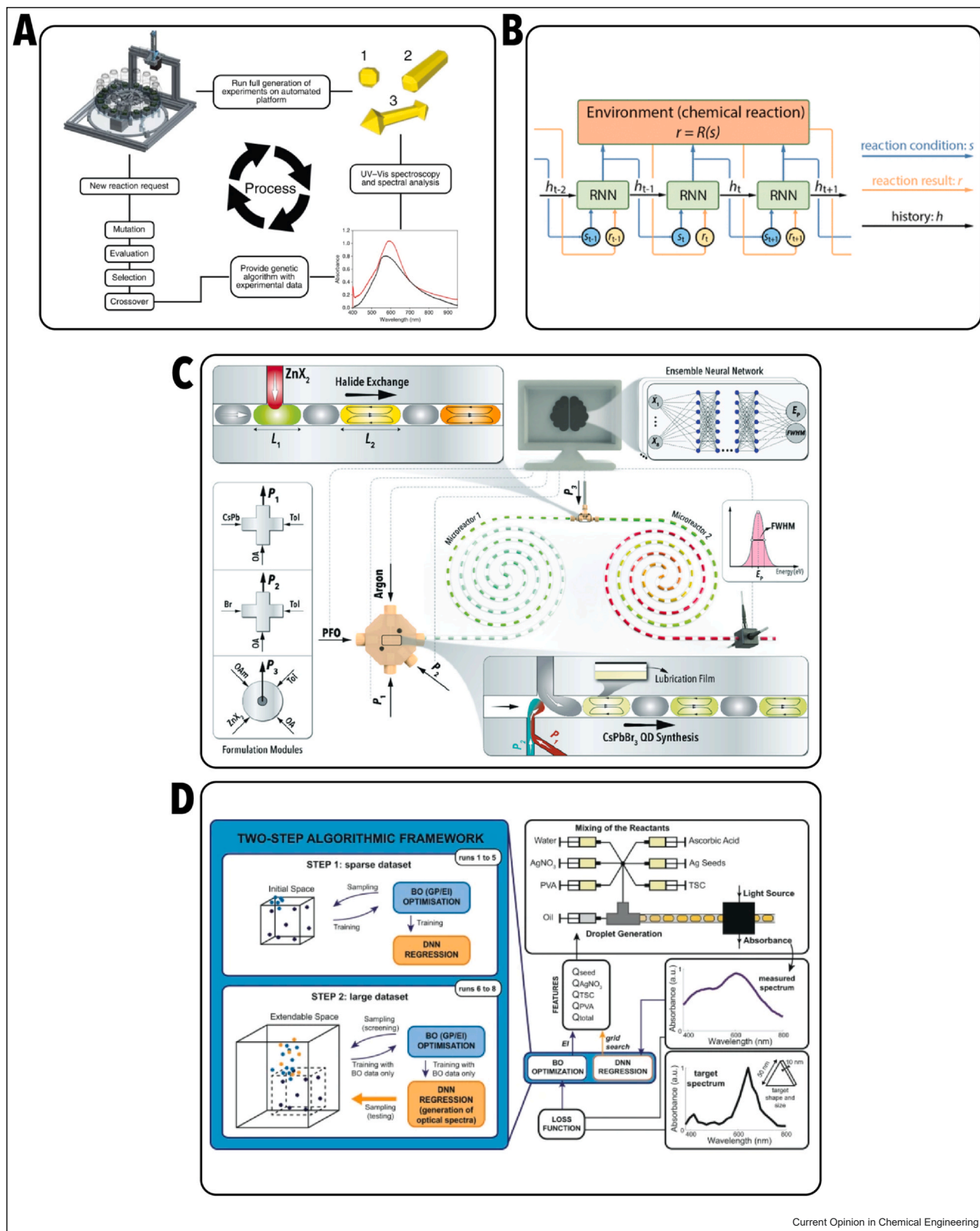
Example algorithms for chemical reaction optimization. **(a)** Optimization flow of an evolutionary algorithm for producing Au nanoparticles by Salley et al. [8]. **(b)** Training of a reinforcement-learning model for reaction optimization by Zhou et al. [40••]. **(c)** Schematic of the modular experimentation platform integrated with an ensemble NN-guided ML algorithm for two-stage reaction optimization by Abdel-Latif et al. [32•]. **(d)** Two-step Bayesian optimization and deep NN absorbance spectra prediction model by Mekki-Berrada et al. [43•].
Images reproduced with permissions from Springer Nature, 2020 **(a)**, American Chemical Society, 2017 **(b)**, Creative Commons License, 2021 **(c)**, and Springer Nature, 2021 **(d)**.

reaction conditions, results, and previous history to predict reaction outcomes. RL uses deep NNs to fit the policy function of the RL algorithm for condition prediction, as shown in Figure 2b. In one of the few examples of RL in reaction engineering, Zhou et al. [40••] utilized a deep RL for the yield optimization of the Pomeranz–Fritsch synthesis of isoquinoline, Friedlaender synthesis of substituted quinoline, synthesis of ribose phosphate, the reaction of 2,6-dichlorophenolindophenol and ascorbic acid, and silver nanoparticle synthesis. The performance of the RL optimization strategy was benchmarked against common black-box optimization methods, such as covariance matrix adaptation– evolution strategy (CMA–ES) [41], Nelder–Mead simplex [42], and SNOBFIT. The RL strategy succeeded in reducing the optimization cost (i.e. total number of experiments required for each optimization) by 71% (32 experiments vs. 111, 187 and no convergence for CMA–ES, SNOBFIT, and Nelder–Mead, respectively). It was also demonstrated that applying a probability distribution for experiment selection around the predicted values results in an improved final function value at the cost of a slightly longer optimization campaign. Additionally, the RL strategy showed improved performance after training on prior reactions, regardless of similar or different reaction mechanisms.

Bayesian optimization (BO) uses prior belief about a system coupled with experimental conditions to predict an updated ground-truth model and associated uncertainty. The belief about the system is commonly modeled by a GPR, but NNs can also be utilized as long as they can provide an estimate of uncertainty either directly or through an ensemble of NNs. One example of a self-driving laboratory using BO-based approach with ensemble NNs by Epps et al. [33••], compared the BO approach to SNOBFIT and CMA–ES for the bandgap tuning of inorganic lead halide perovskite nanocrystals. Decision policies that were either exploration-heavy (SNOBFIT, NN — maximum variance) or exploitation-heavy (NN — exploitation) optimized more slowly and to worse final values than the balanced policies (NN— upper confidence bound, NN — expected improvement, and CMA–ES). Pretraining of the NN ensemble with prior experimental conditions (knowledge transfer) dramatically improved the optimization time for reaching the bandgap within 1 meV under five experiments. The benefits of pretraining diminished near the edges of the trained output space. A second example by Abdel-Latif et al. [32•] used a pretraining stage (NN — maximum variance) of 200 experiments on the halide-exchange reaction of lead halide perovskite nanocrystals, which allowed the ML model to quickly optimize to a selected peak emission energy in five further exploitation experiments (NN — exploitation) per material target from 1.9 to 2.9 eV, shown in Figure 2c. The pretraining becomes more time and resource efficient as multiple optimization campaigns are required. This study further highlights the power of ML models coupled with a closed-loop experimentation strategy to tackle the batch-to-batch variation problem, commonly observed in colloidal synthesis of nanomaterials. In another example of autonomous experimentation using BO, Tao et al. [36•] utilized a Bayesian NN to direct the optimization of a gold nanoparticle synthesis by balancing selection of an experiment for maximum reaction-space exploration followed by an experiment selected for maximum exploitation based on the current belief of the reaction system. The optimization algorithm used in this study, *Chimera*, allowed for optimizing multiple desired properties, moving down the list in order of importance, without sacrificing higher-ranked properties. In another example of BO in autonomous reaction optimization, Mekki-Berrada et al. [43•] combined a Gaussian process-based BO with a deep NN to optimize a droplet-based microfluidic synthesis of silver nanoparticles. The BO algorithm was provided with a loss function by comparing the nanoparticle UV–Vis absorption spectra with a target spectra, while the deep NN was directly fed the full spectra, as shown in Figure 2d. This approach allowed the deep NN to produce simulated UV–Vis absorption spectra for the BO-suggested sampling conditions, allowing for ranking the predicted conditions. After full training of the ML model, the effect of the reaction conditions on the features of the UV–Vis absorption spectra could be investigated. The optimization campaign required 8 sets of 15 experimental reaction conditions and the parameter-space boundary was allowed to be flexible, enabling the expansion of the available reaction space if the predicted optimum was located too close to the edge of the parameter space to be fully explored.

While a variety of different ML modeling and AI-assisted experiment-selection strategies can be utilized for a self-driving lab, the resource-efficient and time-efficient nature of the closed-loop experimental campaigns mainly relies on the characteristics of the system (e.g. reaction time, sampling time, and computing cost). A quickly sampled or low experimental cost system may opt for an approach that requires more experimental data points such as black-box or genetic algorithms, while systems with slow or costly sampling may tend to utilize methods that can handle more sparse datasets. We direct the reader to recent comprehensive reviews further detailing the multitude types of ML algorithms for organic [44,45], and nanomaterial [46] synthesis.

Autonomous reaction modeling

Although ML models have significant power for performing optimizations on chemical reactions, their true benefit lies in the development of the surrogate model of the system. After the training and cross-validation of the ML model with the initial learning dataset, it becomes possible to use the surrogate model to explore the underlying behavior of the system and extract fundamental insight into reaction kinetics and mechanism. Parameter analysis can identify the most critical parameter(s) affecting reaction performance. For example, careful selection of the NN model architecture can also enable extraction of relevant reaction information from the weights and biases of the final trained NN model. Design of experiment models can also be trained and fit to the data if some mechanistic information is known.

In one example of data-driven reaction modeling, Ji et al. [47•] developed a chemical reaction neural network (CRNN) with node architecture based on fundamental rate equations and the Arrhenius law for temperature dependence, as shown in Figure 3a. Through training the CRNN model with time-resolved concentration data, it became possible to extract fundamental reaction pathways and rate constants from the weights and biases of the trained model. Overall, the CRNN modeling approach is very powerful for identifying elementary reaction steps, assuming time-resolved concentration profiles for reagents, intermediates, and products are available. The large amount of data required for this autonomous reaction modeling approach may limit full applicability for reactions with catalytic cycles and reactive intermediate species.

In another example of ML-guided reaction engineering, Vikram et al. [48] developed an ML model for colloidal synthesis of indium phosphide (InP) nanocrystals, using a 25-member ensemble NN with randomized internal architecture of hidden layers and nodes. The ensemble NN was trained and cross-validated starting with an initial 16 random experiments until it could successfully predict a set of additional InP synthesis experiments

with high accuracy (> 90%), as shown in Figure 3b. The model then predicted a parameter-space map and evaluated the feasibility of the target parameters. If the model believed the desired target was feasible, it would then run the experiment at the optimized condition selected based on the ensemble NN model to achieve the desired properties. Once the ensemble NN produces the parameter-space map, it can be used as a surrogate model to effectively examine the effect of the input parameters on the system.

Epps et al. [26••] utilized a GPR-based surrogate model (or system ground truth) for a complex reaction system (halide-exchange reactions of all-inorganic metal halide perovskite nanocrystals) to investigate the role of ML modeling and decision-making under uncertainty on the closed-loop reaction optimization. With the trained surrogate model, simulated experiments can be performed rapidly, enabling the full exploration of the parameter space predicted by the model and the discovery of the Pareto front of the multiobjective optimization. The digital twin of the nanocrystal synthesizer was then utilized to identify and improve metaparameters, such as objective function, NN architecture, and decision policies that resulted in improved autonomous optimization performance (Figure 3c). The optimized ensemble NN model outperformed the previous work and several commonly selected optimization algorithms (SNOBFIT, CMA–ES, NSGA-II [49], and exploration). Benchmarking various choices of metaparameters can have a significant impact on both the speed and the final value achieved for the closed-loop reaction optimization campaigns.

Interpreting machine learning-guided reaction modeling

ML models and performance can be difficult to interpret, and thus, there has been a significant amount of work around visualizing the outputs and interpreting a trained ML model. In the case of the CRNN in Ji et al. [47•], the NN architecture was designed to produce rate constants and power-law coefficients as the weights and biases of the fully trained NN model, however, for many NN models, it is difficult to extract physical meaning from the weights and biases of the hidden layers of the model. For some NN models, such as the one presented in Gao et al. for the prediction of organic synthesis reaction conditions [27,29,31], similarities in the weight matrices can be interpreted as similarities in the output-activation nodes for solvent similarity and reagent similarity. The high-dimensional matrices can also be projected to a two-dimensional visualization by t-distributed stochastic neighbor embedding, to show clustering of similar species, as shown in Figure 4a.

ML-guided optimization campaigns are typically assessed by the speed at which the model can converge to

Figure 3
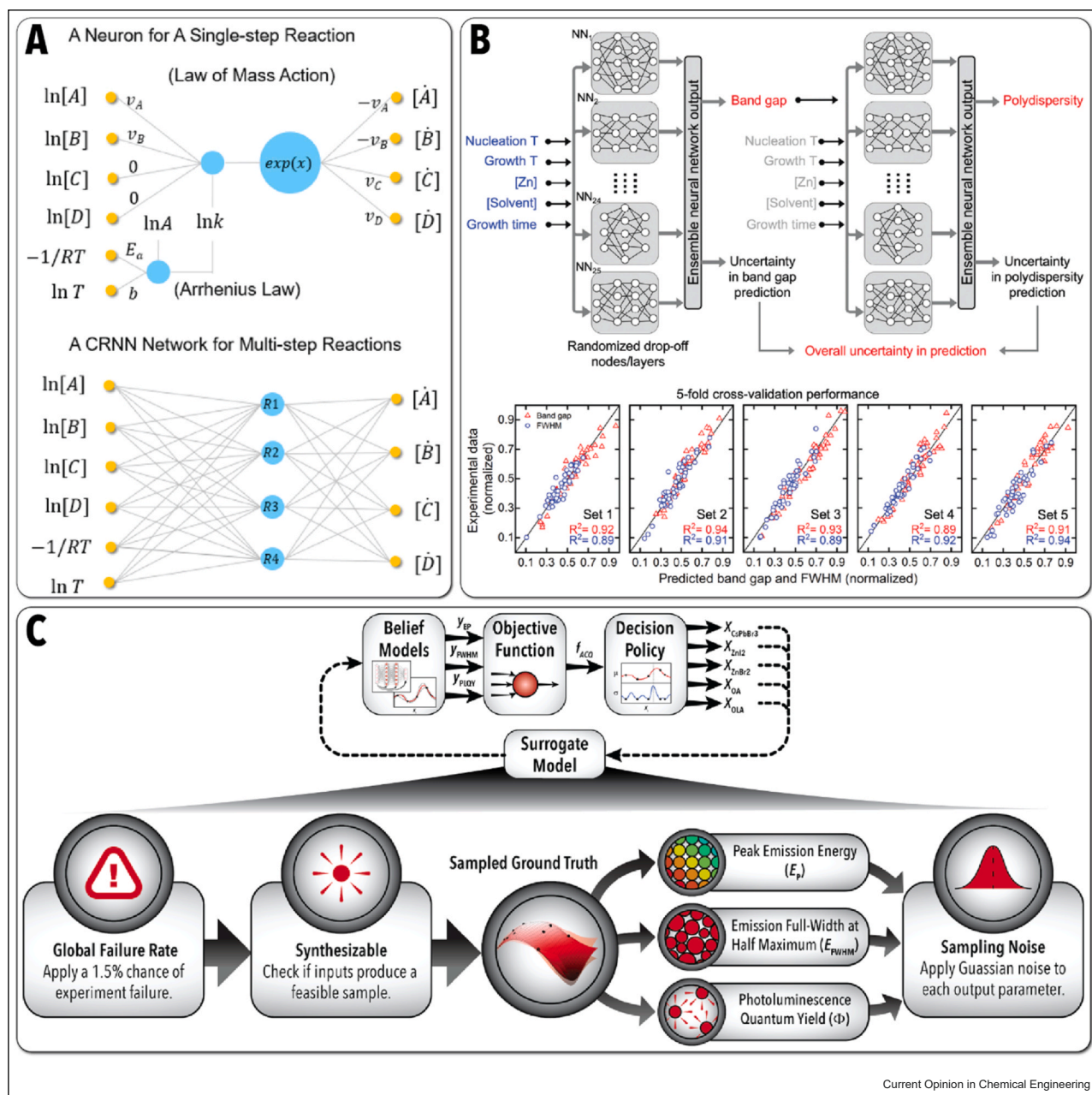
Example surrogate model architectures. **(a)** Chemical reaction NN with weights and biases corresponding to mass action and Arrhenius behavior of elementary reactions in the reaction network by Ji et al. [47•]. **(b)** Ensemble NN and fivefold cross-validation for modeling InP nanocrystal synthesis in a flow reactor by Vikram et al. [48]. **(c)** Surrogate model development for model and decision-policy optimization by Epps et al. [26••]. Images reproduced with permissions from American Chemical Society, 2021 **(a)**, Royal Society of Chemistry, 2021 **(b)**, and Royal Society of Chemistry, 2021 **(c)**.

an optimum in the fitness function (i.e. total experimental cost), as well as the final value that can be obtained. This metric shows an estimate of how many experimental conditions are required for performing an optimization campaign using a specific ML model on a similar system (i.e. benchmarking), as shown by Zhou et al. [40••] (Figure 4b). Similarly, the fitness function or individual outputs can be plotted as a function of sets of input parameters to visualize lower-dimensional slices of the reaction space as response surfaces to better understand the role of different reaction parameters on the reaction outcome (Figure 4c, Li et al. [38]). The

Figure 4
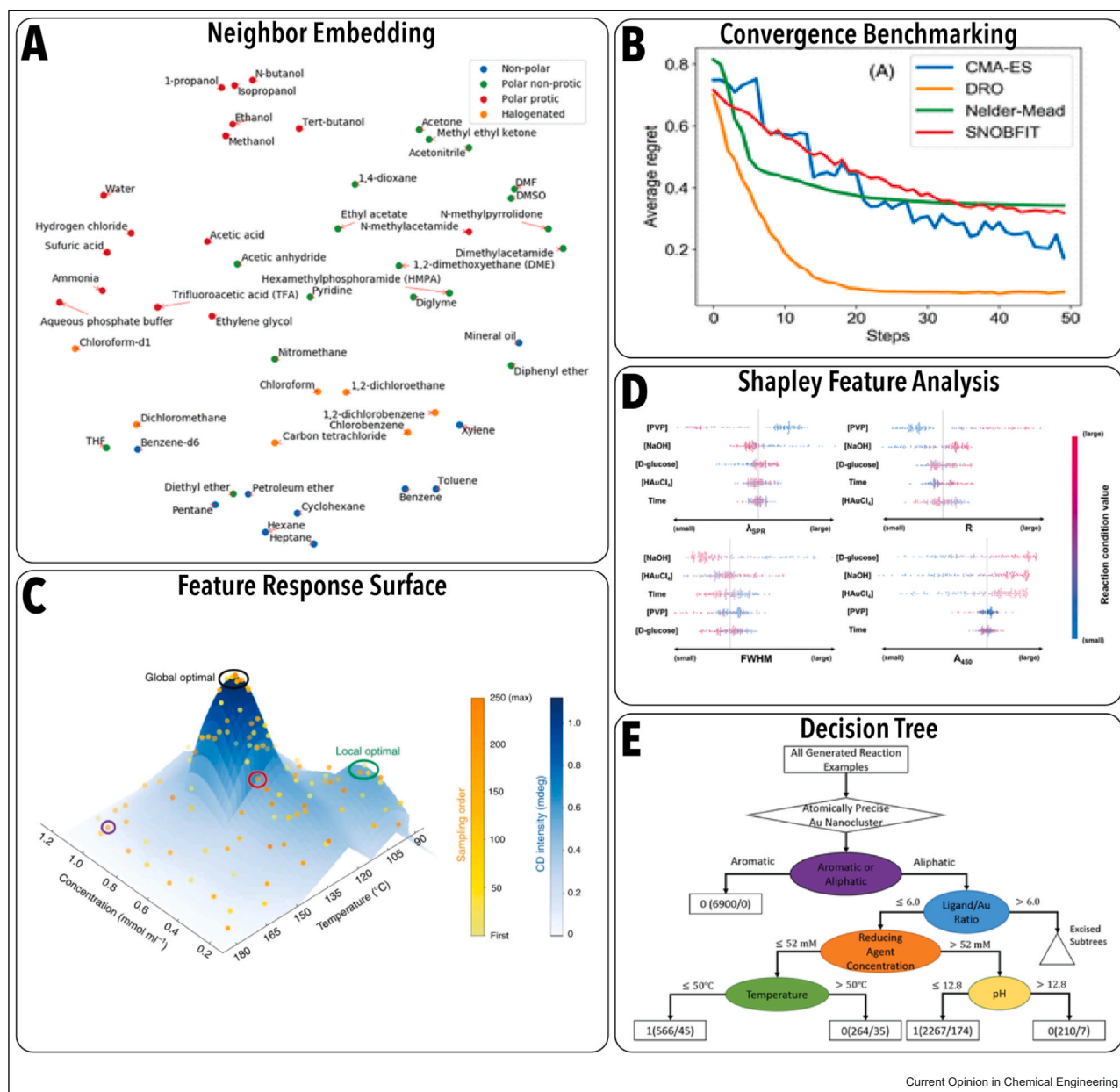
Fundamental knowledge extraction from machine learning models. **(a)** Two-dimensional visualization by t-distributed stochastic neighbor embedding, to visualize clustering of similar chemical species by Gao et al. [27]. **(b)** Model fitness benchmarking by Zhou et al. [40••]. **(c)** Concentration and temperature-response surface for circular dichroism of chiral perovskite nanocrystals by Li et al. [38]. **(d)** Shapely feature analysis of reaction input parameters for Au nanoparticle synthesis by Tao et al. [36•]. **(e)** Atomically specific Au nanoparticle decision tree by Li et al. [50].
Images reproduced with permissions from American Chemical Society, 2018 **(a)**, American Chemical Society, 2017 **(b)**, Creative Commons License, 2020 **(c)**, Wiley, 2021 **(d)**, and Wiley, 2019 **(e)**.

importance of individual input parameters can be determined by a relative contribution or Shapley feature analysis to either highlight important input parameters or remove inputs with little to no effect on the output (dimensionality reduction), as shown in the work by Tao et al. [36•] (Figure 4d). After dimensional reduction, the NN can be retrained with the smaller input space or the

physics-based model [28] can be refactored to remove dependence to improve the time needed for optimization. Similarly, a decision tree-based approach can be utilized as demonstrated by Li et al. [50] to identify key changes that have a dramatic effect in the model output (Figure 4e). Decision trees are primarily used when classifying outputs into discrete bins (atomically precise

AuNP vs. nonprecise AuNP in this case) and not continuous variables.

Outlook

Autonomous chemical science and engineering strategies for optimization and reaction engineering are fast becoming powerful tools for performing time-efficient and material-efficient reaction universe exploration of emerging advanced materials and molecules. Selection of the correct ML model and the experiment-selection algorithm can have dramatic effects on the overall performance of the closed-loop experimentation systems. In this regard, access to data and different ML algorithms for benchmarking becomes increasingly important to identify the most cost-effective ML modeling and decision-making framework before integration within an experimentation platform. Work is being done to extend ML-guided reaction exploration and optimization methods from a single reaction system with a well-defined input space, to input spaces containing discrete variables and flexible boundaries, and even further to predict novel reactions, including chemical species *a priori*. As surrogate modeling approaches gain widespread adoption, these higher-order prediction and optimization problems will have an increasingly robust well of data to pull from in order to improve and refine the more general models. With this wealth of data, it will become possible to benchmark the performance of the different modeling and optimization methods with relevant hyperparameters for a given class of reaction systems [51].

Declaration of Competing Interest

Nothing declared.

Acknowledgements

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest.

1. Bezinge L, Maceiczyk RM, Lignos I, Kovalenko MV, deMello AJ: **Pick a color MARIA: adaptive sampling enables the rapid identification of complex perovskite nanocrystal compositions with defined emission characteristics**. *ACS Appl Mater Interfaces* 2018, **10**:18869-18878.

2. Fitzpatrick DE, Battilocchio C, Ley SV, Novel A: **Internet-based reaction monitoring, control and autonomous self-optimization platform for chemical synthesis**. *Org Process Res Dev* 2016, **20**:386-394.

3. Krishnadasan S, Brown RJC, deMello AJ, deMello JC: **Intelligent routes to the controlled synthesis of nanoparticles**. *Lab Chip* 2007, **7**:1434-1441.

4. Bédard A-C, Adamo A, Aroh KC, Russell MG, Bedermann AA, Torosian J, Yue B, Jensen KF, Jamison TF: **Reconfigurable system for automated optimization of diverse chemical reactions**. *Science* 2018, **361**:1220-1225.

5. Reizman BJ, Wang Y-M, Buchwald SL, Jensen KF: **Suzuki–Miyaura cross-coupling optimization enabled by automated feedback**. *React Chem Eng* 2016, **1**:658-666.

6. Voznyy O, Levina L, Fan JZ, Askerka M, Jain A, Choi M-J, Ouellette O, Todorović P, Sagar LK, Sargent EH: **Machine learning accelerates discovery of optimal colloidal quantum dot synthesis**. *ACS Nano* 2019, **13**:11122-11128.

7. Silitonga AS, Shamsuddin AH, Mahlia TMI, Milano J, Kusumo F, Siswantoro J, Dharma S, Sebayang AH, Masjuki HH, Ong HC: **Biodiesel synthesis from Ceiba pentandra oil by microwave irradiation-assisted transesterification: ELM modeling and optimization**. *Renew Energy* 2020, **146**:1278-1291.

8. 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.

9. Steiner S, Wolf J, Glatzel S, Andreou A, Granda JM, Keenan G, Hinkley T, Aragon-Camarasa G, Kitson PJ, Angelone D, *et al.*: **Organic synthesis in a modular robotic system driven by a chemical programming language**. *Science* 2019, **363**:eaav2211.

10. Han Y, Tang B, Wang L, Bao H, Lu Y, Guan C, Zhang L, Le M, Liu Z, Wu M: **Machine-learning-driven synthesis of carbon dots with enhanced quantum yields**. *ACS Nano* 2020, **14**:14761-14768.

11. Li X-T, Chen L, Wei G-F, Shang C, Liu Z-P: **Sharp increase in catalytic selectivity in acetylene semihydrogenation on Pd achieved by a machine learning simulation-guided experiment**. *ACS Catal* 2020, **10**:9694-9705.

12. Godfrey AG, Masquelin T, Hemmerle H: **A remote-controlled adaptive medchem lab: an innovative approach to enable drug discovery in the 21st Century**. *Drug Discov Today* 2013, **18**:795-802.

13. Coley CW, Thomas DA, Lummiss JAM, Jaworski JN, Breen CP, Schultz V, Hart T, Fishman JS, Rogers L, Gao H, *et al*: **A robotic platform for flow synthesis of organic compounds informed by AI planning**. *Science* 2019, **365**:eaax1566.

14. Rubens M, Vrijsen JH, Laun J, Junkers T: **Precise polymer synthesis by autonomous self-optimizing flow reactors**. *Angew Chem Int Ed* 2019, **58**:3183-3187.

15. Nguyen TN, Nhat TTP, Takimoto K, Thakur A, Nishimura S, Ohyama J, Miyazato I, Takahashi L, Fujima J, Takahashi K, *et al.*: **High-throughput experimentation and catalyst informatics for oxidative coupling of methane**. *ACS Catal* 2020, **10**:921-932.

16. Clayton AD, Schweidtmann AM, Clemens G, Manson JA, Taylor CJ, Niño CG, Chamberlain TW, Kapur N, Blacker AJ, Lapkin AA, *et al.*: **Automated self-optimisation of multi-step reaction and separation processes using machine learning**. *Chem Eng J* 2020, **384**:123340.

17. Ammar Y, Cognet P, Cabassud M: **ANN for hybrid modelling of batch and fed-batch chemical reactors**. *Chem Eng Sci* 2021, **237**:116522.

18. Vaucher AC, Zipoli F, Geluykens J, Nair VH, Schwaller P, Laino T: **Automated extraction of chemical synthesis actions from experimental procedures**. *Nat Commun* 2020, **11**:3601.

19. Rizkin BA, Shkolnik AS, Ferraro NJ, Hartman RL: **Combining automated microfluidic experimentation with machine learning for efficient polymerization design**. *Nat Mach Intell* 2020, **2**:200-209.

20. Bateni F, Epps RW, Antami K, Dargis R, Bennett JA, Reyes KG, Abolhasani M: **Autonomous nanocrystal doping by self-driving fluidic micro-processors**. *Adv Intell Syst* 2022,2200017, , https://doi.org/10.1002/aisy.202200017

21. 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**:1623-1636.

22. Echtermeyer A, Amar Y, Zakrzewski J, Lapkin A: **Self-optimisation and model-based design of experiments for developing a C–H activation flow process**. *Beilstein J Org Chem* 2017, **13**:150-163.

23. Hone CA, Holmes N, Akien GR, Bourne RA, Muller FL: **Rapid multistep kinetic model generation from transient flow data**. *React Chem Eng* 2017, **2**:103-108.

24. Frey D, Shin JH, Musco C, Modestino MA: **Chemically-informed data-driven optimization (ChIDDO): leveraging physical models and Bayesian learning to accelerate chemical research**. *React Chem Eng* 2022, **7**:855-865.

25. 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**:277-282.

26. Epps RW, Volk AA, Reyes KG, Abolhasani M: **Accelerated AI
•• development for autonomous materials synthesis in flow**. *Chem Sci* 2021, **12**:6025-6036.
A trained ensemble NN was utilized as the ground truth for a surrogate model of lead halide perovskite nanocrystal synthesis. The digital reaction system enabled the rapid development and optimization of the architecture and hyperparameters of NN models for closed-loop nanocrystal synthesis optimization.

27. Gao H, Struble TJ, Coley CW, Wang Y, Green WH, Jensen KF:
•• **Using machine learning to predict suitable conditions for organic reactions**. *ACS Cent Sci* 2018, **4**:1465-1476.
An NN for predicting catalysts, solvents, and reagents was created by training on 10 million reaction examples from the Reaxys database. The trained network was successful in identifying a close match to the reaction conditions in 70% of the testing set over a wide range of common reaction classes.

28. Quaglio M, Roberts L, Bin Jaapar MS, Fraga ES, Dua V, Galvanin F: **An artificial neural network approach to recognise kinetic models from experimental data**. *Comput Chem Eng* 2020, **135**:106759.

29. Coley CW, Jin W, Rogers L, Jamison TF, Jaakkola TS, Green WH, Barzilay R, Jensen KF: **A graph-convolutional neural network model for the prediction of chemical reactivity**. *Chem Sci* 2018, **10**:370-377.

30. Wang H, Xie Y, Li D, Deng H, Zhao Y, Xin M, Lin J: **Rapid identification of X-ray diffraction patterns based on very limited data by interpretable convolutional neural networks**. *J Chem Inf Model* 2020, **60**:2004-2011.

31. Wei JN, Duvenaud D, Aspuru-Guzik A: **Neural networks for the prediction of organic chemistry reactions**. *ACS Cent Sci* 2016, **2**:725-732.

32. 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.
Overall optimization campaign cost was reduced in an inorganic lead halide perovskite halide exchange reactions by pretraining an ensemble NN model to improve the speed of the 10 future peak emission energy optimization campaigns.

33. Epps RW, Bowen MS, Volk AA, Abdel-Latif K, Han S, Reyes KG,
•• Amassian A, Abolhasani M: **Artificial chemist: an autonomous quantum dot synthesis bot**. *Adv Mater* 2020, **32**:2001626.
An ensemble NN model was developed that surpassed traditional black-box optimization strategies with a balanced exploration/exploitation policy and Bayesian Optimization methods for halide exchange reaction of lead halide perovskite nanocrystals.

34. Abolhasani M, Coley CW, Xie L, Chen O, Bawendi MG, Jensen KF: **Oscillatory microprocessor for growth and in situ characterization of semiconductor nanocrystals**. *Chem Mater* 2015, **27**:6131-6138.

35. Abolhasani M, Jensen KF: **Oscillatory multiphase flow strategy for chemistry and biology**. *Lab Chip* 2016, **16**:2775-2784.

36. 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, **31**:2106725.
The Chimera objective function was used to sequentially optimize three target parameters of Au nanoparticles, while information about the relative importance of the input parameters to the model output was extracted through Shapely Feature Analysis.

37. Gongora AE, Snapp KL, Whiting E, Riley P, Reyes KG, Morgan EF, Brown KA: **Using simulation to accelerate autonomous experimentation: a case study using mechanics**. *iScience* 2021, **24**:102262.

38. Li J, Li J, Liu R, Tu Y, Li Y, Cheng J, He T, Zhu X: **Autonomous discovery of optically active chiral inorganic perovskite nanocrystals through an intelligent cloud lab**. *Nat Commun* 2020, **11**:2046.

39. Huyer W, Neumaier A: **SNOBFIT–stable noisy optimization by branch and fit**. *ACM Trans Math Softw (TOMS)* 2008, **35**:1-25.

40. Zhou Z, Li X, Zare RN: **Optimizing chemical reactions with deep
•• reinforcement learning**. *ACS Cent Sci* 2017, **3**:1337-1344.
The deep Reinforcement Learning optimizer outperformed state-of-the-art black-box optimization strategies by using 71% fewer steps to converge. The model also had improved performance after training on different reaction systems.

41. Hansen N, Ostermeier A: **Completely derandomized self-adaptation in evolution strategies**. *Evol Comput* 2001, **9**:159-195.

42. Nelder JA, Mead R: **A simplex method for function minimization**. *Comput J* 1965, **7**:308-313.

43. Mekki-Berrada F, Ren Z, Huang T, Wong WK, Zheng F, Xie J, Tian
• IPS, Jayavelu S, Mahfoud Z, Bash D, *et al*.: **Two-step machine learning enables optimized nanoparticle synthesis**. *npj Comput Mater* 2021, **7**:55.
A flexible parameter space boundary was utilized along with a Bayesian optimization strategy to match silver nanoparticle spectra with a target, while a deep NN was used to predict the overall shape of the absorbance spectra for each input point to provide a ranking.

44. Coley CW, Eyke NS, Jensen KF: **Autonomous discovery in the chemical sciences part I: progress**. *Angew Chem Int Ed* 2020, **59**:22858-22893.

45. Coley CW, Eyke NS, Jensen KF: **Autonomous discovery in the chemical sciences part II: outlook**. *Angew Chem Int Ed* 2020, **59**:23414-23436.

46. Epps RW, Abolhasani M: **Modern nanoscience: convergence of AI, robotics, and colloidal synthesis**. *Appl Phys Rev* 2021, **8**:041316.

47. Ji W, Deng S: **Autonomous discovery of unknown reaction
• pathways from data by chemical reaction neural network**. *J Phys Chem A* 2021, **125**:1082-1092.
A chemical reaction NN was developed that used the fundamental law of mass action and Arrhenius behavior to inform the architecture of the network, enabling the extraction of fundamental elementary reaction steps of the reaction system from the weights and biases of the trained model.

48. 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**:17028-17039.

49. Deb K, Pratap A, Agarwal S, Meyarivan T: **A fast and elitist multiobjective genetic algorithm: NSGA-II**. *IEEE Trans Evol Comput* 2002, **6**:182-197.

50. Li J, Chen T, Lim K, Chen L, Khan SA, Xie J, Wang X: **Deep learning accelerated gold nanocluster synthesis**. *Adv Intell Syst* 2019, **1**:1900029.

51. Rohr B, Stein HS, Guevarra D, Wang Y, Haber JA, Aykol M, Suram SK, Gregoire JM: **Benchmarking the acceleration of materials discovery by sequential learning**. *Chem Sci* 2020, **11**:2696-2706.