Chinese Journal of Chemical Engineering 84 (2025) 274–288

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

journal homepage: www.elsevier.com/locate/CJChE

Review

Intelligent chemical synthesis based on microchemical engineering technology

Yongqi Pan [1, #], Yazi Yu [2, #], Lijie Wang [3], Guogang Hu [3], Yujun Wang [1, *], Guangsheng Luo [1]

[1] State Key Laboratory of Chemical Engineering and Low-Carbon Technology, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China
[2] College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China
[3] North China Pharmaceutical Hebei Huamin Pharmaceutical Co., Ltd, Shijiazhuang 052165, China

ARTICLE INFO

Article history:
Received 14 January 2025
Received in revised form
16 April 2025
Accepted 6 May 2025
Available online 4 June 2025

Keywords:
Flow chemistry
Microreactor
Microfluidics
Machine learning
Multi-phase disperse

ABSTRACT

Chemical synthesis is essential in industries such as petrochemicals, fine chemicals, and pharmaceuticals, driving economic and social development. The increasing demand for new molecules and materials calls for novel chemical reactions; however, manual experimental screening is time-consuming. Artificial intelligence (AI) offers a promising solution by leveraging large-scale experimental data to model chemical reactions, although challenges such as the lack of standardization and predictability in chemical synthesis hinder AI applications. Additionally, the multi-scale nature of chemical reactions, along with complex multiphase processes, further complicates the task. Recent advances in microchemical systems, particularly continuous flow methods using microreactors, provide precise control over reaction conditions, enhancing reproducibility and enabling high-throughput experimentation. These systems minimize transport-related inconsistencies and facilitate scalable industrial applications. This review systematically explores recent developments in intelligent synthesis based on microchemical systems, focusing on reaction system design, synthesis robots, closed-loop optimization, and high-throughput experimentation, while identifying key areas for future research.
© 2025 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1. Background

Chemical synthesis plays a crucial role in industries such as petrochemicals [1], fine chemicals [2–4], and pharmaceutical manufacturing [5,6], holding significant importance in economic and social development. The growing demand for a better quality of life calls for the creation of new molecules and materials, which in turn requires novel chemical synthesis reactions. However, due to the complexity of chemical reaction conditions, manual experimental screening often demands considerable time and effort. As a result, the application of artificial intelligence (AI) to explore the chemical reaction space has garnered widespread attention [7–13]. Unlike traditional approaches based on first principles that utilize analytical or numerical methods to characterize chemical reactions, AI methods tend to construct high-parameter numerical models trained on large-scale experimental data [14–22]. Within this research paradigm, the construction of high-quality, structured datasets has become a critical challenge [22].

However, compared to other manufacturing industries, chemical synthesis exhibits lower levels of standardization, reproducibility, and predictability, posing significant challenges to data-driven artificial intelligence (AI) methods [13]. This phenomenon largely arises from the multi-scale nature of chemical synthesis, which spans atomic and molecular-scale chemical reactions, microscale mass and heat transfer, and macroscale equipment design. Interactions between these different scales are highly interdependent and difficult to decouple. Furthermore, due to factors such as cost, solubility, and separation difficulties, many chemical processes in industrial production are multiphase processes [23,24]. These involve complex distributions of concentration [25], temperature [2], and velocity fields [26–29], as well as intricate interphase transfer behaviors, which further complicate the standardization of synthesis processes. Therefore, one critical prerequisite for achieving intelligent chemical synthesis is the development of standardized reaction equipment that enables the decoupling of processes across different scales.

This article is part of a special issue entitled: AI for Chemical Engineering published in Chinese Journal of Chemical Engineering.

* Corresponding author.
E-mail address: wangyujun@mail.tsinghua.edu.cn (Y. Wang).
# These authors contributed equally to this work.

https://doi.org/10.1016/j.cjche.2025.05.010
1004-9541/© 2025 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

In recent years, continuous synthesis methods based on microchemical technologies have been widely applied due to their ability to precisely control chemical reactions at micro-scale spatial and temporal dimensions [30–42]. Microreactors with feature sizes on the order of micrometers enable efficient mass and heat transfer as well as high levels of multiphase dispersion, significantly reducing the impact of non-standardized transport factors on reactions [43]. The enclosed nature of microchannel systems minimizes component evaporation, which is crucial for maintaining reproducible experimental conditions [44]. Additionally, fluidic reactor designs bridge the gap between experimental procedures and industrial-scale operations, facilitating the transition from laboratory research to industrial applications [45]. These features stabilize and render reaction process parameters controllable, allowing for precise regulation of reaction conditions while supporting high-throughput experimentation to generate large volumes of training data. These advantages position microchemical systems as highly promising tools for intelligent synthesis.

This work systematically reviews recent advances in intelligent synthesis based on microchemical systems from four perspectives: reaction system design, synthesis robot construction, closed-loop process optimization, and high-throughput experimentation. Based on an analysis of the existing literature, this work highlights areas that remain insufficiently explored and offers perspectives on future developments in the field.

2. Basic Technology of Intelligent Synthesis

2.1. Introduction to common AI algorithms

To date, numerous machine learning models have been employed to address chemical problems across various scenarios such as the prediction of physical properties [46–50], catalytic performance [51–55], synthetic yield [16,18,20], selectivity [56,57], etc. The primary tasks of these models can be categorized into two types: regression and classification. The main distinction between these tasks lies in the model output: regression tasks typically yield continuous values, whereas classification tasks produce finite, discrete categories. For instance, in catalyst and material design tasks, machine learning models are used for regression to predict activation energies of catalytic reactions and mechanical properties of materials [51]. In the study of liquid-liquid dispersion within microchannels, these models are utilized for classification to differentiate flow patterns such as jetting flow, dripping flow, and squeezing flow [58].

For common regression and classification tasks, tree-based models such as decision trees, random forests, and gradient boosting trees are frequently used machine learning models. Decision Trees operate by constructing a model of decisions based on features of the data, forming a tree-like structure where each node represents a feature, each branch a decision rule, and each leaf an outcome (Fig. 1(a)) [59,60]. They are valued for their simplicity and interpretability, though they can be prone to overfitting. Random Forests (Fig. 1(b)) address this issue by creating an ensemble of multiple Decision Trees, each trained on a random subset of the data, and averaging their predictions to improve accuracy and robustness [61–63]. Gradient Boosting Trees, on the other hand, build trees in a sequential manner, where each new tree corrects the errors of the previous ones, leading to a powerful model that can handle complex relationships in the data [64]. Based on the principles of boosting (Fig. 1(c)), numerous similar models have been developed, including XGBoost [65], AdaBoost [66], and CatBoost [67], which have achieved outstanding results in various tasks. Tree-based models often yield satisfactory predictive performance even on smaller datasets [64]. Given the limited amount of standardized data in the fields of chemistry and materials, these models have been widely applied.

Neural network algorithms, inspired by the human brain's structure and function, are computational models that learn from data through a training process, adjusting the weights of interconnected nodes (neurons) as shown in Fig. 1(d), to minimize prediction errors [68,69]. These algorithms are adaptable, capable of generalizing from training data to make accurate predictions on unseen data, and can model complex, non-linear relationships. Therefore, it is suitable for handling tasks with a large or complex number of input features, such as image recognition [70,71], natural language processing [72,73], and spectral information processing [74,75]. The input to a neural network is typically a multi-dimensional array, and the output is a prediction or classification based on the learned patterns. Common types of neural networks include Convolutional Neural Networks (CNNs), which are particularly effective for image processing tasks due to their ability to capture spatial hierarchies [76], and Recurrent Neural Networks (RNNs), which are suited for sequential data like time series or text, thanks to their memory of previous inputs [77]. The Transformer, a more recent innovation, has revolutionized the field by introducing self-attention mechanisms that allow for more efficient natural language processing, significantly speeding up training [78–80]. Large language models (LLMs) built on Transformer architecture and trained on vast amounts of text data, such as GPT [81] and BERT [82], have set new benchmarks in natural language understanding and generation, capable of producing human-like text and comprehending context across a wide range of language tasks. Based on LLMs, an agent is an intelligent software that uses advanced language understanding to autonomously assist users by performing tasks, providing information, and facilitating interactions [83]. These advancements in neural network algorithms continue to drive progress in artificial intelligence, enabling more sophisticated and efficient solutions across various domains.

By iterating the training-prediction cycle of regression models, optimized workflows can be developed. Bayesian Optimization is a probabilistic model-based optimization method for optimizing black-box functions, i.e., functions whose internal mechanisms are unknown and expensive to evaluate [86–88]. The goal is to find the global optimal solution efficiently with limited computational resources and samples. Bayesian Optimization reduces the search space gradually by constructing a prior distribution model to proxy the objective function and using the posterior distribution of this proxy model to guide the selection of query points, as shown in Fig. 2(a) Due to its ability to optimize based on a very small amount of sample data, Bayesian Optimization is considered promising for chemical synthesis process parameter optimization [89–94]. Multi-Objective Bayesian Optimization (MOBO) optimizes multiple conflicting objectives using Bayesian methods [95,96]. It identifies the Pareto frontier, the set of non-dominated solutions, where no solution is better in all objectives (Fig. 2(b)). The hypervolume shown in Fig. 2(c) measures the quality of the Pareto frontier, calculated as:

$$\text{Hypervolume} = \lambda(\cup_{x\in \boldsymbol{P}}[\boldsymbol{f}(\boldsymbol{x}),\boldsymbol{r}]) \quad (1)$$

Here, $\boldsymbol{P}$ is the Pareto frontier, $\boldsymbol{r}$ is a reference point, and λ is the Lebesgue measure. MOBO iteratively updates a probabilistic model and selects points to evaluate using acquisition functions, balancing exploration and exploitation.

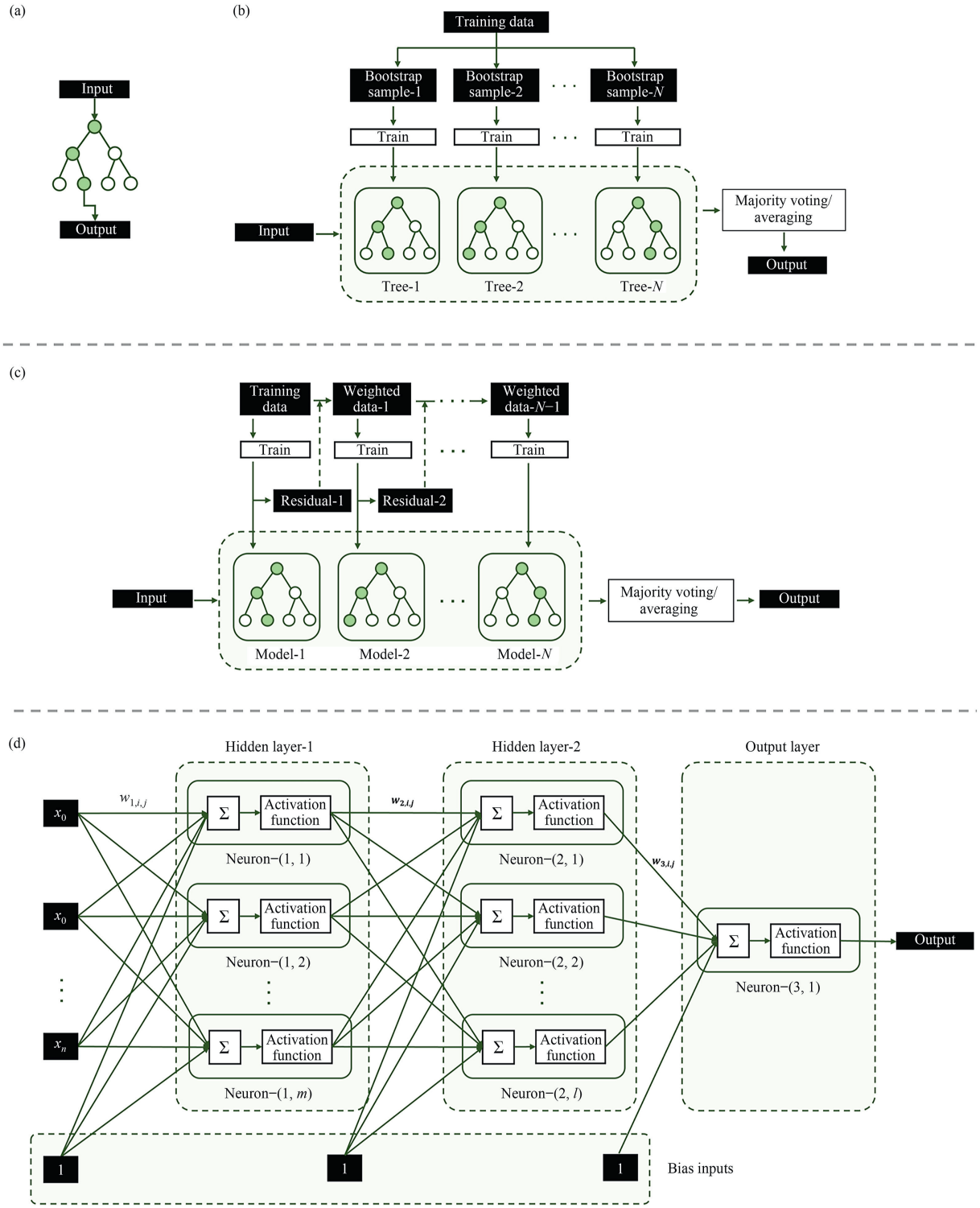

Fig. 1. Schematic diagram of common machine learning models. (a) Decision tree model. (b) Bagging model represented by a random forest model. (c) Boosting model. (d) Neural network model, where the figure shows a simple neural network with two hidden layers.

2.2. Reaction system design based on machine learning

2.2.1. Intelligent design of flow reactor mixers

For the majority of flow synthesis systems, mixing performance is a critical parameter. Excellent mixing performance enables rapid maximization of the contact area between multiple reactant phases, thereby enhancing reaction rate and uniformity. Conversely, poor mixing performance necessitates longer residence times to achieve adequate mixing and may lead to the formation of by-products due to local non-uniformities. To achieve flexible intelligent synthesis systems—capable of performing multiple types of reactions with minimal switching costs—the intelligent design and optimization of flow reactors with high mixing performance is especially critical.

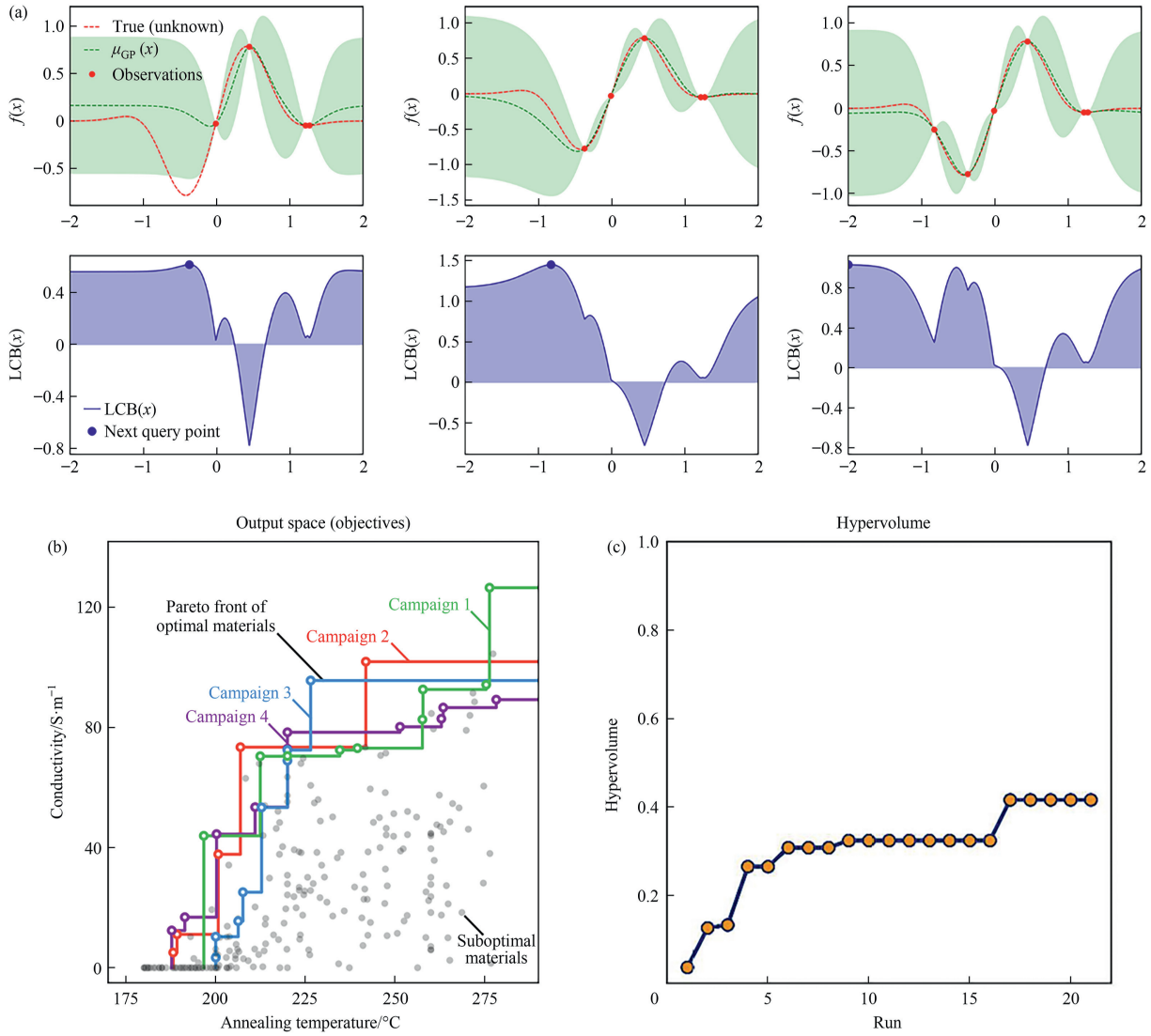

Fig. 2. An example of the Bayesian Optimization process. (a) Changes in the surrogate function and acquisition function during Bayesian Optimization. Reproduced with permission from Ref. [84], © Wu *et al.* 2024. (b) The Pareto front in the multi-objective optimization process. Reproduced with permission from Ref. [49], © MacLeod *et al.* 2022. (c) Changes in the hypervolume during the multi-objective optimization process. Reproduced with permission from Ref. [85]. Copyright 2024, American Association for the Advancement of Science.

In 2023, Nausheen Basha *et al.* [97] at Imperial College investigated the mixing characteristics that improve the performance of spiral tube reactors through a combination of Bayesian Optimization and Computational Fluid Dynamics (CFD). The study employed a data-driven approach, utilizing Bayesian Optimization as a surrogate model method, integrated with CFD as a black-box function, to explore the parameter space of operating conditions, oscillation amplitude, frequency, and net flow rate. The research correlated flow field characteristics with dimensionless numbers such as Strouhal number, oscillatory Dean number, and Reynolds number to evaluate the plug flow performance of the reactor. It was found that, under optimal performance conditions, the oscillating flow was sufficient to limit axial dispersion, which was done by flow reversal and redirection and by promoting Dean vortices. In 2024, the team further advanced their work by publishing another study in *Nature Chemical Engineering*, focusing on the application of machine learning in flow reactor design [98]. This research integrated high-dimensional parameterization, CFD, and multi-fidelity Bayesian Optimization to identify and optimize complex reactor designs. By correlating the hybrid-enhanced vortex flow structure with the reactor performance, the research team utilized advanced fabrication techniques and augmented intelligence methods to successfully develop an experimental plug-flow reactor with a performance improvement of about 60% compared to conventional designs. The results of this research show that by combining advanced fabrication techniques and machine learning methods, more capable reactor designs can be developed, providing new opportunities for discovery, innovation, and advancement in the field of chemical engineering.

2.2.2. *Intelligent optimization of decentralized performance*

In gas-liquid or liquid-liquid reactions, it is often necessary to disperse the phase containing a reactant into another phase in order to avoid localized over-concentration of a reactant that leads to a reduction in utilization or the generation of by-products. In this context, the optimization of microreactors for efficient dispersion is particularly important. In particular, for intelligent synthesis devices, in order to be able to extend the application space of the synthesis device, micro-reactors are often constructed in the form of modularization, so if the intelligent design and optimization of

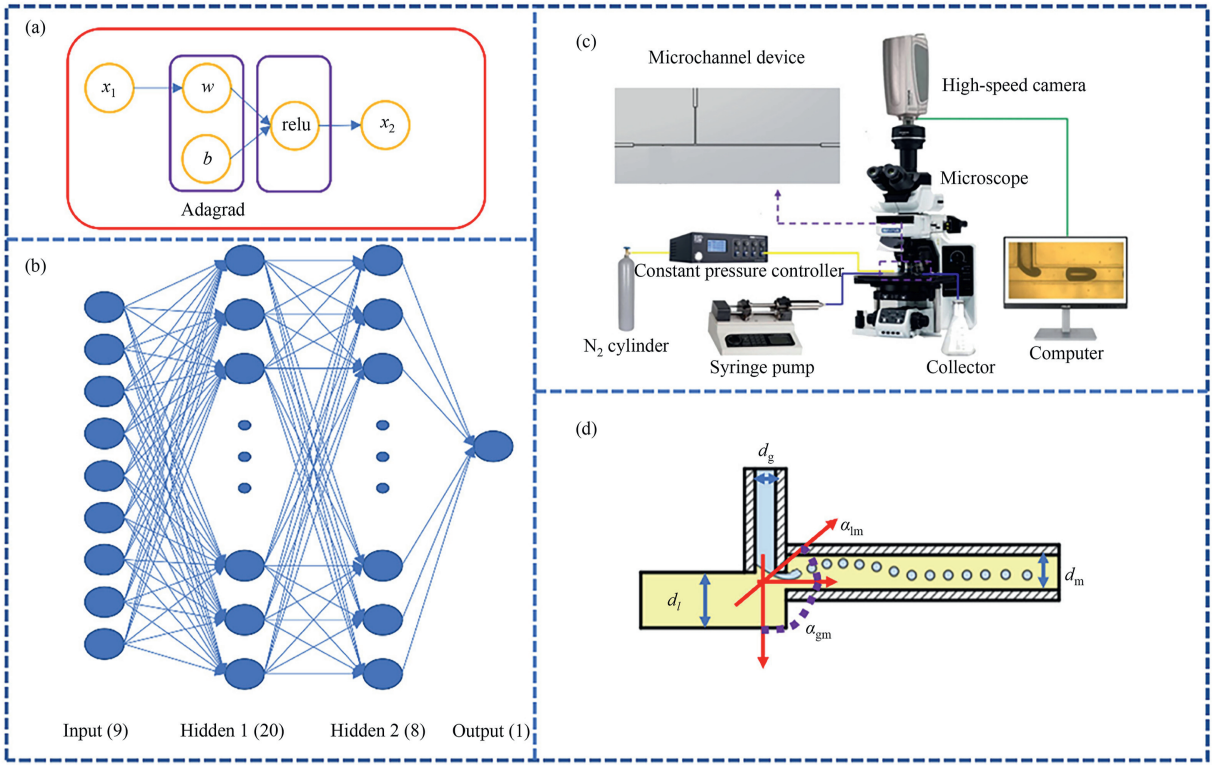

Fig. 3. Neural network-based system for predicting bubble generation size by Luo *et al.* Reproduced from Ref. [101] with permission from the Royal Society of Chemistry.

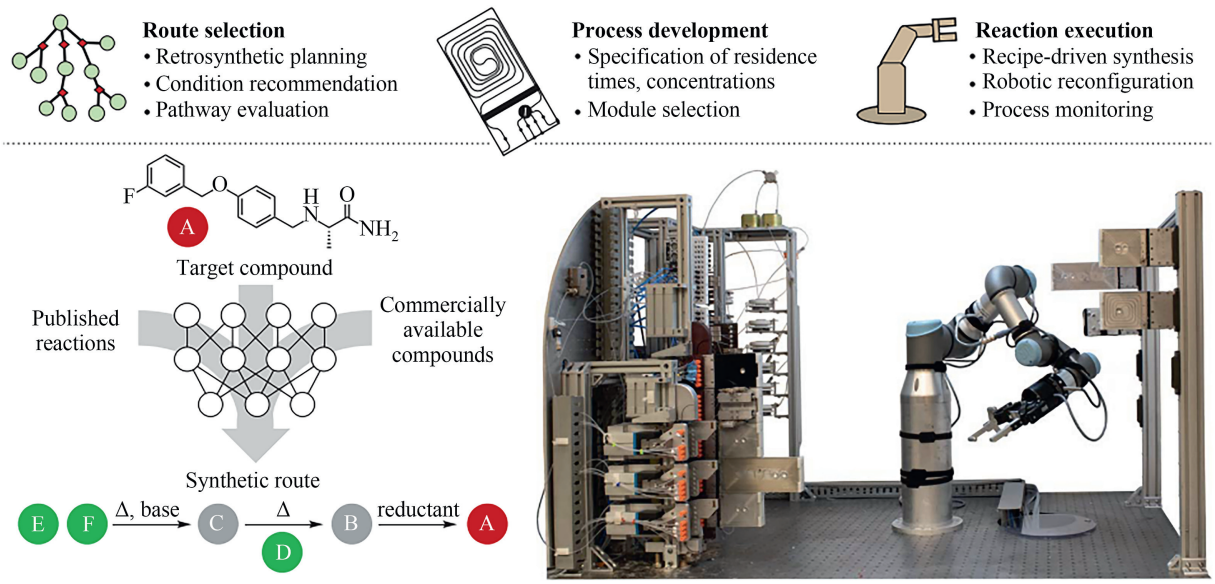

Fig. 4. Modular flow synthesis platform by Jensen *et al.* Reproduced with permission from Ref. [44]. Copyright 2019, American Association for the Advancement of Science.

different micro-dispersed modules are realized for different reactions, the process gap between different reactions will be greatly reduced, and low-cost switching between varieties will be realized [99].

In 2021, Densmore *et al.* [100] developed a machine-learning-based web-based tool called DAFD (Design Automation of Fluid Dynamics) to predict the performance and automate the design of microfluidic flow-focused drop generators. The study utilized machine learning algorithms to predict drop diameters and generation rates with mean absolute errors of less than 10 μm and 20 Hz, respectively. The DAFD tool is capable of controlling user-specified performance to within 4.2% and 11.5% of the expected diameter and rate. The study demonstrates that the DAFD tool can be extended by the community to support additional fluid combinations without the need for extensive machine learning knowledge or large-scale datasets. The study also fabricated 43 flow-focused drop generators through a low-cost rapid prototyping approach and evaluated their performance under a wide range of flow conditions, generating a standardized large-scale dataset containing 998 data points. The development of these datasets and machine learning

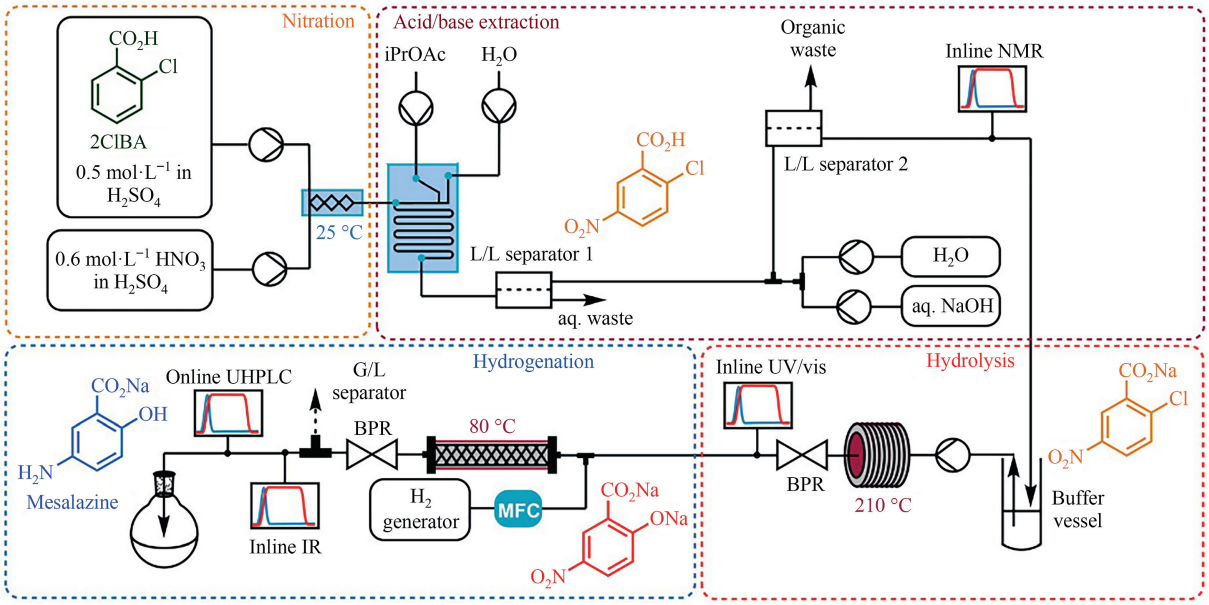

Fig. 5. An automated multi-step reaction optimization system by Kappe *et al.* Reproduced with permission from Ref. [107]. Copyright 2021, John Wiley & Sons.

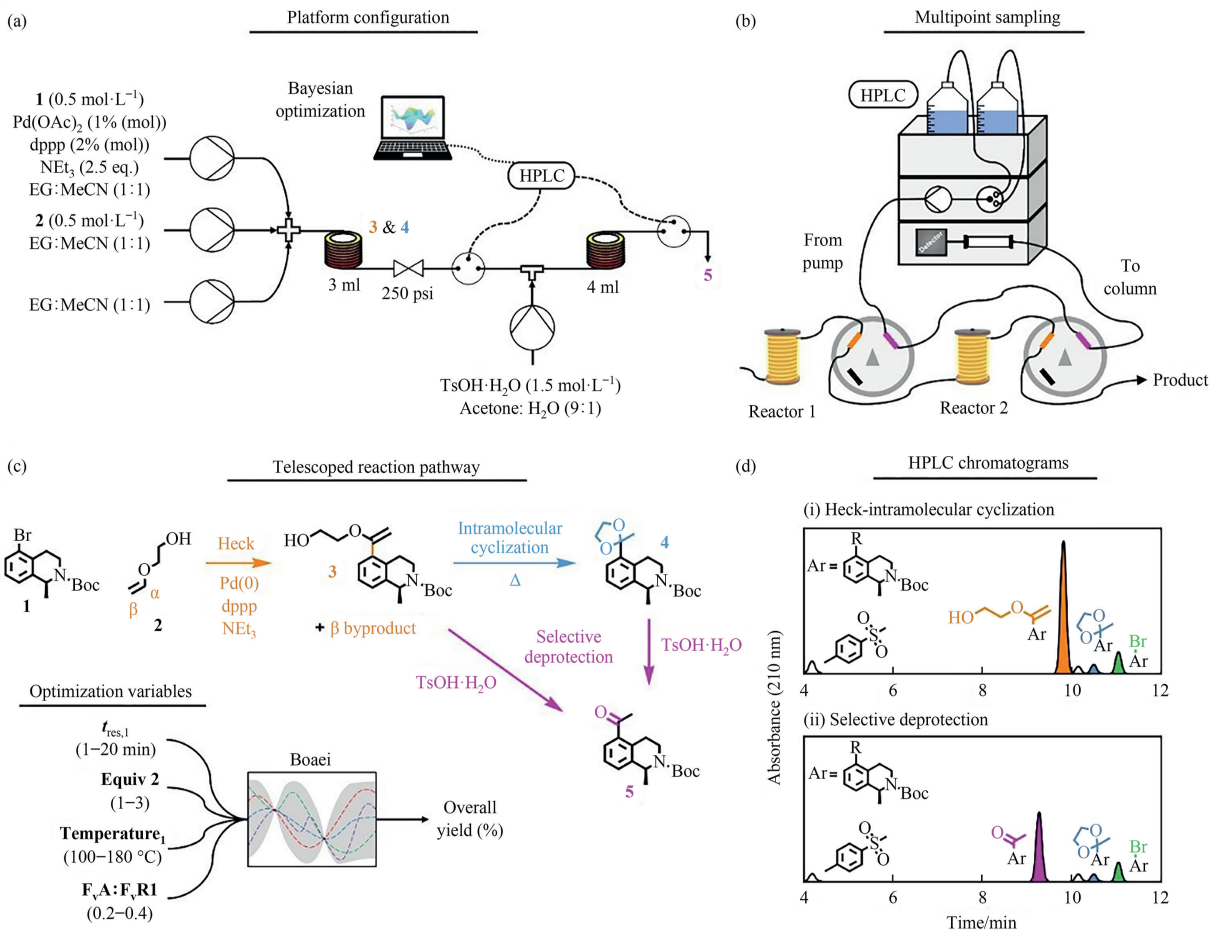

Fig. 6. An automated continuous flow platform for optimizing the Heck phantomization-deprotection reaction by Bourne *et al.* Reproduced with permission from Ref. [108], © Bourne *et al.* 2023.

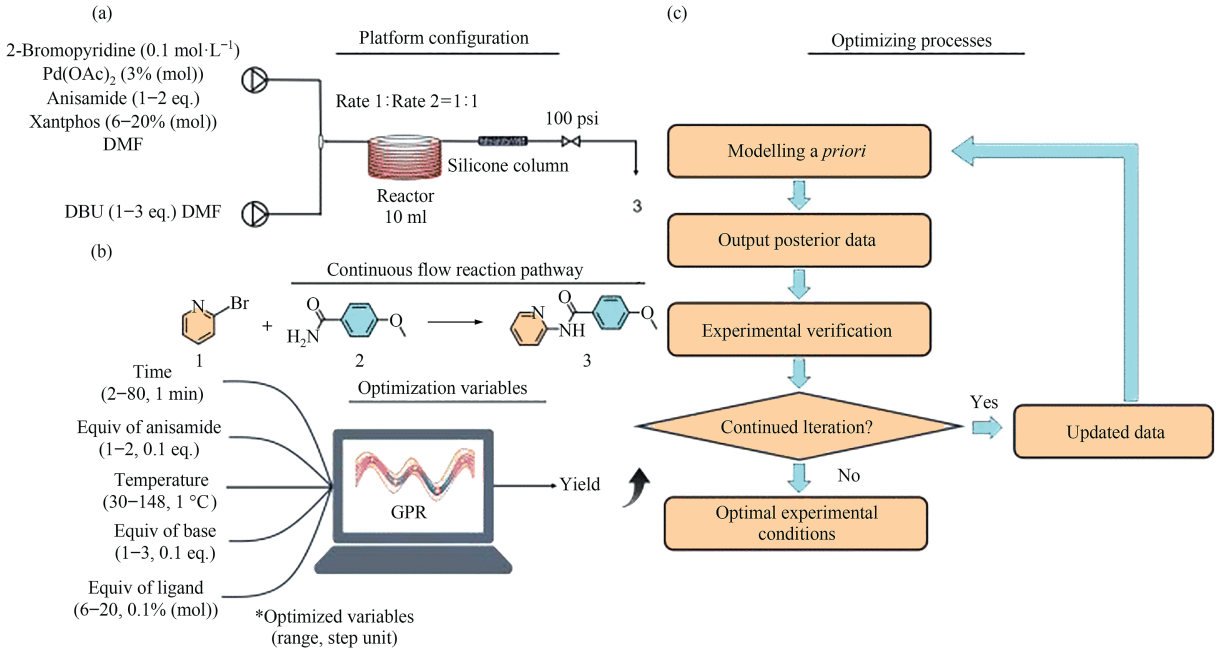

Fig. 7. Online analysis and Bayesian Optimization system by Tao *et al.* Reproduced from Ref. [109] with permission from the Royal Society of Chemistry.

algorithms provides a framework for machine learning-based performance prediction of drop generators and validates a design automation tool that can design drop generators based on user-specified performance.

In 2023, Professor Luo's team [101] published a study in the journal *Lab on a Chip* proposing a generalized neural network model combining mechanism analysis and data modeling for the reliable design of gas-liquid T-joint microdevices (Fig. 3). The study was conducted by constructing a microbubble generation database containing 854 data points and utilizing transfer learning to improve model accuracy. The study also proposed a design method that can achieve a relative deviation of less than 5% in the expected bubble size. The reliability of the method was verified by designing and preparing a novel device that was able to prepare smaller bubbles than other common T-joint devices. This study not only establishes a general and pervasive database and model, but also provides a methodology for the design of gas-liquid T-shaped microreactors.

2.3. Synthesizing robots based on microchemical systems

Similar to other industrial fields, the prerequisite for intelligent and automated synthesis is the digital control of the synthesis equipment. The robot chemist is a classic scenario. In such a scheme, a movable robot is used to connect the various workflows, which was first reported by Copper's group [102], and has been used in several fields. Another option is the cloud lab, which is designed to transform traditional labs into an all-digital control model, thus enabling cloud control, which also establishes the hardware platform technology for AI deployment and application. However, these options often involve the remodeling of entire large laboratories; the ability to build a small, desktop-level automated synthesis workstation would make a lot of sense for research in small labs. In 2018, the Cronin team [103] came up with the idea of the 'Chemputer', a chemical synthesis rig controlled by a programming code language and logic. They have developed a modular robotic system driven by a chemical programming language that abstracts the four key stages of chemical synthesis—reaction, working fluid handling, separation, and purification—and translates them into automated physical operations. For the study, they created the Chempler program, which compiles a high-level description language (ChASM) for chemical synthesis programs into low-level instructions that can be run on modular hardware. The study validated this concept by automating the synthesis of three drug compounds—benadryl hydrochloride, lufenamide, and sildenafil—demonstrating the Chemputer's potential to improve the reproducibility, safety, and collaborative nature of chemical synthesis. In the subsequent research, researchers found that the continuous flow device is more suitable for the construction of Chemputer due to its stable and controllable process parameters and more "solid-state" reaction device, and many useful explorations have been carried out in this direction.

In 2019, Jensen *et al.* [44] at MIT developed an automated, scalable synthesis platform integrating artificial intelligence (AI) planning and robotic execution aimed at improving the efficiency of complex organic molecule synthesis (Fig. 4). This study utilizes published reaction data to computationally design synthetic routes and execute expertly refined chemical recipe files (CRFs) on a robotic flow chemistry platform for scalable and reproducible synthesis. The research team developed an open-source software suite for computer-aided synthetic planning (CASP), which generalizes known chemical reactions to new substrates by learning inverse synthetic transformations, identifying appropriate reaction conditions, and evaluating the likelihood of successful experimental attempts. The software is partially populated with CRFs that require additional details from the chemist user, such as residence time, stoichiometric ratio, and concentration compatible with continuous flow. The robotic arm assembles the modular process cell (reactor and separator) according to the desired process configuration defined in the CRF, connecting the reagent lines and computer-controlled pumps to the reactor inlet. The platform demonstrates the synthesis of 15 drug-related small molecules, including aspirin, secnidazole, lidocaine, valium, (*S*)-warfarin, and safinamide, which are targeted molecules that require a total of eight specific retrosynthetic routes and nine specific process configurations. The platform provides a case study for the modular

synthesis of complex molecules and lays an important foundation for subsequent AI-guided sequential synthesis.

In 2020, Gilmore *et al.* [104] reported an automated synthesis platform that automates the synthesis of small molecules *via* radially arranged continuous flow modules around a central switching station. The method allows small volumes of reagents to be exposed to any reaction conditions desired to achieve the desired transformation. The combination of sequential rather than simultaneous reactions allows for the execution of multi-step processes, enabling the use of different flow rates, the reuse of reactors under different conditions, and the storage of intermediates. This fully automated instrument is capable of performing both linear and convergent synthesis with no need for manual reconfiguration between processes. The study demonstrates the capabilities of this approach by optimizing and multistep synthesis of the target molecules, varying the concentration by on-line dilution, exploring a multistep synthetic strategy for the antiepileptic drug rufinamide, synthesizing eighteen compounds from two derived libraries, and performing a metal photo-oxidative reduction carbon-nitrogen cross-coupling reaction in a photochemical module using the same reagents.

3. Studies on Specific Synthesis Systems

3.1. Optimization of continuous synthesis process parameters using Bayesian algorithm

3.1.1. Single-objective optimization for synthetic yield

For chemical synthesis that favors application scenarios, such as continuous synthesis of pharmaceutical intermediates, fine chemical intermediates, *etc*., the larger the scale of the experiment, the more time and resources are consumed by a single experiment, so how to optimize the target value (*e.g*., yield) to a result as large as possible with as few times as possible is an important topic [105]. In this case, high-throughput experiments will instead cause a surge in cost, and therefore Closed-loop automated optimization based on Bayesian algorithms is a better choice at present [84]. In 2021, Doyle *et al.* [22] reported the Bayesian reaction optimization framework for the optimization of conditions for cross-coupling reactions. The researchers collected a large benchmark dataset of palladium-catalyzed direct arylation reactions, conducted a systematic study of Bayesian Optimization concerning human decision-making in reaction optimization, and applied Bayesian Optimization to the real-world optimization of the Mitsunobu Reaction and the Deoxidative Fluorination Reaction. Using the online platform for benchmarking, the results show that Bayesian Optimization outperforms human decision-making in terms of both average optimization efficiency and consistency.

Achieving closed-loop optimization without human intervention places new demands on the hardware level of the synthesis system. First, the reaction system must be digitally controlled so that key parameters need to be able to be modified in real-time at the control level for process tuning and optimization. Second, for the target quantity of interest (typically yield or conversion), an online monitoring signal is required and this signal can be transformed into a target quantity/alternative variable that forms a monotonic mapping to the target quantity and fed back into the Bayesian Optimization system (Table 1). Digital control is the "synthetic robot" mentioned in Section 2.2. An interesting analogy is that if the synthetic robot is compared to an electric car, then the process of automatic optimization of process parameters is an automatic driver, and the online monitoring instrument is equivalent to the sensing devices such as LIDAR and cameras required for automatic driving, so many literature refer to this kind of system that can automatically optimize reaction parameters as an "automatic driving laboratory" [17,57]. The continuous flow reaction system converts reaction time to flow/dwell time, and stirring mixing to in-channel forced mixing, which makes the process parameters of the reaction stable and controllable, and it is easy to realize the digital control; its flow characteristics make the sampling on-line detection more convenient, so the continuous flow is one of the most suitable reaction processes to be combined with the Bayesian Optimization algorithm. Based on the above characteristics, many researchers engaged in continuous flow reactions have carried out the application of the Bayesian Optimization algorithm to achieve the closed-loop control attempt.

In 2021, Oliver [107] developed a state-of-the-art real-time process analysis platform for the multi-step continuous flow chemical synthesis of the active pharmaceutical ingredient mesalazine (Fig. 5). By integrating four complementary process analytical techniques (NMR, UV/Vis, IR, and UHPLC), they can quantify target products, intermediates, and impurities during synthesis in real time. The platform supports three-step synthesis and three-phase separation with a yield of 1.6 $g \cdot h^{-1}$. This work provides important support for data-driven continuous flow synthesis, demonstrating the potential for the application of self-optimization and model predictive control.

In 2023, Bourne's group [108] developed an automated continuous flow platform for the simultaneous optimization of multi-step chemical reactions (telescoped reactions) (Fig. 6). The research focused on the optimization of the Heck cyclization-deprotection reaction sequence for the synthesis of

Table 1
Characteristics of synthetic reactions involved in various tasks and the online characterization methods employed.

Reaction system	Residence time	Online monitoring equipment
Palladium-catalyzed direct arylation reaction [106]		UHPLC-MS, ^{19}F nuclear magnetic resonance (NMR)
Multistep synthesis of mesalazine [107]	Nitration reaction: 26 s; hydrolysis: 5 min;	NMR, UV/Vis, IR and UHPLC
Heck cyclization-deprotection reaction sequence: Synthesis of a precursor for 1-methyltetrahydroisoquinoline C5 functionalization [108]	1–20 min, longer residence time (>14 min) are more favorable	HPLC
Buchwald–Hartwig reaction (synthesis of pyridinylbenzamide) [109]		HPLC
Synthesis of *N*, *O*-dimethyl-*N*′-nitroisourea [110]	240 min	FTIR
Synthesize the ligand DEHiBA (*N*, *N*-di-(2-ethylhexyl) isobutyramide) [111]	12 s	HPLC
Butylpyridinium bromide synthesis [112]	1–43 min	NMR
Photochemistry and photocatalysis-driven reaction [85]	25 s–40.8 min	NMR
Hydroformylation reaction between syngas (CO and H_2) and olefin (1-octene) in the presence of rhodium and various classes of phosphorus-based ligands [57]	11min	GC-FID

1-methyltetrahydroisoquinoline precursors. Through a simple multipoint sampling method, the group achieved accurate quantification of each reaction using online HPLC and achieved an overall yield of 81% in only 14 h using Bayesian Optimization techniques. This approach takes into account the chemical dependencies between reaction steps, reduces the number of optimization activities, and provides effective technical support for the rapid development of novel multi-step chemical synthesis.

In 2024, Tao's team [109] researched and proposed a self-optimizing Bayesian algorithm (SOBayesian) to optimize the continuous flow synthesis process by introducing Gaussian process regression as a proxy model (Fig. 7). The algorithm improves modeling efficiency by implementing an adaptive strategy during model training rather than adjusting the acquisition function. The algorithm was applied to optimize the Buchwald-Hartwig reaction for the synthesis of pyridinylbenzamides, achieving 79.1% yield after less than 30 rounds of optimization and successfully reducing the number of experiments by 27.6% with less a priori data, thus significantly reducing the experimental cost. This study demonstrated that the reaction process was controlled by kinetics and provided ideas for the optimization of similar reactions. In the same year, Su *et al.* [110] used FTIR online analysis to monitor the continuous flow synthesis process of *O*-methyl-*N*-nitro-*N*-methylisourea and compared the Bayesian Optimization method based on machine learning with the traditional kinetic modeling method. The results showed that both methods obtained about 83% yield with the same reaction parameters. The Bayesian Optimization method found the optimal conditions through only 20 experiments, while the kinetic modeling method required more tedious experiments to formulate and validate the model. This study not only demonstrates the potential of advanced machine learning methods in chemical process optimization but also lays the foundation for further exploration of efficient data-driven chemical synthesis methods.

3.1.2. Finding Pareto Frontiers for multiple optimization objectives

The above work has mainly focused on the optimization of reaction yields in organic synthesis and has achieved efficiency and excellent results beyond manual efforts. However, in the actual production process, in addition to the yield, the reaction rate, cost, impurity content, *etc.*, are also equally important, and then multi-objective optimization (MOO) needs to be achieved. MOO usually employs two main approaches to problem-solving when dealing with multiple conflicting objectives, namely the Weighted Sum Approach and the Pareto Front Approach. The weighting method is a technique to transform a multi-objective problem into a single-objective problem. By assigning different weights to each objective function, the weighted sum of multiple objective functions is optimized as a new single objective function. The Pareto Front method, on the other hand, searches for the set of solutions that makes the improvement of any one objective leads to the deterioration of the other objectives, *i.e.*, Pareto Optimal Solutions (POS). The set of these solutions is called the Pareto Front. In general, the Pareto Front method does not require pre-set objective weights and can provide a variety of trade-off solutions for decision-makers to choose from depending on the context. Therefore, for scenarios involving multiple parameters to be optimized in a real process, researchers often combine Bayesian Optimization and Pareto frontier approaches to iterate the reaction system. For example, Bruce C. Hanson's group developed a self-optimization approach to synthesize the ligand DEHiBA (*N*, *N*-di-(2-ethylhexyl)isobutyramide) for selective uranium coordination and extraction. In the study, the power of machine learning algorithms was harnessed for rapid process development to optimize cost, reagent efficiency, yield, and productivity objective functions [111]. By using an automated flow reactor platform, four synthetic routes were optimized, resulting in a solvent-free, economical pathway at a cost of less than GBP100 per liter and a yield of over 99%.

In 2023, Baldwin's group [112] explored the optimization of butylpyridinium bromide synthesis under continuous flow conditions *via* a human-computer collaborative multi-objective Bayesian Optimization platform (EDBO+). In the study, the group used the EDBO+ algorithm to simultaneously optimize reaction yields and production rates, generating an explicit Pareto frontier. The reaction space was extended by increasing the upper temperature of the reaction and demonstrating the advantages of continuous flow technology over conventional batch chemistry. The study also incorporates semi-automated nuclear magnetic resonance (NMR) spectroscopy to compare the effectiveness of data processing for low-field and high-field NMR and shows that semi-automated processing is effective in low-field NMR spectroscopy, but superior for high-field data. Ultimately, the results demonstrate the integration of human-computer collaborative machine learning methods with continuous flow chemistry, providing new perspectives and tools for chemical synthesis and reaction optimization.

Photocatalytic reactions are often considered to be carried out in continuous flow due to their thermal effects [80,113,114]. However, such continuous flow systems require optimization not only of the flow parameters but also of the light conditions, making manual optimization a tedious task. In 2024, Noël *et al.* [85] developed a robotic platform, called RoboChem, for automated self-optimization, intensification, and amplification of photocatalytic reactions [85]. The platform enables automated and data-rich optimization of photocatalytic transformations by integrating off-the-shelf hardware, custom software, and Bayesian Optimization (BO) algorithms. The RoboChem platform is capable of automated exploration of the parameter space, and self-optimizing not only for a single objective (*e.g.*, yield maximization) but also for considering multiple objectives simultaneously (*e.g.*, optimizing yield and flux at the same time). The platform's versatility is demonstrated by its application to a variety of photocatalytic reactions, including hydrogen atom transfer (HAT) photocatalysis, photoredox catalysis, and metal photocatalysis, which are relevant in pharmaceutical and pesticide chemistry.

Optimization of reaction selectivity is also an important goal in many organic reactions involving complex substrates [115–121]. Optimizing only the yield with reckless disregard for the selectivity of the reaction may result in optimizing the reaction in a direction we do not want. Therefore, in 2024, the group of Abolhasani [57] reported an autopilot catalytic laboratory called Fast-Cat, which is capable of autonomous and efficient parameter space navigation and Pareto frontier mapping, especially in high-temperature, high-pressure, gas-liquid reactions. The Fast-Cat lab enables autonomous ligand benchmarking and multi-objective catalyst performance evaluation with minimal human intervention. In the study, Fast-Cat Lab was utilized to perform rapid Pareto frontier identification of hydroxylation reactions of syngas (CO and H_2) with olefins (1-octene) in the presence of rhodium catalysts and different classes of phosphorus-based ligands. The knowledge scalability of Fast-Cat was demonstrated through reactor benchmarking, which is crucial for the fine and specialty chemical industries. Additionally, the Fast-Cat lab demonstrated that the knowledge it discovered can be transferred directly from a flow reactor to a batch reactor with consistent reaction results, providing a bridge for knowledge transfer and scalability between lab-scale discovery and industrial-scale production.

Despite significant progress in MOO for continuous reaction systems, several challenges remain notable. Bayesian Optimization, being based on black-box models, can only provide numerical

descriptions of the Pareto front for multiple objectives. However, in actual reaction systems, different objectives are closely related. For instance, reaction conversion and selectivity are often associated with the thermodynamic/dynamic properties of the reaction. Understanding the these properties of new reactions is also valuable. However, kinetic experiments require "exploration", which involves considering a wide range of parameter combinations, especially those with lower reaction extents; in contrast, Bayesian Optimization emphasizes "exploitation", focusing experiments around the best parameter combinations. Therefore, balancing "exploration" and "exploitation" to achieve rapid optimization while comprehensively obtaining the kinetic characteristics of the reaction would be beneficial. In single-objective optimization tasks, many strategies have been proven effective in balancing "exploration" and "exploitation", such as Expected Improvement (EI) [122,123] and Knowledge Gradient (KG) [124]. However, for MOO in chemical synthesis systems, there is a lack of sufficient research. Such work could enable the optimization process to achieve engineering goals quickly while accumulating fundamental information about the reaction itself.

3.2. Continuous flow process optimization based on other algorithms

For continuous flow process parameters, there are also different process optimization algorithms for different problems for reference. For example, in 2023, Shih and colleagues [125] proposed a new method that combines the design of experiments and numerical methods to accelerate experimental optimization on digital microfluidics (DMF). The research team combined a one-factor experimental technique (OFAT) with machine learning algorithms to provide recommended optimal conditions without the need to perform a large number of experiments. With OFAT and machine learning techniques, the researchers were able to significantly reduce the number of experiments performed on the DMF device and identify important parameters that affect yield. The study also demonstrates how the OFAT method and machine learning model can be used to improve the synthesis efficiency of [18F]FDG radiotracer, which not only improves the synthesis yield but also reduces the synthesis time, which is important for clinical applications.

Many people have an allergic reaction to penicillin-based or cephalosporin-based antibiotics, which is often due to trace impurities generated during production, storage, and transportation [126–128]. To investigate the factors responsible for the generation of polymer impurities during the synthesis of cefazolin, the synthesis process data were analyzed using a random forest (RF) algorithm by Wang *et al.* [129]. By analyzing the significance of the features, the results pointed out that alkaline conditions have an important influence on sensitizing impurity generation, and this influence was also verified by DFT calculations and experiments. Therefore, the researchers used a membrane dispersion microreactor for effective dispersion of the more alkaline phase of the reaction to reduce the amount of polymer impurities.

The works mentioned in this section did not utilize the conventional Bayesian optimization workflow. In fact, the kernel function of Bayesian Optimization, as a pure black-box model, fails to effectively integrate the empirical knowledge of chemists, resulting in poor transferability between different tasks. Due to the language understanding capabilities of LLMs, their use in optimizing reaction conditions has been anticipated. Gomes *et al.* attempted this for reactions such as Suzuki and Sonogashira couplings [130]. In this work, the large language model was employed to analyze literature, formulate experimental plans, and drive execution-layer hardware. Experimental results were fed back to the model in a "game outcome" format to support its decision-making. The optimization results showed that compared to the standard GPT-4, a GPT-4 incorporating prior knowledge provided better initial guess parameters. Typically, in Bayesian Optimization, these initial guesses are randomly sampled within a reasonable range, often resulting in highly unreasonable values. Although this work was not based on continuous flow or microchannel systems, its conclusions are highly insightful.

3.3. Microchannel-based high-throughput synthesis experiments

The Buchwald-Hartwig reaction for C–N bond coupling is known for its broad applicability [131–139]. However, the reaction involves the use of organic ligands, and the choice of organic ligands is often a difficult one for different substrates. This is due to the complexity of the reaction mechanism and the numerous intermediate influences, which makes it difficult to prescribe targeted ligands through a rational design approach. This leads to the fact that performing the reaction often requires constant trial and error to find the right ligand. To address this issue, in 2018, Doyle *et al.* [22] reported a method for high-throughput experiments based on the well plate droplet method and yield prediction using random forests for the Hartwig-Buchwald reaction for C–N bond coupling. This study utilized a machine learning approach to predict the yields of cross-coupling reactions of aryl halides with 4-methylaniline in the presence of different potential inhibitor additives from data obtained from high-throughput experiments. The researchers developed scripts to compute and extract atomic, molecular, and vibrational descriptors and trained a random forest algorithm model for predicting yields using these descriptors as inputs and reaction yields as outputs. This work laid the groundwork for subsequent high-throughput synthesis and yield prediction methods based on the basic model of "high-throughput experiment-build quantum chemical descriptors-machine learning regression to predict yields." Doyle *et al.* used a droplet experimental setup with a well plate provided by Merck; the advantages of the droplet reaction are that it consumes a small amount of reagents and has a high synthetic throughput, but it also faces the problem of high evaporation. The advantages of the droplet reaction are low reagent consumption and high synthetic throughput, but it also faces the problems of high evaporation (thus necessitating the use of high-boiling solvents) and incompatibility with multiphase reactions, so there is a lot of work on high throughput experiments based on microchemical systems.

In 2024, Fang *et al.* [140] developed a robotic system for ultra-high-throughput chemical synthesis, online characterization, and large-scale conditional screening of photocatalytic reactions based on the liquid core waveguide, microfluidic liquid handling, and artificial intelligence technologies. The system is capable of performing automated reactant mixing preparation, rapid changeover, introduction, ultrafast photocatalytic reactions, online spectroscopic detection, and screening of different reaction conditions. The team applied the system for large-scale screening of photocatalytic [2 + 2] cycloaddition reactions with 12000 reaction conditions, achieving ultra-high throughput of up to 10000 reaction conditions per day. AI-assisted cross-substrate/photocatalyst prediction was also performed based on the collected data. Unlike the work of Doyle *et al.* which achieved the incorporation of reaction conditions into machine learning due to the use of a continuous flow system to standardize the reaction conditions, Doyle *et al.* laid out more than such conditions.

The most significant advantage of microchemical systems is the narrow concentration difference, temperature field, and velocity field distributions, which effectively decouples the transfer and reaction. The results of synthetic reaction experiments performed

in such a case can effectively reflect the kinetics of the intrinsic reaction, thus enabling the study of the reaction mechanism. For example, the study by Wang *et al.* [141] solved the partial differential equations of a one-dimensional axial diffusion model on the open-source platform FEniCS for the first time to explore the reaction mechanism of a photocatalytic reaction. The research team constructed an automated platform including a photochemical microreactor, a continuously controlled pump, a high-power UV-LED light source, an online visible light absorption analysis unit, and a Raspberry Pi-based control unit. The platform is capable of steady-state feeding and sampling functions. Two model reactions, homogeneous photolysis of methylene blue and photo-Favorskii rearrangement synthesis of ibuprofen, were validated by using a genetic algorithm-based symbolic regression approach, which provided insights into the reaction mechanism and benefits for process optimization.

One of the biggest advantages of an intelligent synthesis system over manual labor is the ease with which a large number of permutations can be implemented. For a dataset that already contains a scheme of four attempts of ABCD reactants, adding an E reactant means manually ordering it and adding it to the synthesis system, and thus is not possible to achieve machine closure; however, in the case of a dataset that has already attempted A-B\A-C\D-B, the addition of an attempt at D-C does not require the involvement of a human being, and thus the closed-loop optimization system has great potential for large numbers of permutation screening has great potential. At the same time, since the reacting molecules change, the target value for screening at this point is often no longer the yield, but rather the functional differences between the different molecules, such as optical properties, catalytic properties, and so on. For example, in 2023, Jensen and colleagues [142] reported on an autonomous chemical discovery platform that enables a closed-loop iterative process from prediction to measurement and back again by integrating machine learning generative algorithms, property prediction, computer-aided synthetic planning (CASP), robotics, and automated chemical synthesis, purification, and characterization. In the study, the platform explores the chemical space and optimizes dye properties through an iterative design-make-test-analyze (DMTA) cycle using the molecular framework of organic dyes. The platform can automatically characterize three key properties of the target molecule: maximum absorption wavelength (λ_{max}), octanol-water partition coefficient ($\lg K_{ow}$), and photo-oxidative stability ($\lg K_{deg}$), and feeds back the measurements to re-train the property prediction model. The results demonstrate the platform's ability to explore chemical space and utilize known chemical structures without the need for manual experimentation, providing a new way to accelerate molecular discovery. In 2021, Reis *et al.* [143] proposed a new strategy for computer-guided materials discovery combining automated flow synthesis and machine learning methods. The team developed a software-controlled continuous polymer synthesis platform that synthesized 397 unique copolymer compositions in a six-variable composition space through iterative experimental-computational cycles. Machine learning identified non-intuitive design criteria that explored only <0.9% of the overall composition space to identify over 10 copolymer compositions that outperformed existing materials. This study not only represents a breakthrough in the discovery of 19F magnetic resonance imaging (MRI) agents but also provides new ideas and tools for structure-property relationship prediction in high-dimensional polymer science.

Despite advancements in high-throughput reactions based on continuous flow systems, several challenges remain to be addressed. Currently, the combination of high-throughput experiments with AI analysis has been well-developed. However, these high-throughput experiments are still predominantly based on batch processes rather than continuous flow reactions. The primary reasons may include: (1) Many synthesis reactions of new molecules and materials are unreported chemical reactions. Due to their unknown reaction characteristics, the application of continuous flow may lead to engineering issues such as blockage [43]; (2) For continuous flow reactions, it is crucial to ensure that parallel experiments under different conditions are not affected. Therefore, when switching reaction conditions, it is often necessary to run for 2–3 times or even longer residence times to reach a steady state, which significantly reduces the experimental throughput [113]; (3) Although microchannel reactors theoretically allow for extensive parallel integration to achieve multi-channel reactions, the associated pumps, detectors, and other equipment are difficult to integrate in parallel at a low cost. If these challenges can be effectively resolved, the application of microchannel reactors and continuous flow methods in the high-throughput synthesis of new molecules and materials can be greatly expanded.

4. Summary and Outlook

This work compiles the current main technologies for smart chemical synthesis based on microchemical technology and the current research status for specific synthetic systems. For the basic technologies of smart chemical synthesis, researchers have made many important advances in the design of synthetic equipment, prediction of dispersion state, and digitization of synthetic processes. For intelligent chemical synthesis for specific reaction systems, researchers have explored the chemical reaction space from the perspectives of Bayesian process optimization and high-throughput data-driven synthesis. Although many important results have been carried out by previous researchers, there are still many areas in which researchers should try to realize a wider range of intelligent chemical synthesis.

First, from the perspective of synthetic equipment development, how to expand the design domain of reactor intelligent design is an important direction due to the complexity of chemical reaction parameter space. Because the work of learning how to apply and deploy an intelligent design is time-costly for downstream applicators, the broader its design domain and the wider its application means that the time cost paid by synthetic chemists to learn about the reactor design system is more valuable, and such work can be more widely applied.

The intelligent design of dispersive systems is also a meaningful attempt to incorporate more factors into the system of machine learning in order to build a more generalized model. In practice, the shape of the channel and the force between the fluid and the reactor wall are often difficult to achieve standard parameterization, which often leads to poor homogeneity between different decentralized system design efforts. If empirical data can be combined with analyzed physical parameters, it may be possible to achieve a more general prediction of dispersion effects.

In the optimization of process parameters for reactions, how to achieve non-homogeneous phase reactions, especially those involving solid-phase catalytic reactions, remains an important topic. For continuous reactions, processes with a single mobile phase are often stable and controllable; however, in actual industrial production, a large number of chemical processes are based on solid-phase catalysts due to advantages in cost and separation. Solid-phase catalysts tend to participate in catalytic reactions with their surfaces; however, the surface properties are not easy to quantitatively parameterize, and thus stable regulation is not easily achieved. Although there are many methods to characterize the surface reaction and adsorption properties of solids in the field of catalysis, such as solid-state NMR, XAFS, *etc.*; however, it is still challenging to achieve *in-situ* online monitoring and precise

regulation. Meanwhile, although the reuse of catalysts is often desired in industry, however, the state of reused catalysts can be different from that of the initial use, which introduces new variables to the reaction that are not easy to parameterize.

As for high-throughput reactions combined with machine learning model prediction, how to build template-less or even template-free machine learning models is also a future development direction. Due to the complexity of the multi-scale chemical reaction space, it is impossible to realize the full space search, so the existing machine learning models focus on the subspace for a certain type of reaction. For example, in a C–N bond coupling reaction, the focus is on the catalyst ligand, the type of base, *etc.*, while a photocatalytic reaction might focus on the light intensity and waveband, etc. However, chemical reactions are all recombinations and interactions of atoms and electrons in their microscopic nature, and all follow basic transfer equations in their mesoscopic dimensions, so that migration of information and knowledge between different chemical reactions is theoretically possible, and in fact, this migration has been realized by the chemist's brain. Therefore, how to break the template barriers between the knowledge of different reaction systems, building a more generalized machine learning model, and constructing a more generalized reaction dataset based on microchemical systems that contain transfer letter features is a direction of far-reaching significance to the field of intelligent synthesis.

Recently, LLMs have demonstrated exceptional text processing capabilities. Consequently, numerous efforts have been made to apply LLM-based agents to enhance the efficiency of synthetic work [144,145]. Yaghi *et al.* [146–150] utilized GPT-4 to extract critical information regarding frameworks chemicals synthesis from literature, which was then used to design synthesis conditions for unknown crystals. Gomes *et al.* [15,130] employed LLMs to extract literature, design experimental protocols, and write programs to drive robotic laboratories for cross-coupling reactions. Mo *et al.* [83] designed multiple agents for the entire process of organic synthesis, including experimental design, online detection, kinetic experiment optimization, purification, and scale-up, to accelerate the development of synthetic processes. Wang *et al.* [151,152] used LLMs to extract literature information and applied it to the construction of knowledge graphs in fields such as materials development and biochemistry. Although Wang *et al.* [58] applied LLMs to microchannel reactor design, there is currently a lack of work combining robotic platforms, continuous flow reactions, and LLMs. Gomes *et al.*'s work indicates that LLMs with prior knowledge possess the ability to optimize process conditions more rapidly. Therefore, applying them to the field of continuous flow synthesis conditions, replacing or collaborating with Bayesian Optimization's black-box model, might enable more efficient process development. Additionally, the optimization process of LLMs with prior knowledge may offer better interpretability, aiding in breaking down data barriers between different synthetic reactions.

Finally, whether it is equipment design and manufacturing or intelligent synthesis for specific reaction systems, how to introduce reaction scale information into machine learning models is an important issue for the industrialization of related work. Because the process of reaction from pilot to pilot to scale-up production is a very complex subject, often highly dependent on the design and experimental experience of the engineering team, and at the same time its trial and error is often very expensive. In the process of industrialized production, what needs to be designed is also not just a reactor core, but a whole set of reaction systems. Therefore, if data-driven process amplification can be realized, the distance between laboratory and industrialization will be shortened to a great extent, and it will also greatly accelerate the transformation of new experimental results on the ground, which will make the industrial production more flexible and customized, and better adapt to the growing new demand for chemicals from the people and downstream enterprises.

CRediT Authorship Contribution Statement

Yongqi Pan: Writing – original draft, Investigation, Conceptualization. Yazi Yu: Writing – original draft, Investigation. Lijie Wang: Investigation. Guogang Hu: Investigation. Yujun Wang: Supervision, Resources. Guangsheng Luo: Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (22378227) and Shijiazhuang Science and Technology Bureau (231790163A).

References

[1] Q.Q. Xue, B.H. Yan, Y.J. Wang, G.S. Luo, Continuous synthesis of atomically dispersed Rh supported on $MgAl_2O_4$ using two-stage microreactor, *AIChE J.* 68 (11) (2022) e17841.
[2] Q. Chen, S.T. Xia, Y.J. Wang, G.S. Luo, H.Y. Shang, K. Wang, Continuous synthesis of 1-ethoxy-2, 3-difluoro-4-iodo-benzene in a microreactor system and the Gaussian and computational fluid dynamics simulations, *AIChE J.* 67 (6) (2021) e17217.
[3] S.T. Xia, X.F. Ding, Y.J. Wang, G.S. Luo, Z.T. Wu, Y.F. Cheng, B.Y. Hao, Large-scale synthesis of dihydrostreptomycin *via* hydrogenation of streptomycin in a membrane dispersion microreactor, *Chem. Eng. J.* 334 (2018) 2250–2254.
[4] S.T. Xia, X.F. Ding, Y.J. Wang, G.S. Luo, Continuous-flow synthesis of an important liquid-crystal intermediate using a microreaction system, *Ind. Eng. Chem. Res.* 57 (36) (2018) 12113–12121.
[5] S. Urgaonkar, J. Verkade, Palladium/proazaphosphatrane-catalyzed amination of aryl halides possessing a phenol, alcohol, acetanilide, amide or an enolizable ketone functional group: efficacy of lithium bis(trimethylsilyl) amide as the base, *Adv. Synth. Catal.* 346 (6) (2004) 611–616.
[6] D.J. Koza, Y.A. Nsiah, Palladium catalyzed C-N bond formation in the synthesis of 7-amino-substituted tetracyclines, *J. Org. Chem.* 67 (14) (2002) 5025–5027.
[7] C.W. Coley, W.H. Green, K.F. Jensen, Machine learning in computer-aided synthesis planning, *Acc. Chem. Res.* 51 (5) (2018) 1281–1289.
[8] J.A. Keith, V. Vassilev-Galindo, B.Q. Cheng, S. Chmiela, M. Gastegger, K.R. Müller, A. Tkatchenko, Combining machine learning and computational chemistry for predictive insights into chemical systems, *Chem. Rev.* 121 (16) (2021) 9816–9872.
[9] H.X. Mai, T.C. Le, D.H. Chen, D.A. Winkler, R.A. Caruso, Machine learning for electrocatalyst and photocatalyst design and discovery, *Chem. Rev.* 122 (16) (2022) 13478–13515.
[10] M. Christensen, L.P.E. Yunker, P. Shiri, T. Zepel, P.L. Prieto, S. Grunert, F. Bork, J.E. Hein, Automation isn't automatic, *Chem. Sci.* 12 (47) (2021) 15473–15490.
[11] E.N. Muratov, J. Bajorath, R.P. Sheridan, I.V. Tetko, D. Filimonov, V. Poroikov, T.I. Oprea, I.I. Baskin, A. Varnek, A. Roitberg, O. Isayev, S. Curtarolo, D. Fourches, Y. Cohen, A. Aspuru-Guzik, D.A. Winkler, D. Agrafiotis, A. Cherkasov, A. Tropsha, QSAR without borders, *Chem. Soc. Rev.* 49 (11) (2020) 3525–3564.
[12] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, *Nature* 559 (7715) (2018) 547–555.
[13] X. Hong, Q. Yang, K.B. Liao, J.F. Pei, M. Chen, F.Y. Mo, H. Lu, W.B. Zhang, H.S. Zhou, J.X. Chen, L.B. Su, S.Q. Zhang, S.Y. Liu, X. Huang, Y.Z. Sun, Y.X. Wang, Z. X. Zhang, Z.Z. Yu, S.Z. Luo, X.F. Fu, S.L. You, AI for organic and polymer synthesis, *Sci. China Chem.* 67 (8) (2024) 2461–2496.
[14] M. Christensen, Y. Xu, E.E. Kwan, M.J. Di Maso, Y. Ji, M. Reibarkh, A.C. Sun, A. Liaw, P.S. Fier, S. Grosser, J.E. Hein, Dynamic sampling in autonomous process optimization, *Chem. Sci.* 15 (19) (2024) 7160–7169.
[15] M. Christensen, L.P.E. Yunker, F. Adedeji, F. Häse, L.M. Roch, T. Gensch, G. Dos Passos Gomes, T. Zepel, M.S. Sigman, A. Aspuru-Guzik, J.E. Hein, Data-science driven autonomous process optimization, *Commun. Chem.* 4 (1) (2021) 112.
[16] E. King-Smith, S. Berritt, L. Bernier, X.J. Hou, J.L. Klug-McLeod, J. Mustakis, N. W. Sach, J.W. Tucker, Q.Y. Yang, R.M. Howard, A.A. Lee, Probing the chemical

'reactome' with high-throughput experimentation data, *Nat. Chem.* 16 (4) (2024) 633–643.
[17] J.T. Rapp, B.J. Bremer, P.A. Romero, Self-driving laboratories to autonomously navigate the protein fitness landscape, *Nat. Chem. Eng.* 1 (1) (2024) 97–107.
[18] S.W. Li, L.C. Xu, C. Zhang, S.Q. Zhang, X. Hong, Reaction performance prediction with an extrapolative and interpretable graph model based on chemical knowledge, *Nat. Commun.* 14 (1) (2023) 3569.
[19] Q. Zhu, F. Zhang, Y. Huang, H.Y. Xiao, L.Y. Zhao, X.C. Zhang, T. Song, X.S. Tang, X. Li, G. He, B.C. Chong, J.Y. Zhou, Y.H. Zhang, B.C. Zhang, J.Q. Cao, M. Luo, S. Wang, G.L. Ye, W.J. Zhang, X. Chen, S. Cong, D.L. Zhou, H.R. Li, J.L. Li, G. Zou, W.W. Shang, J. Jiang, Y. Luo, An all-round AI-chemist with a scientific mind, *Natl. Sci. Rev.* 9 (10) (2022) nwac190.
[20] N. Gesmundo, K. Dykstra, J.L. Douthwaite, Y.T. Kao, R.H. Zhao, B. Mahjour, R. Ferguson, S. Dreher, B. Sauvagnat, J. Saurí, T. Cernak, Miniaturization of popular reactions from the medicinal chemists' toolbox for ultrahigh-throughput experimentation, *Nat. Synth.* 2 (2023) 1082–1091.
[21] H.T. Zhao, W. Chen, H. Huang, Z.H. Sun, Z.J. Chen, L.J. Wu, B.C. Zhang, F.M. Lai, Z. Wang, M.L. Adam, C.H. Pang, P.K. Chu, Y. Lu, T. Wu, J. Jiang, Z.Y. Yin, X.F. Yu, A robotic platform for the synthesis of colloidal nanocrystals, *Nat. Synth.* 2 (2023) 505–514.
[22] D.T. Ahneman, J.G. Estrada, S.S. Lin, S.D. Dreher, A.G. Doyle, Predicting reaction performance in C-N cross-coupling using machine learning, *Science* 360 (6385) (2018) 186–190.
[23] Q. Chen, S. Ullah, Y.J. Wang, G.S. Luo, Controllable generation of multi-row bubbly flow in coupled microstructures and its boosting of photocatalytic decarboxylative carbonylation, *Chem. Eng. J.* 464 (2023) 142758.
[24] Q. Chen, Y.J. Wang, G.S. Luo, Visible-light-driven direct decarboxylative carbonylation of carboxylic acids using acridine photocatalysis in oxygen-liquid flow, *Chem. Eng. J.* 461 (2023) 141767.
[25] Z.Y. Yu, C.Y. Guo, X.M. Pang, Y.G. Shen, M.T. Gao, S.Y. Zhao, Y.J. Wang, G.S. Luo, Coprecipitation synthesis of large-pore-volume γ-alumina nanofibers by two serial membrane dispersion microreactors with a circulating continuous phase, *Ind. Eng. Chem. Res.* 62 (3) (2023) 1415–1424.
[26] M.Z. Guo, S.Q. Bai, Y.J. Wang, G.S. Luo, Novel microfabricated nozzle array with grooves for microdroplet generation, *Chem. Eng. J.* 416 (2021) 129103.
[27] L. Sheng, L. Ma, Y.C. Chen, J. Deng, G.S. Luo, A comprehensive study of droplet formation in a capillary embedded step T-junction: from squeezing to jetting, *Chem. Eng. J.* 427 (2022) 132067.
[28] M. Pathak, Numerical simulation of membrane emulsification: effect of flow properties in the transition from dripping to jetting, *J. Membr. Sci.* 382 (1–2) (2011) 166–176.
[29] P.M. Korczyk, V. van Steijn, S. Blonski, D. Zaremba, D.A. Beattie, P. Garstecki, Accounting for corner flow unifies the understanding of droplet formation in microfluidic channels, *Nat. Commun.* 10 (2019) 2528.
[30] S. Ullah, Y.Q. Pan, Q.Q. Xue, T.Y. Huang, Y.Z. Hu, R.P. Ye, Y.J. Wang, G.S. Luo, The structure activity relationship of promoted La doped Ni-CeO_2 catalysts prepared by continuous-flow microreactor for low temperature CO_2 methanation, *Fuel (Guildf.)* 379 (2025) 133034.
[31] X.P. Luo, L.H. Du, F. He, C.H. Zhou, Controllable regioselective acylation of flavonoids catalyzed by lipase in microreactors, *J. Carbohydr. Chem.* 32 (7) (2013) 450–462.
[32] Q.L. Xu, G.S. Li, F.C. Zhu, N.D. Chen, C.W. Chen, Z.Q. Yu, Continuous synthesis of N-(3-Amino-4-methylphenyl)benzamide and its kinetics study in microflow system, *J. Flow Chem.* 12 (3) (2022) 317–325.
[33] H.L. Yi, S.L. Lu, J.J. Wu, Y.J. Wang, G.S. Luo, Parallelized microfluidic droplet generators with improved ladder–tree distributors for production of monodisperse γ-Al_2O_3 microspheres, *Particuology* 62 (2022) 47–54.
[34] F.Y. Zhao, R.H. Yang, J.X. Ma, Y. Gao, Y.J. Wang, G.S. Luo, Alumina microspheres for the adsorption of fatty alcohols containing oxygenates in Fischer-Tropsch synthetic oils, *Sep. Purif. Technol.* 326 (2023) 124593.
[35] J.R. Naber, S.L. Buchwald, Packed-bed reactors for continuous-flow C-N cross-coupling, *Angew. Chem., Int. Ed.* 49 (49) (2010) 9469–9474.
[36] M. Chen, S.L. Buchwald, Continuous-flow synthesis of 1-substituted benzotriazoles from chloronitrobenzenes and amines in a C-N bond formation/hydrogenation/diazotization/cyclization sequence, *Angew. Chem., Int. Ed.* 52 (15) (2013) 4247–4250.
[37] S. Roesner, S.L. Buchwald, Continuous-flow synthesis of biaryls by negishi cross-coupling of fluoro- and trifluoromethyl-substituted (hetero)arenes, *Angew. Chem., Int. Ed.* 55 (35) (2016) 10463–10467.
[38] T. Noël, S.L. Buchwald, Cross-coupling in flow, *Chem. Soc. Rev.* 40 (10) (2011) 5010–5029.
[39] W.C. Fu, T.F. Jamison, Modular continuous flow synthesis of imatinib and analogues, *Org. Lett.* 21 (15) (2019) 6112–6116.
[40] J. Vicente, I. Saura-Llamas, M.J. Oliva-Madrid, J.A. García-López, D. Bautista, A new method for high-yield cyclopalladation of primary and secondary amines. Atom-efficient open-to-air inexpensive synthesis of Buchwald-type precatalysts, *Organometallics* 30 (17) (2011) 4624–4631.
[41] P. Yaseneva, P. Hodgson, J. Zakrzewski, S. Falß, R.E. Meadows, A.A. Lapkin, Continuous flow Buchwald–Hartwig amination of a pharmaceutical intermediate, *React. Chem. Eng.* 1 (2) (2016) 229–238.
[42] S. Ullah, T.Y. Huang, Y.Q. Pan, Q.Q. Xue, Z.Y. Yu, Y.Z. Hu, S.M. Ahmed, R.P. Ye, Y.J. Wang, G.S. Luo, Ceria-modified high pore volume Ni/Al_2O_3 spheres for enhanced low-temperature CO_2 methanation, *Fuel (Guildf.)* 390 (2025) 134736.
[43] G.M. Whitesides, The origins and the future of microfluidics, *Nature* 442 (7101) (2006) 368–373.
[44] C.W. Coley, D.A. Thomas 3rd, J.A.M. Lummiss, J.N. Jaworski, C.P. Breen, V. Schultz, T. Hart, J.S. Fishman, L. Rogers, H.Y. Gao, R.W. Hicklin, P.P. Plehiers, J. Byington, J.S. Piotti, W.H. Green, A. John Hart, T.F. Jamison, K.F. Jensen, A robotic platform for flow synthesis of organic compounds informed by AI planning, *Science* 365 (6453) (2019) eaax1566.
[45] Y. Xie, G.M. Huang, Y.J. Wang, Z.R. Yan, X. Wang, J. Huang, M.T. Gao, W.Y. Fei, G.S. Luo, Synthesis of piperacillin with low impurity content using a new three-feed membrane dispersion microreactor, *Chem. Eng. J.* 387 (2020) 124178.
[46] H.C. Shi, Y.M. Shi, Z.Q. Liang, S.L. Zhao, B. Qiao, Z. Xu, L.J. Wang, D.D. Song, Machine learning-enabled discovery of multi-resonance TADF molecules: unraveling PLQY predictions from molecular structures, *Chem. Eng. J.* 494 (2024) 153150.
[47] H.X. Guo, G.S. Jiang, B.W. Diao, J.J. Du, W. Sun, J.L. Fan, X.J. Peng, An integrated screening approach for designing efficient thermally activated delayed fluorescent materials for OLEDs, *J. Mater. Chem. C* 12 (36) (2024) 14515–14522.
[48] R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T.D. Hirzel, D. Duvenaud, D. Maclaurin, M.A. Blood-Forsythe, H.S. Chae, M. Einzinger, D.G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S.I. Hong, M. Baldo, R.P. Adams, A. Aspuru-Guzik, Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach, *Nat. Mater.* 15 (10) (2016) 1120–1127.
[49] B.P. MacLeod, F.G.L. Parlane, C.C. Rupnow, K.E. Dettelbach, M.S. Elliott, T.D. Morrissey, T.H. Haley, O. Proskurin, M.B. Rooney, N. Taherimakhsousi, D.J. Dvorak, H.N. Chiu, C.E.B. Waizenegger, K. Ocean, M. Mokhtari, C.P. Berlinguette, A self-driving laboratory advances the Pareto front for material properties, *Nat. Commun.* 13 (1) (2022) 995.
[50] S.D. Xu, J.L. Li, P.F. Cai, X.L. Liu, B. Liu, X.N. Wang, Self-improving photosensitizer discovery system *via* Bayesian search with first-principle simulations, *J. Am. Chem. Soc.* 143 (47) (2021) 19769–19777.
[51] J.K. Sun, R. Tu, Y.C. Xu, H.Y. Yang, T. Yu, D. Zhai, X.Q. Ci, W.Q. Deng, Machine learning aided design of single-atom alloy catalysts for methane cracking, *Nat. Commun.* 15 (1) (2024) 6036.
[52] M. Zhong, K. Tran, Y.M. Min, C.H. Wang, Z.Y. Wang, C.T. Dinh, P. De Luna, Z.Q. Yu, A.S. Rasouli, P. Brodersen, S. Sun, O. Voznyy, C.S. Tan, M. Askerka, F.L. Che, M. Liu, A. Seifitokaldani, Y.J. Pang, S.C. Lo, A. Ip, Z. Ulissi, E.H. Sargent, Accelerated discovery of CO_2 electrocatalysts using active machine learning, *Nature* 581 (7807) (2020) 178–183.
[53] X. Song, P.X. Pu, H.S. Feng, H. Ding, Y. Deng, Z. Ge, S.Q. Zhao, T.Y. Liu, Y.S. Yang, M. Wei, X. Zhang, Integrating active learning and DFT for fast-tracking single-atom alloy catalysts in CO_2-to-fuel conversion, *ACS Appl. Mater. Interfaces* 16 (41) (2024) 55416–55428.
[54] K. Tran, Z.W. Ulissi, Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution, *Nat. Catal.* 1 (2018) 696–703.
[55] X.H. Ge, J. Yin, Z.H. Ren, K.L. Yan, Y.D. Jing, Y.Q. Cao, N.N. Fei, X. Liu, X.N. Wang, X.G. Zhou, L.W. Chen, W.K. Yuan, X.Z. Duan, Atomic design of alkyne semihydrogenation catalysts *via* active learning, *J. Am. Chem. Soc.* 146 (7) (2024) 4993–5004.
[56] P.P. Zhang, J. Eun, M. Elkin, Y.Z. Zhao, R.L. Cantrell, T.R. Newhouse, A neural network model informs the total synthesis of clovane sesquiterpenoids, *Nat. Synth.* 2 (2023) 527–534.
[57] J.A. Bennett, N. Orouji, M. Khan, S. Sadeghi, J. Rodgers, M. Abolhasani, Autonomous reaction Pareto-front mapping with a self-driving catalysis laboratory, *Nat. Chem. Eng.* 1 (2024) 240–250.
[58] Y.Q. Pan, Q.F. Xiao, F.Y. Zhao, Z.H. Li, J.Y. Liu, S. Ullah, K.H. Lim, T.Y. Huang, Z. Y. Yu, C. Li, D.Y. Zhang, Q.Q. Xue, Q. Chen, S. Kawi, Y.J. Wang, G.S. Luo, Chat-microreactor: a large-language-model-based assistant for designing continuous flow systems, *Chem. Eng. Sci.* 311 (2025) 121567.
[59] Y.F. Shi, Z.X. Yang, S.C. Ma, P.L. Kang, C. Shang, P. Hu, Z.P. Liu, Machine learning for chemistry: basics and applications, *Engineering* 27 (2023) 70–83.
[60] A.J. Myles, R.N. Feudale, Y. Liu, N.A. Woody, S.D. Brown, An introduction to decision tree modeling, *J. Chemom.* 18 (6) (2004) 275–285.
[61] Y.C. Lo, S.E. Rensi, W. Torng, R.B. Altman, Machine learning in chemoinformatics and drug discovery, *Drug Discov. Today* 23 (8) (2018) 1538–1546.
[62] A. Kadiyala, A. Kumar, Applications of Python to evaluate the performance of bagging methods, *Environ. Prog. Sustain. Energy* 37 (5) (2018) 1555–1559.
[63] J.B.O. Mitchell, Machine learning methods in chemoinformatics, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4 (5) (2014) 468–481.
[64] B.Z. Dou, Z.L. Zhu, E. Merkurjev, L. Ke, L. Chen, J. Jiang, Y.Y. Zhu, J. Liu, B.G. Zhang, G.W. Wei, Machine learning methods for small data challenges in molecular science, *Chem. Rev.* 123 (13) (2023) 8736–8780.
[65] A. Jetybayeva, N. Borodinov, A.V. Ievlev, M.I.U. Haque, J. Hinkle, W.A. Lamberti, J.C. Meredith, D. Abmayr, O.S. Ovchinnikova, A review on recent machine learning applications for imaging mass spectrometry studies, *J. Appl. Phys.* 133 (2) (2023) 020702.
[66] B. Niu, Y.H. Jin, L. Lu, K.Y. Fen, L. Gu, Z.S. He, W.C. Lu, Y.X. Li, Y.D. Cai, Prediction of interaction between small molecule and enzyme using AdaBoost, *Mol. Divers.* 13 (3) (2009) 313–320.
[67] Z.W. Ge, S. Feng, C.C. Ma, K. Wei, K. Hu, W.J. Zhang, X.J. Dai, L.F. Fan, J.H. Hua, Quantifying and comparing the effects of key chemical descriptors on

metal–organic frameworks water stability with CatBoost and SHAP, *Microchem. J.* 196 (2024) 109625.
[68] G.B. Goh, N.O. Hodas, A. Vishnu, Deep learning for computational chemistry, *J. Comput. Chem.* 38 (16) (2017) 1291–1307.
[69] M. Pirdashti, S. Curteanu, M.H. Kamangar, M.H. Hassim, M.A. Khatami, Artificial neural networks: applications in chemical engineering, *Rev. Chem. Eng.* 29 (4) (2013) 205–239.
[70] J. Bernal, K. Kushibar, D.S. Asfaw, S. Valverde, A. Oliver, R. Martí, X. Lladó, Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review, *Artif. Intell. Med.* 95 (2019) 64–81.
[71] L.Y. Chen, S.B. Li, Q. Bai, J. Yang, S.L. Jiang, Y.M. Miao, Review of image classification algorithms based on convolutional neural networks, *Remote Sens.* 13 (22) (2021) 4712.
[72] S. Wu, K. Roberts, S. Datta, J.C. Du, Z.C. Ji, Y.Q. Si, S. Soni, Q. Wang, Q. Wei, Y. Xiang, B. Zhao, H. Xu, Deep learning in clinical natural language processing: a methodical review, *J. Am. Med. Inf. Assoc.* 27 (3) (2020) 457–470.
[73] X.P. Qiu, T.X. Sun, Y.G. Xu, Y.F. Shao, N. Dai, X.J. Huang, Pre-trained models for natural language processing: a survey, *Sci. China Technol. Sci.* 63 (10) (2020) 1872–1897.
[74] L.Q. Huang, R.C. Luo, X. Liu, X. Hao, Spectral imaging with deep learning, *Light Sci. Appl.* 11 (1) (2022) 61.
[75] X.Y. Liu, H.L. An, W.S. Cai, X.G. Shao, Deep learning in spectral analysis: modeling and imaging, *Trac. Trends Anal. Chem.* 172 (2024) 117612.
[76] Q.Y. Xu, W.L. Du, J.J. Xu, J.K. Dong, Neural network-based source tracking of chemical leaks with obstacles, *Chin. J. Chem. Eng.* 33 (2021) 211–220.
[77] X.Y. Li, Y.Q. Xu, H.Q. Yao, K.J. Lin, Chemical space exploration based on recurrent neural networks: applications in discovering kinase inhibitors, *J. Chem. Inf.* 12 (1) (2020) 42.
[78] K.D. Luong, A. Singh, Application of transformers in cheminformatics, *J. Chem. Inf. Model.* 64 (11) (2024) 4392–4409.
[79] J. Jiang, L. Ke, L. Chen, B.Z. Dou, Y.Y. Zhu, J. Liu, B.G. Zhang, T.S. Zhou, G.W. Wei, Transformer technology in molecular science, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 14 (4) (2024) e1725.
[80] Y.Q. Han, X.Y. Xu, C.Y. Hsieh, K.Y. Ding, H.X. Xu, R.J. Xu, T.J. Hou, Q. Zhang, H.J. Chen, Retrosynthesis prediction with an iterative string editing model, *Nat. Commun.* 15 (1) (2024) 6404.
[81] J. Achiam OpenAi, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F.L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, R. Avila, I. Babuschkin, S. Balaji, V. Balcom, P. Baltescu, H. Bao, M. Bavarian, J. Belgum, I. Bello, J. Berdine, G. Bernadett-Shapiro, C. Berner, L. Bogdonoff, O. Boiko, M. Boyd, A.-L. Brakman, G. Brockman, T. Brooks, M. Brundage, K. Button, T. Cai, R. Campbell, A. Cann, B. Carey, C. Carlson, R. Carmichael, B. Chan, C. Chang, F. Chantzis, D. Chen, S. Chen, R. Chen, J. Chen, M. Chen, B. Chess, C. Cho, C. Chu, H.W. Chung, D. Cummings, J. Currier, Y. Dai, C. Decareaux, T. Degry, N. Deutsch, D. Deville, A. Dhar, D. Dohan, S. Dowling, S. Dunning, A. Ecoffet, A. Eleti, T. Eloundou, D. Farhi, L. Fedus, N. Felix, S.P. Fishman, J. Forte, I. Fulford, L. Gao, E. Georges, C. Gibson, V. Goel, T. Gogineni, G. Goh, R. Gontijo-Lopes, J. Gordon, M. Grafstein, S. Gray, R. Greene, J. Gross, S.S. Gu, Y. Guo, C. Hallacy, J. Han, J. Harris, Y. He, M. Heaton, J. Heidecke, C. Hesse, A. Hickey, W. Hickey, P. Hoeschele, B. Houghton, K. Hsu, S. Hu, X. Hu, J. Huizinga, S. Jain, S. Jain, J. Jang, A. Jiang, R. Jiang, H. Jin, D. Jin, S. Jomoto, B. Jonn, H. Jun, T. Kaftan, Ł. Kaiser, A. Kamali, I. Kanitscheider, N.S. Keskar, T. Khan, L. Kilpatrick, J.W. Kim, C. Kim, Y. Kim, J.H. Kirchner, J. Kiros, M. Knight, D. Kokotajlo, Ł. Kondraciuk, A. Kondrich, A. Konstantinidis, K. Kosic, G. Krueger, V. Kuo, M. Lampe, I. Lan, T. Lee, J. Leike, J. Leung, D. Levy, C.M. Li, R. Lim, M. Lin, S. Lin, M. Litwin, T. Lopez, R. Lowe, P. Lue, A. Makanju, K. Malfacini, S. Manning, T. Markov, Y. Markovski, B. Martin, K. Mayer, A. Mayne, B. McGrew, S.M. McKinney, C. McLeavey, P. McMillan, J. McNeil, D. Medina, A. Mehta, J. Menick, L. Metz, A. Mishchenko, P. Mishkin, V. Monaco, E. Morikawa, D. Mossing, T. Mu, M. Murati, O. Murk, D. Mély, A. Nair, R. Nakano, R. Nayak, A. Neelakantan, R. Ngo, H. Noh, L. Ouyang, C. O'Keefe, J. Pachocki, A. Paino, J. Palermo, A. Pantuliano, G. Parascandolo, J. Parish, E. Parparita, A. Passos, M. Pavlov, A. Peng, A. Perelman, F. de Avila Belbute Peres, M. Petrov, H.P. de Oliveira Pinto, Michael Pokorny, M. Pokrass, V.H. Pong, T. Powell, A. Power, B. Power, E. Proehl, R. Puri, A. Radford, J. Rae, A. Ramesh, C. Raymond, F. Real, K. Rimbach, C. Ross, B. Rotsted, H. Roussez, N. Ryder, M. Saltarelli, T. Sanders, S. Santurkar, G. Sastry, H. Schmidt, D. Schnurr, J. Schulman, D. Selsam, K. Sheppard, T. Sherbakov, J. Shieh, S. Shoker, P. Shyam, S. Sidor, E. Sigler, M. Simens, J. Sitkin, K. Slama, I. Sohl, B. Sokolowsky, Y. Song, N. Staudacher, F.P. Such, N. Summers, I. Sutskever, J. Tang, N. Tezak, M.B. Thompson, P. Tillet, A. Tootoonchian, E. Tseng, P. Tuggle, N. Turley, J. Tworek, J.F.C. Uribe, A. Vallone, A. Vijayvergiya, C. Voss, C. Wainwright, J.J. Wang, A. Wang, B. Wang, J. Ward, J. Wei, C.J. Weinmann, A. Welihinda, P. Welinder, J. Weng, L. Weng, M. Wiethoff, D. Willner, C. Winter, S. Wolrich, H. Wong, L. Workman, S. Wu, J. Wu, M. Wu, K. Xiao, T. Xu, S. Yoo, K. Yu, Q. Yuan, W. Zaremba, R. Zellers, C. Zhang, M. Zhang, S. Zhao, T. Zheng, J. Zhuang, W. Zhuk, B. Zoph, GPT-4 Technical Report, 2024.
[82] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, E. Hulburd, D.Y. Liu, M.H. Wang, A.G. Catlin, M.Y. Lei, J.Y. Zhang, Y.Y. Ye, BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding, 2018, https://doi.org/10.48550/arXiv.1810.04805.
[83] Y.X. Ruan, C.Y. Lu, N. Xu, Y.C. He, Y.X. Chen, J. Zhang, J. Xuan, J.Z. Pan, Q. Fang, H.Y. Gao, X.D. Shen, N. Ye, Q. Zhang, Y.M. Mo, An automatic end-to-end chemical synthesis development platform powered by large language models, *Nat. Commun.* 15 (1) (2024) 10160.
[84] Y.F. Wu, A. Walsh, A.M. Ganose, Race to the bottom: bayesian optimisation for chemical problems, *Dig. Dis.* 3 (6) (2024) 1086–1100.
[85] A. Slattery, Z. Wen, P. Tenblad, J. Sanjosé-Orduna, D. Pintossi, T. den Hartog, T. Noël, Automated self-optimization, intensification, and scale-up of photocatalysis in flow, *Science* 383 (6681) (2024) eadj1817.
[86] B. Shahriari, K. Swersky, Z.Y. Wang, R.P. Adams, N. de Freitas, Taking the human out of the loop: a review of Bayesian optimization, *Proc. IEEE* 104 (1) (2016) 148–175.
[87] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, J. Thomas, T. Ullmann, M. Becker, A.L. Boulesteix, D.F. Deng, M. Lindauer, Hyperparameter optimization: foundations, algorithms, best practices, and open challenges, *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 13 (2) (2023) e1484.
[88] S. Greenhill, S. Rana, S. Gupta, P. Vellanki, S. Venkatesh, Bayesian optimization for adaptive experimental design: a review, *IEEE Access* 8 (2020) 13937–13948.
[89] K. Terayama, M. Sumita, R. Tamura, K. Tsuda, Black-box optimization for automated discovery, *Acc. Chem. Res.* 54 (6) (2021) 1334–1346.
[90] M. Meuwly, Machine learning for chemical reactions, *Chem. Rev.* 121 (16) (2021) 10218–10239.
[91] E. del Castillo, M.S. Reis, Bayesian predictive optimization of multiple and profile response systems in the process industry: a review and extensions, *Chemometr. Intell. Lab. Syst.* 206 (2020) 104121.
[92] M. Fernandez, J. Caballero, L. Fernandez, A. Sarai, Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM), *Mol. Divers.* 15 (1) (2011) 269–289.
[93] H.T. Yang, J.L. Li, K.Z. Lim, C.J. Pan, T. Van Truong, Q. Wang, K.R. Li, S. Li, X. Xiao, M. Ding, T.L. Chen, X.L. Liu, Q. Xie, P.V.Y. Alvarado, X.N. Wang, P.Y. Chen, Automatic strain sensor design *via* active learning and data augmentation for soft machines, *Nat. Mach. Intell.* 4 (2022) 84–94.
[94] X. Yuan, M. Suvarna, J.Y. Lim, J. Pérez-Ramírez, X. Wang, Y.S. Ok, Active learning-based guided synthesis of engineered biochar for CO_2 capture, *Environ. Sci. Technol.* 58 (15) (2024) 6628–6636.
[95] X. Liu, C. Wang, Z.M. Yin, X. An, H.X. Meng, Risk-informed multi-objective decision-making of emergency schemes optimization, *Reliab. Eng. Syst. Saf.* 245 (2024) 109979.
[96] C.J. Wu, H.Z. Pan, Z.H. Luo, C. Liu, H. Huang, Multi-objective optimization of residential building energy consumption, daylighting, and thermal comfort based on BO-XGBoost-NSGA-II, *Build. Environ.* 254 (2024) 111386.
[97] N. Basha, T. Savage, J. McDonough, E.A. del Rio Chanona, O.K. Matar, Discovery of mixing characteristics for enhancing coiled reactor performance through a Bayesian optimisation-CFD approach, *Chem. Eng. J.* 473 (2023) 145217.
[98] T. Savage, N. Basha, J. McDonough, J. Krassowski, O. Matar, E.A. del Rio Chanona, Machine learning-assisted discovery of flow reactor designs, *Nat. Chem. Eng.* 1 (8) (2024) 522–531.
[99] M.Q. Liu, H.Y. Hu, Y.J. Cui, J. Song, L. Ma, Z.H. Yuan, K. Wang, G.S. Luo, Modeling and analysis of droplet generation in microchannels using interpretable machine learning methods, *Chem. Eng. J.* 511 (2025) 161972.
[100] A. Lashkaripour, C. Rodriguez, N. Mehdipour, R. Mardian, D. McIntyre, L. Ortiz, J. Campbell, D. Densmore, Machine learning enables design automation of microfluidic flow-focusing droplet generation, *Nat. Commun.* 12 (1) (2021) 25.
[101] Y. Chang, L. Sheng, J.J. Wang, J. Deng, G.S. Luo, A general neural network model co-driven by mechanism and data for the reliable design of gas–liquid T-junction microdevices, *Lab Chip* 23 (22) (2023) 4888–4900.
[102] B. Burger, P.M. Maffettone, V.V. Gusev, C.M. Aitchison, Y. Bai, X. Wang, X. Li, B.M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R.S. Sprick, A.I. Cooper, A mobile robotic chemist, *Nature* 583 (7815) (2020) 237–241.
[103] S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J.M. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P.J. Kitson, D. Angelone, L. Cronin, Organic synthesis in a modular robotic system driven by a chemical programming language, *Science* 363 (6423) (2019) eaav2211.
[104] S. Chatterjee, M. Guidi, P.H. Seeberger, K. Gilmore, Automated radial synthesis of organic molecules, *Nature* 579 (7799) (2020) 379–384.
[105] H.R. Zhang, W. Chen, A. Iyer, D.W. Apley, W. Chen, Uncertainty-aware mixed-variable machine learning for materials design, *Sci. Rep.* 12 (2022) 19760.
[106] B.J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J.I.M. Alvarado, J.M. Janey, R.P. Adams, A.G. Doyle, Bayesian reaction optimization as a tool for chemical synthesis, *Nature* 590 (7844) (2021) 89–96.
[107] P. Sagmeister, R. Lebl, I. Castillo, J. Rehrl, J. Kruisz, M. Sipek, M. Horn, S. Sacher, D. Cantillo, J.D. Williams, C. Oliver Kappe, Advanced real-time process analytics for multistep synthesis in continuous flow, *Angew. Chem., Int. Ed.* 60 (15) (2021) 8139–8148.
[108] A.D. Clayton, E.O. Pyzer-Knapp, M. Purdie, M.F. Jones, A. Barthelme, J. Pavey, N. Kapur, T.W. Chamberlain, A. John Blacker, R.A. Bourne, Bayesian self-optimization for telescoped continuous flow synthesis, *Angew. Chem., Int. Ed.* 62 (3) (2023) e202214511.
[109] R.Z. Liu, Z.H. Wang, W.B. Yang, J.Z. Cao, S.Y. Tao, Self-optimizing Bayesian for continuous flow synthesis process, *Digit. Discov.* 3 (10) (2024) 1958–1966.
[110] J.P. Guo, G.H. Luo, K.J. Chai, W.K. Su, A. Su, Continuous flow synthesis of *N, O*-dimethyl-*N*'-nitroisourea monitored by inline Fourier transform infrared spectroscopy: bayesian optimization and kinetic modeling, *Ind. Eng. Chem. Res.* 63 (23) (2024) 10162–10171.
[111] T. Shaw, A.D. Clayton, R. Labes, T.M. Dixon, S. Boyall, O.J. Kershaw, R.A. Bourne, B.C. Hanson, A self-optimised approach to synthesising DEHiBA for

advanced nuclear reprocessing, exploiting the power of machine-learning, *React. Chem. Eng.* 9 (2) (2024) 426–438.
[112] J.H. Dunlap, J.G. Ethier, A.A. Putnam-Neeb, S. Iyer, S.L. Luo, H.S. Feng, J.A. Garrido Torres, A.G. Doyle, T.M. Swager, R.A. Vaia, P. Mirau, C.A. Crouse, L.A. Baldwin, Continuous flow synthesis of pyridinium salts accelerated by multi-objective Bayesian optimization with active learning, *Chem. Sci.* 14 (30) (2023) 8061–8069.
[113] M. Sayes, G. Benoit, A.B. Charette, Borocyclopropanation of styrenes mediated by UV-light under continuous flow conditions, *Angew. Chem., Int. Ed.* 57 (41) (2018) 13514–13518.
[114] J. Santandrea, C. Minozzi, C. Cruché, S.K. Collins, Photochemical dual-catalytic synthesis of alkynyl sulfides, *Angew. Chem., Int. Ed.* 56 (40) (2017) 12255–12259.
[115] S. Overa, B.H. Ko, Y. Zhao, F. Jiao, Electrochemical approaches for CO_2 conversion to chemicals: a journey toward practical applications, *Acc. Chem. Res.* 55 (5) (2022) 638–648.
[116] B. Kim, D. Kwon, J.O. Baeg, A.P. Muthu, G.H. Gu, J.H. Lee, J. Jeong, W. Kim, W. Choi, Dual-atom-site Sn-Cu/C_3N_4 photocatalyst selectively produces formaldehyde from CO_2 reduction, *Adv. Funct. Mater.* 33 (19) (2023) 2212453.
[117] S.Q. Zhao, H.Q. Liang, X.M. Hu, S.M. Li, K. Daasbjerg, Challenges and prospects in the catalytic conversion of carbon dioxide to formaldehyde, *Angew. Chem., Int. Ed.* 61 (46) (2022) e202204008.
[118] A.U. Pawar, U. Pal, J.Y. Zheng, C.W. Kim, Y.S. Kang, Thermodynamically controlled photo-electrochemical CO_2 reduction at Cu/rGO/PVP/Nafion multi-layered dark cathode for selective production of formaldehyde and acetaldehyde, *Appl. Catal. B Environ.* 303 (2022) 120921.
[119] S. Zeng, E. Vahidzadeh, C.G. VanEssen, P. Kar, R. Kisslinger, A. Goswami, Y. Zhang, N. Mahdi, S. Riddell, A.E. Kobryn, S. Gusarov, P. Kumar, K. Shankar, Optical control of selectivity of high rate CO_2 photoreduction *via* interband- or hot electron Z-scheme reaction pathways in Au-TiO_2 plasmonic photonic crystal photocatalyst, *Appl. Catal. B Environ.* 267 (2020) 118644.
[120] H.H. Cramer, S. Das, M.D. Wodrich, C. Corminboeuf, C. Werlé, W. Leitner, Theory-guided development of homogeneous catalysts for the reduction of CO_2 to formate, formaldehyde, and methanol derivatives, *Chem. Sci.* 14 (11) (2023) 2799–2807.
[121] S. Barman, D.S.S. Sreejith, D. Somnath Garai, D. Ramudu Pochamoni, P. Soumyajit Roy, Selective photocatalytic carbon dioxide reduction by a reduced molybdenum-based polyoxometalate catalyst, *ChemPhotoChem* 3 (2) (2019) 93–100.
[122] K. Omata, Screening of new additives of active-carbon-supported heteropoly acid catalyst for Friedel–Crafts reaction by Gaussian process regression, *Ind. Eng. Chem. Res.* 50 (19) (2011) 10948–10954.
[123] S. Shin, H. Song, Y.S. Shin, J. Lee, T.H. Seo, Bayesian optimization of wet-impregnated co-Mo/Al_2O_3 catalyst for maximizing the yield of carbon nanotube synthesis, *Nanomaterials (Basel)* 14 (1) (2023) 75.
[124] S. Chen, K.G. Reyes, M.K. Gupta, M.C. McAlpine, W.B. Powell, Optimal learning in experimental design using the knowledge gradient policy with application to characterizing nanoemulsion stability, *ASA J. Uncertain. Quantif.* 3 (1) (2015) 320–345.
[125] F. Ahmadi, M. Simchi, J.M. Perry, S. Frenette, H. Benali, J.P. Soucy, G. Massarweh, S.C.C. Shih, Integrating machine learning and digital microfluidics for screening experimental conditions, *Lab Chip* 23 (1) (2022) 81–91.
[126] S.H. Jin, J. Jing, C.L. Zheng, Study on the relation between the high molecular weight impurities in cefotaxime and the allergic reaction ii.characterzation of the high molecular weight impurities in cefotaxime, *Chin*, *J. Antibiot.* 11 (6) (1986) 463–467.
[127] M. Iwata, H. Tokiwa, T. Matuhasi, Detection and characterization of polymers in cephalothin by passive cutaneous anaphylaxis in mice, *Int. Arch. Allergy Appl. Immunol.* 70 (2) (1983) 132–137.
[128] Y.R. Hou, Y.Z. Yuan, M. Zhang, W. Qian, T.J. Hang, Review on the research and detection method of high molecular weight impurities in β-lactam antibiotics, *Pharm. Clin. Res.* 20 (3) (2012) 240–244.
[129] Y.Q. Pan, S.Y. Zhao, L.J. Wang, L.B. Zhang, M.D. Yang, B.J. Wei, G.G. Hu, S. Ullah, Y.J. Wang, G.S. Luo, Large-scale production of cefazolin in a microreactor with a low impurity content and high yield, *Chem. Eng. J.* 497 (2024) 154887.
[130] D.A. Boiko, R. MacKnight, B. Kline, G. Gomes, Autonomous chemical research with large language models, *Nature* 624 (7992) (2023) 570–578.
[131] J.P. Wolfe, H. Tomori, J.P. Sadighi, J. Yin, S.L. Buchwald, Simple, efficient catalyst system for the palladium-catalyzed amination of aryl chlorides, bromides, and triflates, *J. Org. Chem.* 65 (4) (2000) 1158–1174.
[132] G.S. Lemen, J.P. Wolfe, Palladium-catalyzed sp2 C–N bond forming reactions: recent developments and applications. Amination and Formation of sp2 C-N Bonds, Springer, Berlin Heidelberg, 2012, 53.
[133] K.W. Anderson, R.E. Tundel, T. Ikawa, R.A. Altman, S.L. Buchwald, Monodentate phosphines provide highly active catalysts for Pd-catalyzed C-N bond-forming reactions of heteroaromatic halides/amines and (H)N-heterocycles, *Angew. Chem., Int. Ed.* 45 (39) (2006) 6523–6527.
[134] S. Sharif, R.P. Rucker, N. Chandrasoma, D. Mitchell, M.J. Rodriguez, R.D.J. Froese, M.G. Organ, Selective monoarylation of primary amines using the Pd-PEPPSI-IPent(Cl) precatalyst, *Angew. Chem., Int. Ed.* 54 (33) (2015) 9507–9511.
[135] D.S. Surry, S.L. Buchwald, Biaryl phosphane ligands in palladium-catalyzed amination, *Angew. Chem., Int. Ed.* 47 (34) (2008) 6338–6361.
[136] U. Christmann, D.A. Pantazis, J. Benet-Buchholz, J.E. McGrady, F. Maseras, R. Vilar, Experimental and theoretical investigations of new dinuclear palladium complexes as precatalysts for the amination of aryl chlorides, *J. Am. Chem. Soc.* 128 (19) (2006) 6376–6390.
[137] S.D. McCann, E.C. Reichert, P.L. Arrechea, S.L. Buchwald, Development of an aryl amination catalyst with broad scope guided by consideration of catalyst stability, *J. Am. Chem. Soc.* 142 (35) (2020) 15027–15037.
[138] Q.L. Shen, T. Ogata, J.F. Hartwig, Highly reactive, general and long-lived catalysts for palladium-catalyzed amination of heteroaryl and aryl chlorides, bromides, and iodides: scope and structure-activity relationships, *J. Am. Chem. Soc.* 130 (20) (2008) 6586–6596.
[139] J.F. Hartwig, M. Kawatsura, S.I. Hauck, K.H. Shaughnessy, L.M. Alcazar-Roman, Room-temperature palladium-catalyzed amination of aryl bromides and chlorides and extended scope of aromatic C-N bond formation with a commercial ligand, *J. Org. Chem.* 64 (15) (1999) 5575–5580.
[140] J.M. Lu, H.F. Wang, Q.H. Guo, J.W. Wang, T.T. Li, K.X. Chen, M.T. Zhang, J.B. Chen, Q.N. Shi, Y. Huang, S.W. Shi, G.Y. Chen, J.Z. Pan, Z. Lu, Q. Fang, Roboticized AI-assisted microfluidic photocatalytic synthesis and screening up to 10, 000 reactions per day, *Nat. Commun.* 15 (1) (2024) 8826.
[141] Y.H. Wang, C. Shen, M. Qiu, M.J. Shang, Y.H. Su, Kinetic study in an automatic continuous-flow photochemical platform with machine learning, *AIChE. J.* 69 (9) (2023) e18102.
[142] B.A. Koscher, R.B. Canty, M.A. McDonald, K.P. Greenman, C.J. McGill, C.L. Bilodeau, W.G. Jin, H.Y. Wu, F.H. Vermeire, B. Jin, T. Hart, T. Kulesza, S.C. Li, T. S. Jaakkola, R. Barzilay, R. Gómez-Bombarelli, W.H. Green, K.F. Jensen, Autonomous, multiproperty-driven molecular discovery: from predictions to measurements and back, *Science* 382 (6677) (2023) eadi1407.
[143] M. Reis, F. Gusev, N.G. Taylor, S.H. Chung, M.D. Verber, Y.Z. Lee, O. Isayev, F.A. Leibfarth, Machine-learning-guided discovery of ^{19}F MRI agents enabled by automated copolymer synthesis, *J. Am. Chem. Soc.* 143 (42) (2021) 17677–17689.
[144] J. Dagdelen, A. Dunn, S. Lee, N. Walker, A.S. Rosen, G. Ceder, K.A. Persson, A. Jain, Structured information extraction from scientific text with large language models, *Nat. Commun.* 15 (1) (2024) 1418.
[145] Y. Kang, J. Kim, ChatMOF: an artificial intelligence system for predicting and generating metal-organic frameworks using large language models, *Nat. Commun.* 15 (1) (2024) 4705.
[146] Z.L. Zheng, O.F. Zhang, H.L. Nguyen, N. Rampal, A.H. Alawadhi, Z.C. Rong, T. Head-Gordon, C. Borgs, J.T. Chayes, O.M. Yaghi, ChatGPT research group for optimizing the crystallinity of MOFs and COFs, *ACS Cent. Sci.* 9 (11) (2023) 2161–2170.
[147] Z.L. Zheng, Z.C. Rong, N. Rampal, C. Borgs, J.T. Chayes, O.M. Yaghi, A GPT-4 reticular chemist for guiding MOF discovery, *Angew. Chem., Int. Ed.* 62 (46) (2023) e202311983.
[148] Z.L. Zheng, Z.G. He, O. Khattab, N. Rampal, M.A. Zaharia, C. Borgs, J.T. Chayes, O.M. Yaghi, Image and data mining in reticular chemistry powered by GPT-4V, *Dig. Dis.* 3 (3) (2024) 491–501.
[149] Z.L. Zheng, A.H. Alawadhi, S. Chheda, S. Ephraim Neumann, N. Rampal, S.C. Liu, H.L. Nguyen, Y.H. Lin, Z.C. Rong, J. Ilja Siepmann, L. Gagliardi, A. Anandkumar, C. Borgs, J.T. Chayes, O.M. Yaghi, Shaping the water-harvesting behavior of metal-organic frameworks aided by fine-tuned GPT models, *J. Am. Chem. Soc.* 145 (51) (2023) 28284–28295.
[150] Z.L. Zheng, O.F. Zhang, C. Borgs, J.T. Chayes, O.M. Yaghi, ChatGPT chemistry assistant for text mining and the prediction of MOF synthesis, *J. Am. Chem. Soc.* 145 (32) (2023) 18048–18062.
[151] H.X. Liu, H.Y. Yin, Z.Y. Luo, X.N. Wang, Integrating chemistry knowledge in large language models *via* prompt engineering, *Synth. Syst. Biotechnol.* 10 (1) (2024) 23–38.
[152] H. Liu, H. Yin, Z. Luo, X. Wang, Integrating chemistry knowledge in large language models *via* prompt engineering, *Synth. Syst. Biotechnol.* 10 (1) (2025) 23–38.