

Review

A Review on Artificial Intelligence Enabled Design, Synthesis, and Process Optimization of Chemical Products for Industry 4.0

Chasheng He ¹, Chengwei Zhang ², Tengfei Bian ³, Kaixuan Jiao ³, Weike Su ¹ , Ke-Jun Wu ^{4,5,*} and An Su ^{2,*} 

- ¹ National Engineering Research Center for Process Development of Active Pharmaceutical Ingredients, Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang University of Technology, Hangzhou 310014, China
- ² College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, China
- ³ Technology Center, China Tobacco Zhejiang Industrial Co., Ltd., Hangzhou 310009, China
- ⁴ Zhejiang Provincial Key Laboratory of Advanced Chemical Engineering Manufacture Technology, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China
- ⁵ Institute of Zhejiang University-Quzhou, Quzhou 324000, China
- * Correspondence: k.wu@zju.edu.cn (K.-J.W.); ansu@zjut.edu.cn (A.S.)

Abstract: With the development of Industry 4.0, artificial intelligence (AI) is gaining increasing attention for its performance in solving particularly complex problems in industrial chemistry and chemical engineering. Therefore, this review provides an overview of the application of AI techniques, in particular machine learning, in chemical design, synthesis, and process optimization over the past years. In this review, the focus is on the application of AI for structure-function relationship analysis, synthetic route planning, and automated synthesis. Finally, we discuss the challenges and future of AI in making chemical products.

Keywords: artificial intelligence; machine learning; automated synthesis; synthetic route planning; structure-function relationship



Citation: He, C.; Zhang, C.; Bian, T.; Jiao, K.; Su, W.; Wu, K.-J.; Su, A. A Review on Artificial Intelligence Enabled Design, Synthesis, and Process Optimization of Chemical Products for Industry 4.0. *Processes* **2023**, *11*, 330. <https://doi.org/10.3390/pr11020330>

Academic Editor: Blaž Likozar

Received: 11 December 2022

Revised: 16 January 2023

Accepted: 17 January 2023

Published: 19 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Chemists have spent substantial time on repetitive experimental tasks, such as the synthesis of organic compounds, optimization of process parameters, and molecular structure identification. To some extent, these tedious tasks limit the creativity of chemists. As green chemistry continues to evolve, the chemical industry has been working to discover new chemical reactions, catalysts, and equipment to reduce the use of hazardous substances and prepare high-value-added chemicals through sustainable production processes. However, such discoveries are expensive and time-consuming for pure human labor [1].

In the past decade, a growing body of literature and patents attest to AI-driven chemical engineering studies. Baum et al. studied the growth and distribution of artificial intelligence in relevant chemical publications over the past two decades using the CAS Content collection (Figure 1) [2]. As shown in Figure 1, the number of published papers and patents containing AI has increased dramatically since 2015, with the second-highest increase in published papers in industrial chemistry and chemical engineering. Meanwhile, data obtained from SciFinder show that the total number of annual publications on machine learning in the chemical industry exceeded 20,000 in 2021–2022.

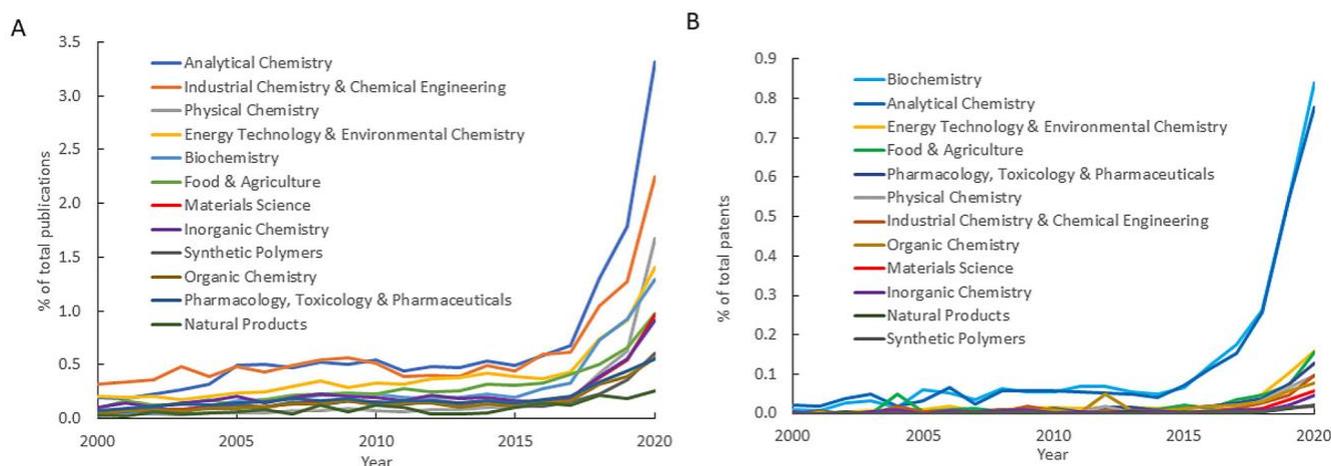


Figure 1. Publication trends for AI in different research areas from 2000 to 2020: (A) journal publications and (B) patent publications. Reproduced with permission from [2].

AI involves several methodological domains, such as reasoning, knowledge representation, solution search, and the basic paradigm of machine learning (ML) among them. In the last few years, especially since the introduction of AlphaGo, ML has been greatly developed in the field of industrial chemistry and chemical engineering, thus greatly helping the development of pharmaceuticals and fine chemicals, thus reducing time and cost [3–5]. So far, much of the literature has summarized the application of machine learning algorithms in the chemical industry (Figure 2) [6]. As shown in Figure 2, supervised learning methods are the most used in the chemical industry, accounting for nearly 70% of the total, while hybrid, unsupervised learning, and combinatorial methods are used much less than supervised learning. Almost all of these machine learning methods are used for data mining and analytics in the chemical industry. The only exception is reinforcement learning, whose applications are currently limited to robotics, gaming, and navigation. Figure 3 depicts in more detail the types of problems solved primarily using supervised methods, namely modeling, optimization, control and monitoring, design and discovery, support to sensorial analysis, and reaction prediction. As for unsupervised methods, they are mostly used for dimensionality reduction, data visualization, and information extraction. Additionally, a subfield of ML is deep learning (DL), which engages deep neural networks (DNNs). DNN constitutes a set of nodes, each of which receives individual inputs and eventually converts them to outputs, either singly or in multiple sessions using algorithms to solve problems. In quantitative structure-activity relationship (QSAR) modeling, deep learning models have achieved state-of-the-art results in molecular property prediction as well as property uncertainty quantification. It is worth noting that the ML-based molecular design approach is different from the mathematical optimization-based approach. Mathematical optimization-based approaches require large amounts of experimental process data, such as reaction rates, which are difficult to obtain with a single form of benchmark. In contrast, machine learning models performing molecular design tasks require only structural information of molecules or simple molecular property information, which are more readily available and accurate than experimental processes. In addition, machine Learning for molecular design has more trainable parameters than mathematical optimization. In general, the more trainable parameters, the higher the accuracy of the trained model given a sufficient amount of data. With the development of Industry 4.0, the successful application of artificial intelligence in areas such as image recognition and text processing has also facilitated its use in drug discovery, including the design optimization of small molecule drugs [7–9]. Key to the development of computer-aided chemistry is the availability of large reaction datasets and high-performance computing, for example, in molecular design, retrosynthetic planning, reaction prediction, and optimization of reaction conditions [10–19].

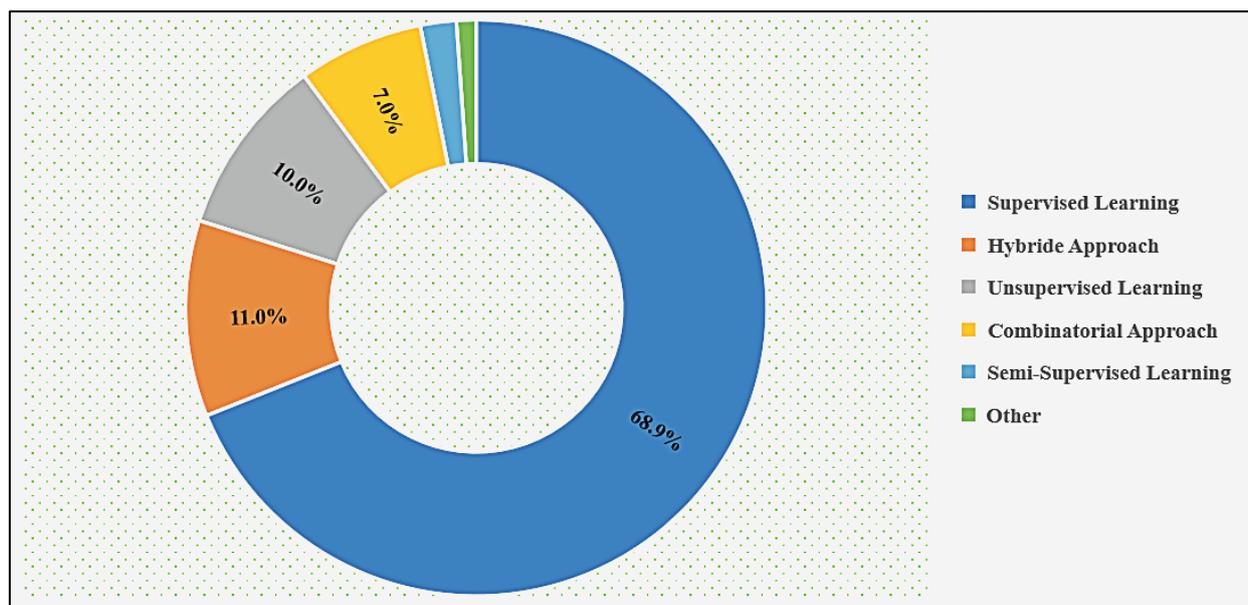


Figure 2. Distribution of the different ML categories in the chemical industry applications.

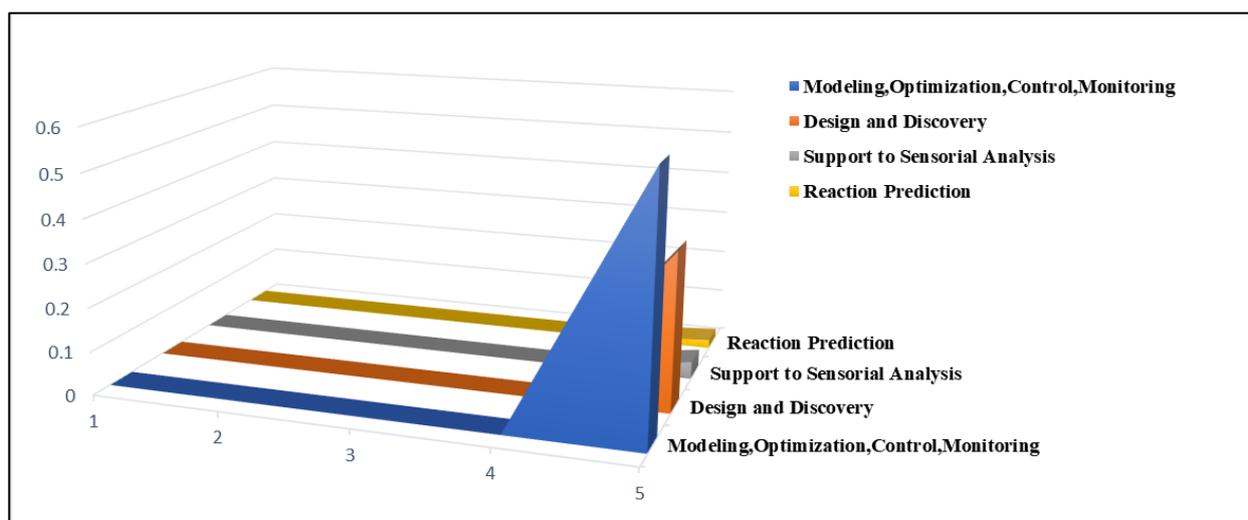


Figure 3. Application of supervised learning in the chemical industry.

This paper reviews the applications of AI in various areas of the chemical industry. First, AI can be used for molecular structure-function relationship analysis. Moreover, applications of AI to chemical reactions include retrosynthetic planning, condition recommendation, and forward reaction prediction. In addition, AI allows the automation of compound synthesis and reduces the repetitive work of laboratory staff.

2. AI Enabled Chemical Process Intensification

2.1. AI for Structure-Function Relationship Analysis

2.1.1. Molecular Property Prediction

Molecular property prediction is an important problem in computer-aided molecular design, and excellent deep-learning models for molecular property prediction can greatly accelerate the progress of experimental studies. Two main types of models are prominent in molecular property prediction—graphical neural networks and sequence-based neural networks, which differ in their representation of different molecules, with the former

requiring molecular graphical information and the latter requiring string representations of molecular structures (Figure 4) [20].

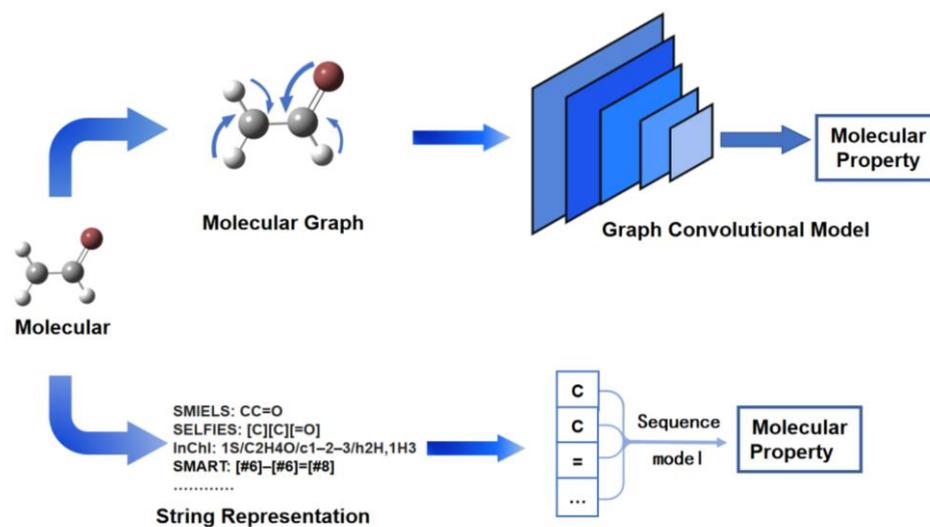


Figure 4. Molecular graph-based and sequence-based models in molecular property prediction.

The direct use of matrices to record molecular structure information is a widely used method of molecular representation known as molecular graphs. Molecular graphs can be trained using graph neural networks. Lu et al. reported the prediction of molecular properties using multilevel Graph Convolutional Neural Networks (MGCN). Different layers of convolutional layers learn the atomic feature information and chemical bond feature information of the molecule and then process the information to predict the molecular properties [21]. In QM9, the MGCN model gains a mean absolute error (MAE) of 0.0642 eV in the HOMO-LUMO gap. The model has excellent predictive performance with generalization capability. Gilmer et al. used a Message Passing Neural Network (MPNN) to predict the QM9 public data set and obtained better performance than any previous model [22,23]. The ratio of the MAE of the MPNN models to the provided chemical accuracy estimate was reported, with a HOMO-LUMO gap of 1.60 eV in QM9. In the framework of the MPNN model, the design of appropriate functions can effectively improve the prediction effect. The directed-MPNN model was used by Yang et al. for the extraction of molecular graph features and predicting the properties of molecules, and the model was tested on 19 public datasets and 16 industry datasets, and the model performance was better than previous models on most tasks [24]. Compared to other papers, the paper gives an MAE of 2.766 ± 0.022 for multi-task prediction of the QM9 database and provides more comparison of model performance.

The recording of molecules using strings is another mainstream molecular representation method, of which the most widely used is SMILES [25]. Deep learning models for natural language processing are well suited to process these sequences, which record molecular information. There is no more effective model for string processing in recent years than the Transformer [26]. Honda et al. reported the use of the Transformer for the prediction of molecular properties in 2019 [27]. Schwaller et al., on the other hand, applied the Transformer model to the prediction of reaction yields [28]. Chithrananda et al. then built several pre-trained models for chemical molecules using the BERT model, which allowed for a significant reduction in training time for later Transformer-based models [29]. Su et al. used these pre-trained models for a transfer learning study to predict the energy gap of metalloporphyrin, spending only one-third of the training time that would have been spent if transfer learning had not been used [30]. Jo et al., on the other hand, used MPNN for processing SMILES information, and the model obtained better results when performing classification tasks on multiple datasets [31]. The molecular graph-based models and sequence-based models, though both perform well in molecular property pre-

diction tasks, have their own advantages. The molecular structure information recorded in molecular graphs is significantly richer than that of sequence methods, and the prediction of molecular properties will be more accurate. The use of sequences to record molecular information has high freedom and can reduce the training cost more easily using transfer learning methods. The two families of models should be selected according to the research content in the next study, or multimodal models can be used to combine their advantages.

2.1.2. Molecular Design

Computer-aided molecular design is another important research direction in cheminformatics, and the design of suitable molecules according to requirements has been a dream function for chemists [32]. Similar to molecular property prediction, both graph generation models and text generation models in the field of deep learning can be used for the molecular design (Figure 5).

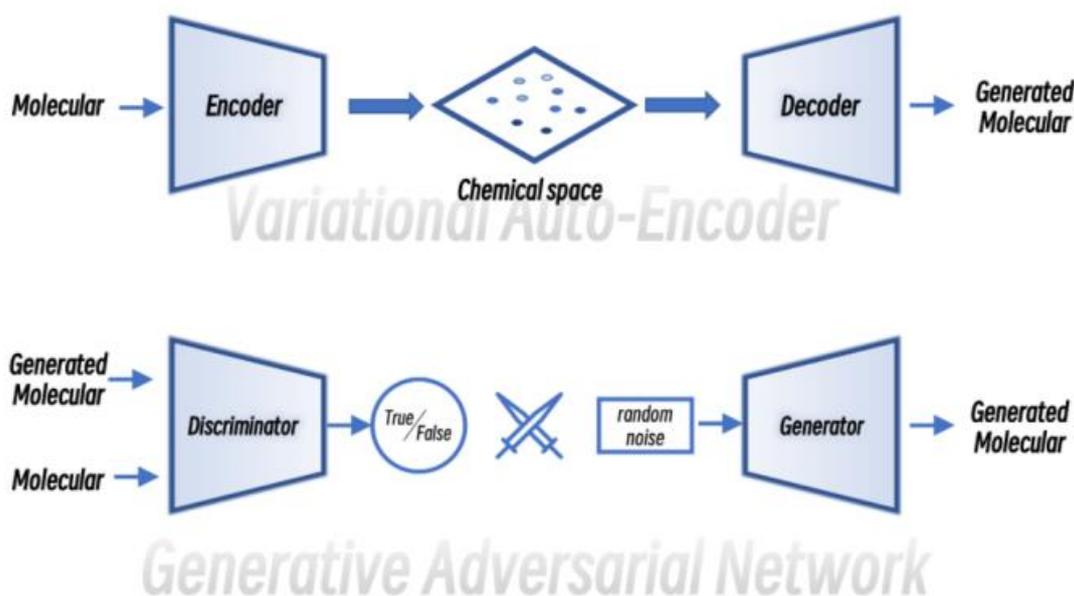


Figure 5. The schematics of Variational Auto-Encoder (VAE) and Generative Adversarial Network (GAN).

In 2018, Gómez-Bombarelli et al. reported the design of new molecules using Variational Auto-Encoder (VAE), a study that will perform molecule generation while mapping the encoded potential chemical space to the corresponding molecular properties, allowing the model to explore the chemical space more efficiently and purposefully [33]. Segler et al., on the other hand, applied recurrent neural networks based on Long Short-Term Memory (LSTM) for ab initio drug design [34]. In this model, transfer learning and reinforcement learning are introduced to improve the validity of the designed new molecules. In the same year, Cao et al. applied Generative Adversarial Network (GAN) to chemical molecule generation, and reinforcement learning also was introduced in the model to score the generated molecules in order to be able to generate molecules that meet the desired target [35]. Flam-Shepherd et al. added MPNN to the decoder and encoder of the VAE model, which greatly improved performance of the VAE model [36].

The two most difficult problems to overcome in computer-aided molecular design are the generation of legitimate chemical molecules and the generation of molecules with target properties or target characteristics, in other words, distribution learning for molecular design and goal-directed molecular optimization [32]. Comparing the performance of molecular design models is not a trivial task. Brown et al. 2019 proposed the GuacaMol platform, which gives different evaluation criteria for the two task models [37]. From current approaches, the use of transfer learning in a separate generative model can improve

the chance of generating valid molecules. On the other hand, the development of novel molecular representation methods with greater robustness, such as SELFIES, can also be effective for the task of distribution-learning of molecular design [38]. In addition, in goal-directed molecular optimization with targets, when the design targets can be quickly computed by computer (e.g., LogP, TPSA, etc.), reinforcement learning can help the model to find the target molecules faster. Furthermore, when the desired property cannot be obtained by simple computation, the potential chemical space in the model can be mapped to the corresponding property before the molecule is designed.

For the design of new molecules, one of the important application areas of AI is interpretable machine learning [39]. For example, Verkhivker et al. developed and implemented interpretable machine learning models for the molecular design of Tyrosine Kinase Inhibitors by combining ChemVAE embedding architecture and cluster decomposition [40]. Recently, a computer-aided molecular design (CAMD) framework for molecular design has been reported. Hatamleh et al. developed a CAMD framework for mosquito repellents to mitigate the drawbacks of currently used repellents [41]. In this framework, a data-driven Hyperbox-based machine learning approach was used to predict the mosquito rejection properties of molecules in the absence of a mechanistic prediction model. Ooi et al. proposed a CAMD-based approach to design fragrance molecules and used a Hyperbox classifier to predict fragrance properties [42]. The resulting model can be interpreted as a parsing decision support rule that establishes a quantitative relationship between the structural parameters of a molecule and its odor characteristics. In addition, a novel data-driven rough set-based machine learning (RSML) model was used as a predictive or diagnostic modeling tool for odor properties to design fragrance molecules [43]. The RSML generates deterministic rules based on the relationship between the topology of fragrant molecules and the odor characteristics from existing odor databases. The generated rules are then integrated into CAMD problems as constraints. The results show that the new method is capable of identifying non-intuitive and promising fragrant molecules that can be used for various applications.

Moreover, in addition to molecular design, several fields are beginning to take advantage of the integration of ML and systems biology, including pathways identification and analysis, modeling of metabolisms and growth, and 3D protein modeling [44]. For example, AI is being used for the dynamic modeling of signaling networks, which helps to understand cellular pathways and facilitate drug discovery. It allows cataloging the changes in gene expression and signaling that occur when cells are exposed to various perturbations, building a network-based understanding of biology [45–47]. For example, in metabolic engineering, ML models, including naive Bayes, decision trees, and logistic regression trained on the pathway information of many organisms, were used in MetaCyc to predict the presence of a novel metabolic pathway in a newly-sequenced organism [44]. In general, the ML models used for pathway prediction showed better performance than standard mathematical and statistical methods. Nevertheless, pathway discovery still relies heavily on traditional approaches such as gene sequence similarity and network analysis. Therefore, better ML algorithms/methods for improving Dynamic and Constraint-based Metabolic Modelling, such as FBA modeling, are needed [44].

2.2. AI for Synthetic Route Planning

AI has been successful in planning synthetic routes performed in the laboratory or evaluated by chemists, including (1) retrosynthetic planning, (2) forward reaction prediction, and (3) condition recommendation. In chemistry, the origins of Computer-assisted synthesis planning (CASP) can be traced back to the translation of retrosynthetic logic into computer code by Corey in the 1960s [48]. Nevertheless, early synthetic route planning relied entirely on the expertise of chemists and did not use statistical learning based on large amounts of data [49–51]. Given the limitation of computational resources, complex algorithms cannot be widely used in synthetic planning. Fortunately, with the growing availability of molecular property datasets, reaction datasets, and increased computational

power, AI for synthetic planning is once again gaining widespread attention [52–56]. In the last 20 years, patterns of reactivity inferred from published response data by AI have become viable alternatives to algorithms based on “expert” rules. It can automate the extraction and training of data, making it easily scalable to merge new responses, which eases the burden on scientists. Today, the retrosynthesis of complex molecules, high-fidelity prediction of reaction outcomes, and automation of chemical reactions are still major research fields.

2.2.1. Retrosynthetic Planning

Rule-based and rule-free methods are the main approaches used for retrosynthesis. The rule-based method is conceptually similar to the process by which an organic chemist selects a known reaction type to apply to a specific synthetic target. It has been well implemented in state-of-the-art detailed synthetic planners, but building an expert-encoded rule is laborious and inherently dependent on the expertise of scientists [57]. Consequently, the automatic generation of reaction rules from accessible reaction databases has attracted the attention of scientists [58]. The reaction rules are generated automatically by extracting reaction templates from the reactions in the database, clustering, and processing them with additional molecules [59,60]. Other methods apply the templates directly to the target where filters, such as similarity-based neural networks, are often used to apply only a chemically relevant subset of the template library to reduce the required computational power [61–67]. Although rule-based approaches are common in the most advanced detailed synthetic planners, the main drawback is the huge computational cost involved in extracting a library of rules or templates. Moreover, the complexity of assessing between all existing rules and new rules increases as the number of codified rules increases, which may ultimately make the problem intractable. In contrast, the rule-free method maps the target compounds directly to potential starting materials, bypassing the need to build a library of reaction rules. It represents molecules as text, such as SMILES strings, thus making the prediction a natural language processing problem [68]. With different types of neural machine translation architectures, forward reaction or retrosynthetic prediction can be achieved. The Molecular Transformer architecture is currently the most popular method of treating chemistry as a language, capable of producing valid SMILE strings more accurately [69,70]. Compared to rule-based methods, rule-free methods are more general and have lower associated computational costs.

Inspired by the use of Molecular Transformers for forward reaction prediction, some retrosynthetic models based on the same architecture have attracted a lot of attention [15,71,72]. Zheng et al. developed a template-free self-correcting retrosynthesis predictor (SCROP) that uses a transformer neural network to predict retrosynthesis [14]. By converting retrosynthesis planning into a molecular linear symbolic problem for machine translation, the method achieves an accuracy of 59.0% on a standard benchmark dataset utilizing a grammar corrector for neural networks. Wang et al. proposed a single-step template-free and Transformer-based method called RetroPrime, which aims to address the issues that the output of the Transformer-based retrosynthesis model tends to suffer from insufficient diversity and high chemical implausibility [73]. What’s more, Tetko et al. investigated the impact of a text-like representation of chemical reactions (SMILES) and the natural language processing (NLP) neural network Transformer architecture on predicting retrosynthetic reactions [74]. Lin et al. used the Transformer architecture to treat each reaction prediction task as a data-driven sequence-to-sequence problem, achieving superior performance for single-step inverse synthesis tasks (Figure 6) [70]. The top-1 accuracy of the retrosynthesis methods discussed above ranged from 41–54% [75]. Even though the increased batch size and training time of the Transformer model by Duan et al. achieved a top-1 accuracy of 54.1% on the 50 k USPTO dataset [76]. In contrast, the RetroTRAE developed by Cernak et al. is free of all SMILES-based translation problems, yielding a top-1 accuracy of 58.3% on the USPTO test dataset [75]. Although the top-1 accuracy is gained using the proprietary training and test sets, it is questionable how models with specific sets

of the same chemical transformations can be used in specific processes. Recently, graph-enhanced transformer and hybrid models were reported, achieving 44.9% top-1 accuracy and more diverse reactant suggestions, respectively, but without substantial improvements over previous work [77,78]. Notably, except for the work of Lin et al., all transformer-based retrosynthesis methods are limited to a single step [69]. Additionally, reagent, catalyst, and solvent conditions were not simultaneously predicted in the retrosynthesis planning. The Molecular Transformer model introduced by Schwaller et al. incorporates a hypergraph exploration strategy for automatic retrosynthesis planning without human intervention [72,79]. Meanwhile, the single-step retrosynthesis model predicts the reactants, reagents, solvents, and catalysts for each retrosynthesis step, bringing the retrosynthesis technology to a new technological level.

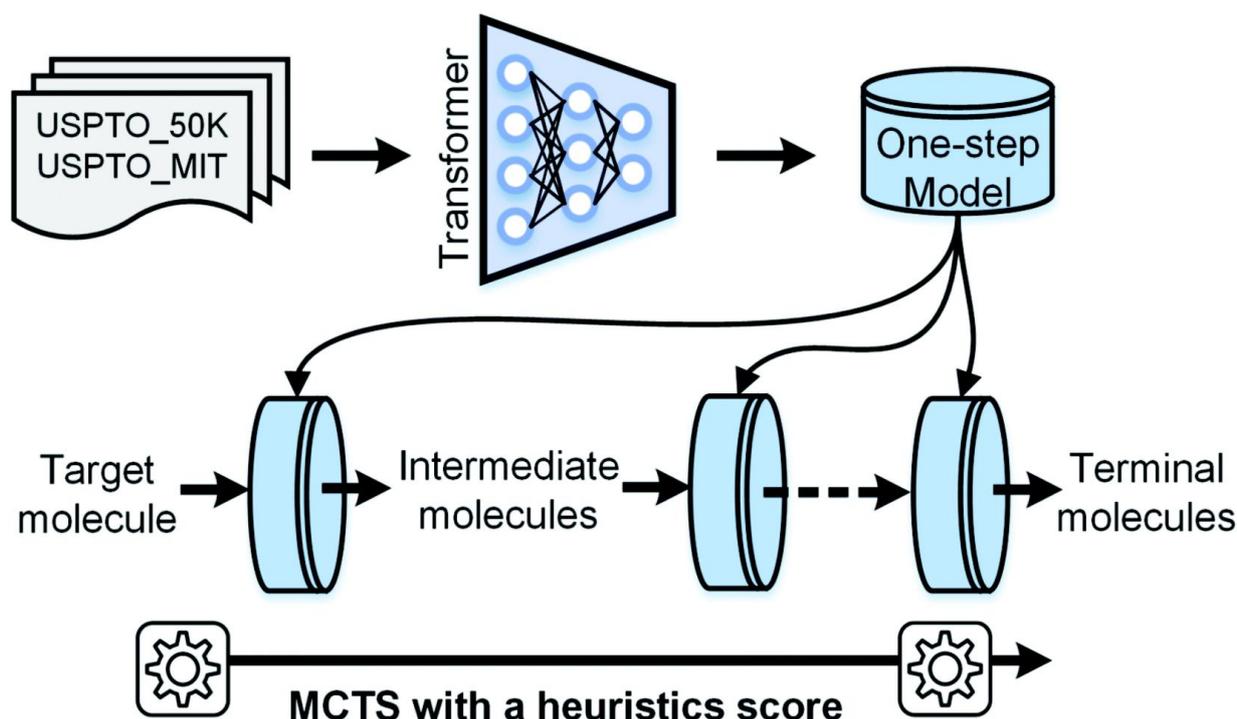


Figure 6. The workflow of AutoSynRoute. Reproduced with permission from [70].

2.2.2. Forward Reaction Prediction

AI is also widely used for forward reaction prediction. Reaction prediction, which predicts possible products from starting materials and conditions, can be used to virtually screen the proposed reactions or to validate the proposed synthesis steps.

Compared to retrosynthesis, the forward reaction prediction has only one answer, leading to a more straightforward quantitative assessment. Currently, AI-based models used for reaction prediction include (1) inferring reaction rules from predefined lists of rules or templates, (2) graph convolutional neural networks that predict changes in atoms and bonds between starting materials to products, and (3) sequence-to-sequence model of the prediction product SMILES. Similar to retrosynthetic planning, reliable data tend to favor the quality of forward reaction prediction results. In the absence of precise data, such as concentration, time, and temperature data, reaction prediction becomes a tricky problem. For example, Lee et al. found that retraining a sequence-to-sequence forward prediction model on its own data did improve the accuracy of company-specific chemistry [15].

Coley et al. combined the traditional use of reaction templates with the flexibility of pattern recognition offered by neural networks to develop a framework for predicting reaction outcomes [80]. In 5-fold cross-validation, it is shown that the trained model is very successful in forward reaction prediction. Similarly, Aspuru-Guzik et al. combined predictor variables with SMARTS transformations to construct a system of predictable

products [81]. Finally, the usability of the system was verified with questions from organic chemistry textbooks. Coley et al. used Weisfeiler-Lehman networks to model higher-order interactions between changes occurring at nodes in a molecule to effectively explore the space of product molecules and predict the outcome of organic reactions [82]. The experimental results show that the accuracy of the model is comparable to the performance of domain experts. Furthermore, given the reactant, reagent, and solvent conditions, Coley's group proposed a supervised learning method to predict the product [16]. By mapping the text sequence representing the reactants to the text sequence representing the products, the reaction prediction can be viewed as a translation problem. For example, Schwaller et al. enabled the prediction of complex organic chemical reactions with template-free sequence-to-sequence models [83]. The model achieves 80.3% top-1 accuracy without relying on auxiliary knowledge, such as response templates or explicit atomic features. Recently, Schwaller's group forward prediction was considered a machine translation problem and developed the Molecular Transformer model [69]. The model can make predictions by inferring correlations between the chemical motifs in the reactants, reagents, and products in the data set, achieving over 90% top-1 accuracy.

Another parameter of interest for the forward reaction prediction is the reaction yield. It can guide the chemist in choosing the route that maximizes the total yield as well as assists in retrosynthetic planning. The model for reaction prediction was mainly built on the high-throughput experiment dataset. Perera et al. studied 15 pairs of electrophilic and nucleophilic reagents for the Suzuki-Miyaura reaction using the HTE technique, and each obtained different products [84]. Doyle et al. trained a random forest algorithm using a high-throughput data set to accurately predict the yield of other Buchwald-Hartwig coupling reactions with multidimensional variables after being trained with thousands of Buchwald-Hartwig coupling reaction data [85]. Similarly, Schwaller et al. used Doyle's high-throughput dataset to predict yields for a total of 3955 Buchwald-Hartwig reactions containing [86]. Andrzej et al. predicted the yield of 16 phosphate ligands for nickel-catalyzed Suzuki cross-coupling by training a linear regression model with two larger data sets obtained by high-throughput experiments (HTE) (Figure 7) [87]. Further, Schwaller's group combined the encoder converter model with a regression layer, and the excellent reaction yield prediction performance of the high model was demonstrated on two high-throughput experimental reaction sets [28]. Although high-throughput experiments are capable of screening multiple reaction variables at the nanomolar level, this technique covers a very narrow chemical space dataset. Structure-based descriptors (molecular fingerprints and molecular maps) are faster and easier to compute for any molecule. Hirst et al. demonstrated the applicability of support vector regression (SVR) in predicting reaction yields using combined data [88]. Schwaller's group treated organic molecules as a language and introduced SMILES strings of reactions into the model to predict reaction yields [68,83]. Additionally, the use of encoder-only transformers, such as Bidirectional Encoder Representations from Transformers (BERT), has led to advances in response yield prediction. The superiority of yield prediction compared to one-hot encoding was demonstrated by Sandfort et al. using a concatenation of multiple molecular "fingerprints" as a representation of alternative reactions [89]. Moreover, Akinori et al. developed a Message Passing Neural Network (MPNN) model for chemical yield prediction for Buchwald-Hartwig cross-coupling yields [90]. Sequence-to-sequence models are not only useful when working with language tokens but also provide high-quality descriptors to predict reaction properties, such as reaction yields.

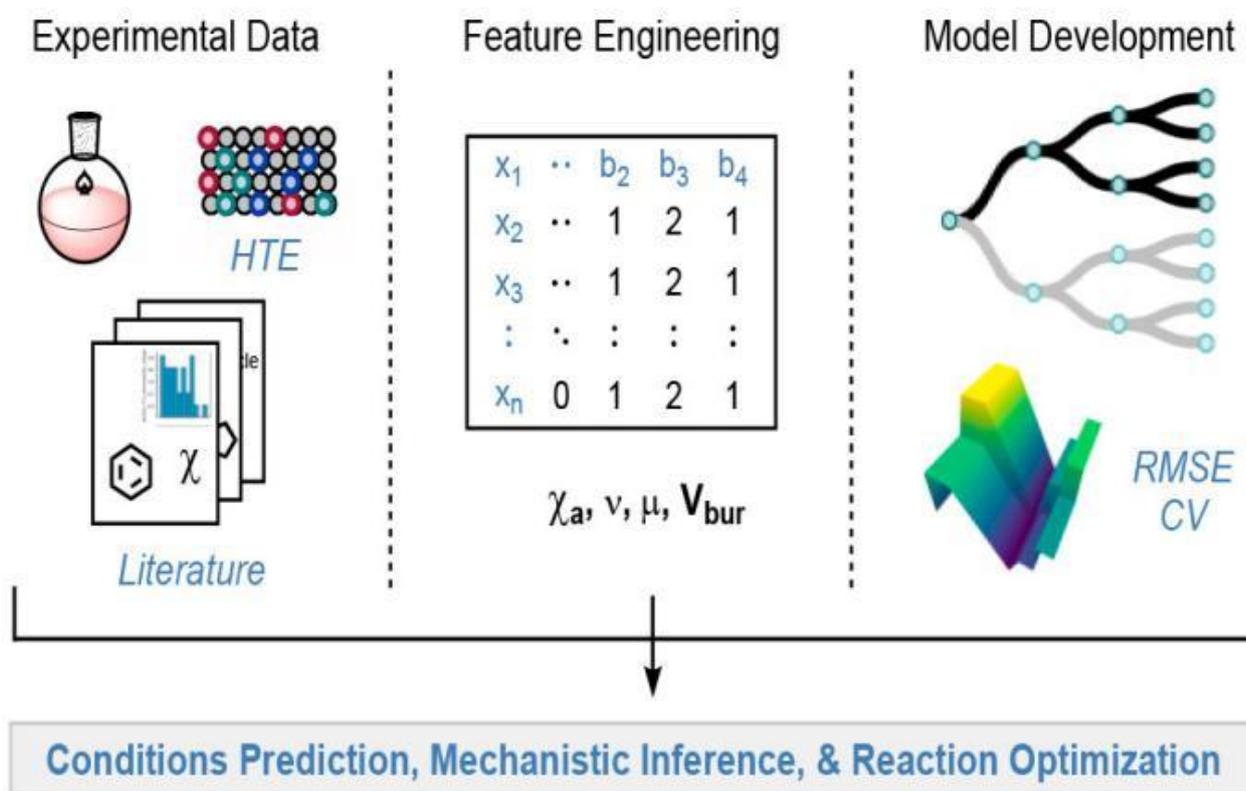


Figure 7. The AI-assisted reaction yield prediction. Reproduced with permission from [87].

2.2.3. Condition Recommendation

For the forward reaction to proceed smoothly, it is necessary to explore the reaction conditions that will achieve the desired transformation. Typically, chemists screen reaction conditions based on their own experience and are biased. Instead, based on a priori knowledge, the AI can more objectively infer the appropriate conditions. However, recommendations for specific reaction conditions were limited to a single reaction class [91,92]. The main reason is the lack of high-quality data, which makes the model difficult to develop. Mainly including (1) quantity, volume, or concentration, (2) reaction time or kinetics, and (3) order of addition of reagents and catalysts. Despite the difficulties, AI has demonstrated the ability to make reaction condition recommendations for more diverse reaction sets. These models provide a strong basis for empirical optimization of reaction conditions but still lack the full details needed for implementation. The discovery of more general reaction conditions requires consideration of a broad region of chemical space derived from a large matrix substrate that intersects with the high-dimensional matrix of reaction conditions. In their optimization of the Suzuki-Miyaura cross-coupling reaction, Aspuru-Guzik et al. identified the phosphine ligand as a classification parameter critical for determining the reaction outcome [93]. Thus, a strategy using computational molecular feature clustering was developed to reveal the conditions for selectively obtaining the desired product isomers in high yields. What's more, Aspuru-Guzik's group reported a simple closed-loop workflow that can be used to discover general reaction conditions using data-guided matrix down-selection, uncertainty minimization machine learning, and robotic experiments [94]. By applying it to the heteroaryl Suzuki-Miyaura cross-coupling reaction, conditions were identified that doubled the average yield relative to a widely used benchmark previously developed using conventional methods. A practical roadmap was provided for solving multidimensional chemical optimization problems with large search spaces.

2.3. AI for Automated Synthesis

Applications of AI in chemical reactions include not only synthetic route planning but also automated synthesis. Traditionally, scientists have been exposed to hazardous, repetitive chemical manipulations for long periods, resulting in a significant waste of resources and time [95,96]. Additionally, cost and condition constraints prevent scientists from conducting too many experiments to obtain desired results. Most importantly, traditional chemical synthesis relies heavily on labor-intensive practices such as scientific training, planning, experience, observation, and interpretation. Fortunately, AI is changing the productivity of modern manufacturing, and modern automation of organic chemistry operations is gradually freeing the hands and minds of organic chemists [85,97,98]. For example, with an auto mated platform, the anti-arrhythmic drug lidocaine, the anti-epileptic drug rufinamide, and the anti-cardiovascular drug sildenafil have been synthesized automatically without human intervention [99,100]. Exactly, AI alleviates the operator from tedious work and manual intervention.

2.3.1. Robotic Lab Platform

Automated chemistry is based on the modularity of common physical operations, such as liquid handling robots, robotic grippers for plate or vial transfer, and computer-controlled heater/shaker blocks, to help scientists reduce labor-intensive laboratory tasks [101]. The platform mainly consists of (1) continuous flow technology combined with process analysis technologies and robotics, (2) automated operation modes combined with hardware for traditional intermittent reactions, and (3) robotics replacing the operator's operation. A simple paradigm for automated chemistry is to automate the operational and sample transfer steps between existing laboratory hardware, such as the mobile robotic chemists of Burger et al. [102]. Mo et al. built a robotic desktop system for the high-throughput collection of TLC data with an image analysis program that automatically calculates compound Rf values. This work reduces the reproducibility of experiments by replacing scientists with robots for repetitive TLC sampling [103]. Currently, the intelligence of chemical synthesis is still in the development stage. Cronin's group developed a modular standard robotic platform to automate laboratory-scale chemical synthesis [99]. With the robotic platform, the authors synthesized three pharmaceutical compounds, Nytol, rufinamide, and sildenafil, without human intervention. The yields and purity of the products and intermediates were comparable to or better than those obtained manually. What's more, the Chemputer synthesis robot can perform many different reactions, including solid-phase synthesis and iterative cross-coupling [104]. Interestingly, the system can simultaneously reuse only 22 different steps in 10 unique modules, and the code can access 17 different reactions, making it possible to link multi-step synthesis to run many different protocols and reactions in a single machine. Although this robotic platform is encouraging, the synthesis of complex organic compounds is still largely artificial. To reduce experimental reproducibility, Jamison et al. developed a plug-and-play continuous flow chemical synthesis system [105]. The system has a flexible robotic arm that can perform all synthetic operations instead of the scientist, automatically synthesizing 15 drugs, including Aspirin, Lidocaine, Diazepam, Warfarin, etc. (Figure 8) [106]. Notably, the system remains insoluble for process intensification (e.g., reducing reaction time), reducing solid formation to avoid blockage, etc. Additionally, predicting the appropriate purification method is challenging, especially for non-column chromatography methods. Moreover, the optimization of multi-step reactions can be complicated by the propagation of parameters. In addition to replacing scientists in labor-intensive laboratory operations, robots can help scientists with other complex tasks. For example, Cronin's group invented robots that automatically read the literature and form a generalized autonomous synthesis workflow [99]. However, manual error correction is still required. Burger et al. used a mobile robot to find photocatalysts that break down water into hydrogen [102]. The robot, driven by a Bayesian search algorithm, performed 688 experiments in an experimental space of 10 variables over eight days. Jamison et al. developed robots that can vary downstream dwell time and control the addition

sequence to minimize undesired reactivity [107]. Robotic reconfigurability and convergent synthesis flexibility play an increasingly important role in assisting with idea generation, experimental design, execution, and optimization to enhance manual experiments.

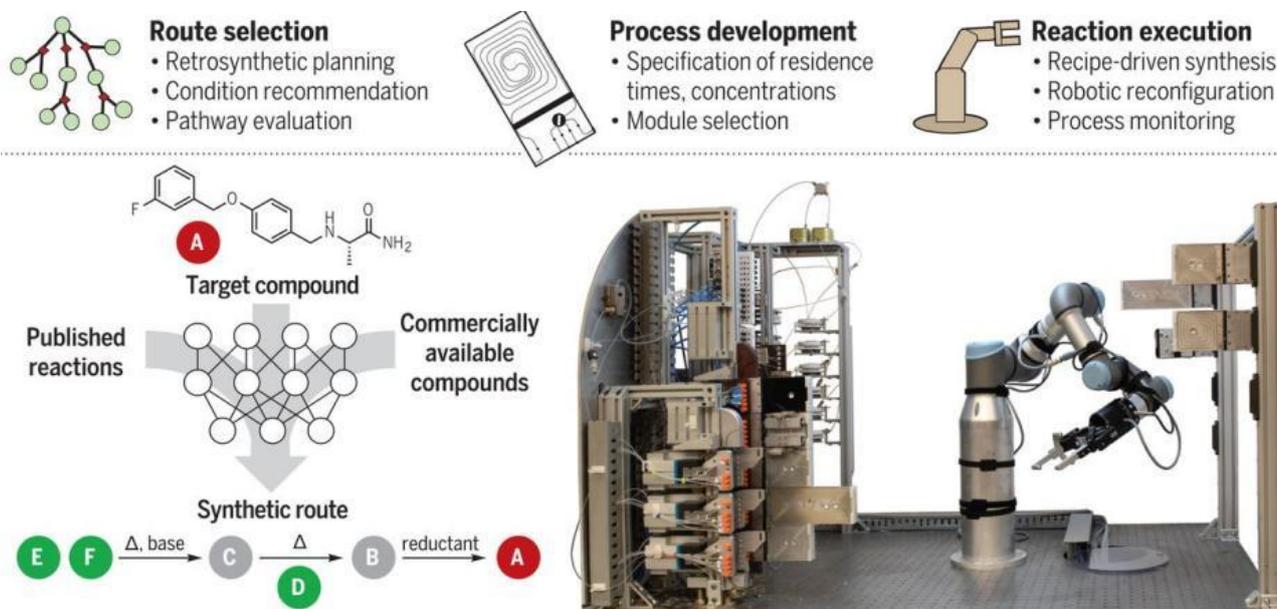


Figure 8. The plug-and-play continuous flow chemical synthesis system. Reproduced with permission from [106].

Robotic lab platforms are used not only in the chemical industry but also in other fields, such as for the automated synthesis of chemical peptides and materials. Peptide Nucleic Acid (PNA) is a synthetic DNA or RNA analog with a peptide chain backbone structure, and traditional synthetic methods require several days to synthesize a biologically active sequence. Bradley et al. have invented a fully automated flow synthesis robot called “Tiny Tides” that allows rapid “one-pot” synthesis of peptide nucleic acid sequences (PPNA) with cell-penetrating peptides [108]. This automated synthesis technology reduces the synthesis time of PPNA from several days to just two hours. Similarly, Aspuru-Guzik’s group reported an algorithm-driven modular robotics-based platform applied to discover thin film materials [109]. Cronin’s group develops an autonomous chemical synthesis robot for exploring, discovering, and optimizing nanostructures driven by real-time spectral feedback, theory, and machine learning algorithms [110]. Additionally, advances in robotics have played a role in precision medicine to improve modern medicine and quality of life, including the delivery of drugs, biologics, genes, and living cells, as detailed in the related review [111].

Fully automated chemical synthesis using AI robots instead of humans is a future trend, with advantages not only in faster and more efficient synthesis but also in the production of compounds that are difficult to synthesize manually. As can be seen, experimental methods based on robotic lab platforms have been used successfully to solve high-dimensional problems in physics, chemistry, and life sciences. It is worth noting that although the necessary hardware units for such tasks are commercially available, cost, standardization, and efficiency issues have made scaling up difficult. Furthermore, current automated multistep synthesis relies on iterative or linear processes and requires compromises in versatility and equipment usage, which means that machines cannot perform multi-step synthesis to run many different protocols and reactions [112–115]. It is believed that in the future, robotic chemists will further change chemical synthesis.

2.3.2. Automated Synthesis

In the field of chemistry, AI can be used to optimize chemical reaction conditions experimentally. In the past, Chemists devote considerable time to evaluating various reaction parameters such as substrates, catalysts, reagents, additives, solvents, concentrations, temperatures, and reactor types. Despite the prevalence of established techniques such as single-objective optimization algorithms, design of experiments (DoE), and other existing techniques, reaction optimization is still often a difficult and time-consuming process for chemists [116–118]. For example, the Single-objective optimization algorithms cannot explore the entire chemical space, thus yielding an overall sub-optimal process. Design of experiments (DOE) methods such as Latin Hypercube Sampling (LHS) typically generate samples that cover the design space as uniformly as possible to improve the accuracy of the overall metamodel. However, it is not the most efficient method if the design goals are clearly defined. Additionally, when a complex model is required to achieve a predefined design goal, the sole experimental methods are inefficient. Consequently, there are often multiple factors to consider during process optimization, such as reaction yield, process cost, impurity levels, and environmental impact. Multi-objective optimization can address multiple (conflicting) objectives encountered in many chemical engineering applications, for example, conversion and selectivity in chemical reactions [119]. Multi-objective optimization techniques such as the parametric approach, epsilon constraint method, or genetic algorithms are used as solution strategies for the multi-objective optimization of chemical reactions [120,121]. However, since they require many functional evaluations and partially derived information is not available, these approaches do not apply to automated chemical reaction systems. Surprisingly, the Bayesian optimization method is a derivative-free global stochastic optimization method for the automatic optimization of multi-objective experimental parameters in chemistry, materials, and other fields [122,123]. For example, both Doyle et al. and Jensen's group used Bayesian algorithms to achieve optimization for performing single or multi-objective reaction parameters [122,124]. Additionally, some excellent multi-objective Bayesian optimization algorithms have been gradually developed, such as Thompson Sampling Efficient Multi-Objective (TS-EMO), Phoénics, Gryffin, and Chimera, etc. (Figure 9) [125–129].

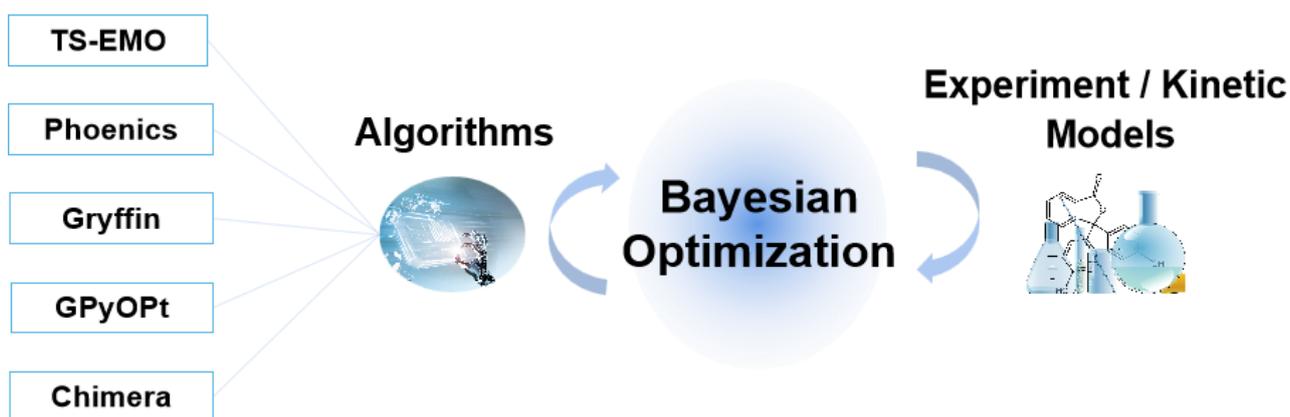


Figure 9. The Bayesian optimization algorithms with a self-optimizing platform.

Bayesian optimization is an iterative response surface-based global optimization algorithm that has shown excellent performance in the optimization of chemical reaction parameters [130–132]. It aims to balance the exploration of areas of uncertainty with the use of available information to obtain high-quality configurations in fewer evaluations. In many cases, Bayesian optimization algorithms outperform expert practitioners and other state-of-the-art global optimization algorithms [133]. Currently, multi-objective Bayesian optimization algorithms, including Thompson Sampling Efficient Multi-Objective (TS-EMO), ParEGO, and Expected Hypervolume Improvement (EHI), aim to approximate the

Pareto front. Lapkin's group study showed that the TS-EMO algorithm has comparable or better data efficiency than both EHI and ParEGO [130]. Further, TS-EMO performs well on a set of mathematical test functions for a given budget compared to the externally commonly used genetic algorithm NSGA-II. Based on the advantages of the TS-EMO algorithm, Lapkin et al. achieved self-optimization for the Sonogashira reaction, Claisen-Schmidt condensation reaction, N-benylation reaction, and N-benylation reaction with flow chemistry systems [130,133]. The optimal conditions corresponding to the trade-off curve (Pareto front) between environmental and economic objectives were successfully identified. The TS-EMO algorithm combined with flow chemistry systems demonstrates the ability to identify optimal reaction conditions and trade-offs (Pareto fronts) between conflicting optimization objectives such as yield, cost, space-time yield, and E-factor in a data-efficient manner [134]. The TS-EMO algorithm applies not only to classical single-step reactions but also to the optimization of multi-step reaction parameters. Lapkin's group combined the TSEMO algorithm with a self-optimizing platform to optimize the Claisen-Schmidt condensation reaction with subsequent liquid-liquid separation, involving three objectives [133]. By optimizing multi-step sequential reactions, AI shows how to re-evaluate optimal reaction conditions with changing downstream post-processing specifications during active learning. The diversity of possible combinations of reagents, solvents, stoichiometry, and temperature for reactions makes the development of new products fraught with difficulties. For example, studies have shown that most catalytic reactions have over 50 million potential conditions, making a robust exploration of the parameter space impractical [135]. Recent work has shown that machine learning and molecular descriptors of a solvent or catalyst can be used to extrapolate performance from a small number of experiments to a large library, but which machine learning strategy to apply in a particular case remains difficult [136,137]. Therefore, Lapkin et al. released an open-source software package called Summit based on Bayesian optimization to optimize the reaction [125]. Summit includes a benchmark that enables the comparison of the performance of different ML strategies, where researchers can test the efficiency of each strategy through virtual experiments. The platform was used to achieve process route development for the SNAr reaction and to screen the optimal catalyst and ligand for the Pd-catalyzed cross-coupling reaction. Furthermore, AI makes it possible to obtain functional molecules with high selectivity from renewable biomaterials and biowastes. The preparation of p-cymene from waste terpene mixtures was reported by Lapkin et al. [138]. This work used the TS-EMO algorithm to optimize the first two steps of the reaction to obtain maximum conversion and selectivity for the production of functional molecules from biomaterials and biowaste. In brief, Bayesian optimization algorithms are tools to develop accurate reaction models without prior knowledge, with a large number of input variables, and with competing objectives. Models developed for individual steps can be used for potential process design and scale-up.

Other examples of multi-objective algorithms developed for the chemical process include Phoenix, Gryffin, and Chimera [127–129]. These algorithms avoid the problem of classical Bayesian algorithms that select data in the order of parameter points. Phoenix uses Bayesian neural networks (BNNs) to construct kernel density estimates of the objective function, and its acquisition function allows the selection of batches of evaluations that run in parallel. Importantly, Phoenix is suitable for the optimization of continuous parameters, such as temperature and concentration, and can be used for the optimization of chemical reaction conditions and material properties. Aspuru-Guzik's group developed Gryffin for optimizing categorical parameters such as solvent selection. The algorithm uses categorical kernel densities that can be relaxed to continuous ones. In addition, it allows the provision of expert knowledge in the form of descriptors for each classification choice and is successfully used for the optimization of chemical reaction conditions. Usually, there are also multiple competing objectives in materials science. Chimera is a generic multi-objective optimization method. It allows for defining a hierarchy of objective preferences that are combined into a single function optimized with any chosen algo-

rithm. Importantly, both the previously mentioned TS-EMO algorithm and the algorithm described here can be combined with an automation platform to automate experiments. For example, Aspuru-Guzik's group deployed ChemOS, together with Phoenix, Gryffin, and Chimera, for the autonomous optimization of manufacturing processes of thin-film materials, multicomponent polymer OPV blends, and reaction conditions of stereoselective Suzuki coupling [93,109,139]. For other excellent work, Aspuru-Guzik's group has described specifically in the review [140]. Other AI-based automation platforms are also being reported. Vlachos's group has developed the NEX Torch platform using state-of-the-art Bayesian optimization algorithms to enable the sampling of continuous variables and discrete values of subtypes [141]. It can help not only chemical synthesis in laboratory experiments but also multi-scale computational tasks from molecular-scale design to reactor-scale optimization.

Nevertheless, AI in automated synthesis still faces many challenges. First, the in-line/online analysis still needs further development, especially in terms of measurement accuracy, instrument response speed, and compatibility with heterogeneous synthesis. In addition, the equipment for automated synthesis is too expensive for research laboratories in developing countries to afford.

3. Conclusions and Outlook

In this review, different aspects of artificial intelligence-enabled chemical process intensification are discussed. In chemistry, AI enables structure-function relationship analysis, including the prediction of molecular properties and the design of molecules. In addition, here is a brief summary of the use of computer-aided synthesis planning (CASP): retrosynthetic planning, condition recommendation, and forward reaction prediction in the Pharmaceutical and Chemical Industry. Moreover, the robotic lab platform enables automated organic synthesis to reduce the repetitive work of laboratory staff. Finally, AI techniques enable the optimization of chemical reaction conditions with multiple objectives, achieving a trade-off between optimal reaction conditions and conflicting optimization objectives (e.g., yield, cost, spacetime yield (STY), and E factor).

Although AI is booming in the chemical industry, it still faces many challenges. Optimal predictions depend on the availability of a stable and high-quality dataset, and the challenge is to obtain sufficient and reliable data. Second, while the arithmetic (quantum and cloud-based approaches) is improving, there are still limitations from the user's perspective. The shortage of data science talent in chemical engineering means that increased collaboration between chemistry and other scientific disciplines may help accelerate the integration of AI with other fields.

Author Contributions: Conceptualization, C.H. and C.Z.; methodology, C.H.; software, C.H. and C.Z.; validation, C.H.; formal analysis, T.B.; investigation, C.H. and C.Z.; resources, K.J.; data curation, T.B.; writing—original draft preparation, C.H. and C.Z.; writing—review and editing, C.H. and C.Z.; visualization, K.J.; supervision, W.S., K.-J.W. and A.S.; project administration, W.S., K.-J.W. and A.S.; funding acquisition, W.S., K.-J.W. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: We gratefully acknowledge the Zhejiang Province Science and Technology Plan Project (No. 2022C01179 and No. 2019-ZJ-JS-03) and the National Natural Science Foundation of China (No. 22108252 and No. U22A20408) for financial support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in [A Review on Artificial Intelligence Enabled Design, Synthesis, and Process Optimization of Chemical Products for Industry 4.0 or processes -2123057-permission here].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Brown, D.G.; Wobst, H.J.; Kapoor, A.; Kenna, L.A.; Southall, N. Clinical development times for innovative drugs. *Nat. Rev. Drug Discov.* **2021**, *21*, 793–794. [[CrossRef](#)] [[PubMed](#)]
2. Baum, Z.J.; Yu, X.; Ayala, P.Y.; Zhao, Y.; Watkins, S.P.; Zhou, Q. Artificial Intelligence in Chemistry: Current Trends and Future Directions. *J. Chem. Inf. Model.* **2021**, *61*, 3197–3212. [[CrossRef](#)] [[PubMed](#)]
3. Mowbray, M.; Vallerio, M.; Perez-Galvan, C.; Zhang, D.; Chanona, A.D.R.; Navarro-Brull, F.J. Industrial data science—a review of machine learning applications for chemical and process industries. *React. Chem. Eng.* **2022**, *7*, 1471–1509. [[CrossRef](#)]
4. Venkatasubramanian, V. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE J.* **2018**, *65*, 466–478. [[CrossRef](#)]
5. Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R.K. Artificial intelligence in drug discovery and development. *Drug Discov. Today* **2020**, *26*, 80–93. [[CrossRef](#)]
6. Trinh, C.; Meimaroglou, D.; Hoppe, S. Machine Learning in Chemical Product Engineering: The State of the Art and a Guide for Newcomers. *Processes* **2021**, *9*, 1456. [[CrossRef](#)]
7. Mitchell, J.B.O. Machine learning methods in chemoinformatics. *WIREs Comput. Mol. Sci.* **2014**, *4*, 468–481. [[CrossRef](#)]
8. Griffen, E.J.; Dossetter, A.G.; Leach, A.G.; Montague, S. Can we accelerate medicinal chemistry by augmenting the chemist with Big Data and artificial intelligence? *Drug Discov. Today* **2018**, *23*, 1373–1384. [[CrossRef](#)]
9. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **2019**, *18*, 463–477. [[CrossRef](#)]
10. Elton, D.C.; Boukouvalas, Z.; Fuge, M.; Chung, P.W. Deep learning for molecular design—A review of the state of the art. *Mol. Syst. Des. Eng.* **2019**, *4*, 828–849. [[CrossRef](#)]
11. Segler, M.H.S.; Pruss, M.; Waller, M.P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018**, *555*, 604–610. [[CrossRef](#)] [[PubMed](#)]
12. Ishida, S.; Terayama, K.; Kojima, R.; Takasu, K.; Okuno, Y. Prediction and Interpretable Visualization of Retrosynthetic Reactions Using Graph Convolutional Networks. *J. Chem. Inf. Model.* **2019**, *59*, 5026–5033. [[PubMed](#)]
13. Gajewska, E.P.; Szymkuc, S.; Dittwald, P.; Startek, M.; Popik, O.; Mlynarski, J.; Grzybowski, B.A. Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses. *Chem* **2020**, *6*, 280–293.
14. Zheng, S.; Rao, J.; Zhang, Z.; Xu, J.; Yang, Y. Predicting Retrosynthetic Reactions Using Self-Corrected Transformer Neural Networks. *J. Chem. Inf. Model.* **2019**, *60*, 47–55. [[CrossRef](#)] [[PubMed](#)]
15. Lee, A.A.; Yang, Q.; Sresht, V.; Bolgar, P.; Hou, X.; Klug-McLeod, J.L.; Butler, C.R. Molecular Transformer unifies reaction prediction and retrosynthesis across pharma chemical space. *Chem. Commun.* **2019**, *55*, 12152–12155. [[CrossRef](#)]
16. Coley, C.W.; Jin, W.; Rogers, L.; Jamison, T.F.; Jaakkola, T.S.; Green, W.H.; Barzilay, R.; Jensen, K.F. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* **2018**, *10*, 370–377. [[CrossRef](#)] [[PubMed](#)]
17. Gao, H.; Struble, T.J.; Coley, C.W.; Wang, Y.; Green, W.H.; Jensen, K.F. Using Machine Learning to Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **2018**, *4*, 1465–1476.
18. Walker, E.; Kammeraad, J.; Goetz, J.; Robo, M.T.; Tewari, A.; Zimmerman, P.M. Learning to Predict Reaction Conditions: Relationships between Solvent, Molecular Structure, and Catalyst. *J. Chem. Inf. Model.* **2019**, *59*, 3645–3654. [[CrossRef](#)]
19. Maser, M.R.; Cui, A.Y.; Ryou, S.; DeLano, T.J.; Yue, Y.; Reisman, S.E. Multilabel Classification Models for the Prediction of Cross-Coupling Reaction Conditions. *J. Chem. Inf. Model.* **2021**, *61*, 156–166. [[CrossRef](#)]
20. Venkatasubramanian, V.; Mann, V. Artificial intelligence in reaction prediction and chemical synthesis. *Curr. Opin. Chem. Eng.* **2022**, *36*, 100749. [[CrossRef](#)]
21. Lu, C.; Liu, Q.; Wang, C.; Huang, Z.; Lin, P.; He, L. Molecular property prediction: A multilevel quantum interactions modeling perspective. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 1052–1060.
22. Gilmer, J.; Schoenholz, S.S.; Riley, P.F.; Vinyals, O.; Dahl, G.E. Neural Message Passing for Quantum Chemistry. In Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research, Sydney, NSW, Australia, 6–11 August 2017; pp. 1263–1272.
23. Blomberg, M.R.A.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P.E.M. Quantum Chemical Studies of Mechanisms for Metalloenzymes. *Chem. Rev.* **2014**, *114*, 3601–3658. [[CrossRef](#)] [[PubMed](#)]
24. Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; et al. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **2019**, *59*, 3370–3388. [[CrossRef](#)] [[PubMed](#)]
25. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* **1988**, *28*, 31–36. [[CrossRef](#)]
26. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *arXiv* **2017**. [[CrossRef](#)]
27. Honda, S.; Shi, S.; Hiroki, R.; Ueda. SMILES Transformer: Pre-trained Molecular Fingerprint for Low Data Drug Discovery. *arXiv* **2019**, arXiv:1911.04738. [[CrossRef](#)]
28. Schwaller, P.; Vaucher, A.C.; Laino, T.; Reymond, J.-L. Prediction of chemical reaction yields using deep learning. *Mach. Learn. Sci. Technol.* **2021**, *2*, 015016. [[CrossRef](#)]

29. Chithrananda, S.; Grand, G.; Ramsundar, B. Chemberta: Large-scale self-supervised pretraining for molecular property prediction. *arXiv* **2020**, arXiv:2010.09885. [[CrossRef](#)]
30. Su, A.; Zhang, C.; She, Y.-B.; Yang, Y.-F. Exploring Deep Learning for Metalloporphyrins: Databases, Molecular Representations, and Model Architectures. *Catalysts* **2022**, *12*, 1485. [[CrossRef](#)]
31. Jo, J.; Kwak, B.; Choi, H.-S.; Yoon, S. The message passing neural networks for chemical property prediction on SMILES. *Methods* **2020**, *179*, 65–72. [[CrossRef](#)]
32. Mouchlis, V.D.; Afantitis, A.; Serra, A.; Fratello, M.; Papadiamantis, A.G.; Aidinis, V.; Lynch, I.; Greco, D.; Melagraki, G. Advances in De Novo Drug Design: From Conventional to Machine Learning Methods. *Int. J. Mol. Sci.* **2021**, *22*, 1676. [[CrossRef](#)]
33. Gómez-Bombarelli, R.; Wei, J.N.; Duvenaud, D.K.; Hernandez-Lobato, J.M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T.D.; Adams, R.P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Sci.* **2018**, *4*, 268–276. [[CrossRef](#)] [[PubMed](#)]
34. Segler, M.H.S.; Kogej, T.; Tyrchan, C.; Waller, M.P. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Central Sci.* **2017**, *4*, 120–131. [[CrossRef](#)] [[PubMed](#)]
35. De Cao, N.; Kipf, T. MolGAN: An implicit generative model for small molecular graphs. *arXiv* **2018**, arXiv:1805.11973. [[CrossRef](#)]
36. Flam-Shepherd, D.; Wu, T.; Aspuru-Guzik, A. Graph deconvolutional generation. *arXiv* **2020**, arXiv:2002.07087. [[CrossRef](#)]
37. Brown, N.; Fiscato, M.; Segler, M.H.; Vaucher, A.C. GuacaMol: Benchmarking Models for de Novo Molecular Design. *J. Chem. Inf. Model.* **2019**, *59*, 1096–1108. [[CrossRef](#)]
38. Krenn, M.; Hase, F.; Nigam, A.K.; Friederich, P.; Aspuru-Guzik, A. Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation. *Mach. Learn. Sci. Technol.* **2020**, *1*, 045024. [[CrossRef](#)]
39. Dybowski, R. Interpretable machine learning as a tool for scientific discovery in chemistry. *New J. Chem.* **2020**, *44*, 20914–20920. [[CrossRef](#)]
40. Krishnan, K.; Kassab, R.; Agajanian, S.; Verkhivker, G. Interpretable Machine Learning Models for Molecular Design of Tyrosine Kinase Inhibitors Using Variational Autoencoders and Perturbation-Based Approach of Chemical Space Exploration. *Int. J. Mol. Sci.* **2022**, *23*, 11262. [[CrossRef](#)]
41. Hatamleh, M.; Chong, J.W.; Tan, R.R.; Aviso, K.B.; Janairo, J.I.B.; Chemmangattuvalappil, N.G. Design of mosquito repellent molecules via the integration of hyperbox machine learning and computer aided molecular design. *Digit. Chem. Eng.* **2022**, *3*, 100018. [[CrossRef](#)]
42. Ooi, Y.J.; Aung, K.N.G.; Chong, J.W.; Tan, R.R.; Aviso, K.B.; Chemmangattuvalappil, N.G. Design of fragrance molecules using computer-aided molecular design with machine learning. *Comput. Chem. Eng.* **2021**, *157*, 107585. [[CrossRef](#)]
43. Radhakrishnapany, K.T.; Wong, C.Y.; Tan, F.K.; Chong, J.W.; Tan, R.R.; Aviso, K.B.; Janairo, J.I.B.; Chemmangattuvalappil, N.G. Design of fragrant molecules through the incorporation of rough sets into computer-aided molecular design. *Mol. Syst. Des. Eng.* **2020**, *5*, 1391–1416. [[CrossRef](#)]
44. Helmy, M.; Smith, D.; Selvarajoo, K. Systems biology approaches integrated with artificial intelligence for optimized food-focused metabolic engineering. *Metab. Eng. Commun.* **2020**, *11*, e00149. [[CrossRef](#)]
45. Ji, Z.; Su, J.; Liu, C.; Wang, H.; Huang, D.; Zhou, X. Integrating Genomics and Proteomics Data to Predict Drug Effects Using Binary Linear Programming. *PLoS ONE* **2014**, *9*, e102798. [[CrossRef](#)] [[PubMed](#)]
46. Ji, Z.; Wu, D.; Zhao, W.; Peng, H.; Zhao, S.; Huang, D.; Zhou, X. Systemic modeling myeloma-osteoclast interactions under normoxic/hypoxic condition using a novel computational approach. *Sci. Rep.* **2015**, *5*, 13291. [[CrossRef](#)]
47. Peng, H.; Zhao, W.; Tan, H.; Ji, Z.; Li, J.; Li, K.; Zhou, X. Prediction of treatment efficacy for prostate cancer using a mathematical model. *Sci. Rep.* **2016**, *6*, 21599. [[CrossRef](#)]
48. Corey, E.J.; Wipke, W.T. Computer-Assisted Design of Complex Organic Syntheses. *Science* **1969**, *166*, 178–192. [[CrossRef](#)] [[PubMed](#)]
49. Cook, A.; Johnson, A.P.; Law, J.; Mirzazadeh, M.; Ravitz, O.; Simon, A. Computer-aided synthesis design: 40 years on. *WIREs Comput. Mol. Sci.* **2011**, *2*, 79–107. [[CrossRef](#)]
50. Ihlenfeldt, W.-D.; Gasteiger, J. Computer-Assisted Planning of Organic Syntheses: The Second Generation of Programs. *Angew. Chem. Int. Ed.* **1996**, *34*, 2613–2633. [[CrossRef](#)]
51. Todd, M.H. Computer-aided organic synthesis. *Chem. Soc. Rev.* **2005**, *34*, 247–266. [[CrossRef](#)] [[PubMed](#)]
52. Ruddigkeit, L.; van Deursen, R.; Blum, L.C.; Reymond, J.-L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864–2875. [[CrossRef](#)]
53. Davies, I.W. The digitization of organic synthesis. *Nature* **2019**, *570*, 175–181. [[CrossRef](#)]
54. Coley, C.W.; Eyke, N.S.; Jensen, K.F. Autonomous Discovery in the Chemical Sciences Part I: Progress. *Angew. Chem. Int. Ed.* **2020**, *59*, 22858–22893. [[CrossRef](#)]
55. Coley, C.W.; Eyke, N.S.; Jensen, K.F. Autonomous Discovery in the Chemical Sciences Part II: Outlook. *Angew. Chem. Int. Ed.* **2019**, *59*, 23414–23436. [[CrossRef](#)]
56. Shen, Y.; Borowski, J.E.; Hardy, M.A.; Sarpong, R.; Doyle, A.G.; Cernak, T. Automation and computer-assisted planning for chemical synthesis. *Nat. Rev. Methods Prim.* **2021**, *1*, 23. [[CrossRef](#)]
57. Klucznik, T.; Mikulak-Klucznik, B.; McCormack, M.P.; Lima, H.; Szymkuć, S.; Bhowmick, M.; Molga, K.; Zhou, Y.; Rickershauser, L.; Gajewska, E.P.; et al. Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. *Chem* **2018**, *4*, 522–532. [[CrossRef](#)]

58. Ravitz, O. Data-driven computer aided synthesis design. *Drug Discov. Today: Technol.* **2013**, *10*, e443–e449. [[CrossRef](#)]
59. Law, J.; Zsoldos, Z.; Simon, A.; Reid, D.; Liu, Y.; Khew, S.Y.; Johnson, A.P.; Major, S.; Wade, R.A.; Ando, H.Y. Route Designer: A Retrosynthetic Analysis Tool Utilizing Au-tomated Retrosynthetic Rule Generation. *J. Chem. Inf. Model.* **2009**, *49*, 593–602. [[CrossRef](#)]
60. Christ, C.D.; Zentgraf, M.; Kriegl, J.M. Mining Electronic Laboratory Notebooks: Analysis, Retrosynthesis, and Reaction Based Enumeration. *J. Chem. Inf. Model.* **2012**, *52*, 1745–1756. [[CrossRef](#)]
61. Segler, M.H.S.; Waller, M.P. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chem. A Eur. J.* **2017**, *23*, 5966–5971. [[CrossRef](#)]
62. Coley, C.W.; Rogers, L.; Green, W.H.; Jensen, K.F. Computer-Assisted Retrosynthesis Based on Molecular Similarity. *ACS Central Sci.* **2017**, *3*, 1237–1245. [[CrossRef](#)]
63. Segler, M.H.S.; Waller, M.P. Modelling Chemical Reasoning to Predict and Invent Reactions. *Chem. A Eur. J.* **2017**, *23*, 6118–6128. [[CrossRef](#)] [[PubMed](#)]
64. Baylon, J.L.; Cilfone, N.A.; Gulcher, J.R.; Chittenden, T.W. Enhancing retro synthetic reaction prediction with deep learning using multiscale reaction classification. *J. Chem. Inf. Model.* **2019**, *59*, 673–688. [[CrossRef](#)] [[PubMed](#)]
65. Thakkar, A.; Kogej, T.; Reymond, J.-L.; Engkvist, O.; Bjerrum, E.J. Datasets and their influence on the development of computer assisted synthesis planning tools in the pharmaceutical domain. *Chem. Sci.* **2019**, *11*, 154–168. [[CrossRef](#)] [[PubMed](#)]
66. Genheden, S.; Thakkar, A.; Chadimová, V.; Reymond, J.-L.; Engkvist, O.; Bjerrum, E. AiZynthFinder: A fast, robust and flexible open-source software for retrosynthetic planning. *J. Cheminformatics* **2020**, *12*, 70. [[CrossRef](#)]
67. Tu, Z.; Coley, C.W. Permutation Invariant Graph-to-Sequence Model for Template-Free Retrosynthesis and Reaction Prediction. *J. Chem. Inf. Model.* **2022**, *62*, 3503–3513. [[CrossRef](#)]
68. Cadeddu, A.; Wylie, E.K.; Jurczak, J.; Wampler-Doty, M.; Grzybowski, B.A. Organic Chemistry as a language and the implications of chemical linguistics for structural and retrosynthetic analyses. *Angew. Chem. Int. Ed.* **2014**, *53*, 8108–8112. [[CrossRef](#)]
69. Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C.A.; Bekas, C.; Lee, A.A. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Central Sci.* **2019**, *5*, 1572–1583. [[CrossRef](#)]
70. Lin, K.; Xu, Y.; Pei, J.; Lai, L. Automatic retrosynthetic route planning using template-free models. *Chem. Sci.* **2020**, *11*, 3355–3364. [[CrossRef](#)]
71. Liu, X.; Li, P.; Song, S. Decomposing Retrosynthesis into Reactive Center Prediction and Molecule Generation. *bioRxiv* **2019**, 677849. [[CrossRef](#)]
72. Schwaller, P.; Petraglia, R.; Zullo, V.; Nair, V.H.; Haeuselmann, R.A.; Pisoni, R.; Bekas, C.; Iuliano, A.; Laino, T. Predicting retrosynthetic pathways using trans-former-based models and a hyper-graph exploration strategy. *Chem. Sci.* **2020**, *11*, 3316–3325. [[CrossRef](#)]
73. Wang, X.; Li, Y.; Qiu, J.; Chen, G.; Liu, H.; Liao, B.; Hsieh, C.-Y.; Yao, X. RetroPrime: A Diverse, plausible and Transformer-based method for Single-Step retrosynthesis predictions. *Chem. Eng. J.* **2021**, *420*, 129845. [[CrossRef](#)]
74. Tetko, I.V.; Karpov, P.; Van Deursen, R.; Godin, G. State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nat. Commun.* **2020**, *11*, 5575. [[CrossRef](#)]
75. Ucak, U.V.; Ashyrmamatov, I.; Ko, J.; Lee, J. Retrosynthetic reaction pathway prediction through neural machine translation of atomic environments. *Nat. Commun.* **2022**, *1*, 1186. [[CrossRef](#)]
76. Duan, H.; Wang, L.; Zhang, C.; Guo, L.; Li, J. Retrosynthesis with attention-based NMT model and chemical analysis of “wrong” predictions. *RSC Adv.* **2020**, *10*, 1371–1378. [[CrossRef](#)]
77. Mao, K.; Xiao, X.; Xu, T.; Rong, Y.; Huang, J.; Zhao, P. Molecular graph enhanced transformer for retrosynthesis prediction. *Neurocomputing* **2021**, *457*, 193–202. [[CrossRef](#)]
78. Chen, B.; Shen, T.; Jaakkola, T.S.; Barzilay, R. Learning to Make Generalizable and Diverse Predictions for Retrosynthesis. *arXiv* **2019**, arXiv:1910.09688. [[CrossRef](#)]
79. Pesciullesi, G.; Schwaller, P.; Laino, T.; Reymond, J.-L. Transfer learning enables the molecular transformer to predict regio- and stereoselective reactions on carbohydrates. *Nat. Commun.* **2020**, *11*, 4874. [[CrossRef](#)]
80. Coley, C.W.; Barzilay, R.; Jaakkola, T.S.; Green, W.H.; Jensen, K.F. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Central Sci.* **2017**, *3*, 434–443. [[CrossRef](#)]
81. Wei, J.N.; Duvenaud, D.; Aspuru-Guzik, A. Neural Networks for the Prediction of Organic Chemistry Reactions. *ACS Central Sci.* **2016**, *2*, 725–732. [[CrossRef](#)]
82. Jin, W.; Coley, C.; Barzilay, R.; Jaakkola, T. Predicting Organic Reaction Outcomes with Weisfeiler-Lehman Network. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 2607–2616.
83. Schwaller, P.; Gaudin, T.; Lányi, D.; Bekas, C.; Laino, T. “Found in Translation”: Predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. *Chem. Sci.* **2018**, *9*, 6091–6098. [[CrossRef](#)]
84. Perera, D.; Tucker, J.W.; Brahmabhatt, S.; Helal, C.J.; Chong, A.; Farrell, W.; Richardson, P.; Sach, N.W. A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science* **2018**, *359*, 429–434. [[CrossRef](#)] [[PubMed](#)]
85. Ahneman, D.T.; Estrada, J.G.; Lin, S.; Dreher, S.D.; Doyle, A.G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science* **2018**, *360*, 186–190. [[CrossRef](#)]

86. Schwaller, P. Learning the Language of Chemical Reactions-Atom by Atom. Linguistics-Inspired Machine Learning Methods for Chemical Reaction Tasks. Ph.D. Thesis, Universität Bern, Bern, Switzerland, 2021. Available online: <https://boristheses.unibe.ch/id/eprint/2736> (accessed on 27 May 2021).
87. Żurański, A.M.; Alvarado, J.I.M.; Shields, B.J.; Doyle, A.G. Predicting Reaction Yields via Supervised Learning. *Acc. Chem. Res.* **2021**, *54*, 1856–1865. [CrossRef]
88. Haywood, A.L.; Redshaw, J.; Hanson-Heine, M.W.D.; Taylor, A.; Brown, A.; Mason, A.M.; Gärtner, T.; Hirst, J.D. Kernel Methods for Predicting Yields of Chemical Reactions. *J. Chem. Inf. Model.* **2021**, *62*, 2077–2092. [CrossRef] [PubMed]
89. Sandfort, F.; Strieth-Kalthoff, F.; Kühnemund, M.; Beecks, C.; Glorius, F. A Structure-Based Platform for Predicting Chemical Reactivity. *Chem* **2020**, *6*, 1379–1390. [CrossRef]
90. Sato, A.; Miyao, T.; Funatsu, K. Prediction of Reaction Yield for Buchwald-Hartwig Cross-coupling Reactions Using Deep Learning. *Mol. Inform.* **2021**, *41*, 2100156. [CrossRef] [PubMed]
91. Nielsen, M.K.; Ahneman, D.T.; Riera, O.; Doyle, A.G. Deoxyfluorination with Sulfonyl Fluorides: Navigating Reaction Space with Machine Learning. *J. Am. Chem. Soc.* **2018**, *140*, 5004–5008. [CrossRef]
92. Li, J.; Eastgate, M.D. Making Better Decisions during Synthetic Route Design: Leveraging Prediction to Achieve Green-ness-by-Design. *React. Chem. Eng.* **2019**, *4*, 1595–1607. [CrossRef]
93. Christensen, M.; Yunker, L.P.E.; Adedeji, F.; Häse, F.; Roch, L.M.; Gensch, T.; Gomes, G.D.P.; Zepel, T.; Sigman, M.S.; Aspuru-Guzik, A.; et al. Data-science driven autonomous process optimization. *Commun. Chem.* **2021**, *4*, 112. [CrossRef]
94. Angello, N.H.; Rathore, V.; Beker, W.; Wołos, A.; Jira, E.R.; Roszak, R.; Wu, T.C.; Schroeder, C.M.; Aspuru-Guzik, A.; Grzybowski, B.A.; et al. Closed-loop optimization of general reaction conditions for heteroaryl Suzuki-Miyaura coupling. *Science* **2022**, *378*, 399–405. [CrossRef]
95. Ley, S.V.; Fitzpatrick, D.E.; Ingham, R.J.; Myers, R.M. Organic Synthesis: March of the Machines. *Angew. Chem. Int. Ed.* **2015**, *54*, 3449–3464. [CrossRef] [PubMed]
96. Ley, S.V.; Fitzpatrick, D.E.; Myers, R.M.; Battilocchio, C.; Ingham, R.J. Machine-Assisted Organic Synthesis. *Angew. Chem. Int. Ed.* **2015**, *54*, 10122–10136. [CrossRef] [PubMed]
97. Li, J.; Ballmer, S.G.; Gillis, E.P.; Fujii, S.; Schmidt, M.J.; Palazzolo, A.M.E.; Lehmann, J.W.; Morehouse, G.F.; Burke, M.D. Synthesis of many different types of organic small molecules using one automated process. *Science* **2015**, *347*, 1221–1226. [CrossRef] [PubMed]
98. Chatterjee, S.; Guidi, M.; Seeberger, P.H.; Gilmore, K. Automated radial synthesis of organic molecules. *Nature* **2020**, *579*, 379–384. [CrossRef]
99. Mehr, S.H.M.; Craven, M.; Leonov, A.I.; Keenan, G.; Cronin, L. A universal system for digitization and automatic execution of the chemical synthesis literature. *Science* **2020**, *370*, 101–108. [CrossRef]
100. Steiner, S.; Wolf, J.; Glatzel, S.; Andreou, A.; Granda, J.M.; Keenan, G.; Hinkley, T.; Aragon-Camarasa, G.; Kitson, P.J.; Angelone, D.; et al. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **2019**, *363*, eaav2211. [CrossRef]
101. Gao, W.; Raghavan, P.; Coley, C.W. Autonomous platforms for data-driven organic synthesis. *Nat. Commun.* **2022**, *13*, 1075. [CrossRef]
102. Burger, B.; Maffettone, P.M.; Gusev, V.V.; Aitchison, C.M.; Bai, Y.; Wang, X.; Li, X.; Alston, B.M.; Li, B.; Clowes, R.; et al. A mobile robotic chemist. *Nature* **2020**, *583*, 237–241. [CrossRef]
103. Xu, H.; Lin, J.; Liu, Q.; Chen, Y.; Zhang, J.; Yang, Y.; Young, M.C.; Xu, Y.; Zhang, D.; Mo, F. High-throughput discovery of chemical structure-polarity relationships combining automation and machine-learning techniques. *Chem* **2022**, *8*, 3202–3214. [CrossRef]
104. Angelone, D.; Hammer, A.J.S.; Rohrbach, S.; Krambeck, S.; Granda, J.M.; Wolf, J.; Zaleskiy, S.; Chisholm, G.; Cronin, L. Convergence of multiple synthetic paradigms in a universally programmable chemical synthesis machine. *Nat. Chem.* **2020**, *13*, 63–69. [CrossRef]
105. Bédard, A.-C.; Adamo, A.; Aroh, K.C.; Russell, M.G.; Bedermann, A.A.; Torosian, J.; Yue, B.; Jensen, K.F.; Jamison, T.F. Reconfigurable system for automated optimization of diverse chemical reactions. *Science* **2018**, *361*, 1220–1225. [CrossRef] [PubMed]
106. Coley, C.W.; Thomas, D.A.; Lummiss, J.A.M.; Jaworski, J.N.; Breen, C.P.; Schultz, V.; Hart, T.; Fishman, J.S.; Rogers, L.; Gao, H.; et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **2019**, *365*, 557–565. [CrossRef]
107. Nambiar, A.M.K.; Breen, C.P.; Hart, T.; Kulesza, T.; Jamison, T.F.; Jensen, K.F. Bayesian Optimization of Computer-Proposed Multistep Synthetic Routes on an Automated Robotic Flow Platform. *ACS Central Sci.* **2022**, *8*, 825–836. [CrossRef] [PubMed]
108. Li, C.; Callahan, A.J.; Phadke, K.S.; Bellaire, B.; Farquhar, C.E.; Zhang, G.; Schissel, C.K.; Mijalis, A.J.; Hartrampf, N.; Loas, A.; et al. Automated Flow Synthesis of Peptide–PNA Conjugates. *ACS Central Sci.* **2021**, *8*, 205–213. [CrossRef]
109. MacLeod, B.P.; Parlane, F.G.L.; Morrissey, T.D.; Häse, F.; Roch, L.M.; Dettelbach, K.E.; Moreira, R.; Yunker, L.P.E.; Rooney, M.B.; Deeth, J.R.; et al. Self-driving laboratory for accelerated discovery of thin-film materials. *Sci. Adv.* **2020**, *6*, eaaz8867. [CrossRef] [PubMed]
110. Jiang, Y.; Salley, D.; Sharma, A.; Keenan, G.; Mullin, M.; Cronin, L. An artificial intelligence enabled chemical synthesis robot for exploration and optimization of nanomaterials. *Sci. Adv.* **2022**, *8*, eabo2626. [CrossRef]
111. Soto, F.; Wang, J.; Ahmed, R.; Demirci, U. Medical Micro/Nanorobots in Precision Medicine. *Adv. Sci.* **2020**, *7*, 2002203. [CrossRef]
112. Plante, O.J.; Palmacci, E.R.; Seeberger, P.H. Automated Solid-Phase Synthesis of Oligosaccharides. *Science* **2001**, *291*, 1523–1527. [CrossRef]

113. Caruthers, M.H. Gene Synthesis Machines: DNA Chemistry and Its Uses. *Science* **1985**, *230*, 281–285. [[CrossRef](#)]
114. Ghislieri, D.; Gilmore, K.; Seeberger, P.H. Chemical Assembly Systems: Layered Control for Divergent, Continuous, Multi-step Syntheses of Active Pharmaceutical Ingredients. *Angew. Chem. Int. Ed.* **2014**, *54*, 678–682. [[CrossRef](#)]
115. Britton, J.; Jamison, T.F. A Unified Continuous Flow Assembly Line Synthesis of Highly Substituted Pyrazoles and Pyrazolines. *Angew. Chem. Int. Ed.* **2017**, *54*, 678–682.
116. Weissman, S.A.; Anderson, N.G. Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. *Org. Process. Res. Dev.* **2014**, *19*, 1605–1633. [[CrossRef](#)]
117. Skilton, R.A.; Bourne, R.; Amara, Z.; Horvath, R.; Jin, J.; Scully, M.J.; Streng, E.S.; Tang, S.L.Y.; Summers, P.A.; Wang, J.; et al. Remote-controlled experiments with cloud chemistry. *Nat. Chem.* **2014**, *7*, 1–5. [[CrossRef](#)] [[PubMed](#)]
118. McMullen, J.P.; Stone, M.T.; Buchwald, S.L.; Jensen, K.F. An Integrated Microreactor System for Self-Optimization of a Heck Reaction: From Micro- to Mesoscale Flow Systems. *Angew. Chem. Int. Ed.* **2010**, *49*, 7076–7080. [[CrossRef](#)]
119. Aworinde, S.M.; Schweidtmann, A.M.; Lapkin, A.A. The concept of selectivity control by simultaneous distribution of the oxygen feed and wall temperature in a microstructured reactor. *Chem. Eng. J.* **2018**, *331*, 765–776. [[CrossRef](#)]
120. Bhaskar, V.; Gupta, S.K.; Ray, A.K. Applications of Multiobjective Optimization in Chemical Engineering. *Rev. Chem. Eng.* **2000**, *16*, 1–54. [[CrossRef](#)]
121. Xu, M.; Bhat, S.; Smith, R.; Stephens, G.; Sadhukhan, J. Multi-objective optimisation of metabolic productivity and thermo-dynamic performance. *Comput. Chem. Eng.* **2009**, *33*, 1438–1450. [[CrossRef](#)]
122. Shields, B.J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J.I.M.; Janey, J.M.; Adams, R.P.; Doyle, A.G. Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **2021**, *590*, 89–96. [[CrossRef](#)]
123. Shambhawi, S.; Csányi, G.; Lapkin, A.A. Active Learning Training Strategy for Predicting O Adsorption Free Energy on Perovskite Catalysts using Inexpensive Catalyst Features. *Chem. Methods* **2021**, *1*, 444–450. [[CrossRef](#)]
124. Nandiwale, K.Y.; Hart, T.; Zahrt, A.F.; Nambiar, A.M.K.; Mahesh, P.T.; Mo, Y.; Nieves-Remacha, M.J.; Johnson, M.D.; García-Losada, P.; Mateos, C.; et al. Continuous stirred-tank reactor cascade platform for self-optimization of reactions involving solids. *React. Chem. Eng.* **2022**, *7*, 1315–1327. [[CrossRef](#)]
125. Felton, K.C.; Rittig, J.G.; Lapkin, A.A. Summit: Benchmarking Machine Learning Methods for Reaction Optimisation. *Chem* **2021**, *1*, 116–122. [[CrossRef](#)]
126. Bradford, E.; Schweidtmann, A.M.; Lapkin, A. Efficient multiobjective optimization employing Gaussian processes, spectral sampling and a genetic algorithm. *J. Glob. Optim.* **2018**, *71*, 407–438. [[CrossRef](#)]
127. Häse, F.; Roch, L.M.; Aspuru-Guzik, A. Chimera: Enabling hierarchy based multi-objective optimization for self-driving laboratories. *Chem. Sci.* **2018**, *9*, 7642–7655. [[CrossRef](#)] [[PubMed](#)]
128. Häse, F.; Aldeghi, M.; Hickman, R.J.; Roch, L.M.; Aspuru-Guzik, A. Gryffin: An algorithm for Bayesian optimization of categorical variables informed by expert knowledge. *Appl. Phys. Rev.* **2021**, *8*, 031406. [[CrossRef](#)]
129. Häse, F.; Roch, L.M.; Kreisbeck, C.; Aspuru-Guzik, A. Phoenix: A Bayesian Optimizer for Chemistry. *ACS Central Sci.* **2018**, *4*, 1134–1145. [[CrossRef](#)]
130. Schweidtmann, A.M.; Clayton, A.D.; Holmes, N.; Bradford, E.; Bourne, R.A.; Lapkin, A.A. Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives. *Chem. Eng. J.* **2018**, *352*, 277–282. [[CrossRef](#)]
131. Felton, K.C.; Wigh, D.S.; Lapkin, A.A. Multi-task Bayesian Optimization of Chemical Reactions. *ChemRxiv* **2020**. [[CrossRef](#)]
132. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian optimization of machine learning algorithms. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 2951–2959.
133. Clayton, A.D.; Schweidtmann, A.M.; Clemens, G.; Manson, J.A.; Taylor, C.; Niño, C.G.; Chamberlain, T.W.; Kapur, N.; Blacker, A.J.; Lapkin, A.A.; et al. Automated self-optimisation of multi-step reaction and separation processes using machine learning. *Chem. Eng. J.* **2019**, *384*, 123340. [[CrossRef](#)]
134. Jeraal, M.I.; Sung, S.; Lapkin, A.A. A Machine Learning- Enabled Autonomous Flow Chemistry Platform for Process Optimization of Multiple Reaction Metrics. *Chem. Methods* **2021**, *1*, 71–77. [[CrossRef](#)]
135. Murray, P.M.; Tyler, S.N.G.; Moseley, J.D. Beyond the Numbers: Charting Chemical Reaction Space. *Org. Process. Res. Dev.* **2013**, *17*, 40–46. [[CrossRef](#)]
136. Amar, Y.; Schweidtmann, A.M.; Deutsch, P.; Cao, L.; Lapkin, A. Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis. *Chem. Sci.* **2019**, *10*, 6697–6706. [[CrossRef](#)] [[PubMed](#)]
137. Zhang, C.; Amar, Y.; Cao, L.; Lapkin, A.A. Solvent Selection for Mitsunobu Reaction Driven by an Active Learning Surrogate Model. *Org. Process. Res. Dev.* **2020**, *24*, 2864–2873. [[CrossRef](#)]
138. Jorayev, P.; Russo, D.; Tibbetts, J.D.; Schweidtmann, A.M.; Deutsch, P.; Bull, S.D.; Lapkin, A.A. Multi-objective Bayesian optimization of a two-step synthesis of p-cymene from crude sulphate turpentine. *Chem. Eng. Sci.* **2022**, *247*, 116938. [[CrossRef](#)]
139. Langner, S.; Häse, F.; Perea, J.D.; Stubhan, T.; Hauch, J.; Roch, L.M.; Heumueller, T.; Aspuru-Guzik, A.; Brabec, C.J. Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems. *Adv. Mater.* **2020**, *32*, e1907801. [[CrossRef](#)] [[PubMed](#)]

140. Pollice, R.; Gomes, G.D.P.; Aldeghi, M.; Hickman, R.J.; Krenn, M.; Lavigne, C.; Lindner-D'Addario, M.; Nigam, A.; Ser, C.T.; Yao, Z.; et al. Data-Driven Strategies for Accelerated Materials Design. *Accounts Chem. Res.* **2021**, *54*, 849–860. [[CrossRef](#)]
141. Wang, Y.; Chen, T.-Y.; Vlachos, D. NEXTorCh: A Design and Bayesian Optimization Toolkit for Chemical Sciences and Engineering. *J. Chem. Inf. Model.* **2021**, *61*, 5312–5319. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.