ISSN 2350-1030

International Journal of Recent Research in Physics and Chemical Sciences (IJRRPCS)
Vol. 10, Issue 2, pp: (113-140), Month: October 2023 – March 2024, Available at: www.paperpublications.org

Catalyst Design: A Narrative Study on Artificial Intelligence

Alexander Sixtus Eko

Chemical Engineering Department, University of Abuja

DOI: https://doi.org/10.5281/zenodo.10874819

Published Date: 26-March-2024

Abstract: **Catalysis represents a very essential aspect of chemical production and industrial growth and development. However, designing ideal green catalysts for various chemical processes can be challenging as the usual theoretical noesis coupled with the customary traditional methods (involving reiterative trials-and-errors), can be exhausting, uneconomical and as well as appear inviable in certain cases. This is owing to the fact that going beyond the stale empirical techniques, as catalytic chemical processes become more and more advanced and complicated, catalysts design have advanced beyond the adopted theoretical studies into deep structural and morphological studies of the chemical space of catalytic materials and even as far as involving smart learning (which engages computer aid in catalysts design). Thus, in contribution to the research efforts made towards transcending the challenges encountered by catalyst design practitioners, this review highlights some noteworthy benefits of inculcating artificial intelligence, a top-down front liner approach which have essentially facilitated present day advancements in catalysts design and promotion of green catalysis.**

Keywords: **Artificial Intelligence, Machine Learning, Model, Reinforcement Learning, Automated Robotic Arm.**

1. INTRODUCTION

Artificial Intelligence in Catalyst Design

Artificial Intelligence (AI) is the science that engages the intelligence of machines, particularly computer systems. Despite their ginormous value, not few commercial catalysts (predominantly heterogeneous) have been discovered using trial-and-error experimental approaches that rely on the chemical intuition of catalysis practicians [1]. The understanding in early times have been obscure and ambiguous. However, in current times research methodologies have switched from notional and theoretical studies to data-driven experimental discoveries. The difficulties associated with deviating from the usual empirical experimental approaches towards catalyst design using **Predictive Models** are multifaceted, including the fact that heterogeneous catalysis spans time and length scales of more than nine orders of magnitude [12], and that the catalyst performance depends on many variables, such as; the catalyst composition, morphology, support material and reaction environment conditions (including temperature, solvent and external potential). This inexactness and extensiveness of the parameter space makes the design and optimization of heterogeneous catalysts challenging. In recent past, research specialists have turned to machine learning (ML) as an advanced technique to accelerate the study and discovery of catalysts, by using its tools to navigate the parameter space more efficiently [3-6].

ML is a part of the computer science field specifically concerned with AI – the use of algorithms for data interpretation in a manner that replicates the human learning method. ML teaches computer systems to make decisions by learning from large data sets. The goal is for the machine to improve its learning accuracy and provide data based on that learning to the user [7]. However, it is recognized that ML is already boosting computational chemistry and chemical engineering, at various levels [8-18]. Although various facets have been influenced, however there are complications arising from inconsistency in summarizing the developments in the enumerated areas of ML application [19]. These are extracted from

recent contributions, that can be regarded as complementary and providing an overall perspective of the applications. In summary, the recent recorded contributions that can be reckoned to as complemental and rendering a general paradigm of its application can be summed up in these different comprehensive and conceptual approaches in:

1. Understanding and controlling chemical systems and related behavior [12, 20-27];

2. Computing, optimizing, or predicting structure-property relationships [28-32], density functional theory (DFT) functionals, and interatomic potentials [24, 33-39];

3. Driving generative models for inverse design (i.e., produce stable molecules from a set of desired properties) [40-51];

4. Screening, synthesizing, and characterizing new compounds and materials [52-56];

5. Improving catalytic technologies and analytical tools [39, 47, 57-60];

6. Developing quantum algorithms for molecular simulations, and progressing quantum sensing [24, 27, 29, 61-65], and so on but to mention a few examples.

The complementary fields of chemical sciences and engineering are data-rich areas, encompassing complex information which is often unstructured and poorly understood. The inculcation of AI has overwhelmingly reduced the challenges in design and the tediousness of experiment by enabling laboratory computerization and mechanization [66], prognosticate bioactivities of new drugs [67-69], optimizing conditions of reactions [70] and proposing synthetic routes to complex target molecules [71].

ML Models in Catalyst Design

ML is a subfield of AI that embraces methods that self-deduce perceptual structures and patterns from related data. Catalysis researchers have taken advantage of these learned structures and patterns to streamline their work in various gray areas and at different levels, including atomistic simulation of reaction conditions [72, 73], catalyst surface phase diagram construction, reaction mechanism prediction [74, 75] and catalyst structure disambiguation [76-78]. Most applications of ML in catalysis thus far have used **Black-Box Models** (such as Gaussian process models or neural networks), widely used in catalysis – of which computational high-throughput catalyst screening has benefitted the most [79-82] in making predictions of computable physical properties (also known as descriptors) such as adsorption or formation energies, that can be related to the performance (that is, activity or stability) of the catalyst material [1-4]. A models' interpretability and accuracy are key to its implementation. However, black-box models are accuracy biased over interpretability/explainability. The challenge with extracting meaningful physical insights from black-box models as a result of the high degree of complexity of the models is as a result of the cumbersomeness of its internal logic making it not readily explainable. However, in order to bypass the critical challenge associated with black-box models in interpreting the internal logic leading to the conclusive results of the model, there exists a class of methods for indirectly extracting interpretable information from black-box ML models after training. These approaches are termed post-hoc analysis methods, referred to as **Grey-Box Methods**.

A number of grey-box methods are model-agnostic and thence usable with any class of ML model. The information from grey-box methods can take variety of forms but is usually a set of visualizations or sensitivity measures called feature-importance scores [71]. Grey-box methods can generate explanations that are either global or local. Global explanations allow interpretation of the dataset-level relationships and patterns learned by black-box models, whereas local explanations allow practitioners to understand why black-box models make a specific prediction for a single data point. Not all ML methods require a grey-box method to interpret the relationships they have uncovered. Some ML methods yield such insights directly, referred to herein as **Glass-Box Models**.

Glass-box models are used to find simple analytical expressions that relate input variables to target properties, to identify hidden or underlying structures in the data, to make predictions under enforced modularity or causal structure, or to suggest causal structure directly. Generally, these glass-box methods have constraints, such as enforced simplicity, that make direct interpretation of glass-box ML results possible. However, glass-box methods are preferred over others if extracting scientific insight is the central objective. Notwithstanding, in spite of the accomplishments of explainable ML application in catalysis, hitherto the field is still incipient, with significant challenges yet to be surmounted for it to reach its full potential. Esterhuizen *el al*. [1] accordingly highlighted the critical challenges that are related to ML's desegregation with

experimentation, including; desegregating explainable ML with experimental data, capturing the intricacies of catalysts design with ML-deduced descriptors, magnitude of dataset, model accessibility and recyclability, explicating interpretations. In other to address these challenges, researchers (both theorists and experimentalists) must cooperate attentively.

There are higher chances based on speculations of an increase in research interest on the application of ML in the field of heterogeneous catalysis in the future. It is near certain considering the complexity of the challenges (such as uncovering structure–mechanism–reactivity relationships for multi-component catalysts; explicating the physicochemical properties that govern photocatalytic, plasma-catalytic and electro catalytic reactions; analyzing microscopic and spectroscopic data; etc.,) encountered in the field are advanced for simple experimentations to rede. ML offers great prospects in tackling such problems.

Records of its early success for progressively accelerating research and development in different fields of application invigorates the sureness in its ability to fast track tackling these challenges. Nevertheless, applications of ML to generate novel cognition and hypotheses remain few and far between. However, Strieth *et al.* [71] points out that this is chiefly due to the fact that most applications of ML in catalysis up till present have utilized black-box models. Black-box models can be advantageous from the pragmatic standpoint of high its predictive accuracy. Nonetheless, directly interpreting their behavior is intractable. Several interpretable ML methods have been employed in recent heterogeneous catalysis studies, which have been broadly grouped into grey-box and glass-box methods. Grey-box methods allows for the interpretation of black box-models (like glass-box models which have this built-in feature). However, sole reliance on grey-box methods is risky as a result of a potential information gap between the black-box model and the interpretation. On the other hand, glass-box models exhibit superiority over grey-box models for modal applications which entails that the development of scientific insight is the principal objective.

In this current dispensation, it is ultimately rational that further applications of ML methods in interpretation of catalysis (especially heterogeneous catalysis) will accelerate noetic generation in the field of study coupled with researchers continued demand for novel and advanced evolutionary ways to bridging the gap that exists between man and computers [83, 84]. Over the years, AI increasingly penetrated and recorded notable successes when employed at various levels in the scientific and technological space where there are important challenges in the deployment of an intelligent framework for designing new catalytic materials [85-88].

Review Studies on Application of Computer Programming and AI in Catalysts Design

A Trend on Challenges and Opportunities

Traditionally, the method employed in the discovery of novel catalysts is by trial-and-error coupled with chemical intuition. Nonetheless, further research has made it evident that this process can be automated. An automatic ML framework development capable of guiding itself to discovering intermetallic surfaces with desired catalytic properties is desirable [85]. Tran and Ulissi [83] reported of the creation of an automated machine-learning framework for catalysts screening in their work by building up a small dataset of the reactivity properties of metal surfaces. Contrary to the conventional computational method involving hard-coded instructions provided by technical ingenious, the platform teaches itself by progressively seeing examples using algorithms. Tran and Ulissi [83] screened a combination of 31 elements in the chemical compound from materials project [89], of which the machine-learning platform identified 130 candidates surface across 54 intermetallics for CO_2 reduction reaction (CO_2RR) and 258 surfaces across 102 intermetallics for H_2 evolution reaction (HER). Furthermore, they devised a fingerprinting method utilizing atomic numbers, Pauling negativity, coordination numbers of the adsorbate with surface atoms, and average adsorption energies on pure metal surfaces for each element type. However, the models output target was the reactivity descriptors, i.e. *CO and *H adsorption energies for CO_2RR and HER, respectively. Such simplification of complex reaction networks is enabled by the scaling relations between adsorption energies of key reaction intermediates on metal surfaces [90, 91]. In simpler terms, this approach offers the possibility of training computers with the known catalytic properties of materials and predicts the top potential catalysts for a reaction of interest (specifically electrochemical CO_2RR and HER from water powered by green electricity) were used as test systems in this study. The traditional catalysts are either too expensive to procure on a large scale and/or require higher energy penalty than the prescribed thermodynamics. This can be a handful in the studies such as that requiring the design of viable catalysts for the contending half reactions simultaneously occurring at the cathodic end, conjugated with the balancing

reaction at the anode for water oxidation is of primal interest, with high demand for sustainable production of chemicals and fuels while mitigating (GHG) emissions which has in recent times been propelled by societal clamor as a driving force. The framework is designed to be fully automatic, which allows researchers from diverse backgrounds to have a hands-on experience of catalyst exploration without diving into the details of quantum chemical simulations of catalytic systems and machine learning. Whilst the work of Trans and Ulissi [83] presents an exciting step towards leapfrogging the development from its current stage, yet it should be stressed that the development of a strong, versatile and intelligent ML framework for catalysts design is in its early stages. The development is largely attributed to the collaborative efforts amongst materials scientists and engineers, machine-learning practitioners and algorithms developers to make practical use of AI in catalysis and catalysts design.

Although machine learning offers the promise of being able to explore a vast chemical space automatically, there are still a few challenges toward the implementation of AI in catalysis:

1. Firstly, additional measures – including thermal/electrochemical stability, adsorbate-induced surface segregation, and cost of precious metals and the simplicity of synthesis – are essentialities to farther specify the candidate choice of catalysts before synthesis and experimental testing.

2. Secondly, in the case where DFTs apply the uncertainties associated with the involvement of machine-learning models needs to be defined in quantity. This is due to the approximations in DFT functionals, reaction mechanisms, energy scaling relations, along with intrinsic errors in model regression [92, 93] (which will be later addressed in this review study).

3. Thirdly, ML models typically are regarded as black-box models due to their unease of interpretation. However, here importance is be placed emphatically on identification of the active materials whilst deducing the factors fundamental to optimum characteristics, which is critical for increasing the confidence of model users, and in the best case extracting novel scientific/technological design principles for catalysts.

4. Lastly, to keep pace with industries' ever growing demands for more complex and multi-functional catalytic materials in chemical/energy transformations, catalysis practitioners are tasked to develop fingerprints of more sophisticated active sites with targeted functionalities [93].

Over the years, predictive models developed (such as models developed for oil production and transformation [94-98]), did not achieve the uttermost success in series of their applications [99, 100]. However, in the course of evolvement and search for improved and advanced techniques inculcating AI, successes have been recorded in diverse fields of study, including catalyst designs [101-104]. Xim Bokhimi [101] described the use of AI techniques in industrial processes related to heterogeneous catalysis in hydrodesulfurization. The work aimed at giving a predictive approach that involves the use of machine learning in creating a database suitable and applicable in future researches and industrial process relating to hydrodesulfurization processes. In their approach, two variables (X and Y) were defined in such a way that these two descriptors were mutual to one another; mutual in the sense that there is much emphasis on a peculiar feature that spotlights these variables which will aid in describing them when one of both works on a new application. These descriptions presented were relative to the construction of learning machines that could:

1. Deduce adsorption energies (with mean absolute errors of 0.15eV for a diverse chemical space);

2. Predict novel catalysts through learning from the catalytic behavior of materials synthesized via atomic substitutions;

3. Predict more accurately surface areas of porous matters than the usual Brunauer-Emmett-Teller (BET) method;

4. Define the adsorption isotherms of nanoporous materials more speedily than the Monte Carlo mathematical simulation technique used for predictive modeling; and

5. Predict the sulfur content in a terminal product after hydrodesulfurization by learning from the variables associated with the process.

These predictions prompted the authors to synthesize several catalysts through the substitution of Ru with the following 19 elements: Cu, Ni, Cr, W, Hf, Zn, Bi, Pd, Mo, Y, Sc, Sr, Mg, Os, Pt, Au, Nb, Fe, and Rh. Furthermore, in a series of recent papers, inspired by the classic Sabatier principle established in heterogeneous catalysis [105, 106], volcano plots were

introduced as effective tool to performing high throughput virtual screening in homogeneous catalysis, and perform insightful result analysis at the same time, to find optimal catalysts [107-110].

Likewise in recent studies, computer-supported catalyst design approach has received heightened attention with the basic application of *ab initio* simulations (i.e. calculations from first principles in a software/code based on DFT) for chemical reactions and their potential energy surfaces in predicting the practicability of thermodynamic and kinetic specific transformations [111-113]. *Ab initio* simulations perhaps may seem too basic since they handle schematic and simplified or not so complex physical systems with ideal interfaces, and transitions at the atomic scale. Nonetheless, their results converge to an exact solution, but only when approximations are small in sufficient magnitude. In times of recent, this method viable for calculating electron-density distribution around moving ions provides the most accurate modelling techniques. In the course of catalysts design, they take into cognizance and account for local chemical and magnetic effects thereby providing significant potential for prediction of catalytic characteristics/properties. The systematic application of machine learning in homogeneous catalysis design is becoming more and more of a predominant research topic which has experienced an upsurge with the invention of novel schemes for optimizing processes via strategies such as statistical modelling of experimental reactivity and selectivity data accompanied with chemical descriptors [114-116]. Quintessentially utilizing multivariate linear regression models for experimental drifts modelling joined with the development of new computational descriptors ultimately gave way for fast catalyst design in addition to improved yield of products, improved rate of reactivity and improved selectivities, such as *enantio*-selectivity of chiral molecules (enantiomers) [117-119]. A hybrid design method (using machine learning coupled with computational transition state modelling) has experimentally proven to be highly accurate in reproducing Gibbs free activation energies for nucleophilic aromatic substitution reactions, of which in principle this workflow is applicable to catalytic reactions [120]. The primary challenge of catalyst optimization is to a large extent catalyst design [121] and therefore it is not only necessary to streamline catalytic experiments themselves and their analysis, but the catalysts method of synthesis needs to be rethought from the bottom-up for efficient closed-loop optimization studies [122]. Improved catalytic reactions will streamline the synthesis pathways of new catalysts making novel developments possible [124]. In that sense, catalysis is autocatalytic for the development of new catalysts. As noted earlier, the challenge of predictive ML models represents one part of the challenges, and furthermore the interpretation of the ML models another. Realizing simple explainable artificial intelligence models can be dissatisfactory (such as the contemporary black-box models of many ML methods) [121, 124-127].

Another route toward achieving interpretable artificial intelligence could be via involvement of symbolic artificial intelligence [128]. Chemistry is a vast field of study that largely depends on symbols for representations and interpretation of processes and related phenomenon. However, interpretability is highly dependent on virtual representations of given problems (which are computational in this case) [129]. Therefore, reevaluating utilized model representations will naturally guide against complexities while encouraging interpretable artificial intelligence models development. Nonetheless, in that regard the panoptic use of descriptors is more or less dissatisfactory as it can lead to unintended negligence/exploitation of essential hidden correlations.

From the perspective of quantum mechanics, the approach of direct use of molecular wave functions or molecular electron densities, or systematic simplifications thenceforth, would be best appealing. On that note, the ML models represent the operators acting on the wave functions, i.e., the use of molecular representations to deliver the accompanying observables [130-132].

Another noteworthy development in ML that are progressively infiltrating the center stage in chemical catalysis design are **Generative Models** [133-135]. Generative models are smart concepts that suggests new molecular structures with peculiar targeted characteristics. This is perfectly suited for exploratory optimization problems without restriction to the known molecular space. This kind of workflow has enormous potential when applied systematically to catalyst design, with significant advances concisely in the field.

Presently, the whole of scientific and patent literature on homogeneous catalysis is an immense data swamp that is prohibitively tiresome to excavate as a result of uncommon, rather than the basic criteria for reporting results being implemented. However, the need for a factual database of catalysts has been recognized decades ago [136]. At the current rates of advancement in computational chemistry, distinctively scientists have been able to simulate nanomaterials and reaction systems with comparatively higher precision. It is much easier to execute mathematical operations for a model

system with over 1000 atoms by utilizing optimized DFT methods. Despite this, for catalytic reactions, *ab initio* (from first principle) computations of ample systems interactive with their chemical environ remains a major challenge. The development of fresh hypothetical techniques for anatomizing individually the principal/elementary catalytic steps pertaining to adsorption, desorption, diffusion, reaction and final conversion, elucidating the evolution of electronic characteristics at every step, would be extremely invaluable. Ultimately, it all points out to the fact that dialogue between theorists and experimentalists is indispensable. Existent catalytic systems are complex and operate in a fashion that contemporary computing methods cannot sufficiently capture. Experimental methods and techniques can be thus complementary with theoretical insights to improvements in catalytic designs. Conversely, experimentalists can optimize their methods by exploiting insights from theoretical calculations. Currently, predictive modelling in catalysis using solely theoretical simulations stands as unapproachable [137].

Furthermore, in current time's autonomous atomistic-scale (AAS) computations, have been engaged as excellent tools in revolutionizing heterogeneous (electro-catalytic) processes [138]. Heterogeneous (electro-catalysis) have been widely and essentially employed in the industrial production of chemicals, from refineries and petrochemicals, manufacture of ammonia and sulfuric acid. Nonetheless, some opportunities and challenges are still being encountered, including the fact that:

- AAS computations can significantly accelerate improvements in catalysts design. However, the essential (software) substructures have received marginalized degree of attention and acceptance;

- Peculiarly, reinforcement learning has been better modified to meet with the demands of targeted properties in catalyst development in addition to minimized human cost. Notwithstanding, accounting for catalyst stability under operating conditions is rather challenging but accomplishable by a combination of high-throughput computations and machine learning. High-throughput computations are very powerful.

- AAS computations could be complemented by autonomous laboratories, given chance for high-throughput experimentations and catalyst optimization via ML. Withal, the large capital investment required for procurement and maintenance of research facilities could be a major hindrance [139].

The atomistic computations which is of relevance to comprehending and designing of heterogeneous catalysts engages various tools. For example, mass and charge transport have significant impact on the overall kinetics and are contingent upon the meso-scale (intermediate) characteristics of the catalytic layer as exemplified for the reduction of CO_2 [140] and production of hydrogen from solar (photovoltaic) cells [141]. However, the intrinsic activity of a catalyst is primarily ascertained by its structure at the atomic-scale and determines the highest peak of its performance. In heterogeneous (electro-catalysis), this intrinsic activity, being the focus of the present perspective, is closely connected to the complex interface between the solid catalyst and either the gas phase (exemplary for petrochemistry) or liquid (or aqueous) phase, typical for electro-catalysis and biomass conversion.

Another efficient tool to consider are **Surrogate Models**. To efficiently search through the vast chemical space of possible catalysts, screening can, and should be accelerated by surrogate models [142]. Surrogate models like predictive models are a class of mathematical functions that assay a system to forecast the outcome of a costly computation established on descriptors much easier to source, such as geometric fingerprints, graphs, or intrinsic material properties like the d-band center. However, these days surrogate models predominantly are reliant on ML [142], gradually replacing their ancestor (which is a linear scaling relation) [143]. In spite of this trend toward sophisticated ML, the prediction of transition-state energies for hydrodeoxygenation (particularly relevant for biomass conversion) demonstrated that linear models lead to essentially the same quality as more advanced, non-linear models [144]. Nevertheless, in general, machine learning holds high promise for accelerating molecule, material [145, 146], drug [147], and most importantly catalyst [148, 149] development based on a data-driven paradigm. In addition and in a short remark, it is noteworthy that surrogate models in the context of catalysis design can be classified under two broad categories [139]:

1. Interatomic potentials (also called machine-learning potentials (MLPs)) that are used as an alternative to DFT; and
2. Effective models that are used to circumvent energy evaluations of atomistic models altogether.

Featuring surrogate models conjugated with costly DFT computations brings about the attainment of autonomous computational workflows whereby the surrogate models are used to determine additional computations that are required either to (in)-validate the predictions of the model or to buffer the model into a more robust one.

In order to deliver on the promise of accelerating catalyst design, research and developments would have to shift focus, transitioning from simpler theoretical ‘‘proof-of-principle’’ demo, where the ultimate objective was to showcase the power of ML, to applications in solving problems, (such as the use of OC20 in training graph-convolutional neural networks for replacement of DFTs energy evaluations by MLPs41 to actually develop novel catalysts, such as demonstrated for CO_2 electro reduction) [150]. In this context, again generative models become a cornerstone. A generative model is a function proposing structures outside of the training set. In the absence of these ‘‘universal’’ generative models, we are advocating increased reliance on reinforcement learning (RL) techniques when performing ‘‘problem driven’’ machine-learning studies. As noted before, RL is yet to be widely adopted by the scientific and technological community. Nonetheless, RL has already proven to be highly efficient when applied in optimizing a vast number of chemistry-relevant functions. For instance, RL has been trained to efficiently optimize the geometries of organic molecules [151] and in addition to determining the threshold energy pathway in the complex Haber-Bosch reaction over Fe(III), efficaciously learning their chemical kinetics [152].

Active learning minimizes the amount of futile computations in order to limit, comparably only a small range of corresponding training sets that are needful. RL has more often than not been instrumental as an active learning framework where training set built in accordance to the demands of the model being trained and the promising or target locations being explored thence cutting down on the amount of ‘‘useless’’ computations so that only comparably small training sets are needed. Ideally, this is well suited for transfer learning: having learned an optimal policy for one problem might be an excellent starting point for learning an optimal policy on a related problem making predictions less enigmatic for similar problems. Transfer learning is also closely linked to combining information from different sources. For instance, a combination of three sources of information (such as surrogate models, DFT computations, and experiments) in an optimum way might not be transpicuous, but under the condition that they exhibit somewhat similar trends, the algorithms on how to minimize the overall cost incurred for solving a given optimization problem have therefore been developed [153]. Furthermore, given that RL is, by construction, problem and system acquainted, formulating generative models is more tractable: within a class of systems it is quite natural to screen out for selection the sites of interest, both for surface modifications (typical for substitutional doping) and adsorption sites.

Ability to train surrogate models with small training sets is very significant and important when it comes to the application of ML in solving novel problems for which the required training set was initially or naturally not constituted beforehand. The autonomous high-throughput experimentation which was developed some decades back to explore a vast chemical compound space, (e.g., metals, organics, organometallics, and inorganic solids [154]) was advantageous to researchers who were able to unravel chemical compounds then unknown, exhibiting more exotic characteristics than those conventionally synthesized thereby acquiring a larger data base on catalysts at a rapid and unprecedented scale [155]. However, since high-throughput experimentation is more of a cost intensive approach compared to the conventional ways, experimental designs involving catalyst synthesis, characterization, and testing would have to be prudently organized to maximize information output with limited number of experimentations. As earlier stated, with rapid progress in automation control, integrative high-throughput synthesis of catalysts can be made fully autonomous with the support of a guided robotic arm attached to the control system – as employed in catalysts synthesis by the sol-gel method. Therefore, optimized parameter conditions, e.g. catalyst composition, mixing sequence, reaction treatments, and so on can be obtained with minimal human intervention [156-158]. For instance, sputtering is a versatile process for catalyst synthesis at both laboratory and industrial scale, enabling deposition of a thin film of catalyst with controllable thickness and hence might be laborious for human intervention. This, however, might be convolutional but if well planned will be socioeconomically viable and functional in achieving some of the objectives of green chemistry (such as safer chemistry). High-throughput depositions allow creation of a library of catalyst materials with a controllable composition gradient and a large range of film thickness [159].

Pulsed-laser deposition is another phenomenal technique that promotes fast and massive deposition of numerous homogeneous materials, executed by an ablation from a high-energy UV laser. This technique has likewise been adopted for the combinatory catalyst library, which uses a typical series of quaternary masks in a so-called multi-plume pulsed-laser deposition system [160]. Sample holders are being rotated by autonomous robotic arm which are designed usually to house pellet precursors and the consequent transfer for further treatment and characterization. In the sol-gel synthesis of catalysts, a library of catalysts as usual is autonomously prepared by a robotic arm with a pipette to convey the precursor solutions, transferring them to small phials (between 2–5 mL in capacity) which serves as micro-reactors for wherein sol-gel reaction

takes place [161]. In recent times, jet dispensing (equipt in automatic printing technology) was used for high-throughput synthesis of a library of co-crystals [162]. In setting up the gradient library in parallel, precursor ink was formulated with a preset concentration. This technique assures a more rapid flow in fluid dispensing and also gives a compositional gradient of higher degree of accuracy and conformity, thence cutting down on the number of experimentations to be carried out and economizing the time spent in production. However, according to Cong *et al.* [162] the record has attained about 1,000,000 formulations within an operating hour.

Although combinatory synthesis involves the formulation of vast arrays of the gradient materials, nonetheless, the discovery of structure property relationships is accelerated by high-throughput characterization – which can be made amply autonomous by the robotic arm [163]. An example is the development of automated rotating sample changer for X-ray diffraction to identify crystallographic features of catalysts. In D8 ADVANCE, developed by Bruker, which can measure up to 90 samples in parallel, the robot arm transfers the sample to the rotation sample stage, allowing permanent rotation and automatic positioning adjacent to the X-ray beam [164]. Raman spectroscopy is a powerful and non-destructive tool to obtain surface properties of catalysts and elucidate reaction mechanisms [165]. In a modern high throughput Raman technology setup, a robotic system is employed to move samples and acquire data. This is typically done by the deposition of molecular or solid catalysts onto the multi-well plate attached to an automated sample stage [166]. Achieving laser beam focus is one of biggest challenges in measuring high throughput Raman spectra for non-experts, and for this reason autofocus technology has been developed to allow laser beam refocusing during sample holder rotation [167]. Several advanced Raman technologies such as UV resonance Raman spectroscopy, surface-enhanced Raman spectroscopy, and time-resolved and spatially resolved Raman spectroscopy can also gather information on how catalytic mechanisms occur by probing the solutions or reaction intermediates for catalytic CO_2 reduction, water splitting, or water purification [168].

High-throughput catalyst experimentation is very significant in accelerating catalyst formulation [169]. For example, automated analysis of catalytic products can be carried out by gas chromatography-mass spectrometry (GC-MS) and high-performance liquid chromatography utilizing a robotic handling pipette designed to render reliable and accurate liquid injection, sample preparation, and pretreatment. Having a miniaturized electrochemical workstation is convenient for conducting parallel catalyst testing, whose end target is to shrivel chemical laboratories to a lab-on-a-chip system. Microfluidic reactors are sophisticated setups used to test catalyst activity. Their advantages are their versatility, small volumes, fast operation speeds, and capability of parallelization, as well as well-controlled parameters (e.g., temperature, pressure) [170]. For instance, researchers have studied a gradient catalyst consisting of Cu, Pd, and Au (Cu_x Pd_y $Au_{(1-x-y)}$ alloy) connected to individual microfluidic channels, where each end of the channel is accessible by a programmable and movable liquid-handling robot-equipped GC-MS nozzle which rapidly screens 100 H_2/D_2 exchange products within 10 min [171].

Noteworthy, according to Stephan *et al.* [140], catalyst development can be accelerated by high throughput, autonomous computations that identify promising (active) and realistic (under given reaction conditions) catalyst surfaces. Taking into account the inordinate complexities associated with predicting the practicability and (long-term) stability of a given catalyst from the application of first-principles atomistic computations, the involvement of high-throughput autonomous computations in optimal theory-guided catalyst design consisting of *in silico* screening of the chemical space vies to ascertain promising components and active-site patterns. This screening of the chemical space is followed by experimental machine-learning enhanced optimization of the synthesis protocol and the reaction conditions to achieve active and stable catalysts within the computationally identified family. This experimental optimization can integrate any user-defined cost function, for example a trade-off between price, activity, and stability. With time, the autonomous laboratories might become as available as supercomputing facilities, opening a new branch of catalysis research, requiring skills somewhere between experimental and computational sciences.

Furthermore, with the seemingly uncontrollable diminution of fossil-fuel reserves and the global call to mitigate environmental pollution and greenhouse gas accumulation, the demand for alternative source of energy in the world today has even the more in this present dispensation stirred up the need to convert stable molecules (such as CO_2, CH_4, H_2O) into fuels and useful chemicals [172-174]. In the catalytic conversion of CO_2 activation of other molecular reactants is needful as well, molecular reactants such as CH_4, H_2O, and H_2 in particular, because of its environmentally benign nature can serve as a reagent which is produced by H_2O electrolysis or photo splitting thereby eluding the production of redundant CO_2 [175-177]. Oxygen vacancies have been proposed as active sites for CO_2 conversion on some materials [178]. Consequently, this

gives rise to a new idea that predictions of catalytic activity of materials for CO_2 conversion can be refined based on analysis of activation of other reactants and defects.

It is not new that catalysts design involves predictive modeling of the catalyst-reactants interaction which is challenging as a result of the complexity and multifariousness of structure-property relationships across the vast chemical space. Mazheika *et al.* [179] reported a systematic scheme for a noetic catalysts design using the AI (viz. decision tree regression (DTR) and subgroup discovery (SDG) analysis) in identifying catalyst genes (features) that match with chemical mechanisms that initiates, facilitate, or impede the activation of CO_2 leading to a chemical conversion. SDG permits the identification of one or more distinct combinations of catalyst materials features (genes) that encourages catalyst activation. On the other hand, DTR analysis is performed using Python scikitlearn libraries with fitting of model done with regard to the cost function (i.e. mean squared error (MSE) and mean absolute error (MAE)), enclosing the deviation of fitted values of a target property from the actual values [179]. The AI model is trained on first-principles data for a wide category of oxides. Surfaces of good catalysts experimentally established have been shown to systematically exhibit gene combinations which is consequential to the strengthened extension of a C-O bond. These same gene combinations minimized the OCO-angle, which was formally suggested as an indicator of activation, even though under the constraint of the satisfied Paul Sabatier principle. The Sabatier principle is a qualitative way of predicting the activity of heterogeneous catalyst and in this case is taken into account in order to ensure that the adsorption energy between reactants and catalysts is neither too strong nor too weak. However, for this principle to be satisfied the transfer of valence electron to CO_2 has to be intermediate or moderate. This phenomenon is attainable by the charge density delocalization about the oxygen (O) sites and/or by defacement of the adsorbed molecule resulting from covalent bonds formation between O-atoms in CO_2 and surface cations. It was based on these findings that Mazheika *et al.* [179] not only developed the subgroup-discovery strategy for finding improved oxide-based catalysts for the conversion of chemically inert molecules such as CO_2 into useful chemicals or fuels but also proposed a set of new promising catalyst materials for CO_2 conversion. On account of the number of active surface cuts and Paul Sabatier principle, Mazheika *et al.* [179] proposed $NaSbO_3$ to be the most promising yet to be explored catalyst for temperatures ranging as high as 340°C and for pressures of about 1 atmosphere of CO_2. Going further, Mazheika *et al.* [179] proposed some $A^{+1}B^{+5}O_3$ type promising materials namely: $KSbO_3$ (for temperatures up to 110°C) and $RbNbO_3$ (up to 360°C) belonging to the two subgroups, and $LiSbO_3$ (230°C), $CsNbO_3$ (260°C), $CsVO_3$ (110°C), $NaVO_3$ (130°C) all of which are classified under either of the subgroups.

Furthermore, the ability to routinely employ ML models periodically in the execution of research practices will have a drastically be impactful on R&D of polymeric materials. Advancements in ML and automated experimentation are poised to vastly accelerate research in polymer sciences. Studies have recurrently shown that AI systems and ML models features immense prospects in achieving a much more rapid development of polymeric materials through labor costs economization following antecedent time reduction in running processes for further identification of optimized material surfaces [180-184]. Another vital facet for enabling ML integration in research workflows is data representation. However, a lot of data models are inflexible and rigid making it herculean to fit-in a wide matrix of experiment and experimental data types found in field of polymer sciences. This inflexibility presents a significant barrier for researchers to leverage their historical data in ML development. Park *et al.* [185] showed that a domain specific language, termed Chemical Markdown Language (CMDL), provides flexible, extensible, and consistent representation of disparate experiment types and polymer structures. CMDL enables seamless use of historical experimental data to fine-tune regression transformer (RT) models for generative molecular design tasks. Park *et al.* [185] demonstrated the utility of this approach through the generation and the experimental validation of catalysts and polymers in the context of ring-opening polymerization (ROP), critically showing how the CMDL tuned model preserves key functional groups within the polymer structure, allowing for experimental validation.

Although open source repositories, data models, and polymer representations have significantly advanced the development of ML models for polymer chemistry, there exists a need for software tools which provide flexibility in experimental data representations and their translation into ML training sets. Such tools are able to invalidate important barriers and enables research groups to set out leveraging their own historical data sets in ML applications as well as provide an interface to the broader ecosystem of open-source tools, databases, and models being developed for polymer informatics. To create a highly adaptable software toolkit for data representation and demonstrate its utility in ML workflows for catalyst and materials design, first identified were three critical features [185]:

- Extensibility — such that new data or experiment types can be readily accommodated;
- Support for definition of polymer representations; and
- Support for representation of continuous-flow experiments.

Datasets developed by the use of CMDL facilitated the development of very efficient RT models for ROP catalysts design and structurally valid co-polymers. The conventional R&D pattern ordinarily carried out by experimental trial-and-error in the research laboratory [186, 187, 188], would demand researchers to determine the latent target sites of catalyst materials and synthesis paths based on prior knowledge, and then move on to carrying out experimentations to optimize the synthesis conditions. The extensive development cycle and performance unpredictability hinders discovery and application of new catalytic materials. However, DFTs computations have been broadly applied in examining and/or predicting theoretical performance of catalytic materials and also explore deep into the mechanism of their reaction [189-191]. In spite this, the unexplored chemical space might seem infinite since DFT calculations are limited when it comes to predicting them. However, the computational power cost of DFT calculations increases significantly when taking into cognizance the increase in actual environmental factors [189, 192]. Therefore, more strategies are requisite to speed up discovery of new materials.

AI is no new technology, however in various fields and at various stages in R&D it has recorded successes when systemically applied:

- To theoretical studies;
- In the development of scientific methods and techniques; and also
- In systems advancement used to simulate, broaden and flesh out human intelligence [193].

Since the inculcation of AI via ML and intelligent robot technology over the years there has been widespread advancement in practical and theoretical studies [194-196]. The application of ML in structuring predictive models is consolidated on the availability of qualitative and quantitative variety of algorithms to predict the unknown chemical space – which is the soul of its applicability in catalyst design [194, 197-199], material characterization [200-203] and a host of other fields of study interest [204, 205]. AI technology (such as ML, deep learning, robotics) have with great precision improved the screening and synthesis rates, accuracies of energy catalytic materials [196]. Han and Xiang [206] on "the applications of AI in intelligent design and synthesis of energy catalytic materials, summarizing and introducing AI techniques applied to materials science, gave detailed description of the workflow of ML, data sourcing and assemblage, as well as providing information regarding the algorithms applicable in ML models structuring. Han and Xiang [206] extensively, elaborated on the aspects of: ML models applications in intelligent design of materials, intelligent synthesis technology of materials, intelligent characterization of materials, intelligent design of energy catalytic materials, and so on. Catalytic materials usually have different theoretical performance indexes in diverse fields of applications, such as theoretical *overpotential* in electrocatalysis field [207], adsorption energy of intermediates associated with *overpotential* [208], band gap in photocatalysis field [209], etc. These performance evaluation parameters are commonly the simulated energy barrier of a catalytic reaction or the materials' *chemophysics*, otherwise obtainable by calculations [192, 210]. In addition to electrocatalysis, metal oxide (Me-O) materials screened with the aid of ML also have good application prospects in other catalytic fields. In material screening, the structural framework of Me-O coupled with the complex nature and diverseness of doped metals provides a suitable stage for ML to effective play its role. In a study, Xie *et al.* [211] introduced Lewis acid strength, which is strongly related to the kinetic reaction rate of oxygen reduction reaction (ORR) of perovskite oxides (ABO_3) at high temperature, as a descriptor and verified the validity of eight different regression models. According to Xie *et al.* [211], among all regression methods, the mean square error (MSE) values corresponds to the ML model-based training set (being 0.009 Ωcm^2) and the test set (being 0.013 Ωcm^2), attaining the best fitting effect among all other regression models. The authors discovered that Lewis acid strength at sites ultimately contributed to the effective performance recorded with recorded results indicating strong correlation between ionic Lewis acid strength and the intrinsic ORR activity at elevated temperatures of the cathode, verified in the electrochemical characterization.

Recently, Xu *et al.* [208] had found through computational and experimental ML assisted studies that the Oxygen Evolution Reaction (OER) activity of spinel oxides is basically ascertained by covalent competition between tetrahedral and octahedral

points of positioning – of which the authors, however concluded was the determinant of the cationic site exposure, and hence its activity. Driven by this finding, Xu *et al.* [208] computed data sets for over 300 spinel oxides via theoretical computations which were subsequently applied (with ML models) in screening for covalent competition of spinel oxides giving an average absolute error of 0.05 eV. Wang *et al.* [209] developed a targeted drive method based on ML techniques and DFT calculations to find stable lead-free organic-inorganic hybrid perovskites (HOIPs). Also, Smit *et al.* [212] trained a set of ML models automated to designate oxidation states to metal ions in a metal-organic framework (MOF). The research centered on prediction of oxidation states of metal centers in MOF, particularly in mixed valence MOF and flexible MOF. Furthermore, the study provided an application for the distribution of oxidation states to metal centers in MOF on a materials' chemical space. The research of authors [209, 212] presents a more concise and accurate technique for self-prediction of the materials properties. ML is used to assign values and record data for properties such as valence states of material structures.

For a variety of metal combinations (such as Bimetallics, alloys, etc.) in catalyst design it is however very difficult to elucidate comprehensibly the synergies that exist between metals, because of the complexity of their chemical structure [213-215]. While quantum chemical computations shows there exist a possibility of assaying the chemical relationship between alloys structural composition and properties, high computational and time costs limit the development of metal combination catalytic materials. Fortunately, ML in its capacity to handle complex problems has proven to be very effective in solving the problems of metal combination systems. Now, a basic ML model can be designed to predict the properties and behavior of materials, which in this case bring about speedy breakthrough in discovery of high-performance alloy materials.

To backup this claim, Sargent *et al.* [213] were able to develop a highly efficient Farady Copper-Aluminum (Cu-Al) electrocatalyst by employing theoretical computation and ML. These researchers were able to construct 244 different Cu-containing intermetallic crystals, identifying 12,229 surfaces and 228,969 adsorption sites and the CO adsorption energies at various sites was computed by DFT to produce a data set for ML. The ML prediction outcomes showed that Cu-Al as a (most) promising material with high activity and selectivity for CO_2 reduction. Experimentations and computations revealed that Cu-Al alloy provides multiple sites to achieve the best binding with CO, to effectively reduce carbon dioxide. In addition, the Faraday efficiency exceeded 80% with a high current density ($400mA/cm^2$) – the highest Faraday efficiency so far obtained. The data utilized in this research work was drawn from a collection of databases and DFT computations, producing a comprehensive yet efficient data collection technique. Data collection has to be rigorous for efficient screening and synthesis of new high-performance materials.

Also, for electrochemical CO_2RR, Wang *et al.* [214] reported the discovery and optimization of additives using ML in the preparation of Cu catalysts for electrochemical CO_2 reduction. Synthesis of Cu catalyst was by electrochemical deposition using copper salt as staple, and various metal salts and organic molecules acting as additives. Following three iterative experimental tests, ML analysis, prediction and redesign, respectively it was discovered that salt of tin (Sn) was an essential additive for CO and HCOOH extraction, and fatty alcohol an essential additive favorable for C^{2+} formation. Farther catalysts characterization synthesized using various additives showed that fatty alcohols may encourage the formation of Cu_2O cube in the course of electro-deposition. In addition, alloy catalysts selected by ML also show exciting potential performance in ORR as earlier discussed. Han *et al.* [215] combined first-principles DFT with ML technology and based on neural network potential algorithm, systematically calculated, simulated and screened the composition, element distribution and ORR properties of terpolymer PtFeCu nanoparticles.

For application of ML-assisted screening in alloy materials, Vivek B. Shenoy *et al.* [216] in their research work trained a graph neural network to predict the adsorption energy of catalyst/adsorption system based on the influence of tensile effect of alloy catalyst on adsorption energy of reactive species. The Cu-based binary alloy catalyst in the Open catalyst Project [216] was utilized as the source for data collection in computing the adsorption energy at various tensile conditions. The established ML model successfully predicted the adsorption energy of 85% of the unknown test data. Taking ammonia synthesis as an example, potential catalysts for Cu-S alloy under tensile strain were selected. In this work, the impact of the alloy structural tensile properties on the adsorption of various species was discussed in-depth by employing ML. While the research study however was exclusive of experimental components, it is of extreme importance that the relationship between the theoretical adsorption of various species and metal-combination (alloy materials) structure be ideally established. In such cases as this, ML is ideal for optimizing the engineering characterization, accelerating selection and for enumeration

of the core descriptors for alloy materials. Linic *et al.* [217] provided a crosscut to building accurate and explainable electron-structure descriptors of characterized catalytic materials by utilizing unsupervised ML principal component analysis. The authors rebuilt the acquired descriptors of the electronic structure of the material so as to account for the electronic structure effect captured by every one of the principal component descriptors, and also the local changes in the geometric structure of a region these effects map to. Furthermore, this technique is established by searching for principal component descriptors responsible chemisorption on transition metal/alloys surfaces and is analogized with the observed outcome of chemisorption descriptors on physical bases in order to affirm the technique's accuracy.

In other practice, for atomically dispersed catalyst (i.e. catalyst which constitute of single atoms sparsely distributed over the surface of the support [218, 219]), the heterogeneity of the support is consequent to metal atoms exposure to a variety of chemical environs on the support surface, hence bringing about possible dissimilarities in the catalytic activity of different metal atoms at different regions of the support surface. However, atomically dispersed catalysts can be supported on diverse surfaces, such as surfaces of: metals, metal oxides, carbon materials, polymer materials, and so on [220]. The structural characteristics of atomic-level dispersion catalysts is fundamental to ascertaining what the interaction between the metal and the coordination atoms (C, N, O, S, P) on the support will be. This phenomenal interaction is highly essential for single metal atoms' stability and activity. In recent pasts, researchers have suggested using diatomic catalysts as better substitute to atomically dispersed catalyst at the nascence of their discovery. This being as a result of their characteristic superior metal loading capacity, highly complex and their possession of more flexible active sites [187, 221]: desirable characteristic properties which are responsible for their superior catalytic performance and offers better chances for electrocatalysis. It is hard to discover atomic-level dispersion catalysts having higher stability and activity by engaging the conventional techniques of theoretical computations and experimental synthesis on account of the wide variety of coordination atoms and bimetal combinations of support. On account of this, inculcating ML is a most worthy approach to tackling and resolving such challenges [222-224].

In times of recent, atomic-scale dispersion catalysts (otherwise known as atomically dispersed catalyst) have drawn much attraction to researchers in electrocatalysis studies. Being conversant with significant electrode reactions (such as ORR, OER), there is a pressing need for design of electrocatalysts with superior performance characteristics. Li *et al*. [225] devised a single-atom catalyst design strategy which was based on first-principles computing and topological ML in developing an efficient OER catalyst. Employing the DFT method, the OER characteristics of 15 metal atoms were computed at single-vacancy and double-vacancy defects respectively and afterwards the theoretical overpotential was obtained by calculating the adsorption energy of oxygen species. Furthermore, the topological structures around metal atoms were analyzed based on the topological learning algorithm, extracting the node information and the link information between metal atoms and the substrate, and a miniature collection of DFT calculation data were compounded in training the prediction model to predict the OER catalytic performance of different other transition metals on carbon substrates with variety of structures. In order to screen, monoatomic catalysts in a more effective way, the team suggested the use of a volcanic-type curve description technique, and that the catalyst screening rate be increased (by a factor of 13 x 10^4). Huang *et al.* [226] in designing a high-performance dual-function OER/ORR catalyst, based on C_2N structure in a combination with DFT calculation and ML, discovered that the adsorption energy of a single oxygen atom has a volcanic relationship with the catalytic activity, and hence combined with the normalized Fermi abundance, formed a fresh electronic structure descriptor. In addition, Ding *et al.* [227] based on a single transition metal AlP system, discovered that in substituting two P atoms with two N atoms improved the catalytic activity. First using DFT computations they were able to affirm the electrocatalytic performance of bifocal oxygen. Then they going farther to use the ML method based on gradient lifting regression model to explore for other possible sources of catalytic activity. Observations indicated that the d electron number, the radius and charge transfer of the atom of transition metal are also important descriptors that corresponds and explains the adsorption activity. As regards to the development of catalyst for lithium-sulfur batteries, Li *et al.* [228] consistently studying the adsorption mode of polysulfide based on ML technique computed by high-flux DFT computations, screened thousands of transition metal monoatomic catalysts on nitrogen-doped carbon materials support. Based on a classifier trained by the convolutional neural network of crystal graphs, the team of researchers successfully characterized the sorbent with S-S bond breaking distinguishing them from other types of sorbent. The ML-trained regression model was also efficient in adsorption energy prediction, having an average absolute error of about 0.14 eV, and could also predict a series of actively performing catalysts.

DFT computations can be effectively applied while studying the material structure-activity relationship of atomic-scale dispersion catalyst, due to the structure of their active center. Thence, ML has been very much applicable by researchers in dealing with complex metal combinations in the material screening of atomically dispersed catalysts. Considering, ML models are capable of aiding researchers in studying the dynamic processes of materials to in order to acquire a more vivid reaction mechanism, its application in studying the active structure of atomically dispersed materials is accurate since it is comparatively simple and comprehensible. These findings are principal to advancements of and gives a novel directive for research and development of ML-assisted atomically dispersed techniques.

The conventional techniques of designing metal composite materials ordinarily engages numerous parametric quantities, and requires iterative trials to obtain the ideal synthesis conditions. Moreover, owing to the complex nature of the electronic structures of metal complexes, the often strong correlation effects, DFT will be inaccurate in predicting the characteristic features of strongly correlated systems, therefore rendering DFT-based screening unreliable. Nonetheless, Liu *et al.* [229] was able to successfully (and for the first time) apply classification algorithm in ML-optimized model to channel the synthesis of two-dimensional materials by chemical vapor deposition (CVD) and regression algorithm in guiding the hydrothermal synthesis of sulfur-nitrogen-doped blue fluorescent quantum dots with high fluorescence yield, based on which a carbon quantum dot solution (yielding about 55.5%) was synthesized successfully. Kulik *et al.* [230] devised an inexpensive technique capable of enhancing the dependability of DFT high-throughput computational screening. A ML model was trained on the basis of diagnostic parameters of the strong association effects of 5,000 transition-state metal complexes, to speedily recognize the strong association effects of electrons in large-scale (>100,000 molecules) high-throughput screening. This technique is favorable for application in the design and development of functional molecules such as catalysts. Although, most current researches have focused on application of ML in metal complex catalytic materials design. Nonetheless, ML has been instrumental in the design of some inorganic catalytic materials as well. However, there exist a vast number and class of energy catalytic materials (such as non-metal materials). Metal-based materials have definite crystal structures and apparent active centers in energy catalytic reactions making it much easier for feature extraction. This is not so with non-metal materials. Moreover, with the vast variety of metal doping types and consequently the complexity of material design becoming more complicated there is need for a more credible approach to attacking challenges. ML offers great prospects in dealing with and solving these challenges.

Presently, the end target of chemical researches are intricate and high-dimensional. However, the conventional research techniques are principally rigorous and by trial-and-error – methods, which although objective are however subject to prejudice and shallow preconceptions since it prevents the objective study of peculiar and intricate phenomenon that are key to understanding catalyst mechanisms, which is essential to unlocking hidden facts about them. Exploring, discovering and ultimately synthesizing ideal catalysts comes with challenges such as confronting a vast unmapped chemical space. In material synthesis, the challenge of optimizing synthesis conditions and vividly establishing the relationship between processes, characteristic properties of materials, and most promising synthesis path can be overcome by inculcation of computer-assisted retro-synthesis analytics. The progress made so far has been impactful especially in terms of high-throughput synthesis; which has been time economical as a result of allowing multiple experimental processes been run concomitantly, yet the analysis of experimental outcome is difficult to carryout. In attempt to discover more efficient novel synthesis paths Cooper *et al.* [231] successfully developed a self-operative AI robot chemist having android attributes capable of handling various lab equipment in a standard lab like a human. From the initial test, 668 experiments were independently accomplished within the span of 8 days by building an adaptive automated experimental environment and developing a new chemical catalyst. Furthermore, the optimization of synthetic variables were studied by way of the developed algorithm. The developed intelligent synthesis system was able to survey the variables of 10 dimensions based on the outcomes of the preceding experimentation and also could ascertain the successive optimum experimentation to carry out from a set of more than 100 million prospective chemical experimentations in the lab. Cronin *et al.* [232] had suggested an autonomous chemical synthesis robot to search, detect and recognize, as well as optimize nanostructures goaded by real-time spectral feedback, theory, and ML algorithms. In the chemical space, five categories of NPs were able to be identified in about 1,000 experimentations in a multiple steps open exploration involving the synthesis of nanoparticulate gold (Au-NPs) by visible characterization. Moreover, in an attempt to optimize the nanostructures of Au-NPs in other to obtain choice optical properties a yield of about 95% was achieved, through a combination of experimental and spectral simulations. Although, a good number of researchers on account of the use of intelligent synthesis platforms for novel materials have given irrefutable verdicts based on their respective systems of research. Nonetheless, the approach of intelligent systems is

material bias, as it can only be used for peculiar category of materials. Therefore, there is need to establish a research database incorporating material information for researchers. Based on this, Luo and Jiang *et al.* [233] built a data-intelligent-driven whole-process robotic chemist by building and integrating mobile robots, chemical workstations, intelligent operating systems, and scientific databases. The designed intelligent platform could automatically search and read literature, analyze literature data, propose scientific hypotheses and formulate experimental schemes. Using a Bayesian optimization program, the team was able to discover the optimal catalyst from a class of 550,000 attainable metal ratios. In comparison with the conventional 'trial-and-error' laboratory technique of determined catalyst materials, the intelligently synthesized materials exhibited superior characteristics. In addition to the use of intelligent synthesis platforms, the authors incorporated literature reading and analysis modules, thereby modifying the entire system into a comparatively more intelligent workflow system. Currently, AI technological approach (especially in terms of robotics, ML, etc.) has greatly drawn a lot of attention in the research space with its rapid development schemes and therefore is very crucial for advancements in the chemical field. However, developing intelligent platforms befitting of various species of chemical materials around intelligent robots is challenging. The possibility of a convergence of ideas from robotics technological experts and researchers in designing intelligent synthesis platform will be a convenient move counteractive to some of these challenges.

Intelligent Characterization and Reaction Kinetics of Materials

Prior to the selection of target catalytic material via intelligent design and synthesis, there is need for characterization of target materials' reaction process and fine structure in the atomic spectrum [233-236]. AI has been an incentive which on account of its strategic conjugation with basic scientific experimentation, has birthed advanced schemes and directives towards improving research prospects [234-236]. The dynamic simulation of catalytic materials is essential in their synthesis and application. Knowledge of dynamic mechanism of catalysts materials can be reformative to researchers' approach in discovering and designing not just novel but also superior catalysts materials. Ordinarily, simulating the dynamics of materials via first principles is often demanding; it involves high computational power, time uneconomical because oftentimes materials dynamics very complex and hard to decipher. Thus, the use of ML becomes essential in order to speed up the dynamic simulation process of materials. In an attempt to synthesize a standard transferable organic molecule, A. E. Roitberg *et al.* [237] trained on quantum computing through deep neural networks (DNN). After suggesting the ANI method, the research team built the ANI-1 model based on the database constructed from the network. The model trained was capable of predicting the total energy of an organic molecule comprising of four different atom types (i.e. hydrogen, carbon, nitrogen and oxygen). On account of a series of case studies, the research team showed that ANI-1 had superior chemical accuracy than DFT computations on larger molecular systems. While the study did not investigate on the chemical reaction dynamics of the system, the predictions for the molecular energy model proposed by the study renders a rather more efficient algorithm for molecular dynamics simulation. Leveraging on such improvements will bring about rapid and efficient research predictions and energy computations for kinetic simulation processes. The work of Gong *et al.* [233], in attempt to solve the challenge of the active site of oxide-derived copper catalyst in generating multiple carbon (C^{2+}) products in the process of CO_2 electro-reduction (CO_2RR) was ill-defined, until by using neural networks (NN) in conjugate with DFT computation and simulation the author was able to describe the real catalyst surface model. The molecular dynamics of the material under the reaction potential were simulated by ML model, and over 150 surface sites were discovered, proving that the reaction sites were the planar-square and convex-square sites in its structure. The authors further used ML to modify the complex process of CO_2RR for better understanding giving a more vivid account of the reactions activity. The study went further to prove the importance of ML in the simulation of reaction dynamics because compared with the use of the conventional DFT computations, explanations on the complexity of microscopic reaction processes were better established. While in contemporary times research in the field of materials sciences and precisely on the scope of energy catalysis has focused on material performance, researchers have often neglected the 'unseen' material characteristics (such as the microscopic reaction mechanisms, in-situ fine structure) that are principal to the evident performance; which are essential for effective simulation of the dynamic process of materials and for rapid prediction of the in-situ characterization data [208]. On the basis of AI inculcation, researchers are no more restricted to the use of only contemporary standards and techniques of material characterization, and of high-computing-power costs equipment, but are capable of acquiring more precise predictions with the availability of credible database for material characterization beneficial for accurate interpretation of dynamic simulations.

2. SUMMARY

The traditional ways of catalysts discovery and synthesis come with a lot of challenges (such as: the cost inefficiencies and high energy demands, the uncertainty of the large unexplored chemical space, challenges of repeated experimental trials, the uncertainty of a successful discovery, environmental implication of catalyst utilization and so on). AI has offered great prospects to tackling these challenges. The inculcation of ML has advanced the catalysts design field, with the development of models and smart active learning systems that engages computational and theoretical techniques in predicting the characteristic properties and behavior of chemical systems, their anticipated activity and selectivity (i.e. predictive models) and can possibly go as far as suggesting new molecular structures with peculiar targeted characteristics (i.e. generative models). These represents the key goals of AI in catalysts design. Predictive models are primarily database builders, while generative models are secondary database generators. However, in the application of predictive modelling another challenge ensues from the high dependence on black-box models by researchers from earliest times and even in contemporary times. This has been a limitation to understanding catalysts design profoundly. Black-box models are somewhat unreliable and problematic despite their high level of accuracy, their inexplicability and complexities – which may infer that new undesirable features and descriptors begin to come into play. A conceptual combination of systems knowledge (white-box models) with statistical information from dataset (black-box models) have birthed the grey-box models. Interpretability is of key concern in catalysts design. Therefore, grey-box models and glass-box models which are recommended as better and best easier to understand better predictive models respectively. Glass-box models are easily explainable compared to black-box models where explanations are generally approximate, but comparative in terms of accuracy.

With the discussions so far, it is very clear that an accruing number of researchers place more focus on the use of ML for material screening and verification of DFT calculations. However, advancements in material sciences is inseparable from researchers' comprehensive knowledge of the relationship between the structural properties of materials and their activity. Also, featurization – a process in material screening that involves the collection of varied data forms and their conversion into numerical data explainable by basic ML algorithms, is very vital. In addition, predictions accuracy is very much affected by extraction of descriptor and essential characteristics. All these are constituted amongst researchers on the material structure and the target activity based on the understanding. Beyond model development, research has been in recent past very keen about the development of fully automated systems, robotic arms and chemists to ease the task of carrying out laboratory works and also bypass unnecessary human involvement with chemical systems. AI has great potential to unearth characteristics features which catalysts practitioners maybe alien to, make theoretical proposals that are key to maneuvering the challenges encountered in catalysts design. All of these developments have been advantageous to advancements in the discovery and implementation of cognitive intuitions principal to the exposure of some less intricate and obscure concepts that have advanced the field of study.

3. CONCLUSION AND RECOMMENDATIONS

To determine the best catalyst design involving artificial intelligence (machine learning), the inherent physicochemical properties of catalysts (such as morphology/crystal structure, valency, etc., even at the atomic and subatomic levels) that make them ideal choices for a particular process has been revealed, and altered to choice catalysts. The involvement of nanocatalysis helps tackle to a large extent a good number of the challenges encountered by AI in creating a database to resolve issues of persistent iterative trial-and-error. As a result of their pronounced features such as increased surface area, activity and selectivity, etc., the sensitivity of AI tools in their applications is increased as atomic characteristics are made more macroscopic to detection.

In addition, it will be advantageous if all experimental results, including the raw data and metadata containing the experimental parameters, are transferred automatically to cloud-based servers, which can afterwards be analyzed by automated data analysis and visualization tools. The most critical role of high-throughput experimentation is to find the structure-activity relationship of the catalyst. Thus, the algorithm developed for automated data analysis should estimate or predict the optimum synthesis condition as feedback to the high-throughput experimentation. This is an ideal concept of data-guided combinatory synthesis and data-driven catalyst discovery.

Finally, the development of explainable artificial intelligence in the field of chemical studies will lead to advancements in human comprehension, cut down on cost as evident in the development of reinforcement and active learning, inspire hybrid human- and computer-guided catalyst design, and in the end improve machine learning models for catalysts design.

REFERENCES

[1] Mahmoodi, K., Babak Alinejad, B., "Enhancing of Hydrogen Generation Rate in Reaction of Aluminium in Water". *International Journal of Hydrogen Energy*, Volume 35 Issue 11, Pages 5227–5232, 2010.

[2] Jacques A. Esterhuizen, Bryan R. Goldsmith, Suljo Linic (2022). Interpretable machine learning for knowledge generation in heterogeneous catalysis. *Nature Catalysis.* Vol. 5:175–184. https://doi.org/10.1038/s41929-022-00744-z

[3] Dionisios G. Vlachos, "A Review o Multiscale Analysis: Examples from Systems Biology, Materials Engineering, and Other Fluid-Surface Interacting Systems". *Advances in Chemical Engineering*, Volume 30, Pages 1-61, 2005.

[4] Goldsmith, B. R., Esterhuizen, J., Liu, J.-X., Bartel, C. J. & Sutton, C. Machine learning for heterogeneous catalyst design and discovery. *AlChE J*. 64, 2311–2323 (2018).

[5] Schlexer Lamoureux, P. et al. Machine learning for computational heterogeneous catalysis. ChemCatChem 11, 3581–3601 (2019).

[6] Kitchin, J. R. Machine learning in catalysis. *Nat. Catal*. 1, 230–232 (2018).

[7] Toyao, T. et al. Machine learning for catalysis informatics: recent applications and prospects. ACS Catal. 10, 2260–2297 (2020).

[8] IBM. "Machine Learning." Accessed October 16, 2023 https://www.ibm.com/cloud/learn/machinelearning

[9] Baiyu Lu (2023). The Application of Machine Learning in Chemical Engineering: A Literature Review. R. B. B. M. Hussain et al. (Eds.): ICHSSR 2023, ASSEHR 765, pp. 57–66, 2023. https://doi.org/10.2991/978-2-38476-092-3_9

[10] Schweidtmann, Schweidtmann, Artur M., et al."Machine Learning in Chemical Engineering: A Perspective." Chemie Ingenieur Technik, vol. 93, no. 12, 2021, pp. 2029–2039.

[11] Dobbelaere, Maarten R., et al. "Machine Learning in Chemical Engineering: Strengths, Weaknesses, Opportunities, and Threats." Engineering, vol. 7, no. 9, 2021, pp. 1201–1211.

[12] McBride, Kevin, and Kai Sundmacher."Overview of Surrogate Modeling in Chemical Process Engineering." Chemie Ingenieur Technik, vol. 91, no. 3, 2019, pp. 228–239.

[13] Venkatasubramanian, Venkat. "The Promise of Artificial Intelligence in Chemical Engineering: Is It Here, Finally?" AIChE Journal, vol. 65, no. 2, 2018, pp. 466–478.

[14] Lee, Jay H., et al. "Machine Learning: Overview of the Recent Progresses and Implications for the Process Systems Engineering Field." Computers & Chemical Engineering, vol. 114, 2018, pp. 111–121.

[15] Obaidi, Martin, and Jil Klünder. "Development and Application of Sentiment Analysis Tools in Software Engineering: A Systematic Literature Review." Evaluation and Assessment in Software Engineering, 2021.

[16] Glass, Robert L., et al. "An Analysis of Research in Computing Disciplines." Communications of the ACM, vol. 47, no. 6, 2004, pp. 89–94.

[17] Kitchenham, Barbara, et al. "Systematic Literature Reviews in Software Engineering – A Systematic Literature Review." Information and Software Technology, vol. 51, no. 1, 2009, pp. 7–15.

[18] Babar, M. A., & He, Z. (2009). Systematic literature reviews in software engineering: Preliminary results from interviews with researchers. Proceedings of the Third International Symposium on Empirical Software Engineering and Measurement, ESEM 2009, October 15–16, 2009, Lake Buena Vista, Florida, USA. IEEE.

[19] Technology, I. S., Editor-In-Chief, C. W., & Contributor, R. P. G. (2013). Systematic literature reviews in software engineering. Information and Software Technology, 55(6), 919–920.

[20] Cova TFGG and Pais AACC (2019) Deep Learning for Deep Chemistry: Optimizing the Prediction of Chemical Patterns. Front. Chem. 7:809. DOI:10.3389/fchem.2019.00809

[21] Chakravarti, S. K. (2018). Distributed representation of chemical fragments. ACS Omega 3, 2825–2836. doi: 10.1021/acsomega.7b02045

[22] Fuchs, J.-A., Grisoni, F., Kossenjans, M., Hiss, J. A., and Schneider, G. (2018). Lipophilicity prediction of peptides and peptide derivatives by consensus machine learning. Medchemcomm 9, 1538–1546. doi: 10.1039/C8MD00370J

[23] Janet, J. P., Chan, L., and Kulik, H. J. (2018). Accelerating chemical discovery with machine learning: simulated evolution of spin crossover complexes with an artificial neural network. J. Phys. Chem. Lett. 9, 1064–1071. doi: 10.1021/acs.jpclett.8b00170

[24] Elton, D. C., Boukouvalas, Z., Fuge, M.D., and Chung, P.W. (2019). Deeplearning for molecular design—a review of the state of the art. Mol. Syst. Design Eng. 4, 828–849. doi:10.1039/C9ME00039A

[25] Mezei, P. D., and Von Lilienfeld, O. A. (2019). Non-covalent quantum machine learning corrections to density functionals. arXiv[preprint].arXiv:1903.09010

[26] Sanchez-Lengeling, B., Roch, L. M., Perea, J. D., Langner, S., Brabec, C. J., and Aspuru-Guzik, A. (2019). A Bayesian approach to predict solubility parameters. Adv. Theory Simul. 2:1800069. https://doi:10.1002/adts. 201800069

[27] Xu, Y., Lin, K., Wang, S., Wang, L., Cai, C., Song, C., et al. (2019). Deep learning for molecular generation. Future Med. Chem. 11, 567–597. https://doi:10.4155/fmc-2018-0358

[28] Zhang, P., Shen, L., and Yang, W. (2019). Solvation free energy calculations with quantum mechanics/molecular mechanics and machine learning models. J. Phys. Chem. B 123, 901–908. https://doi:10.1021/acs.jpcb.8b11905

[29] Varnek, A., and Baskin, I. (2012). Machine learning methods for property prediction in chemoinformatics: quo vadis? J. Chem. Inf. Model. 52, 1413–1437. https://doi:10.1021/ci200409x

[30] Ramakrishnan, R., Dral, P. O., Rupp, M., and Von Lilienfeld, O. A. (2014). Quantum chemistry structures and properties of 134 kilo molecules. Sci. Data 1:140022. https://doi:10.1038/sdata.2014.22

[31] Goh, G. B., Hodas, N. O., and Vishnu, A. (2017). Deep learning for computational chemistry. J. Comput. Chem. 38, 1291–1307. https://doi:10.1002/jcc.24764

[32] Simões,R.S.,Maltarollo,V.G.,Oliveira,P.R.,andHonorio,K.M.(2018).Transfer and multi-task learning in QSAR modeling: advances and challenges. Front. Pharmacol. 9:74. https://doi:10.3389/fphar.2018.00074

[33] Chandrasekaran, A., Kamal, D., Batra, R., Kim, C., Chen, L., and Ramprasad, R. (2019). Solving the electronic structure problem with machine learning. NPJ Comput. Mater. 5:22. https://doi:10.1038/s41524-019-0162-7

[34] Snyder, J. C., Rupp, M., Hansen, K., Müller, K.-R., and Burke, K. (2012). Finding density functionals with machine learning. Phys. Rev. Lett. 108:253002. https://doi:10.1103/PhysRevLett.108.253002

[35] Ramakrishnan, R., Dral, P. O., Rupp, M., and Von Lilienfeld, O. A. (2015). Big data meets quantum chemistry approximations: the 1-machine learning approach. J. Chem. Theory Comput. 11, 2087–2096. https://doi:10.1021/ acs.jctc.5b00099

[36] Faber, F. A., Hutchison, L., Huang, B., Gilmer, J., Schoenholz, S. S., Dahl, G. E., et al. (2017). Prediction errors of molecular machine learning models lower than hybrid DFT error. J. Chem. Theory Comput. 13, 5255–5264. https://doi:10.1021/acs.jctc.7b00577

[37] Hegde, G., and Bowen, R. C. (2017). Machine-learned approximations to density functional theory hamiltonians. Sci. Rep. 7:42669. https://doi:10.1038/srep42669

[38] Smith, J. S., Isayev, O., and Roitberg, A. E. (2017). ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. Chem. Sci. 8, 3192–3203. https://doi:10.1039/C6SC05720A

[39] Pronobis, W., Schütt, K. T., Tkatchenko, A., and Müller, K.-R. (2018). Capturing intensive and extensive DFT/TDDFT molecular properties with machine learning. Eur. Phys. J. B 91:178. https://doi:10.1140/epjb/e2018-90148-y

[40] Schleder, G. R., Padilha, A. C. M., Acosta, C. M., Costa, M., and Fazzio, A. (2019). From DFT to machine learning: recent approaches to materials science–a review. J. Phys. Mater. 2:032001. https://doi:10.1088/2515-7639/ab084b

[41] White, D., and Wilson, R. C. (2010). Generative models for chemical structures. J. Chem. Inf. Model. 50, 1257–1274. https://doi:10.1021/ci9004089

[42] Benjamin, S.-L., Carlos, O., Gabriel, L., G., and Alan, A.-G. (2017). Optimizing Distributions Over Molecular Space. An Objective-Reinforced Generative Adversarial Network for Inverse-design Chemistry (ORGANIC). ChemRxiv [Preprint]. https://doi:10.26434/chemrxiv.5309668.v3

[43] Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A., and Zhavoronkov, A. (2017). druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. Mol. Pharm. 14, 3098–3104. https://doi:10.1021/acs.molpharmaceut.7b00346

[44] Harel, S., and Radinsky, K. (2018). Prototype-based compound discovery using deep generative models. Mol. Pharm. 15, 4406–4416. https://doi:10.1021/acs.molpharmaceut.8b00474

[45] Jørgensen, P. B., Schmidt, M. N., and Winther, O. (2018b).Deep generative models for molecular science. Mol. Inform. 37:1700133. https://doi:10.1002/minf.201700133

[46] Kang, S., and Cho, K. (2018). Conditional molecular design with deep generative models. J. Chem. Inf. Model. 59, 43–52. https://doi:10.1021/acs.jcim.8b00263

[47] Li, Y., Zhang, L., and Liu, Z. (2018b). Multi-objective de novo drug design with conditional graph generative model. J. Cheminform. 10:33. https://doi:10.1186/s13321-018-0287-6

[48] Freeze, J. G., Kelly, H. R., and Batista, V. S. (2019). Search for catalysts by inverse design: artificial intelligence, mountain climbers, and alchemists. Chem. Rev. 119, 6595–6612. https://doi:10.1021/acs.chemrev.8b00759

[49] Sanchez-Lengeling, B., and Aspuru-Guzik, A. (2018). Inverse molecular design using machine learning: generative models for matter engineering. Science 361, 360–365. https://doi:10.1126/science.aat2663

[50] Schneider, G. (2018). Generative models for artificially-intelligent molecular design. Mol. Inform. 37:1880131. https://doi:10.1002/minf.201880131

[51] Arús-Pous, J., Blaschke, T., Ulander, S., Reymond, J.-L., Chen, H., and Engkvist, O. (2019). Exploring the GDB-13 chemical space using deep generative models. J. Cheminform. 11:20. https://doi:10.1186/s13321-019-0341-z

[52] Jensen, J. H. (2019). A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space. Chem. Sci. 10, 3567–3572. https://doi:10.1039/C8SC05372C

[53] Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D., and Doyle, A. G. (2018). Predicting reaction performance in C–N cross-coupling using machine learning. Science 360, 186–190. https://doi:10.1126/science.aar5169

[54] Coley, C. W., Green, W. H., and Jensen, K. F. (2018a). Machine learning in computer-aided synthesis planning. Acc. Chem. Res. 51, 1281–1289. https://doi:10.1021/acs.accounts.8b00087

[55] Granda, J. M., Donina, L., Dragone, V., Long, D.-L., and Cronin, L. (2018). Controlling an organic synthesis robot with machine learning to search for new reactivity. Nature 559, 377–381. https://doi:10.1038/s41586-018-0307-8

[56] Segler, M. H. S., Preuss, M., and Waller, M. P. (2018). Planning chemical syntheses with deep neural networks and symbolic AI. Nature 555, 604–610. https://doi:10.1038/nature25978

[57] Li, J., and Eastgate, M. D. (2019). Making better decisions during synthetic route design: leveraging prediction to achieve greenness by design. React. Chem. Eng. 4, 1595–1607. https://doi:10.1039/C9RE00019D

[58] Li, H., Zhang, Z., and Liu, Z. (2017). Application of artificial neural networks for catalysis: a review. Catalysts 7:306. https://doi:10.3390/catal7100306

[59] Gao, H., Struble, T. J., Coley, C. W., Wang, Y., Green, W. H., and Jensen, K. F. (2018). Using machine learning to predict suitable conditions for organic reactions. ACS Central Sci. 4, 1465–1476. https://doi:10.1021/acscentsci.8b00357

[60] Huang, S.-D., Shang, C., Kang, P.-L., and Liu, Z.-P. (2018). Atomic structure of boron resolved using machine learning and global sampling. Chem. Sci. 9, 8644–8655. https://doi:10.1039/C8SC03427C

[61] Durand, D. J., and Fey, N. (2019). Computational ligand descriptors for catalyst design. Chem. Rev. 119, 6561–6594. https://doi:10.1021/acs.chemrev.8b00588

[62] Ramakrishnan, R., and Von Lilienfeld, O. A. (2017). "Machine learning, quantum chemistry, and chemical space," in Reviews in Computational Chemistry, Vol. 30, eds A. L. Parrill and K. B. Lipkowitz (Wiley), 225–256. https://doi:10.1002/9781119356059.ch5

[63] Xia, R., and Kais, S. (2018). Quantum machine learning for electronic structure calculations. Nat. Commun. 9:4195. https://doi:10.1038/s41467-018-06598-z

[64] Ahn, S., Hong, M., Sundararajan, M., Ess, D. H., and Baik, M.-H. (2019). Design and optimization of catalysts based on mechanistic insights derived from quantum chemical reaction modeling. Chem. Rev. 119, 6509–6560. https://doi:10.1021/acs.chemrev.9b00073

[65] Christensen, A. S., Faber, F. A., and von Lilienfeld, O.A. (2019). Operators in quantum machine learning: Response properties in chemical space. J. Chem. Phys. 150:064105. https://doi:10.1063/1.5053562

[66] Zaspel, P., Huang, B., Harbrecht, H., and Von Lilienfeld, O. A. (2019). Boosting quantum machine learning models with a multilevel combination technique: pople diagrams revisited. J. Chem. Theory Comput. 15, 1546–1559. https://doi:10.1021/acs.jctc.8b00832

[67] Mullin, R. The Lab of the Future Is Now. Chem. Eng. News 2021, 28.

[68] Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W. Deep Learning for Molecular Design: a Review of the State of the Art. Molecular Systems Design & Engineering 2019, 4, 828−849.

[69] Bender, A.; Cortés-Ciriano, I. Artificial Intelligence in Drug Discovery: What Is Realistic, What Are Illusions? Part 1: Ways to Make an Impact, and Why We Are Not There Yet. Drug Discovery Today 2021, 26, 511−524.

[70] Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. Qsar without Borders. Chem. Soc. Rev. 2020, 49, 3525− 3564.

[71] Strieth-Kalthoff, F.; Sandfort, F.; Segler, M. H. S.; Glorius, F. Machine Learning the Ropes: Principles, Applications and Directions in Synthetic Chemistry. Chem. Soc. Rev. 2020, 49, 6154−6168.

[72] Struble, T. J.; Alvarez, J. C.; Brown, S. P.; Chytil, M.; Cisar, J.; DesJarlais, R. L.; Engkvist, O.; Frank, S. A.; Greve, D. R.; Griffin, D. J.; Hou, X.; Johannes, J. W.; Kreatsoulas, C.; Lahue, B.; Mathea, M.; Mogk, G.; Nicolaou, C. A.; Palmer, A. D.; Price, D. J.; Robinson, R. I.; Salentin, S.; Xing, L.; Jaakkola, T.; Green, W. H.; Barzilay, R.; Coley, C. W.; Jensen, K. F. Current and Future Roles of Artificial Intelligence in Medicinal Chemistry Synthesis. J. Med. Chem. 2020, 63, 8667− 8682.

[73] Artrith, N. & Kolpak, A. M. Understanding the composition and activity of electrocatalytic nanoalloys in aqueous solvents: a combination of DFT and accurate neural network potentials. Nano Lett. 14, 2670–2676 (2014).

[74] Boes, J. R. & Kitchin, J. R. Modeling segregation on AuPd(111) surfaces with density functional theory and Monte Carlo simulations. J. Phys. Chem. C 121, 3479–3487 (2017).

[75] Ulissi, Z. W., Singh, A. R., Tsai, C. & Nørskov, J. K. Automated discovery and construction of surface phase diagrams using machine learning. J. Phys. Chem. Lett. 7, 3931–3935 (2016).

[76] Peterson, A. A. Acceleration of saddle-point searches with machine learning. J. Chem. Phys. 145, 074106 (2016).

[77] Ulissi, Z. W., Medford, A. J., Bligaard, T. & Nørskov, J. K. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. Nat. Commun. 8, 14621 (2017).

[78] Kolsbjerg, E. L., Peterson, A. A. & Hammer, B. Neural-network-enhanced evolutionary algorithm applied to supported metal nanoparticles. Phys. Rev. B 97, 195424 (2018).

[79] Jennings, P. C., Lysgaard, S., Hummelshøj, J. S., Vegge, T. & Bligaard, T. Genetic algorithms for computational materials discovery accelerated by machine learning. NPJ Comput. Mater. 5, 46 (2019).

[80] Ma, X., Li, Z., Achenie, L. E. K. & Xin, H. Machine-learning-augmented chemisorption model for CO2 electroreduction catalyst screening. J. Phys. Chem. Lett. 6, 3528–3533 (2015).

[81] Li, Z., Wang, S., Chin, W. S., Achenie, L. E. & Xin, H. High-throughput screening of bimetallic catalysts enabled by machine learning. J. Mater. Chem. A 5, 24131–24138 (2017).

[82] Zhong, M. et al. Accelerated discovery of CO2 electrocatalysts using active machine learning. Nature 581, 178–183 (2020).

[83] Tran, K. & Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO2 reduction and H2 evolution. Nat. Catal. 1, 696–703 (2018).

[84] Ben Dickson (2017), How Artificial Intelligence is revolutionizing human-computer interaction https://thenextweb.com/artificialintelligence/2017/05/10/artificial-intelligence-revolutionizinghuman-computer-interaction/

[85] Okwori Anthony Okpe, Odey Adinya John, Edward Onyebueke Agu, Otiko Anthony Obogo (2019). The Impacts of Artificial Intelligence and Nanotechnology on Human Computer Interaction. *International Journal of Innovative Science and Research Technology*. Volume 4, Issue 11. ISSN No:-2456-2165.

[86] Zheng Li, Siwen Wang and Hongliang Xin (2018). Toward artificial intelligence in catalysis. *NAture CAtAlysis.* Vol. 1, pp. 641–642. https://doi.org/10.1038/s41929-018-0150-1

[87] John R. Kitchin, "Machine Learning in Catalysis", *Nature Catalysis*, Volume 1, Issue 4, Pages 230–232 (2018).

[88] Andrew J. Medford, M. Ross Kunz, Saah m. Ewing, Tammie Borders, and Rebecca Fushimi, "Extracting Knowledge from Data through Catalysis Informatics", *ACS Catal,* Volume 8, Issue 8, Pages 7403-7429, 2018.

[89] Bryan R. Goldsmith, Jacques Esterhuizen, Jin-Xun Liu, Christopher J. Bartel, Christopher Sutton, "machine Learning for Heterogeneous Catalyst Design and Discovery", *AlChe Journal,* Volume 64, Issue 7, Pages 2311-2323, 2018.

[90] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, Gerbrand Ceder, " Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis", *Computational Materials Science,* Volume 68, Pages 314-319, 2013.

[91] F. Abild-Pedersen, J. Greeley, F. Studt, J. Rossmeisl, T. R. Munter, P. G. Moses, E. Skúlason, T. Bligaard, and J. K Nørskov, "Scaling Properties of Adsorption Energies for Hydrogen-Containing Molecules on Transition-Metal Surfaces", *Phys. Rev. Lett.,* Volume 99, Issue1, Pages 1-6, 2007

[92] Peterson A. Andrew & Nørskov K. Jens, "Activity Descriptors for CO_2 Electroreduction to Methane on Transition-Metal Catalysts", *Journal of Physical Chemistry Letters.* Volume 3,, Issue 2, Pages 251–258, 2012.

[93] Andrew J Medford, Jess Wellendorff, Aleksandra Vojvodic, Felix Studt, frank Abild-Pedersen, Karsten W Jacobsen, Thomas Bligaard, Jens K Nørskov, "Assessing the Reliability of Calculated Catalytic Ammonia Synthesis Rates", *Science,* Volume 345, Issue, 6193, Pages 197–200, 2014.

[94] Eric A Walker, Donald Mitchell, Gabriel A Terejanu, Andreas Heyden, "Identifying Active Sites of the Water-Gas Shift Reaction over Titania Supported Platinum Catalysts under Uncertainty". Volume 8, Issue 5, Pages 3990–3998, 2018.

[95] Li, K., and Horne, R. N. (2003). A Decline Curve Analysis Model Based on Fluid Flow Mechanisms. Proc. Li2003adc, Spe-83470-ms. https://doi:10.2118/83470-MS

[96] Irisarri, E., García, M. V., Pérez, F., Estévez, E., and Marcos, M. (2016). "A ModelBased Approach for Process Monitoring in Oil Production Industry," in IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), Berlin, Germany, 6-9 Sept. 2016 (IEEE), 1–4. https://doi:10.1109/ETFA.2016.7733633

[97] Hutahaean, J., Demyanov, V., and Christie, M. A. (2017). On Optimal Selection of Objective Grouping for Multi objective History Matching. SPEJ.22, 1296–1312. doi:10.2118/185957-PA

[98] Farrusseng, D., Baumes, L., and Mirodatos, C. (2003). "Data Management for Combinatorial Heterogeneous Catalysis: Methodology and Development of Advanced Tools," in High-Throughput Analysis. Editors R. A. Potyrailo and E. J. Amis (Boston, MA: Springer). https://doi:10.1007/978-1-4419-8989-525

[99] Wang, Y., Shang, D., Yuan, X., Xue, Y., and Sun, J. (2019). Modeling and Simulation of Reaction and Fractionation Systems for the Industrial Residue Hydrotreating Process. Processes 8, 32. https://doi:10.3390/pr8010032

[100] Cai, G., Liu, Z., Zhang, L., Shi, Q., Zhao, S., and Xu, C. (2021). Systematic Performance Evaluation of Gasoline Molecules Based on Quantitative Structure-Property Relationship Models. Chem. Eng. Sci. 229, 116077. https://doi:10.1016/j.ces.2020.116077

[101] Bokhimi X (2021) Learning the Use of Artificial Intelligence in Heterogeneous Catalysis. Front. Chem. Eng. 3:740270. https://doi:10.3389/fceng.2021.740270

[102] Erdem Günay, M., and Yıldırım, R. (2021). Recent Advances in Knowledge Discovery for Heterogeneous Catalysis Using Machine Learning. Catal. Rev. 63, 120–164. https://doi:10.1080/01614940.2020.1770402

[103] Goldsmith, B. R., Esterhuizen, J., Liu, J. X., Bartel, C. J., and Sutton, C. (2018). Machine Learning for Heterogeneous Catalyst Design and Discovery. Aiche J. 64, 2311–2323. https://doi:10.1002/aic.16198

[104] Schlexer Lamoureux, P., Winther, K. T., Garrido Torres, J. A., Streibel, V., Zhao, M., Bajdich, M., et al. (2019). Machine Learning for Computational Heterogeneous Catalysis. ChemCatChem 11, 3581–3601. https://doi:10.1002/cctc.201900595

[105] Yang, W., Fidelis, T. T., and Sun, W. H. (2020). Machine Learning in Catalysis, from Proposal to Practicing. ACS Omega 5, 83–88. https://doi:10.1021/acsomega.9b03673

[106] Sabatier, P. Hydrogénations et Déshydrogénations Par Catalyse. Berichte der deutschen chemischen Gesellschaft 1911, 44, 1984–2001.

[107] Balandin, A. A. Modern State of the Multiplet Theory of Heterogeneous Catalysis. In Advances in Catalysis; Eley, D. D., Pines, H., Weisz, P. B., Eds.; Academic Press, 1969; Vol. 19, pp 1–210.

[108] Busch, M.; Wodrich, M. D.; Corminboeuf, C. A Generalized Picture of C–C Cross-Coupling. ACS Catal. 2017, 7, 5643–5653.

[109] Wodrich, M. D.; Sawatlon, B.; Busch, M.; Corminboeuf, C. On the Generality of Molecular Volcano Plots. ChemCatChem 2018, 10, 1586–1591.

[110] Meyer, B.; Sawatlon, B.; Heinen, S.; Lilienfeld, O. A. von; Corminboeuf, C. Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts. Chem. Sci. 2018, 9, 7069– 7077.

[111] Wodrich, M. D.; Sawatlon, B.; Solel, E.; Kozuch, S.; Corminboeuf, C. Activity-Based Screening of Homogeneous Catalysts through the Rapid Assessment of Theoretically Derived Turnover Frequencies. ACS Catal. 2019, 9, 5716–5725.

[112] Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. Acc. Chem. Res. 2017, 50, 539–543.

[113] Houk, K. N.; Cheong, P. H.-Y. Computational Prediction of Small-Molecule Catalysts. Nature 2008, 455, 309–313.

[114] Bahmanyar, S.; Houk, K. N. Transition States of Amine-Catalyzed Aldol Reactions Involving Enamine Intermediates: Theoretical Studies of Mechanism, Reactivity, and Stereoselectivity. J. Am. Chem. Soc. 2001, 123, 11273–11283.

[115] Santiago, C. B.; Guo, J.-Y.; Sigman, M. S. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. Chem. Sci. 2018, 9, 2398–2412.

[116] Reid, J. P.; Sigman, M. S. Comparing Quantitative Prediction Methods for the Discovery of SmallMolecule Chiral Catalysts. Nature Reviews Chemistry 2018, 2, 290–305.

[117] Strieth-Kalthoff, F.; Sandfort, F.; Segler, M. H. S.; Glorius, F. Machine Learning the Ropes: Principles, Applications and Directions in Synthetic Chemistry. Chem. Soc. Rev. 2020.

[118] Harper, K. C.; Bess, E. N.; Sigman, M. S. Multidimensional Steric Parameters in the Analysis of Asymmetric Catalytic Reactions. Nature Chemistry 2012, 4, 366–374.

[119] Milo, A.; Bess, E. N.; Sigman, M. S. Interrogating Selectivity in Catalysis Using Molecular Vibrations. Nature 2014, 507, 210–214.

[120] Orlandi, M.; Coelho, J. A. S.; Hilton, M. J.; Toste, F. D.; Sigman, M. S. Parametrization of NonCovalent Interactions for Transition State Interrogation Applied to Asymmetric Catalysis. J. Am. Chem. Soc. 2017, 139, 6803–6806.

[121] Gabriel dos Passos Gomes, Robert Pollice and Alán Aspuru-Guzik (2019). "Navigating through the Maze of Homogeneous Catalyst Design with Machine Learning". aspuru@utoronto.ca

[122] Jorner, K.; Brinck, T.; Norrby, P.-O.; Buttar, D. Machine Learning Meets Mechanistic Modelling for Accurate Prediction of Experimental Activation Energies. 2020.

[123] Renom-Carrasco, M.; Lefort, L. Ligand Libraries for High Throughput Screening of Homogeneous Catalysts. Chem. Soc. Rev. 2018, 47, 5038–5060.

[124] Schultz, D.; Campeau, L.-C. Harder, Better, Faster. Nature Chemistry 2020, 1–4.

[125] Rai, A. Explainable AI: From Black Box to Glass Box. J. of the Acad. Mark. Sci. 2020, 48, 137–141.

[126] Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. Information Fusion 2020, 58, 82–115.

[127] Carvalho, D. V.; Pereira, E. M.; Cardoso, J. S. Machine Learning Interpretability: A Survey on Methods and Metrics. Electronics 2019, 8, 832.

[128] Heller, E. J.; Tomsovic, S. Postmodern Quantum Mechanics. Physics Today 1993, 46, 38.

[129] Garnelo, M.; Shanahan, M. Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations. Current Opinion in Behavioral Sciences 2019, 29, 17–23.

[130] Lambard, G.; Gracheva, E. SMILES-X: Autonomous Molecular Compounds Characterization for Small Datasets without Descriptors. Mach. Learn.: Sci. Technol. 2020, 1, 025004.

[131] Ramakrishnan, R.; von Lilienfeld, O. A. Many Molecular Properties from One Kernel in Chemical Space. CHIMIA International Journal for Chemistry 2015, 69, 182–186.

[132] Ramakrishnan, R.; Lilienfeld, O. A. von. Machine Learning, Quantum Chemistry, and Chemical Space. In Reviews in Computational Chemistry; John Wiley & Sons, Ltd, 2017; pp 225–256.

[133] Christensen, A. S.; Faber, F. A.; von Lilienfeld, O. A. Operators in Quantum Machine Learning: Response Properties in Chemical Space. J. Chem. Phys. 2019, 150, 064105.

[134] Lim, J.; Hwang, S.-Y.; Moon, S.; Kim, S.; Kim, W. Y. Scaffold-Based Molecular Design with a Graph Generative Model. Chem. Sci. 2020, 11, 1153–1164.

[135] Maziarka, Ł.; Pocha, A.; Kaczmarczyk, J.; Rataj, K.; Danel, T.; Warchoł, M. Mol-CycleGAN: A Generative Model for Molecular Optimization. Journal of Cheminformatics 2020, 12, 2.

[136] Schwalbe-Koda, D.; Gómez-Bombarelli, R. Generative Models for Automatic Chemical Design. In Machine Learning Meets Quantum Physics; Schütt, K. T., Chmiela, S., von Lilienfeld, O. A., Tkatchenko, A., Tsuda, K., Müller, K.-R., Eds.; Lecture Notes in Physics; Springer International Publishing: Cham, 2020; pp 445–467.

[137] FanYang, Dehui Deng, Xiulian Pan, Qiang Fuand Xinhe Bao. "Understanding nano effects in catalysis". *National Science Review*, 2015 2:183–201. https://doi:10.1093/nsr/nwv024

[138] Ito, T.; Hamada, H.; Kintaichi, Y.; Sasaki, M. Database for Catalysis Design. Catalysis Today 1991, 10, 223–232.

[139] Hu, X., and Yip, A.C.K. (2021). Heterogeneous catalysis: enabling a sustainable future. Front. Catal. 1,3.

[140] Stephan N. Steinmann, Angga Hermawan, Mohammed Bin Jassar, and Zhi Wei She (2022). Autonomous high-throughput computations in catalysis. *Chem Catalysis* 2, 940–956.

[141] Kas, R., Yang, K., Bohra, D., Kortlever, R., Burdyny, T., and Smith, W.A. (2020). Electrochemical CO2 reduction on nanostructured metal electrodes: fact or defect? Chem. Sci. 11, 1738–1749.

[142] Modestino, M.A., Hashemi, S.M.H., and Haussener, S. (2016). Mass transport aspects of electrochemical solar-hydrogen generation. EnergyEnviron.Sci.9,1533–1551.

[143] Chen, C., Zuo, Y., Ye, W., Li, X., Deng, Z., and Ong, S.P. (2020). A critical review of machine learning of energy materials. Adv. Energy Mater. 10, 1903242.

[144] Bligaard, T., Norskov, J.K., Dahl, S., Matthiesen, J., Christensen, C.H., and Sehested, J. (2004). The Bronsted-EvansPolanyi relation and the volcano curve in heterogeneous catalysis. J. Catal. 224, 206–217.

[145] Abdelfatah, K., Yang, W., Vijay Solomon, R., Rajbanshi, B., Chowdhury, A., Zare, M., Kundu, S.K., Yonge, A., Heyden, A., and Terejanu, G. (2019). Prediction of transitionstate energies of hydrodeoxygenation reactions on transition-metal surfaces based on machine learning. J. Phys. Chem. C 123, 29804–29810.

[146] Moosavi, S.M., Jablonka, K.M., and Smit, B. (2020). The role of machine learning in the understandingand designofmaterials.J.Am. Chem. Soc. 142, 20273–20287.

[147] Pollice, R., dos Passos Gomes, G., Aldeghi, M., Hickman, R.J., Krenn, M., Lavigne, C., Lindner-D'Addario, M., Nigam, A., Ser, C.T., Yao,Z., et al. (2021). Data-driven strategies for accelerated materials design. Acc. Chem. Res. 54, 849–860.

[148] Bannigan, P., Aldeghi, M., Bao, Z., Ha ¨se, F., Aspuru-Guzik, A., and Allen, C. (2021). Machine learning directed drug formulation development. Adv. Drug Deliv. Rev. 175, 113806.

[149] Schlexer Lamoureux, P., Winther, K.T., Garrido Torres, J.A., Streibel, V., Zhao, M., Bajdich, M., Abild-Pedersen, F., and Bligaard, T. (2019). Machine learning for computational heterogeneous catalysis. ChemCatChem 11, 3581–3601.

[150] Toyao,T.,Maeno,Z.,Takakusagi,S.,Kamachi, T., Takigawa, I., and Shimizu, K. (2020). Machine learning for catalysis informatics: recentapplicationsandprospects.ACSCatal. 10, 2260–2297.

[151] Zhong, M., Tran, K., Min, Y., Wang, C., Wang, Z., Dinh, C.-T., De Luna, P., Yu, Z., Rasouli, A.S., Brodersen, P., et al. (2020). Accelerated discovery of CO2 electrocatalysts using active machine learning. Nature 581, 178–183.

[152] Ahuja, K., Green, W.H., and Li, Y.-P. (2021). Learning to optimize molecular geometries using reinforcement learning. J. Chem. Theory Comput. 17, 818–825.

[153] Lan, T., and An, Q. (2021). Discovering catalytic reaction networks using deep reinforcement learning from first-principles. J. Am. Chem. Soc. 143, 16804–16812.

[154] Herbol, H.C., Poloczek, M., and Clancy, P. (2020). Cost-effective materials discovery: Bayesian optimization across multiple information sources. Mater. Horiz. 7, 2113– 2123.

[155] Turner, H.W., Volpe, A.F., and Weinberg, W.H. (2009). High-throughput heterogeneous catalyst research. Surf. Sci. 603, 1763–1769.

[156] Nursam, N.M., Wang, X., and Caruso, R.A. (2015). High-throughput synthesis and screening of titania-based photocatalysts. ACS Comb. Sci. 17, 548–569.

[157] Potgieter, K., Aimon, A., Smit, E., von Delft, F., and Meijboom, R. (2021). Robotic catalysis: a high-throughput method for miniature screening of mesoporous metal oxides. Chem. Methods 1, 192–200.

[158] Stroyuk, O., Raievska, O., Langner, S., Kupfer, C., Barabash, A., Solonenko, D., Azhniuk, Y., Hauch, J., Osvet, A., Batentschuk, M., et al. (2021). High-throughput robotic synthesisand photoluminescence characterization of aqueous multinary copper–silver indium chalcogenide quantum dots. Part. Part. Syst. Charact. 38, 2100169.

[159] Krska, S.W., DiRocco, D.A., Dreher, S.D., and Shevlin, M. (2017). The evolution of chemical high-throughput experimentation to address challenging problems in pharmaceutical synthesis. Acc. Chem. Res. 50, 2976–2985.

[160] Shi,Y.,Yang, B.,Rack,P.D.,Guo, S.,Liaw,P.K., and Zhao, Y. (2020). High-throughput synthesis and corrosion behavior of sputterdeposited nanocrystalline Alx(CoCrFeNi)100x combinatorial high-entropy alloys. Mater. Des. 195, 109018.

[161] Mao, S.S., and Zhang, X. (2015). Highthroughput multi-plume pulsed-laser deposition for materials exploration and optimization. Engineering 1, 367–371.

[162] Cong, P., Doolen, R.D., Fan, Q., Giaquinta, D.M., Guan, S., McFarland, E.W., Poojary, D.M., Self, K., Turner, H.W., and Weinberg, W.H. (1999). High-throughput synthesis and screening of combinatorial heterogeneous catalyst libraries. Angew. Chem. Int. Ed. Engl. 38, 483–488.

[163] Scoutaris, N., Nion, A., Hurt, A., and Douroumis,D.(2016). Jetdispensingasahigh throughput method for rapid screening and manufacturing of cocrystals. CrystEngComm 18, 5079–5082.

[164] Ortega, C., Otyuskaya, D., Ras, E.-J., Virla, L.D., Patience, G.S., and Dathe, H. (2021). Experimental methods in chemical engineering: high throughput catalyst testing — HTCT. Can. J. Chem. Eng.99, 1288–1306.

[165] Bruker. D8 ADVANCE. https://www.bruker.com/en/products-and-solutions/diffractometers-and-scattering-systems/x-ray-diffractometers/d8-advance-family/d8-advance.html

[166] Goldrick, S., Umprecht, A., Tang, A., Zakrzewski,R.,Cheeks,M.,Turner,R.,Charles, A., Les, K., Hulley, M., Spencer, C., et al. (2020). High-throughput Raman spectroscopy combined with innovate data analysis workflow to enhance biopharmaceutical process development. Processes 8, 1179.

[167] Mondol, A.S., Patel, M.D., Ru ¨ger, J., Stiebing, C., Kleiber, A., Henkel, T., Popp, J., and Schie, I.W. (2019). Application of high-throughput screening Raman spectroscopy (HTS-RS) for label-free identification and molecular characterization of pollen. Sensors 19, 4428.

[168] Coffey, P., Smith, N., Lennox, B., Kijne, G., Bowen, B., Davis-Johnston, A., and Martin, P.A. (2021). Robotic arm material characterisation using LIBS and Raman in a nuclear hot cell decommissioning environment. J. Hazard. Mater. 412, 125193.

[169] Westley, C., Xu, Y., Carnell, A.J., Turner, N.J., and Goodacre, R. (2016). Label-free surface enhanced Raman scattering approach for high-throughput screening of biocatalysts. Anal. Chem. 88, 5898–5903.

[170] Nguyen, T.N., Nhat, T.T.P., Takimoto, K., Thakur, A., Nishimura, S., Ohyama, J., Miyazato, I., Takahashi, L., Fujima, J., Takahashi, K., et al. (2020). High-throughput experimentation and catalyst informatics for oxidative coupling ofmethane.ACS Catal.10, 921–932.

[171] Roberts, E.J., Habas, S.E., Wang, L., Ruddy, D.A., White, E.A., Baddour, F.G., Griffin, M.B., Schaidle, J.A., Malmstadt, N., and Brutchey, R.L. (2017). High-throughput continuous flow synthesis of nickel nanoparticles for the catalytic hydrodeoxygenation of guaiacol. ACS Sustain. Chem. Eng. 5, 632–639.

[172] Kondratyuk, P., Gumuslu, G., Shukla, S., Miller, J.B., Morreale, B.D., and Gellman, A.J. (2013). A microreactor array for spatially resolved measurement of catalytic activity for hig hthroughput catalysis science. J. Catal. 300, 55–62.

[173] Arakawa, H. et al. Catalysis research of relevance to carbon management: progress, challenges, and opportunities. Chem. Rev. 101, 953–996 (2001).

[174] Olah, G. A. Beyond oil and gas: the methanol economy. Angew. Chem. Int. Ed. 44, 2636–2639 (2005).

[175] Olah, G. A., Goeppert, A. & Surya Prakash, G. K. Chemical recycling of carbon dioxide to methanol and dimethyl ether: from greenhouse gas to renewable, environmentally carbon neutral fuels and synthetic hydrocarbons. J. Org. Chem. 74, 487–498 (2009).

[176] Kondratenko, E. V., Mul, G., Baltrusaitis, J., Larrazábal, G. O. & PérezRamírez, J. Status and perspectives of CO2 conversion into fuels and chemicals by catalytic, photocatalytic and electrocatalytic processes. Energy Environ. Sci. 6, 3112 (2013).

[177] Li, J. et al. Volcano trend in electrocatalytic CO2 reduction activity over atomically dispersed metal sites on nitrogen-doped carbon. ACS Catal. 9, 10426 (2019).

[178] Frei, M. S., Mondelli, C., Short, M. I. M. & Pérez-Ramírez, J. Methanol as a hydrogen carrier: kinetic and thermodynamic drivers for its CO2-based synthesis and reforming over heterogeneous catalysts. ChemSusChem. 13, 6330 (2020).

[179] Martin, O. et al. Indium oxide as a superior catalyst for methanol synthesis by CO2 hydrogenation. Angew. Chem. Int. Ed. 55, 6261 (2016).

[180] Aliaksei Mazheika, Yang-Gang Wang, Rosendo Valero, Francesc Viñes, Francesc Illas, Luca M. Ghiringhelli, Sergey V. Levchenko, Matthias Scheffler (2022). Artificial-intelligence-driven discovery of catalyst genes with application to CO2 activation on semiconductor oxides. *Nature Communications*. 13:419 https://doi.org/10.1038/s41467-022-28042-z

[181] Sattari, K., Xie, Y. & Lin, J. Data-driven algorithms for inverse design of polymers. Soft Matter 17, 7607 –7622 (2021).

[182] Audus, D.J. & de Pablo, J. J. Polymer informatics: opportunities and challenges. ACS Macro Lett. 6, 1078 –1082 (2017).

[183] Chen, G. et al. Machine-learning-assisted denovo design of organic molecules and polymers: opportunities and challenges. Polymers 12, 163 (2020).

[184] Himanen, L., Geurts, A., Foster, A. S. & Rinke, P. Data-driven materials science: status, challenges, and perspectives. Adv. Sci. 6, 1900808 (2019).

[185] Hatakeyama-Sato, K. (2022). Recent advances and challenges in experiment-oriented polymer informatics. Polym.J.1–15 https://doi.org/10.1038/s41428-022-00734-9

[186] Nathaniel H. Park, Matteo Manica, Jannis Born, James L. Hedrick, Tim Erdmann, Dmitry Yu. Zubarev, Nil Adell-Mill, Pedro L. Arrechea (2023). Artificial intelligence driven design of catalysts and materials for ring opening polymerization using a domain-specific language. *Nature Communications*. 14:3686. https://doi.org/10.1038/s41467-023-39396-3

[187] Xiangyu Guo, Shiru Lin, Jinxing Gu, Shengli Zhang. "Simultaneously achieving high activity and selectivity toward two-electron O2 electroreduction: The power of single-atom catalysts". *ACS Catal.*, 2019, 9 (12) 11042-11054.

[188] Shi Chen, Hao Huang, Peng Jiang, Kang Yang, Jiefeng Diao, Shipeng Gong, Shuai Liu, Minxue Huang, Hui Wang, Qianwang Chen. "Mn-doped RuO2 nanocrystals as highly active electrocatalysts for enhanced oxygen evolution in acidic media". *ACS Catal.* 10 (2) (2019) 1152-1160

[189] E. Czech, T. Troczynski, "Hydrogen Generation Through Massive Corrosion of Deformed Aluminium in Water", *International Journal of Hydrogen Energy*, Volume 35, Issue 3, Pages 1029-1037, 2010.

[190] A. Jain, Y. Shin, K. A. Persson, Computational predictions of energy materials using density functional theory, Nat. Rev. Mater 1 (1) (2016) 15004

[191] H. Niu, X. Wang, C. Shao, et al., Revealing the oxygen reduction reaction activity origin of single atoms supported on g-C3N4 monolayers: A first-principles study, J. Mater. Chem. A 8 (14) (2020) 6555-6563

[192] H. Xu, D. Cheng, D. Cao, et al., A universal principle for a rational design of single-atom electrocatalysts, Nat. Catal. 1 (5) (2018) 339-348

[193] A. Mohajeri, N. L. Dashti, Cooperativity in bimetallic sacs: An efficient strategy for designing bifunctional catalysts for overall water splitting, J. Phys. Chem. C 123 (51) (2019) 30972-30980

[194] M. Krenn, J. Landgraf, T. Foesel, et al., Artificial intelligence and machine learning for quantum technologies, Phys. Rev. A 107 (1) (2023) 010101

[195] H. Masood, C. Y. Toe, W. Y. Teoh, et al., Machine learning for accelerated discovery of solar photocatalysts, ACS Catal. 9 (12) (2019) 11774-11787

[196] Q. Zhu, F. Zhang, Y. Huang, et al., An all-round ai-chemist with a scientific mind, Natl. Sci. Rev. 9 (10) (2022) nwac190.

[197] E. O. Pyzer-Knapp, J. W. Pitera, P. W. J. Staar, et al., Accelerating materials discovery using artificial intelligence, high performance computing and robotics, NPJ Comput. Mater. 8 (1) (2022) 84

[198] C. Fare, P. Fenner, M. Benatan, et al., A multi-fidelity machine learning approach to high throughput materials screening, NPJ Comput. Mater. 8 (1) (2022) 257

[199] H. Mai, T. C. Le, D. Chen, et al., Machine learning for electrocatalyst and photocatalyst design and discovery, Chem. Rev. 122 (16) (2022) 13478-13515

[200] P. Xu, X. Ji, M. Li, et al., Small data machine learning in materials science, NPJ Comput. Mater. 9 (1) (2023) 42

[201] B. J. Shields, J. Stevens, J. Li, et al., Bayesian reaction optimization as a tool for chemical synthesis, Nature 590 (7844) (2021) 89-96

[202] K. M. Jablonka, D. Ongari, S. M. Moosavi, et al., Big-data science in porous materials: Materials genomics and machine learning, Chem. Rev. 120 (16) (2020) 8066-8129

[203] Z. Fu, W. Liu, C. Huang, et al., A review of performance prediction based on machine learning in materials science, Nanomaterials 12 (17) (2022) 2957

[204] B. Han, Y. Lin, Y. Yang, et al., Deep-learning-enabled fast optical identification and characterization of 2d materials, Adv. Mater. 32 (29) (2020) e2000953

[205] Y. Jiang, Y. Yu, M. Kong, et al., Artificial intelligence for retrosynthesis prediction, Engineering 25 (2022) 32-50

[206] Y. Shen, J. E. Borowski, M. A. Hardy, et al., Automation and computer-assisted planning for chemical synthesis, Nat. Rev. Methods Primers 1 (1) (2021) 23

[207] Linkai Han , Zhonghua Xiang , Intelligent design and synthesis of energy catalytic materials, Fundamental Research (2023), doi: https://doi.org/10.1016/j.fmre.2023.10.012

[208] X. Han, W. Zhang, X. Ma, et al., Identifying the activation of bimetallic sites in NiCo2S4@g-C3N4-CNT hybrid electrocatalysts for synergistic oxygen reduction and evolution, Adv. Mater. 31 (18) (2019) 1808281

[209] Y. Sun, H. Liao, J. Wang, et al., Covalency competition dominates the water oxidation structure–activity relationship on spinel oxides, Nat. Catal. 3 (7) (2020) 554-563

[210] S. Lu, Q. Zhou, Y. Ouyang, et al., Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning, Nat. Commun. 9 (1) (2018) 3405

[211] I. C. Man, H. Y. Su, F. Calle-Vallejo, et al., Universality in oxygen evolution electrocatalysis on oxide surfaces, ChemCatChem 3 (7) (2011) 1159-1165

[212] S. Zhai, H. Xie, P. Cui, et al., A combined ionic lewis acid descriptor and machine-learning approach to prediction of efficient oxygen reduction electrodes for ceramic fuel cells, Nat. Energy 7 (9) (2022) 866-875

[213] K. M. Jablonka, D. Ongari, S. M. Moosavi, et al., Using collective knowledge to assign oxidation states of metal cations in metal-organic frameworks, Nat. Chem. 13 (8) (2021) 771-777

[214] Z. Wang, Y. Gu, L. Zheng, et al., Machine learning guided dopant selection for metal oxide-based photoelectrochemical water splitting: The case study of Fe2O3 and CuO, Adv. Mater. 34 (10) (2022) e2106776

[215] M. Zhong, K. Tran, Y. Min, et al., Accelerated discovery of CO2 electrocatalysts using active machine learning, Nature 581 (7807) (2020) 178-183

[216] B. Tang, Y. Lu, J. Zhou, et al., Machine learning-guided synthesis of advanced inorganic materials, Mater. Today 41 (2020) 72-80

[217] Y. Guo, X. He, Y. Su, et al., Machine-learning-guided discovery and optimization of additives in preparing Cu catalysts for CO2 reduction, J. Am. Chem. Soc. 143 (15) (2021) 5755-5762

[218] H. Chun, E. Lee, K. Nam, et al., First-principle-data-integrated machine-learning approach for high-throughput searching of ternary electrocatalyst toward oxygen reduction reaction, Chem. Catal 1 (4) (2021) 855-869

[219] C. C. Price, A. Singh, N. C. Frey, et al., Efficient catalyst screening using graph neural networks to predict strain effects on adsorption energy, Sci. Adv. 8 (2022) eabq5944

[220] J. A. Esterhuizen, B. R. Goldsmith, S. Linic, Uncovering electronic and geometric descriptors of chemical activity for metal alloys and oxides using unsupervised machine learning, Chem. Catal 1 (4) (2021) 923-940

[221] W. Xu, H. Tang, H. Gu, et al., Research progress of asymmetrically coordinated single-atom catalysts for electrocatalytic reactions, J. Mater. Chem. A 10 (28) (2022) 14732-14746

[222] M. Babucci, A. Guntida, B. C. Gates, Atomically dispersed metals on well-defined supports including zeolites and metal-organic frameworks: Structure, bonding, reactivity, and catalysis, Chem. Rev. 120 (21) (2020) 11956-11985

[223] S. Ji, Y. Chen, X. Wang, et al., Chemical synthesis of single atomic site catalysts, Chem. Rev. 120 (21) (2020) 11900-11955

[224] X. Zhao, X. Liu, B. Huang, et al., Hydroxyl group modification improves the electrocatalytic ORR and OER activity of graphene supported single and bi-metal atomic catalysts (Ni, Co, and Fe), J. Mater. Chem. A 7 (42) (2019) 24583-24593

[225] B. M. Hunter, H. B. Gray, A. M. Müller, Earth-abundant heterogeneous water oxidation catalysts, Chem. Rev. 116 (22) (2016) 14120-14136

[226] Z. Lian, M. Yang, F. Jan, et al., Machine learning derived blueprint for rational design of the effective single-atom cathode catalyst of the lithium-sulfur battery, J. Phys. Chem. Lett. 12 (29) (2021) 7053-7059

[227] L. Wu, T. Guo, T. Li, Machine learning-accelerated prediction of overpotential of oxygen evolution reaction of single-atom catalysts, iScience 24 (5) (2021) 102398

[228] Y. Ying, K. Fan, X. Luo, et al., Unravelling the origin of bifunctional OER/ORR activity for single-atom catalysts supported on C2N by dft and machine learning, J. Mater. Chem. A 9 (31) (2021) 16860-16867

[229] X. Liu, Y. Zhang, W. Wang, et al., Transition metal and N doping on alp monolayers for bifunctional oxygen electrocatalysts: Density functional theory study assisted by machine learning description, ACS Appl. Mater. Interfaces 14 (1) (2022) 1249-1259

[230] F. Liu, C. Duan, H. J. Kulik, Rapid detection of strong correlation with machine learning for transition-metal complex high-throughput screening, J. Phys. Chem. Lett. 11 (19) (2020) 8067-8076

[231] B. Burger, P. M. Maffettone, V. V. Gusev, et al., A mobile robotic chemist, Nature 583 (7815) (2020) 237-241

[232] Y. Jiang, D. Salley, A. Sharma, et al., An artificial intelligence enabled chemical synthesis robot for exploration and optimization of nanomaterials, Sci. Adv. 8 (2022) eabo2626

[233] D. Cheng, Z. J. Zhao, G. Zhang, et al., The nature of active sites for carbon dioxide electroreduction over oxide-derived copper catalysts, Nat. Commun. 12 (1) (2021) 395

[234] L. Xing, Z. Chen, W. Chen, et al., Robotic platform for accelerating the high-throughput study of silver nanocrystals in sensitive/selective hg2+ detection, Chem. Eng. J. 466 (2023) 143225

[235] H. Zhao, W. Chen, H. Huang, et al., A robotic platform for the synthesis of colloidal nanocrystals, Nat. Synth 2 (6) (2023) 505-514

[236] C. Zheng, K. Mathew, C. Chen, et al., Automated generation and ensemble-learned matching of x-ray absorption spectra, NPJ Comput. Mater. 4 (1) (2018) 12

[237] J. S. Smith, O. Isayev, A. E. Roitberg, Ani-1: An extensible neural network potential with dft accuracy at force field computational cost, Chem. Sci. 8 (4) (2017) 3192-3203.