American Journal of Advanced Multidisciplinary Research and Innovation (AJAMRI)

E-ISSN: XXXX-XXXX ● Website: www.ajamri.com ● Email: editor@ajamri.com

AI-Driven Discovery of Low-Carbon Chemical Pathways for Sustainable Industrial Transformation

Prof. Richard West

Professor, Northeastern University, USA

Abstract

Industrial chemical production provides materials, fuels, fertilizers, polymers, pharmaceuticals, and intermediates that support modern economies. However, many established chemical pathways depend on fossil-derived feedstocks, energy-intensive reaction conditions, carbon-intensive electricity, hazardous solvents, and processes that generate substantial quantities of waste. Reducing these environmental burdens requires more than incremental plant-level efficiency improvements. It demands the discovery and implementation of alternative catalysts, feedstocks, solvents, reaction sequences, separation methods, and process configurations capable of delivering chemically viable products with lower life-cycle greenhouse gas emissions.

Artificial intelligence offers a new approach to this discovery challenge by learning relationships among molecular structures, reaction conditions, catalyst properties, energy requirements, product yields, selectivity, and environmental performance. Machine learning, active learning, reaction-prediction systems, generative models, Bayesian optimization, and autonomous experimentation can help researchers navigate chemical spaces that are too extensive for conventional trial-and-error investigation. This paper examines how these capabilities can support low-carbon chemical-pathway discovery and sustainable industrial transformation. A conceptual review is integrated with an illustrative screening simulation comparing unguided candidate selection with AI-guided active learning. Within the simulated evaluation budget of 100 candidates, the AI-guided workflow identifies 30 promising pathways compared with 13 under unguided screening. These values are methodological illustrations and do not represent industrial experimental findings.

The analysis indicates that AI can improve candidate prioritization, connect molecular discovery with process-level assessment, and support multi-objective decisions involving emissions, energy, yield, selectivity, toxicity, cost, and scalability. Its usefulness remains dependent on reliable chemical data, mechanistic validation, uncertainty quantification, life-cycle boundaries, experimental confirmation, and human oversight. The study proposes that responsible AI-guided pathway discovery should combine data-driven prediction with chemical theory, high-throughput experimentation, process simulation, techno-economic assessment, and life-cycle analysis. Such an integrated framework can accelerate industrial decarbonization while reducing the risk that apparently efficient reactions transfer emissions or environmental burdens to other stages of the value chain.

American Journal of Advanced Multidisciplinary Research and Innovation (AJAMRI)

E-ISSN: XXXX-XXXX ● Website: www.ajamri.com ● Email: editor@ajamri.com

Keywords: artificial intelligence; low-carbon chemistry; sustainable chemical pathways; catalyst discovery; active learning; industrial decarbonization; green chemistry; process systems engineering.

1. Introduction

The chemical industry occupies a central position within global manufacturing because it supplies essential inputs to agriculture, construction, health care, transportation, electronics, energy systems, and consumer products. Its products enable economic and social development, yet their manufacture frequently requires high temperatures, elevated pressures, fossil-based feedstocks, carbon-intensive hydrogen, and complex separation operations. Emissions arise not only from the energy used to operate chemical plants but also from the carbon incorporated into feedstocks and released during reactions, product use, waste treatment, or end-of-life processing.

Conventional industrial innovation has improved efficiency through better catalysts, process integration, heat recovery, reactor design, and plant control. These measures remain necessary, but the depth of decarbonization required in many chemical value chains may demand more fundamental changes. Alternative pathways can involve renewable electricity, green hydrogen, captured carbon dioxide, biomass-derived intermediates, biocatalysis, electrochemistry, photocatalysis, solvent substitution, circular feedstocks, and lower-temperature synthesis. Identifying a technically plausible pathway is only the first stage. The pathway must also satisfy requirements relating to selectivity, yield, durability, product quality, safety, cost, availability of inputs, and industrial scalability.

The number of possible combinations is exceptionally large. Researchers may need to compare thousands of catalyst compositions, molecular structures, solvents, reaction conditions, and process configurations. Traditional experimentation examines only a small portion of this design space because laboratory trials require time, specialist labor, materials, analytical equipment, and energy. Computational chemistry can expand the search but may also become expensive when high-fidelity calculations are applied to large molecular or catalytic systems. Artificial intelligence can contribute by learning surrogate relationships from experimental and computational data and using them to rank candidates before costly validation.

AI-based chemical discovery includes several complementary techniques. Supervised-learning models predict properties such as reaction yield, activation energy, adsorption energy, catalyst stability, and selectivity. Generative systems propose molecules, catalysts, or reaction pathways that satisfy specified objectives. Active learning selects the next experiment by balancing exploitation of promising regions against exploration of uncertain ones. Bayesian optimization identifies favorable reaction conditions with relatively few evaluations. Robotic platforms can perform selected reactions, measure outcomes, and return the results to an algorithm for iterative improvement.

The scientific opportunity is accompanied by an important methodological risk. A pathway may be described as low carbon because it operates at a lower temperature, even though its feedstock, electricity, solvent, catalyst, purification requirement, or waste treatment carries a substantial emissions burden. AI models trained only to maximize yield may recommend processes that conflict with broader sustainability goals. Meaningful discovery therefore requires multi-objective evaluation and explicit life-cycle boundaries. This paper investigates how AI can be integrated with chemical knowledge, process

engineering, and environmental assessment to discover pathways that contribute credibly to sustainable industrial transformation.

2. Review of Literature

2.1 Data-Driven Representation of Chemical Reactions

Early applications of machine learning in chemistry concentrated on predicting molecular properties from descriptors derived from composition, bonding, geometry, or electronic structure. These approaches reduced the need to perform a high-cost calculation for every candidate. The development of graph-based representations allowed models to treat atoms as nodes and chemical bonds as edges, enabling them to learn structural relationships more directly.

Coley and colleagues demonstrated that machine-learning methods can support reaction prediction and synthesis planning. Reaction-prediction models learn transformations from collections of previously reported reactions and estimate likely products or precursors. Schwaller and colleagues advanced sequence-based chemical modelling through transformer architectures capable of learning reaction representations from text-like molecular notation. These systems showed that data-driven methods could recognize regularities across large reaction databases.

Stocker and colleagues examined machine learning within chemical reaction space rather than limiting prediction to isolated compounds. Their analysis demonstrated that the topology and composition of a reaction network affect model performance. Molecules that function as highly connected reaction hubs may be particularly important in training-set selection. This finding is relevant to low-carbon pathway discovery because industrial pathways are networks of connected transformations, not independent molecular predictions.

Data quality remains a persistent concern. Published chemical literature often emphasizes successful reactions while unsuccessful experiments, unstable catalysts, unfavorable conditions, and scale-up failures remain underreported. A model trained primarily on successful reactions may develop an incomplete understanding of chemical feasibility. Variations in units, measurement conditions, catalyst descriptions, analytical methods, and reporting conventions can further weaken the reliability of combined datasets.

2.2 AI-Assisted Catalyst and Reaction-Condition Discovery

Catalyst discovery is a promising AI application because catalyst performance depends on complex relationships among composition, surface structure, adsorption, reaction intermediates, temperature, pressure, and surrounding chemical conditions. Density functional theory can calculate relevant properties but may be too computationally demanding for unrestricted screening. Machine-learning surrogate models can approximate selected calculations and prioritize candidates for more accurate evaluation.

Chen and colleagues developed a machine-learning approach using comparatively simple features to examine electrocatalysts for carbon dioxide reduction. Such studies illustrate how data-driven models may connect material descriptors with catalytic outcomes. The predictions remain conditional on the relevance of the training data, the quality of the selected descriptors, and confirmation under experimentally realistic operating conditions.

Bayesian optimization and active learning are particularly useful when experiments are expensive. Instead of screening candidates randomly, the model uses existing observations to estimate both predicted performance and uncertainty. The next experiment is selected because it is expected either to perform well or to improve the model's knowledge. This feedback process can reduce the number of laboratory evaluations required to locate promising operating regions.

Coley and collaborators connected AI-based planning with robotic flow synthesis. Their platform illustrated a pathway toward automated chemical production in which computational planning and laboratory execution become integrated. Autonomous systems can improve experimental consistency and generate machine-readable data, but they still require human decisions concerning safety, acceptable objectives, chemical plausibility, and interpretation of unexpected results.

2.3 Green Chemistry and Industrial Decarbonization

Green chemistry provides a conceptual foundation for evaluating chemical pathways. Anastas and Warner emphasized pollution prevention, atom economy, safer synthesis, renewable feedstocks, energy efficiency, reduced derivatization, catalysis, degradation, and accident prevention. These principles show that sustainability cannot be represented by a single measure.

Industrial decarbonization further requires life-cycle analysis and process systems engineering. A reaction with favorable laboratory yield may require energy-intensive separation, rare catalytic materials, or a solvent that is difficult to recover. Conversely, a reaction with a moderate yield may be attractive if it uses renewable feedstocks, operates under mild conditions, and permits efficient recycling.

AI-guided pathway discovery should therefore incorporate multiple levels of evidence. Molecular calculations can identify thermodynamic or kinetic plausibility. Experiments can validate yield, selectivity, and catalyst stability. Process simulation can estimate material and energy balances. Techno-economic analysis can evaluate commercial feasibility, while life-cycle assessment can examine environmental burdens across the wider supply chain.

3. Objectives of the Study

3.1 Scientific Discovery Objectives

The primary objective is to examine how artificial intelligence can accelerate the discovery of chemical pathways that reduce greenhouse gas emissions while remaining chemically and industrially credible. The study evaluates applications involving reaction prediction, catalyst screening, pathway generation, process optimization, and autonomous experimentation.

A related objective is to clarify how machine learning can complement rather than replace chemical theory and experimental evidence. AI predictions are treated as decision-support outputs that require mechanistic assessment and laboratory validation.

3.2 Sustainability Assessment Objectives

The study aims to identify the environmental variables that should be included in AI-guided pathway selection. These include process emissions, electricity demand, reaction temperature, feedstock origin, solvent burden, catalyst durability, waste generation, material circularity, and separation requirements.

The analysis also examines the risk of burden shifting. A model that minimizes reactor energy without considering upstream electricity, feedstock production, downstream purification, or waste treatment may produce misleading recommendations.

3.3 Industrial Transformation Objectives

The third objective is to develop an implementation perspective for integrating AI discovery with industrial research and process development. The paper considers how active learning, laboratory automation, process simulation, life-cycle assessment, and techno-economic analysis can be organized into a coherent workflow.

The study further seeks to identify governance requirements involving data quality, uncertainty, safety, reproducibility, intellectual property, workforce capability, and environmental claim verification.

4. Research Methodology

4.1 Research Design and Information Sources

The paper adopts a conceptual and simulation-based design. It synthesizes scholarly work from artificial intelligence, computational chemistry, catalysis, green chemistry, industrial ecology, and process systems engineering. No proprietary industrial data, human participants, confidential chemical formulations, or physical laboratory experiments were used.

Relevant literature was identified through multidisciplinary academic databases and publisher platforms. Search concepts included “machine learning reaction prediction,” “AI catalyst discovery,” “active learning chemistry,” “autonomous chemical synthesis,” “low-carbon chemical processes,” “green chemistry,” and “industrial decarbonization.” Peer-reviewed publications and established foundational works available within the defined reference period were prioritized.

4.2 Analytical Framework

Candidate pathway discovery was evaluated through five connected dimensions: chemical performance, carbon performance, resource efficiency, industrial feasibility, and decision reliability. Each dimension contains variables that should be assessed before a pathway is characterized as promising.

Table 1: Multi-Dimensional Evaluation Framework for AI-Discovered Pathways

Dimension	Representative indicators	Analytical purpose
Chemical performance	Yield, selectivity, conversion and stability	Determines whether the pathway performs its intended chemical function
Carbon performance	Direct and life-cycle emissions, renewable-carbon use	Examines whether emissions are genuinely reduced
Resource efficiency	Energy, solvent, water, catalyst and feedstock requirements	Identifies material and utility burdens
Industrial	Scalability, process integration,	Evaluates translation beyond laboratory

American Journal of Advanced Multidisciplinary Research and Innovation (AJAMRI)

E-ISSN: XXXX-XXXX ● Website: www.ajamri.com ● Email: editor@ajamri.com

Dimension	Representative indicators	Analytical purpose
feasibility	safety and cost	conditions
Decision reliability	Uncertainty, validation, reproducibility and domain coverage	Determines confidence in AI-supported selection

A pathway should not receive a favorable overall assessment solely because it performs strongly on one dimension. Multi-objective selection is required because chemical performance, emissions reduction, cost, and safety may conflict.

4.3 Illustrative Screening Simulation

The simulation compares two hypothetical discovery strategies. The unguided strategy represents candidate selection without systematic use of predictive uncertainty or active-learning prioritization. The AI-guided strategy represents an idealized active-learning workflow that uses existing observations to select subsequent candidates.

A promising candidate within the illustrative model represents a pathway meeting a predefined combination of chemical and sustainability thresholds. The simulation does not assign a real catalyst, reaction, product, or emissions value to any candidate. Its purpose is methodological rather than predictive.

5. Analysis

5.1 AI-Enabled Pathway Discovery Opportunities

AI can contribute at the earliest stage by mapping reaction networks and identifying plausible transformations from alternative feedstocks. Retrosynthetic systems may propose routes that use captured carbon, renewable intermediates, recycled molecules, or biogenic inputs. Chemists can then evaluate whether the proposed transformations are mechanistically credible and compatible with available catalysts and process conditions.

Catalyst screening provides another opportunity. Machine-learning models can use composition, electronic properties, adsorption energies, structural descriptors, and experimental observations to rank materials. High-cost quantum-chemical calculations can then be reserved for candidates with favorable predictions or high information value. This staged process does not eliminate computation; it directs computation toward the most informative regions.

Reaction-condition optimization can examine temperature, pressure, solvent, catalyst loading, residence time, reactant ratio, and electrochemical potential. Bayesian optimization may identify efficient operating regions using fewer experiments than exhaustive factorial designs. For decarbonization, the objective function should incorporate more than product yield. It may include energy demand, selectivity, waste, solvent recovery, catalyst lifetime, and estimated life-cycle emissions.

American Journal of Advanced Multidisciplinary Research and Innovation (AJAMRI)
E-ISSN: XXXX-XXXX ● Website: www.ajamri.com ● Email: editor@ajamri.com

5.2 Simulated Screening Results

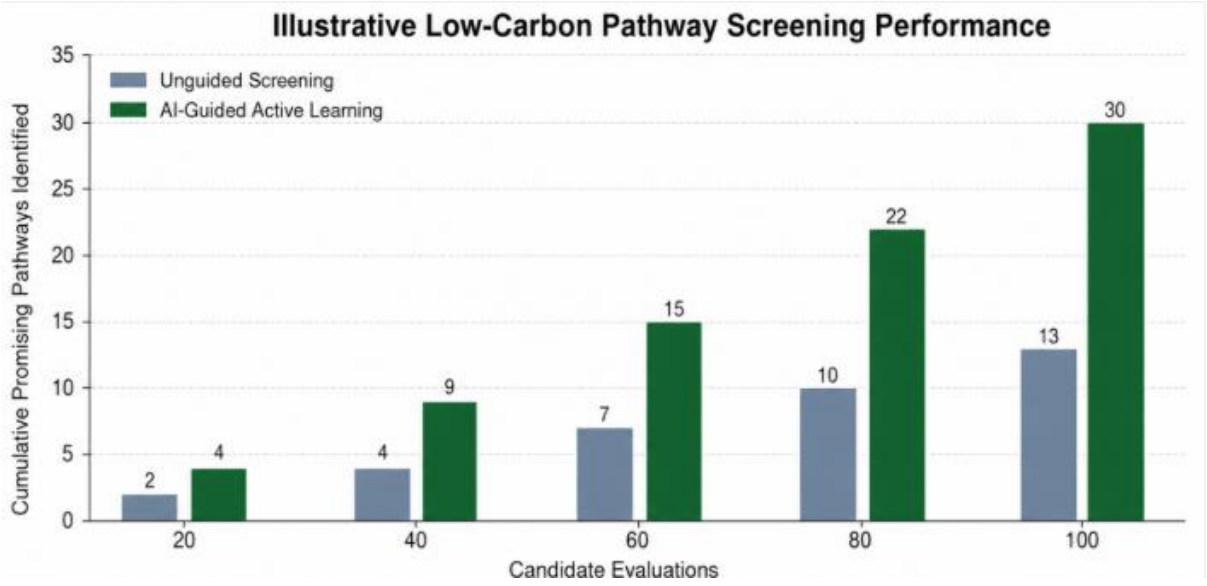
Table 2: Illustrative Comparison of Candidate-Screening Strategies

Candidate evaluations	Unguided screening	AI-guided active learning	Additional candidates identified by AI-guided search
20	2	4	2
40	4	9	5
60	7	15	8
80	10	22	12
100	13	30	17

At the smallest evaluation budget, the AI-guided strategy identifies twice as many promising candidates as unguided screening. The difference expands as additional evaluations are completed. At 100 evaluations, the illustrative AI-guided workflow identifies 30 promising pathways, compared with 13 under unguided screening.

The result reflects the assumed learning mechanism. Each completed evaluation provides information that improves later selection. Unguided screening does not use this information systematically and therefore allocates more evaluations to comparatively unproductive regions. This pattern represents an illustrative expectation rather than a verified performance level.

Graph 1: Illustrative Low-Carbon Pathway Screening Performance



Supporting data: Candidate-evaluation budgets = 20, 40, 60, 80, and 100; unguided screening = 2, 4, 7, 10, and 13; AI-guided active learning = 4, 9, 15, 22, and 30.

Interpretation: The separation between the two screening strategies increases with the evaluation budget. Within the assumptions of the simulation, active learning uses information from earlier evaluations to prioritize candidates with higher predicted value or greater capacity to reduce uncertainty.
Key Finding: AI-guided screening may improve the productivity of chemical-pathway discovery, but its actual effectiveness must be established through prospective comparisons using identical candidate spaces, laboratory resources, performance thresholds, and stopping rules.

5.3 Integration with Process and Life-Cycle Assessment

A reaction-level model may predict yield or activation energy without accounting for the wider production system. Industrial decisions require mass and energy balances, separation modelling, utility demand, heat integration, recycling, catalyst replacement, and waste treatment. These variables can change the relative ranking of pathways.

Process simulation should therefore be connected to AI discovery. When the model proposes a reaction, a preliminary flowsheet can estimate equipment requirements, energy consumption, recycle loops, and product purification. The resulting process information can return to the discovery model as an additional objective or constraint.

Life-cycle assessment provides a wider boundary by including upstream feedstocks, electricity generation, transport, infrastructure, coproduct allocation, product use, and end-of-life treatment. AI can accelerate parts of this analysis, but its results depend on system boundaries and background datasets. Every low-carbon claim should specify which emissions are included and which are excluded.

6. Challenges

6.1 Data Scarcity, Bias, and Chemical Reliability

Chemical data are distributed across journal articles, patents, corporate records, laboratory notebooks, and instrument systems. These sources frequently use inconsistent terminology, measurement conditions, and reporting standards. Failed experiments and negative results are particularly underrepresented.

Dataset bias may cause a model to favor familiar reaction classes, common catalysts, or heavily studied materials. A model can appear accurate when tested on data similar to its training set but perform poorly when applied to novel feedstocks or operating conditions. Discovery requires extrapolation, while many machine-learning systems are strongest at interpolation.

Uncertainty estimation is therefore essential. Researchers should distinguish predictions located within a well-supported chemical domain from those produced far beyond the training data. High-uncertainty proposals may still be valuable, but they require more cautious interpretation and stronger validation.

6.2 Multi-Objective Optimization and Burden Shifting

Low-carbon pathway discovery involves objectives that may conflict. A catalyst may increase conversion while relying on a scarce material. A solvent may improve selectivity while increasing toxicity or recovery energy. Electrification may reduce on-site fuel combustion while increasing upstream emissions if the electricity supply remains carbon intensive.

AI systems must not collapse these trade-offs into a single unexplained score. Weighting decisions should be transparent, sensitivity-tested, and reviewed by chemists, process engineers, environmental specialists, safety professionals, and affected decision-makers.

Burden shifting can also occur geographically. A chemical plant may report lower direct emissions while relying on carbon-intensive imported feedstocks or outsourced waste processing. Life-cycle and supply-chain data are necessary to evaluate whether environmental improvement is genuine.

6.3 Experimental Translation, Safety, and Scale-Up

A prediction is not equivalent to a chemical discovery, and a laboratory result is not equivalent to an industrial pathway. Candidate materials may be difficult to synthesize, unstable under operating conditions, sensitive to impurities, or unsuitable for regeneration. Reaction performance at a small scale may change when heat transfer, mixing, pressure, residence-time distribution, and separation constraints become significant.

AI-guided autonomous laboratories require strong safety controls. Algorithms should not be permitted to explore unrestricted reaction conditions without chemical compatibility checks, pressure limits, hazard screening, emergency controls, and human authorization.

Industrial translation also depends on existing assets. A low-carbon process may require equipment, infrastructure, or feedstocks unavailable at the intended location. Retrofit compatibility and transitional cost should be assessed alongside theoretical environmental performance.

7. Discussion

7.1 Reframing Chemical Discovery as a Systems Problem

The analysis suggests that low-carbon pathway discovery should be understood as a systems problem rather than a narrow search for high-performing reactions. Molecular structures, catalysts, process equipment, energy sources, supply chains, markets, and environmental impacts are interdependent.

AI is especially useful for coordinating information across these levels. A model can screen molecular candidates, another can estimate process conditions, and a third can approximate environmental performance. The outputs should be integrated through a common decision framework rather than optimized independently.

This systems perspective reduces the risk that a pathway is selected because of an isolated laboratory metric. It also encourages early identification of scale-up constraints, resource limitations, and downstream separation burdens.

7.2 Human–AI Collaboration in Sustainable Chemistry

Chemical expertise remains essential because reaction data do not represent every mechanistic or operational constraint. Chemists can identify implausible transformations, recognize unsafe combinations, interpret unexpected behavior, and determine which predictions deserve experimentation.

AI contributes speed, pattern recognition, and structured exploration. It can compare large candidate spaces and identify relationships that may not be obvious through manual inspection. The most productive arrangement is therefore collaborative: the model prioritizes possibilities, while researchers define objectives, review uncertainty, validate mechanisms, and interpret results.

Human oversight is also necessary for sustainability decisions. The relative importance of emissions, toxicity, cost, water consumption, land use, circularity, and social impact cannot be determined by an algorithm alone. These priorities reflect scientific evidence, industrial strategy, regulation, and public values.

7.3 Governance for Credible Industrial Transformation

Organizations adopting AI-guided discovery should maintain traceable datasets, documented model versions, predefined validation procedures, and records of human decisions. Predictions should be reproducible to the extent permitted by data confidentiality and intellectual-property restrictions.

Environmental claims require particular care. Organizations should distinguish predicted reductions from measured performance and laboratory results from commercial-scale outcomes. A pathway should not be marketed as low carbon before its material and energy balances, system boundaries, and life-cycle assumptions have been reviewed.

Collaborative data standards could improve the field by making chemical conditions, failures, measurement uncertainty, and environmental variables more consistently available. Governance should protect legitimate confidential information while supporting scientific verification and cumulative learning.

8. Conclusion

Artificial intelligence can expand the search for low-carbon chemical pathways by improving reaction prediction, catalyst screening, experimental prioritization, process optimization, and integration of diverse chemical information. Its principal contribution is not the automatic replacement of chemical research but the more strategic allocation of computational and experimental resources.

The illustrative simulation demonstrates how an AI-guided active-learning workflow may identify a larger number of promising candidates within a fixed evaluation budget. Because the values are simulated, they do not demonstrate actual industrial performance. Prospective laboratory studies are required to measure screening efficiency under controlled and comparable conditions.

Sustainable industrial transformation requires that AI discovery be connected with process engineering, techno-economic assessment, safety analysis, and life-cycle evaluation. A pathway cannot be considered low carbon solely because it achieves a high reaction yield or operates at a reduced temperature. Emissions and resource burdens must be examined across feedstock production, energy supply, reaction, separation, recycling, transport, use, and end-of-life treatment.

The most defensible development model combines AI prediction, chemical reasoning, uncertainty quantification, controlled experimentation, and transparent environmental assessment. Under these conditions, AI can become a meaningful tool for industrial decarbonization while avoiding exaggerated claims, unsafe experimentation, and the transfer of environmental burdens to less visible parts of the chemical value chain.

References

1. Anastas, P. T., & Warner, J. C. (1998). *Green chemistry: Theory and practice*. Oxford University Press.
2. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. *Nature, 559*(7715), 547–555. https://doi.org/10.1038/s41586-018-0337-2
3. Chen, A., Zhang, X., Chen, L., Yao, S., & Zhou, Z. (2020). A machine learning model on simple features for CO_2 reduction electrocatalysts. *The Journal of Physical Chemistry C, 124*(41), 22471–22478. https://doi.org/10.1021/acs.jpcc.0c05964
4. Coley, C. W., Barzilay, R., Green, W. H., Jaakkola, T. S., & Jensen, K. F. (2017). Convolutional embedding of attributed molecular graphs for physical property prediction. *Journal of Chemical Information and Modeling, 57*(8), 1757–1772. https://doi.org/10.1021/acs.jcim.6b00601
5. Coley, C. W., Green, W. H., & Jensen, K. F. (2018). Machine learning in computer-aided synthesis planning. *Accounts of Chemical Research, 51*(5), 1281–1289. https://doi.org/10.1021/acs.accounts.8b00087
6. Coley, C. W., Thomas, D. A., Lummiss, J. A. M., Jaworski, J. N., Breen, C. P., Schultz, V., Hart, T., Fishman, J. S., Rogers, L., Gao, H., Hicklin, R. W., Plehiers, P. P., Byington, J., Piotti, J. S., Green, W. H., Hart, A. J., Jamison, T. F., & Jensen, K. F. (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science, 365*(6453), eaax1566. https://doi.org/10.1126/science.aax1566
7. Garcia-Martinez, J. (Ed.). (2016). *The chemical element: Chemistry's contribution to our global future*. Wiley-VCH.
8. Guo, Y., Guo, Z., Zhou, J., & Ye, Z. G. (2021). Machine-learning-guided discovery and optimization of additives in preparing Cu catalysts for CO_2 reduction. *Journal of the American Chemical Society, 143*(14), 5515–5522. https://doi.org/10.1021/jacs.1c00339
9. Häse, F., Roch, L. M., & Aspuru-Guzik, A. (2019). Next-generation experimentation with self-driving laboratories. *Trends in Chemistry, 1*(3), 282–291. https://doi.org/10.1016/j.trechm.2019.02.007
10. Janet, J. P., & Kulik, H. J. (2020). *Machine learning in chemistry*. American Chemical Society. https://doi.org/10.1021/acs.infocus.7e4001
11. Jensen, Z., Kim, E., Kwon, S., Gani, T. Z. H., Román-Leshkov, Y., Moliner, M., Corma, A., & Olivetti, E. A. (2019). A machine learning approach to zeolite synthesis enabled by automatic literature data extraction. *ACS Central Science, 5*(5), 892–899. https://doi.org/10.1021/acscentsci.9b00193
12. Mater, A. C., & Coote, M. L. (2019). Deep learning in chemistry. *Journal of Chemical Information and Modeling, 59*(6), 2545–2559. https://doi.org/10.1021/acs.jcim.9b00266
13. Schwaller, P., Gaudin, T., Lányi, D., Bekas, C., & Laino, T. (2018). "Found in translation": Predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. *Chemical Science, 9*(28), 6091–6098. https://doi.org/10.1039/C8SC02339E

14. Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Hunter, C. A., Bekas, C., & Lee, A. A. (2019). Molecular Transformer: A model for uncertainty-calibrated chemical reaction prediction. *ACS Central Science, 5*(9), 1572–1583. https://doi.org/10.1021/acscentsci.9b00576
15. Stocker, S., Csányi, G., Reuter, K., & Margraf, J. T. (2020). Machine learning in chemical reaction space. *Nature Communications, 11*, Article 5505. https://doi.org/10.1038/s41467-020-19267-x
16. Wan, X., Liang, X., & Mar, A. (2021). Machine-learning-accelerated catalytic activity predictions of transition metal phthalocyanine dual-metal-site catalysts for CO_2 reduction. *The Journal of Physical Chemistry Letters, 12*(23), 5594–5600. https://doi.org/10.1021/acs.jpclett.1c01526