International Journal of Physics, Chemistry, Mathematics and Biology
International Open Access, Peer reviewed and Refereed Journal
(Multidisciplinary, Quaterly, Multilanguage)
Volume: 1 Issue: 1 October-December 2025
www.ijpcmb.com ISSN (Online): Applied

Prediction of Chemical Reactions and Product Yields Using Artificial Intelligence

Dr. Ajay Kumar Singh
Assistant professor, Department of Chemistry, Shivpati P. G. College Shoharatgarh, Shiddharthnagar
E-mail ID- ajay.neelam78@gmail.com

Abstract

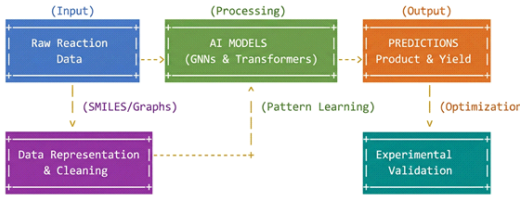
Artificial intelligence (AI) has emerged as a transformative tool for predicting chemical reactions and estimating product yields, offering substantial benefits for sustainable and efficient chemical synthesis. This review synthesizes the theoretical foundations, data-representation schemes, and machine-learning paradigms that enable accurate reaction and yield prediction. Modern AI models—including graph neural networks, template-free architectures, and transformer-based sequence models—learn reactivity rules directly from large reaction datasets, overcoming several limitations of early expert-system and quantum-chemistry-driven approaches. Data quality and curation remain central challenges, as public reaction databases often contain inconsistencies, structural biases, and incomplete yield information. Strategies such as standardized cleaning pipelines, benchmarked data splits, and uncertainty-quantification methods (e.g., ensembles, conformal prediction) improve model reliability and generalization. Recent advances demonstrate promising cross-domain transfer, enabling models trained on one reaction class to perform well on others. Incorporating reaction conditions, catalysts, and multimodal inputs further enhances predictive performance and selectivity estimation. Prospective experimental validation, transparent reporting standards, and open-science frameworks are essential to ensure reproducibility and trust in AI-assisted chemical design. Overall, AI-guided prediction of reaction outcomes and yields provides a powerful route toward accelerated discovery, greener synthesis pathways, and improved decision-making in chemical research and industry.

Keywords: Artificial Intelligence; Chemical Reaction Prediction; Yield Estimation; Graph Neural Networks; Reaction Databases; Transformer Models; Uncertainty Quantification.

Introduction

The chemical and related industries devote considerable effort to developing safer, greener, and more efficient practices. Artificial intelligence (AI) offers a powerful means of extracting value from extensive chemistry datasets to identify and exploit opportunities while addressing safety and environmental concerns. One promising AI application is the prediction of chemical reactions, which can improve the efficiency and accelerate the discovery of novel compounds. Similarly, generating accurate estimates of product yields enables a rapid evaluation of candidate reactions and helps optimize conditions to maximize output. This work surveys the available AI tools for predicting both chemical reactions and associated yields, focusing on the data and models used, the state of validation, and the practical implications for chemists (N. Wei et al., 2016).

Despite extensive literature addressing reaction prediction, many studies concentrate on predicting just the product or the reaction mechanism but do not estimate yield or treat the two problems simultaneously. Consequently, reaction-prediction systems that accomplish both and yield-estimation frameworks that do not depend on explicit reaction specification remain rare. Moreover, the need to delineate chemical structures in input data or generate explicit SMILES strings in output precludes many approaches from being considered template-free. Applying supervised multimodal-learning techniques enabling simultaneous conversion, yield, and selectivity prediction on both molecular-graph and reaction-SMILES representations can help overcome these limitations.



Theoretical Foundations of AI in Chemistry

The development of automated synthetic planning systems necessitates accurate prediction of the outcome of a chemical reaction when only the starting materials are known. Consequently, reaction outcome prediction has gained prominence in machine-learning-driven synthetic planning frameworks. Recently, deep-learning techniques have been applied to predict reaction outcomes for a diverse range of draw- and drop-type transformations, exploiting a mixture of reaction databases that include raw images of chemical structures. AI software is now available that can predict reaction outcomes for a broad spectrum of organic reactions with extensive user support to aid in algorithm selection and justification of predictions. Despite the pioneering works that have applied AI to inorganic and organometallic reactions, research into these transformative technologies is still in its infancy. Permutation-invariant graphical representations can capture fundamental reactivity principles, enabling the prediction of whether a candidate transformation occurs in selected chemical spaces such as reaction classes, elements, organometallics, and inorganic compounds. Training on large organic datasets improves cross-domain transferability, facilitating the search for novel inorganic or organometallic reactions without the need for costly re-engineering. The support of large pre-trained models by the scientific community nurtures rapid advancement (Wang et al., 2022).

The emergence of Reaction Network Explorer (RxNE) marks a further pivotal development. Current software platforms for reaction prediction predominantly utilise deep neural networks trained on curated public datasets, employing techniques such as molecular fingerprints, template matching, and molecular graphs. Opportunities exist to conduct retrosynthetic analyses on reaction networks and catalysis datasets that have yet to receive widespread scrutiny, using non-deep-learning and chemistry-agnostic schemes to express reactants, solvents, and catalysts as multi-modal mixtures. These assets, designed to anticipate complete reactions between organic reagents, provide extensive preliminary exploration in the prediction of reaction pathways (Mercado et al., 2023).

Machine Learning Paradigms for Reaction Prediction: Machine learning provides three paradigms applicable to reaction prediction: supervised, semi-supervised, and reinforcement learning. Supervised approaches leverage paired reactant–product information present in reaction databases, whereas semi-supervised methods utilize unlabelled reactant-only data to further train a previously supervised-reactant-only model. Reinforcement learning operates without existing databases of chemical reactions by formulating the synthesis problem as a sequential decision-making task, generating candidate reaction sequences while exploring

reaction rules through trial-and-error (Meuwly, 2021).

Data Representation and Molecular Descriptors: Data representation constitutes a crucial aspect of the molecular design cycle thatwidely varies according to the information sought and the nature of the target molecules (Guan et al., 2020). Molecular graphs and molecular descriptor approaches are typically employed in the analysis and design of organic molecules without pertaining to any specific molecule that might serve as a prototype. Two classes of molecular data representation form the basis of multiscale designs and analyses. In the first class, fully-connected molecular structures, represented by undirected molecular graphs, divide the molecules into atoms, bonds, and connectivity, disregarding any specific structure relevant to existing molecules. As commonly implemented, the data structure serves as a base for connectivity analysis and virtual screening (N. Wei et al., 2016). Molecular fingerprints and molecular descriptors constitute the second category. Molecular fingerprint approaches convert a chemical structure into a binary string of fixed length rank-based on a predefined criterion, whereas molecular descriptor methods characterize the structure via a set of numerical parameters related to functional/structural properties. The former approach constitutes the basis for structure search and reaction prediction; the latter forms a foundation for quantitative structure–activity relationship work, cycle predict, and yield prediction.

Evaluation Metrics for Predictive Models: For quantitative predictive models, key performance indicators include RMSE, MAE, R2, categorical calibration metrics, and AUROC for classification tasks. An unadulterated and representative train/test split must be established, accompanied by baseline performance assessments across the dataset. RMSE and MAE quantify average absolute deviation, with the former weighting larger residuals more heavily. R2, the coefficient of determination, measures how well model predictions approximate ground truth. For conformally calibrated models, the PiT calibration metric quantifies the ratio of to-be-specified thresholds covering the ground-truth value, facilitating inception of reactivity-guided reaction predictions (Fitzner et al., 2023). AUROC assesses classification prediction quality on binary labels.

Data Sources and Curation

A multitude of public reaction databases have been compiled by academic research groups, providing a wide range of data on different homologous reaction classes (S. Wigh et al., 2024). Reaction databases vary in terms of cover, format, and degree of difficulty for access (Mercado et al., 2023). The importance of a thorough assessment of data in public repositories is emphasized, including discussions of data quality and potential biases. Specific sources of imposed errors are identified and strategies to mitigate these errors through careful curation and extraction are documented. Descriptions of the data-cleaning steps which are executed to filter out unwanted entries and of procedures performed to eliminate duplicates, standardize the nomenclature, or remap missing attributes are also supplied.

Molecular conversion, product yield, and selectivity constitute the three central metrics used by chemists to express the performance of a synthesis plan. Selections of the yield and selectivity targets, along with various other chemical descriptors, enable the evaluation of the realism of synthetic proposals. These definitions clarify the chemical meaning of the trained networks and allow relevant conditions that influence the outcome of a reaction to be integrated into the model. The influence of repressors and enhancers on a target chemical quantity varies sensibly across compounds, multiproduct reactions, and transformations belonging to different areas of chemistry.

Public Reaction Databases: Among the myriad of possible chemical reactions, public reactions databases focus on expected reactions addressing expert users' needs at critical aspects of design (Ghiandoni et al., 2019). Starting from a generalized overview of current solutions dedicated to the prediction of chemical reactions, the sequence is narrowed down to encompass the selected datasets, laid out with the aim of supporting either reaction-type or

reaction-structure-based design. To highlight the community's effort, datasets are classified according to the specific aspects covered, ranging from bulk annotation of reaction type-without structural information-reaction structure at lab level or reaction class stipulating likely solvents or catalysts (N. Wei et al., 2016).

With their large-scale and open-access collection of organic reaction data, public databases constitute critical enablers of automated reaction prediction targets and help formulate standard benchmarks for evaluation. Among these resources, what, where, how, and even why chemically-proven reaction hypotheses occurs, are answered with unique utility, helps devise innovative reaction pathways—retrosynthetic design—and supports further development that anticipates not only reaction targets, but also predicted conversion and/or yield of products. Since, reactions of such nature are typically envisioned under a series of predefined conditions, datasets also exist that enforce strict curation and annotation.

Data Quality, Bias, and Curation Strategies: Chemical reactions in liquid-phase organic chemistry are central to the production of chemicals, materials, and pharmaceuticals. Improving the sustainability of these synthetic routes can have a considerable impact on global emissions, as carbon dioxide is currently released due to the combustion of fossil fuels to power chemical production and manufacturing processes. Machine learning approaches aim to assist chemists by predicting reaction yields based on large databases of experimental data (S. Wigh et al., 2024). However, current public databases of chemical reactions are limited in scope and suffer from quality issues, including systematic biases that have been shown to affect model performance (Mercado et al., 2023).

Data Splitting and Benchmarking Protocols: Data splitting and benchmarking protocols are essential for evaluating machine learning (ML) models aimed toward chemical reaction predictions, retrosynthesis planning and reaction condition optimization (Mercado et al., 2023). To assess the robustness and generalization capabilities of these models, a well-defined set of splits must be established; literature examples highlight the importance of various evaluation criteria, such as prospective validation, cross-domain generalization and incorporation of expert feedback during training (Raghavan et al., 2024).

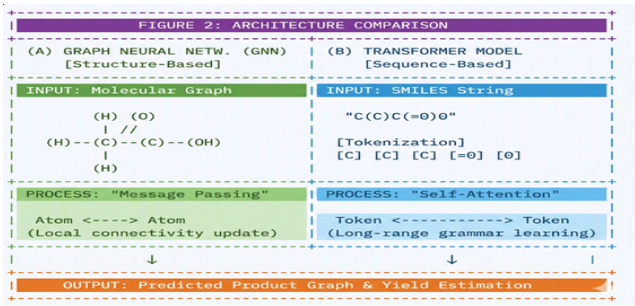
A classification of datasets into training, validation and test sets is fundamental for the performance assessment of ML models. Specifically, only the test set should be used for model evaluation, while the training and validation sets serve to train the model and perform hyperparameter tuning, respectively. In addition, a semi-automated data curation pipeline allows for the selection of datasets that align with different modelling and evaluation tasks complementary to training, testing and validating multiple ML models. Following widely adapted benchmarking methodologies proposed in the literature, the overall data curation process currently achieves the separation of preparations into the larger training set and smaller validation, learning-centered, and final sets of 50000, 10000, 2000 and 1810 unique reactions, respectively.

Model Architectures for Reaction Prediction

The oldest prediction tools were developed in the late 1980s and early 1990s, typified by CAMEO, EROS, and SOPHIA. These used a dedicated core program with hand-crafted rules encoding expert knowledge to suggest possible reaction outcomes (N. Wei et al., 2016). They correctly envisaged products for reactions included in their knowledge base, but struggled with transformations never before encountered. Knowledge-free approaches began to emerge in the 1990s. Quantum-chemistry-based methods leverage first-principles calculations to explore wider sets but are computationally prohibitive for comprehensive screening of compounds and reaction types. Machine learning provides an alternative that requires neither predetermined reaction pathways, rules, nor extensive quantum-mechanical calculations. Models trained on data about past transformations learn to infer the outcome chemistry—

from general reaction types to specific products—based only on features of the initial compounds.

The second category of approaches computes a representation of the whole molecular graph. These are applicable to arbitrarily sized reactants, circumventing the complications involved in 2D-to-3D geometry conversion. State-of-the-art models use graph neural networks that exploit connections between structures, conserve atomic features, and respect structure–property relationships. Such techniques can optimize to minimize a loss function that reflects the preservation of both schematic and atom-level information during message passing, hence enhancing generalization performance on unseen compounds (H. S. Segler & P. Waller, 2016).



Template-Based and Template-Free Approaches: The possible approaches to predicting chemical reactions include template-based and template-free methods. Template-based prediction methods currently dominate by providing the best trade-offs between prediction quality and computational efficiency. A reaction prediction problem can be reformulated as a template-matching task, which can be solved with templates constructed from historical reaction data (Plehiers et al., 2018). The classical methods to extract templates rely on the enumeration of reaction types or reaction centers of molecules. Early approaches and modern graph-based approaches represent reaction templates as simply connected reaction graphs. Template-based approaches capture salient chemical features. Nevertheless, template-based methods cannot serve a reaction that does not match any stored templates in the training set.

Template-free models formulate the prediction task in the form of structure-to-structure transformation. They model a reaction as a graph transformation and can represent less frequent reactions without templates. Template-free approaches do not require preprocessing reaction data to extract scenarios. Graph Neural Networks can model reaction generation directly on undirected molecular graphs and have been applied to reaction-based molecular generation tasks (N. Wei et al., 2016). Sequence-based models deal with chemical structures by treating them as sequences and focus on graphs as intermediate representations. By training a pre-trained model on arbitrarily large sequence datasets from similar domains, the models can achieve superior characteristics with a small quantity of data, and the Transformer architecture is a competitive framework for these tasks.

Neural Networks for Molecular Graphs: Reaction datasets, including ReactDB, Jin et al. and USPTO, are primarily described in SMILES format. SMILES, however, do not provide an intuitive representation of actual reactions. Reaction graphs, by contrast, capture the transformation of reactants into products in a symmetrical representation in which each atom maps onto the graph. Residual-based graph neural networks with uncertainty quantification are therefore a natural choice for reaction yield prediction. A lightweight hypertuning method and a meta-learning framework also ease parameter estimation for different data subsets. Hyperparameter values thus defined maintain good generalization across chemical domains, and model performance can be rigorously evaluated on out-of-distribution datasets covering different reaction types. This composition-based fine-tuning addresses the challenges of domain adaptability and reaction chemistry diversity associated with large pretrained models (Kwon et al., 2022).

The best approaches to prediction of reaction outcomes generally adopt template-free models

that rely on higher-level reaction graph representations for greater generalization. However, the direct conversion of SMILES strings yields wider coverage than template-free alternatives. Graph-based models such as machine attention networks further mitigate the data inefficiency of template-based frameworks and exhibit superior generalization across datasets and prediction tasks (Tavakoli et al., 2022).

Sequence-Based Models and Transformer Architectures: Chemical reactions can be represented as text strings, such as those in SMILES notation, enabling their treatment as natural-language sequences. Sequence-based models can therefore tokenize reactants, reagents, and products into text representations and apply natural-language processing approaches such as attention mechanisms to study chemical problems. Training on datasets of reaction data facilitates unsupervised pretraining on text corpora and domain-specific fine-tuning on chemical information and reaction coordinates. Studies have proposed sequence-based models specifically for reaction prediction, enabled by the availability of large reaction databases, and a molecular transformer model adapts an architecture from natural-language processing to chemical tasks. The molecular transformer predicts transformations, selectivity, and uncertainty, rivaling graph-based techniques (Schwaller et al., 2018). Multitask learning extends the approach to low-resource reaction types such as Baeyer–Villiger, Chan–Lam, and Heck reactions, and shares parameters and latent spaces while conditioning separate heads (Qiao et al., 2022).

Comparison of AI Architectures for Reaction Prediction

Architecture Type	Input Representation	Key Advantage	Limitation
Template-Based	Reaction Rules / SMARTS	High interpretability; chemically valid outputs.	Cannot predict unknown/new reaction types.
Graph Neural Networks (GNN)	Molecular Graphs (Nodes & Edges)	Captures 3D structure & atom mapping well.	computationally expensive for large datasets.
Transformers (Seq2Seq)	SMILES Strings (Text)	Excellent at learning "grammar" of chemistry; highly scalable.	May generate invalid SMILES strings occasionally.
Quantum-ML Hybrid	Quantum Descriptors	High accuracy for transition states.	Very slow; limited to small molecules.

Prediction of Reaction Outcomes and Yields

The task of correctly predicting organic reaction outcomes and yield is a complex problem that normally necessitates a high degree of chemical knowledge and expertise. Several early reaction algorithms attempted a straightforward extension of reaction prediction to include products of a given transformation by modeling the reaction in specified conditions. The CAMEO system represents this approach (N. Wei et al., 2016) , accepting organic reactants as input, using a knowledge base of reagents, solvents and workup procedures, and providing prediction and annotation of the product. EROS followed a different route by deriving reaction conditions from the reactants and employing transformation rules to predict likely products. A great body of knowledge about bond polarization and resonance in organic reactivity was incorporated in the system. By marking reactive sites of molecules with scalar values, sites of electrophilic and nucleophilic preference could be distinguished, giving an obvious initial orientation for chemical transformation. SOPHIA models knowledge about molecular perceptions directly in the graphic representation of chemical structures, accepting a minimum number of organic radicals as input and producing an accurate product structure in full deference to required conditions. This system is acceptable for estimating the final product of transformation only when standard organic chemical reactions take place. All these methods rely upon expert knowledge, expert rules, and databases of known transformations, so they can only predict the known reaction—not entirely unseen reactions with analogous reactants. Quantum chemistry methods can explore the chemical hypothesis space in a much wider region than expert systems, but they always need an enormous computation cost on any medium or large set of reactions. When reaction prediction moves to the next step of providing possible products, machine learning seems to be more appropriate, since no expert knowledge

is necessary for characterization. ReactionPredictor, for example, models the reactive system in terms of electron source and sink locations and pairs them according to a small list of fundamental reaction types; these pairs can be combined into a complete electronic hypothesis. By the complete electronic hypothesis as well as the initial input molecular structure at hand, a neural network provides scores for all possible transformation mechanisms. The framework can, therefore, characterize the molecules exclusively via 2D information without biomolecular descriptors.

Yield estimation in contrast usually demands a rebut cycle. Desired precursors to a molecule can be sought and a directory of transformations built by automation, each with associated diagrams of reactants and products. After this stage a replacement of information within the reaction system is activated. It is noteworthy that any reaction still has its particular conditions for good yield; many targets need such conditions to fulfil. As a consequence at least part of elemental specifications must therefore be maintained in the activity allowing transferability. Sometimes no target molecule is available and the entire dataset in a certain domain becomes of significant scale. In this situation yield estimation becomes a subject of sole interest. Transferability is regaled with modelling prior to yield interpolation at every so often.

Besides representation of other chemical formulations added into reactions, catalyst differing configuration may also constitute a reaction just as in homogeneous catalysis. Whenever a method begins with both reactants and extra specifications focusing conditions as only the party it should be accompanied always covers the specification of catalyst and the approach over the area of reaction focuses solely upon different types of reactions not into specific conditions. It is broadly known that precisely this catalyst tends idiotically fittings flexible adjusting becomes normal at the domain with either cyclic or chain involved. Containing or excluding reaction conditions may thus actively alter interaction effects.

Regression Frameworks for Yield Estimation: As one of the most challenging tasks in machine learning and cheminformatics, yield estimation constitutes a key bottleneck for reaction prediction. Several regression frameworks, as detailed below, have been successfully adapted to this end. These methods leverage shared molecular descriptors between reactant and product species, enabling model calibration without requiring complete knowledge of the specific reaction mechanism or a diverse training dataset. Pursuing a yield prediction formulation permits the separate consideration of both reaction conditioning and chemical identity within the feature set, thereby streamlining the introduction of conditioning variables and interaction effects into the model (Fitzner et al., 2023) ; (Raghavan et al., 2024).

Uncertainty Quantification and Confidence Calibration: Predictive models in chemoinformatics commonly employ a single point estimate for chemical quantities, neglecting uncertainty due to inherent data noise, model approximation, or concept drift. Extrapolation away from the training distribution, a frequent yet often overlooked scenario, compounds this uncertainty. Consequently, quantifying this uncertainty is paramount to prevent overreliance on predictions under such conditions (Gruich et al., 2023). Existing approaches include Bayesian network models and Monte Carlo Dropout (MC-Dropout). More advanced alternatives comprise deep ensembles, variational dropout, conformal prediction, and other strategies.

Incorporating Reaction Conditions and Catalysts: Conditioning reagents or catalyst selection as auxiliary inputs enables the model to assess the influence of reaction conditions on yield. Interaction effects among multiple variables can further enhance prediction power when the data set includes compounds without reaction conditions or catalyst information. Incorporating knowledge from catalytic or synthetic literature, such as rules for metal–ligand or acid–base combinations, may directly improve reaction–yield predictions (Kang et al., 2022).

Validation, Generalization, and Benchmarking

Machine learning (ML) models trained on synthetic data have been shown to be able to predict properties for molecules not present in the training set. Similar approaches can be

used for reactions, such as leveraging molecular fragment similarity to transfer knowledge accumulated on different starting materials. Two chemical collections are considered and complement the original training database by testing whether the model can predict properties for data that are neither part of the training nor validation datasets. The performance is evaluated on the previously unseen transfer set and the quality of the predictions is discussed. The final two datasets are tested on models with no further training after it was completed on the original training dataset and on models that were re-trained to ensure a fair comparison. In the last five databases, candidate transformations introduced different perspectives on the reaction space through the fragmentation of reactants, which added decomposition knowledge to products and increased expected increases from fragments. Models might be able to correctly predict one transformation despite being unable to predict the other. Moreover, reactions can also be mapped to transformations that maintain or add atoms, while the final public database only considers atom-addition reactions without including any removal or modification, observable in one candidate dataset on which the transformations become fission only.

Cross-Domain Generalization: Cross-domain generalization is a desirable property for reactions trained on data from one chemistry domain and applied to another but remains poorly studied. An examination of the reaction space investigates the potential for transferability across different classes. Three datasets are investigated: a set of approximately 26,000 Diels-Alder reactions, a set of over 500,000 amide reactions covering a range of catalysts, and a set of approximately 270,000 hydrogenation reactions, both of which impose strong thermodynamic and kinetic restraints to prevent overfitting the experimental conditions. A transfer study on a Diels-Alder dataset, where the primitive reaction networks are characterized for given reactants takes place. The simple re-routing of reaction networks permits generalization beyond a single reaction class. Amide coupling successfully demonstrates that models trained on one dataset generalize to different classes characterized by other elementary reactions and that knowledge of the intermediate state improves generalization. Using adam with a learning rate of 0.001, the network reaches 60% of its maximum 10 intervals earlier when trained on Diels-Alder reaction mechanisms compared to amide coupling. The estimation of a hydrogenation reaction using the prior model illustrates the capacity to extract kinetic parameters of molecules the model has never encountered. An amide coupling reaction trained on the amide dataset can also produce feasible Diels-Alder alkene candidates when it is used in the reaction network extraction stage. Thus, models gleaned from one reaction class can enhance when applied to another, that candidate molecules mapped as pseudo-intermediate actions learned in the model can also assist generalization (Shi et al., 2024).

Prospective Validation and Experimental Collaboration: Artificial intelligence (AI) can be employed to make predictions about chemical reactions, including reaction outcomes and yield. Because chemical reaction prediction itself is a multi-faceted topic, each of the major components can be treated independently, allowing flexible exploration of their applicability to prediction in machine learning environments. Consequently, predictive overlap can also be established through collaborative processes in which working experimental data from one computer is used to further refine models situated on another experimental machine. Such ongoing dialogue enables another company to produce greater ownership of the work while also benefiting from the large high-quality dataset captured in a challenging environment.

To establish assurance in AI-generated predictions of organic reaction outcomes and product yield, prospective validation using an experimental setup that is kept deliberately distinct from the training data enables evaluation of transferability and model robustness. Owing to increasing collaboration with the chemical community for developing machine-learning models, design activities can also be utilized to provide specifications and directions that grant greater control of AI-level optimization processes. Reagent selection, for instance, represents an element

of reaction design that can be influenced to facilitate regulatory fit. Such wider engagement that promotes prospective assessment of efficiency and generalization subsequently aids adaptation of designs undertaken in yet greater domains while retaining access to other chemical models (Mercado et al., 2023).

Assessment of transferability is further aided by the arrangement of specific monitoring collections aimed at retroactive prospects that stand alone from prevailing design operations (Fitzner et al., 2023). These aspects of machine-learning modeling not only contribute to established understanding of modern systems but also equip experimentalists with unique feedback loops—presenting opportunities for acceleration when amenable data are introduced. Data-donation initiatives are also on offer in various existing forums and at established events to encourage industry input and support AI-guided enhancement in organic synthesis.

Reproducibility and Reporting Standards: Reproducibility and reporting standards are essential in chemical reaction prediction research. To facilitate validation and further progress, the following requirements are advocated: (1) publication of the entire dataset or provision of a permanent link to download it, ensuring adherence to the Química Abierta (Open Chemistry) Policy and obligation to avoid the distribution of copyrighted materials; (2) archiving of the code enabling data preparation, model training, and prediction, in line with the Software Sustainability Institute recommendations; (3) implementation of model cards detailing the trained models, compatible with both the TensorFlow Model Cards and Hugging Face Model Cards formats, and at a minimum covering the tasks performed, training environment specifications, a readme file for installation, a description of dataset preparation and structure, performance metrics on train/test splits, and relevant training and fine-tuning hyperparameters; (4) use of version control for the dataset, code, and model cards, and a dedicated audit trail documenting every modification made and the dates of those changes.

Challenges and Limitations

Artificial intelligence in chemistry must deal with challenges, limitations, and ethical considerations, including data deficiencies, model unreliability, and constrained interpretability concerning safety, security, and responsible AI use. Reaction data remain scarce due to production costs and time constraints; even widely adhered-to standards such as the International Union of Pure and Applied Chemistry (IUPAC) recommendations leave much unreliability and subjectivity unaddressed, and chemical reaction and yield data between domains exhibit differing statistical distributions (Mercado et al., 2023).

Model Reliability: Data is central to any AI application, particularly for reaction prediction and yield estimation. The larger and more diverse the training set, the more reliable the model's generalization. Despite recent increases in available experimental data, reactions are often reported with limited or no yield, hindering progress (Fitzner et al., 2023). Reactions yielding less than 50% are frequently overlooked by synthesis chemists, yet models trained on such limited data cannot adequately represent reactions with yields above this percentage. Furthermore, commonly used datasets can exhibit cyclic biases that originate from the synthetic community's collective preferences or interests. Such biases present significant challenges to accurate prediction, yet they have not been systematically assessed in the literature.

Optimizing chemical reactions typically requires precise control over numerous parameters, yet few studies focus on reaction condition estimation (Jorner et al., 2020). Data representing ideal conditions is rarely available, and models trained on large annotated datasets often neglect critical conditioning variables, jeopardizing their utility in real-world scenarios. Several datasets specify valid reaction candidates or provide extensive high-throughput experimentation results, but conditioning information remains largely absent. The utility and transferability of existing datasets may be further impaired by variations in formats, fragmentation, bias, and the presence of erroneous or misguided information.

Interpretability and Explainability in AI-Driven Predictions: The successful application of chemical reaction prediction and related areas in AI hinges on developing accurate models that generate useful and trustworthy predictions. It is essential, therefore, to increase the transparency and interpretability of such models. In many cases, users want to understand model decisions to identify weaknesses and ensure predictions are reliable. Establishing a clear rationale behind AI-driven predictions is also valuable when reporting to stakeholders without AI expertise, as it can persuade them to act on the model's recommendations. Consequently, trust grows in both the technology itself and the predictions it generates (Mercado et al., 2023).

Explanations may focus on the model's overall operation, at the level of a single prediction, or on the influence of specific input features. Methods that attribute importance to features in interpretable chemical systems, such as molecular graphs, reaction conditions, and catalyst identities, are particularly relevant. A substantial amount of research on explainable artificial intelligence for reactions addresses rationales for predictions rather than the underlying reaction mechanisms (Lamens & Bajorath, 2023). Depending on the design of the predictive model, potential explanatory tools range from feature attribution techniques to counterfactual generation, where adjustments to the input are sought that yield different outcomes.

Safety, Security, and Responsible Use of AI in Chemistry: AI impacts chemistry and chemical engineering by emulating catalytic mechanisms. The reaction prediction module of software packages like REINVENT, GraphINVENT, ASKCOS, IBM RXN, Chemical.AI, and Molecule.one predicts optimal pathways. AI performs both retrosynthetic and forward-synthetic analysis. Retrosynthesis helps find reaction routes from product to reactants while forward synthesis aids mechanistic inference (Wang et al., 2022). Training datasets comprise proprietary chemical reactions from the ChemSpider database. Users lack full awareness of limitations, potentially disregarding side effects, toxic byproducts, gas emissions, and experimental considerations. Programs can mislead with incorrect proposals or mechanisms (He et al., 2023).

Conclusion

Much progress has been made in the past five years concerning the application of artificial intelligence to the prediction of chemical reactions and product yields. Despite the discrepancies between estimated and actual yields, especially for complex chemistries, and despite the data and model constraints that underlie the cross-domain generalization of existing methods, several use cases demonstrate the practical relevance of AI technologies in this field. As possibilities widen, the potential for aiding experimentalists increases commensurately. Innovations in generative AI, synthesising data from multiple modalities, and other generic-learning strategies may lift current specifications set arbitrarily by chemists, while establishing standards for information paper trails will further boost accessibility and reproducibility. Many challenges remain—from data volume, structure, or quality, to model uncertainty or irregularity, security against alteration to psychological safety. Yet forthcoming framework and service developments promise to accelerate accessibility for resource- and expertise-poor users and diffuse innovations more widely across the sector.

References:

1. Wei, N., Duvenaud, D., & Aspuru-Guzik, A. (2016). Neural Networks for the Prediction of Organic Chemistry Reactions. *ACS Central Science*, 2(10), 725–732.
2. Wang, W., Liu, Y., Wang, Z., Hao, G., & Song, B. (2022). The way to AI-controlled synthesis: how far do we need to go?. *Briefings in Bioinformatics*, 23(6), bbac465.
3. Mercado, R., Kearnes, S. M., & Coley, C. W. (2023). Data Sharing in Chemistry: Lessons Learned and a Case for Mandating Structured Reaction Data. *Journal of Chemical Information and Modeling*, 63(13), 3953–3965.
4. Meuwly, M. (2021). Transformative Applications of Machine Learning for Chemical Reactions. *Chemical Reviews*, 121(15), 10218–10239.

5. Guan, Y., Coley, C. W., Wu, H., Ranasinghe, D., Heid, E., Struble, T. J., Pattanaik, L., Green, W. H., & Jensen, K. F. (2020). Regio-selectivity prediction with a machine-learned reaction representation and on-the-fly quantum mechanical descriptors. *Chemical Science*, 12(6), 2198–2208.
6. Fitzner, M., Wuitschik, G., Koller, R., Adam, J. M., & Schindler, T. (2023). Machine Learning C–N Couplings: Obstacles for a General-Purpose Reaction Yield Prediction. *Chemistry – A European Journal*, 29(10), e202203126.
7. Wigh, D. S., Arrowsmith, J., Pomberger, A., Felton, K. C., & Lapkin, A. A. (2024). ORDerly: Data Sets and Benchmarks for Chemical Reaction Data. *Journal of Chemical Information and Modeling*, 64(9), 3585–3597.
8. Ghiandoni, G. M., Bodkin, M. J., Chen, B., Hristozov, D., Wallace, J. E. A., Webster, J., & Gillet, V. (2019). Development and application of a data-driven reaction classification model: comparison of an electronic lab notebook and the medicinal chemistry literature. *Journal of Cheminformatics*, 11, Article number: 59.
9. Raghavan, P., Rago, A. J., Verma, P., Hassan, M. M., Goshu, G. M., Dombrowski, A. W., Pandey, A., Coley, C. W., & Wang, Y. (2024). Incorporating Synthetic Accessibility in Drug Design: Predicting Reaction Yields of Suzuki Cross-Couplings by Leveraging AbbVie's 15-Year Parallel Library Data Set. *Journal of Medicinal Chemistry*, 67(11), 9032–9047.
10. Segler, M. H. S., & Waller, M. P. (2016). Modelling Chemical Reasoning to Predict Reactions. *Chemistry – A European Journal*, 23(25), 5966–5971.
11. Plehiers, P., Marin, G., Stevens, C., & Van Geem, K. (2018). Automated reaction database and reaction network analysis: extraction of reaction templates using cheminformatics. *Journal of Cheminformatics*, 10, Article number: 11.
12. Kwon, Y., Lee, D., Choi, Y. S., & Kang, S. (2022). Uncertainty-aware prediction of chemical reaction yields with graph neural networks. *Journal of Cheminformatics*, 14, Article number: 3.
13. Tavakoli, M., Shmakov, A., Ceccarelli, F., & Baldi, P. (2022). Rxn Hypergraph: a Hypergraph Attention Model for Chemical Reaction Representation. *Machine Learning: Science and Technology*, 3(2), 025006.
14. Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Bekas, C., & Lee, A. (2018). Molecular Transformer - A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Central Science*, 5(9), 1572–1583.
15. Qiao, H., Wu, Y., Zhang, Y., Zhang, C., Wu, X., Wu, Z., Zhao, Q., Wang, X., Li, H., & Duan, H. (2022). Transformer-based multitask learning for reaction prediction under low-resource circumstances. *Briefings in Bioinformatics*, 23(6), bbac433.
16. Gruich, C., Madhavan, V., Wang, Y., & Goldsmith, B. (2023). Clarifying Trust of Materials Property Predictions using Neural Networks with Distribution-Specific Uncertainty Quantification. *Digital Discovery*, 2, 1025-1037.
17. Kang, P. L., Shi, Y. F., Shang, C., & Liu, Z. P. (2022). Artificial intelligence pathway search to resolve catalytic glycerol hydrogenolysis selectivity. *Nature Communications*, 13, Article number: 3991.
18. Shi, R., Yu, G., Huo, X., & Yang, Y. (2024). Prediction of chemical reaction yields with large-scale multi-view pre-training. *Nature Communications*, 15, Article number: 1689.
19. Meyer, B., Sawatlon, B., Corminboeuf, C., Anatole von Lilienfeld, O., & Heinen, S. (2018). Machine learning meets volcano plots: Computational discovery of cross-coupling catalysts. *Chemical Science*, 9(21), 4752–4761.
20. Schilter, O., Pacheco Gutierrez, D., Folkmann, L. M., Castrogiovanni, A., García-Durán, A., Zipoli, F., Roch, L. M., & Laino, T. (2024). Combining Bayesian optimization and automation to simultaneously optimize reaction conditions and routes. *Nature Communications*, 15, Article number: 4333.
21. Lin, K., Xu, Y., Pei, J., & Lai, L. (2020). Automatic retrosynthetic route planning using template-free models. *Chemical Science*, 11(12), 3355–3364.
22. Latendresse, M., Malerich, J. P., Herson, J., Krummenacker, M., Szeto, J., Vu, V. A., Collins, N., & Madrid, P. B. (2023). SynRoute: A Retrosynthetic Planning Software. *Journal of Chemical Information and Modeling*, 63(17), 5433–5442.
23. Jorner, K., Brinck, T., Norrby, P. O., & Buttar, D. (2020). Machine learning meets mechanistic modelling for accurate prediction of experimental activation energies. *Chemical Science*, 12(3), 1163–1175.
24. Lamens, A., & Bajorath, J. (2023). Explaining Multiclass Compound Activity Predictions Using Counterfactuals and Shapley Values. *Journal of Chemical Information and Modeling*, 63(14), 4287–4296.
25. He, J., Feng, W., Min, Y., Yi, J., Tang, K., Li, S., Zhang, J., Chen, K., Zhou, W., Xie, X., Zhang, W., Yu, N., & Zheng, S. (2023). Control Risk for Potential Misuse of Artificial Intelligence in Science. *arXiv preprint arXiv:2312.06632*.