International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1(s) (2026)



ijerst.editor@gmail.com

editor@ijerst.com

www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

Artificial Intelligence in Chemistry: A Comprehensive Review of Methods, Applications, and Future Perspectives

Pallavi B. Mote[1], Mahesh S. Jangale[2], Kiran B. Kharad[3*]

[1,3]Jijamata College of Science and Arts, Dnyaneshwarnagar, Bhende(Bk), Tal.-Newasa, Dist.-Ahilyanagar, Maharashtra, India 414605. Affiliated to SPPU Pune

[2]Mula Education Society's Arts Commerce and Science College, Sonai, Tal. - Newasa, Dist.-Ahilyanagar, Maharashtra, India 414105. Affiliated to SPPU Pune

Corresponding Author Email

kharadkiran@gmail.com

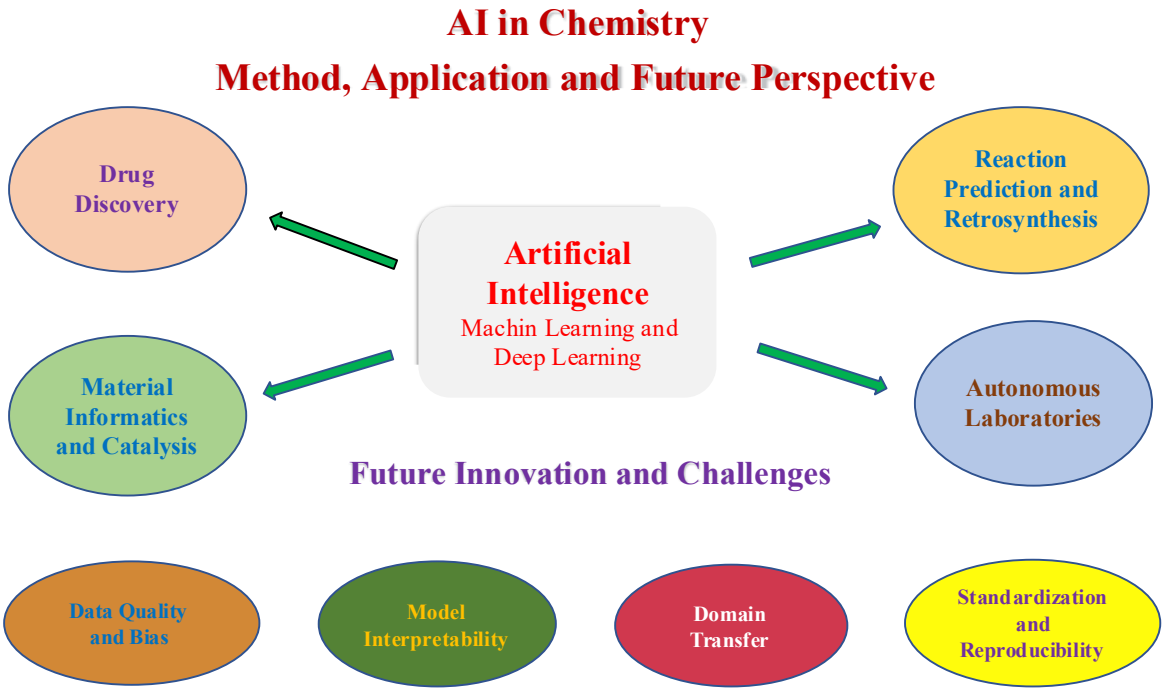
Abstract:

Artificial intelligence (AI) is increasingly reshaping the chemical sciences by supporting data-centric discovery, predictive simulations, and automated laboratory workflows. Contemporary AI approaches such as machine learning, deep neural networks, graph-based models, and natural language processing enable the systematic interpretation of intricate chemical datasets and molecular structures. These technologies have found extensive use in areas including pharmaceutical development, reaction forecasting, computer-assisted retrosynthesis, materials engineering, catalyst design, and spectral data analysis. The combination of AI with high-throughput screening techniques and robotic platforms has accelerated the emergence of autonomous, self-operating laboratories. Nevertheless, important limitations persist, particularly regarding data reliability, transparency of model decisions, reproducibility of results, and the adaptability of models across different chemical domains. Ongoing progress in algorithm development, dataset standardization, and collaborative interaction between human expertise and intelligent systems is anticipated to foster more sustainable, efficient, and innovative advancements in chemical research.

Keywords:

Artificial Intelligence (AI); Machine Learning; Deep Learning; Chemical Informatics; Molecular Modelling; Reaction Prediction; Retrosynthesis; Drug Discovery; Materials Informatics; Autonomous Laboratories; Data-Driven Chemistry; Computational Chemistry.

Graphical Abstract:



Introduction:

Artificial intelligence (AI) has become a powerful driving force in the field of chemical sciences, significantly altering the strategies used to design, evaluate, and refine molecular systems. The exponential growth of chemical information arising from high-throughput experiments, advanced spectroscopic techniques, computational modeling, and published research has created a strong demand for sophisticated computational approaches that can interpret large and complex datasets efficiently and uncover valuable insights.[1] Conventional chemoinformatics methods and quantitative structure–activity relationship (QSAR) models established the foundation for predictive analysis in chemistry. However, the advent of advanced machine learning (ML) and deep learning (DL) approaches has substantially enhanced predictive precision, computational efficiency, and the capacity for large-scale automated modeling.[2] Recent developments in neural network methodologies, particularly graph-oriented frameworks and transformer-based architectures, have

significantly strengthened the ability to represent molecular structures and forecast chemical reactions. Graph neural networks (GNNs) are especially effective because they capture atomic relationships and local electronic characteristics directly from molecular graph structures, eliminating the need for manually engineered descriptors.[3] At the same time, transformer architectures built on attention mechanisms have shown exceptional capability in predicting reaction products and performing retrosynthetic planning, frequently reaching performance levels comparable to those of experienced chemists.[4] These developments illustrate the shift from rule-based systems toward data-driven and self-improving algorithms in chemical research.

Artificial intelligence has been widely implemented in multiple branches of chemistry. Within pharmaceutical research, machine learning techniques support virtual compound screening, novel molecule design, and lead refinement, significantly expediting the discovery of potential drug candidates.[5] In the field of synthetic chemistry, data-centric methodologies assist in optimizing reaction parameters, forecasting product yields, and enabling computer-guided retrosynthetic route design.[6] In a similar manner, materials science has advanced through the use of materials informatics strategies, which enable the prediction of electronic, mechanical, and thermodynamic characteristics, thereby speeding up the development of innovative functional materials.[7] Furthermore, the combination of artificial intelligence with robotic systems and continuous-flow chemistry technologies has enabled the development of self-operating laboratories that perform closed-loop experiments and automatically refine experimental conditions. [8] Although considerable advancements have been achieved, important challenges still persist. Issues related to model transparency, reliability and consistency of data, potential biases, reproducibility of results, and the ability to generalize models across different chemical domains remain major concerns.[9] Overcoming these challenges demands the establishment of uniform benchmarking protocols, clear and transparent reporting standards, and strong interdisciplinary cooperation among chemists, computer scientists, and engineers. This review presents an extensive survey of artificial intelligence techniques applied in chemistry, explores major areas of implementation, and analyzes current advancements along with associated challenges. Through a critical assessment of recent progress, we seek to outline prospective pathways for harmonizing AI with both experimental and theoretical chemistry, ultimately promoting sustainable, efficient, and intelligent innovation in the chemical sciences.

Review:

Artificial intelligence (AI) has swiftly developed into a powerful paradigm in the chemical sciences, significantly redefining the ways molecular systems are interpreted, forecasted, and engineered. The dramatic increase in chemical information originating from high throughput experiments, spectroscopic analyses, computational modeling, and digital scientific publications has exceeded the capabilities of conventional statistical techniques and rule-based methodologies to manage and extract meaningful insights. [10] As a result, machine learning (ML) and deep learning (DL) methods have become highly effective approaches for identifying significant patterns within complex, nonlinear, and high dimensional data structures. [11] This review provides an in-depth examination of fundamental AI techniques in chemistry, highlights their key applications across various subfields, and addresses present limitations along with future prospects.

1.Foundational Methods in Artificial Intelligence for Chemistry

1.1 Machine Learning Paradigms

Machine learning encompasses computational methods that enhance their predictive capability through data-driven learning rather than relying on explicitly programmed rules. Within the chemical sciences, supervised learning continues to be the predominant approach used for modeling and prediction tasks. [12] In supervised learning, algorithms are developed using annotated datasets, enabling them to predict target variables such as physicochemical characteristics, biological responses, or reaction efficiencies. Frequently employed techniques in this context include linear regression, support vector machines (SVM), random forest (RF) models, k-nearest neighbors (kNN), and various gradient boosting approaches.[13] These techniques are widely utilized in quantitative structure–activity relationship (QSAR) and quantitative structure property relationship (QSPR) studies, where molecular structural features are systematically linked to experimentally measured results. [14] Unsupervised learning approaches, such as clustering methods and principal component analysis (PCA), support the identification of hidden patterns, reduction of data dimensionality, and systematic exploration of chemical space.[15] Such approaches allow the detection of underlying patterns within molecular data collections and are especially valuable for assessing chemical diversity and designing compound libraries. In addition, semi-supervised and active learning techniques are progressively being incorporated into experimental pipelines to minimize the amount of labeled data needed while maintaining or enhancing predictive accuracy. [16] Reinforcement learning (RL), an approach in which models improve through repeated

interactions guided by reward-driven feedback, has become increasingly prominent in tasks such as de novo molecule generation and the optimization of chemical reactions.[17] By defining objective functions such as binding affinity or synthetic accessibility, RL frameworks iteratively generate improved chemical structures aligned with predefined targets.

1.2 Deep Learning and Representation learning

Deep learning represents an advanced branch of machine learning that employs multi layered neural network architectures to automatically learn hierarchical and abstract features from data.[18] In contrast to traditional machine learning methods that depend on manually engineered molecular descriptors, deep learning models derive feature representations directly from primary inputs, including molecular graphs and text-based encodings. Convolutional neural networks (CNNs) have been tailored for chemical applications by utilizing grid-like molecular formats and three-dimensional voxel-based structural representations.[19] Recurrent neural networks (RNNs) and long short-term memory (LSTM) architectures are designed to handle sequential data formats, including SMILES-based molecular representations, facilitating the prediction of reaction products as well as various molecular properties.[20] Graph neural networks (GNNs) have emerged as a highly influential advancement in this field. Since molecules can be inherently described as graphs where atoms serve as vertices and chemical bonds act as connecting edges GNNs leverage this structure by enabling information exchange between nodes. Through this message-passing mechanism, they effectively capture both local interactions and overall molecular architecture, which are crucial for reliable property estimation.[21] These network designs often surpass conventional fingerprint-based approaches in various cheminformatics applications because they effectively capture relational patterns and electronic interactions within molecular structures. Meanwhile, transformer-based models initially introduced for natural language processing utilize attention mechanisms to represent long-distance relationships within data sequences.[22] In chemical applications, transformer models frame reaction outcome prediction and retrosynthetic planning as sequence-to-sequence translation problems, extracting underlying chemical rules from extensive reaction datasets. These approaches have achieved performance comparable to that of skilled chemists in forecasting reaction products and proposing feasible synthetic routes.[23]

1.3 Molecular Representation and Chemical data

The effectiveness of AI-driven approaches in chemistry largely depends on how molecular structures are encoded for computational analysis. Frequently used representation formats include SMILES (Simplified Molecular Input Line Entry System), InChI identifiers, molecular fingerprints such as extended-connectivity fingerprints (ECFP), Coulomb matrices, and graph-based embedding techniques.[24] SMILES notation expresses molecular structures as linear text sequences, making them compatible with sequence-oriented neural network models. Molecular fingerprints translate structural motifs into binary or numerical feature vectors that are commonly applied in traditional machine learning methods. In contrast, advanced representation techniques such as embeddings generated through graph neural networks and autoencoders are capable of learning intricate structural patterns and electronic characteristics that extend beyond manually defined descriptors.[21] Large publicly available databases including PubChem, CHEMBL, ZINC, and the Materials Project provide extensive datasets for training and benchmarking AI models.[25] However, challenges such as incomplete annotations, measurement variability, and dataset bias necessitate careful curation and validation before model development.

2.Applications of AI in Chemical Science

2.1 Drug Discovery and Pharmaceutical Chemistry

Pharmaceutical research and development have become one of the most significant areas to benefit from the adoption of artificial intelligence. Conventional drug development processes are lengthy and expensive, frequently requiring more than ten years of investigation and substantial financial investment reaching billions of dollars.[26] Artificial intelligence based virtual screening speeds up the identification of promising early-stage drug candidates by estimating binding strengths and biological activity across extensive chemical libraries.[27] Machine learning algorithms are additionally employed to estimate pharmacokinetic and safety-related properties, encompassing absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiles.[28] Detecting undesirable characteristics at an early phase helps minimize failure rates during advanced stages of development. Furthermore, generative approaches including variational autoencoders (VAEs), generative adversarial networks (GANs), and reinforcement learning based strategies facilitate the creation of novel molecular structures from scratch.[29] Such approaches are capable of designing new chemical entities that are concurrently tailored to satisfy several criteria, including biological potency, target selectivity, and practical synthetic accessibility. Their combination with structure-guided

computational methods and advanced protein structure prediction techniques further strengthens the efficiency of lead refinement and optimization.

2.2 Reaction Prediction and Retrosynthesis

Forecasting the results of chemical reactions and planning efficient synthetic pathways are fundamental tasks in organic chemistry. Traditionally, rule driven expert systems depended on reaction templates that were explicitly crafted and encoded by specialists. [30] Although such template-based approaches performed well for established reaction types, they were limited in their ability to generalize to unfamiliar chemical transformations. In contrast, modern data-driven neural network models can infer reaction products directly from encoded reactant inputs. Sequence-to-sequence architectures built on transformer frameworks have demonstrated strong top-k predictive performance on standard benchmark reaction datasets. [23] Retrosynthetic design platforms combine neural network–based reaction prediction tools with strategic search methods, including Monte Carlo tree search, to systematically and efficiently navigate complex multistep synthesis pathways.[31] Hybrid approaches combining symbolic chemistry knowledge with neural models provide improved reliability and interpretability. These AI-driven systems significantly reduce the time required for route planning and may suggest innovative synthetic strategies beyond human intuition.

2.3 Materials Informatics and Catalysis

The vast combinatorial diversity inherent in material compositions provides a highly suitable domain for artificial intelligence driven investigation. Machine learning approaches are employed to estimate properties such as electronic band gaps, formation energies, mechanical behavior, and thermodynamic stability using composition-based descriptors.[32] When integrated with high-throughput density functional theory (DFT) computations, artificial intelligence significantly speeds up the evaluation and selection of promising materials for applications such as energy storage systems, photovoltaic devices, and catalytic processes. [33] Catalyst development is enhanced through machine learning assisted investigation of material composition and operating conditions. In this context, surrogate modeling techniques emulate computationally intensive quantum mechanical simulations, thereby lowering computational expenses while maintaining reliable predictive performance.[34] Materials informatics frameworks integrate predictive modeling with optimization algorithms, guiding

experimental synthesis toward promising compositional spaces and accelerating the discovery cycle.

2.4 Spectroscopy and Analytical Chemistry

Artificial intelligence has significantly improved the analysis of spectroscopic and analytical measurements. Neural network models trained on extensive spectral databases enable automated structural identification from techniques such as NMR, IR, Raman, and mass spectrometry.[35] Chemometric modeling techniques enhance the extraction of meaningful signals from noisy data, thereby facilitating real-time process supervision and quality assurance in industrial chemical operations. Additionally, image driven machine learning methods interpret microscopic and imaging data used in materials characterization, enabling rapid and precise detection of phase changes and morphological features.

2.5 Autonomous and Self-Driving Laboratories

The convergence of artificial intelligence with robotic technologies and automated synthesis systems has led to the development of self-driving laboratories that can perform iterative, closed-loop optimization without continuous human intervention.[36] Within these platforms, machine learning models recommend optimal experimental settings, robotic units carry out the reactions, analytical tools assess the outcomes, and the resulting data are continuously returned to the algorithm for progressive improvement. Active learning methodologies select experiments that are likely to provide the greatest informational value, thereby minimizing the total number of trials needed. Such autonomous laboratory systems improve consistency, increase experimental throughput, and accelerate discovery processes, marking a significant transition toward smart, automated frameworks in chemical research.

3. Challenges and Limitations

Despite significant progress, several limitations hinder widespread AI adoption in chemistry.

1) Data Quality and Bias:
   The reliability of AI systems strongly depends on the integrity of the data used for training. Inaccurate measurements, non-uniform reporting practices, and imbalanced

datasets can introduce systematic bias, ultimately limiting model robustness and its ability to generalize across diverse chemical scenarios.[37]

2) Interpretability:

Deep learning models frequently function as “black box” systems, making it difficult to extract clear mechanistic understanding or rational explanations for their predictions.[38] Advancing explainable artificial intelligence (XAI) techniques is essential to enhance transparency, strengthen confidence, and promote acceptance among chemists as well as regulatory authorities.

3) Domain Transferability:

Algorithms developed using data from limited chemical spaces often struggle to perform reliably when applied to unfamiliar or broader domains. Approaches such as transfer learning and domain adaptation are being explored to address and reduce this constraint.[39].

4) Standardization and Reproducibility:

The absence of uniform benchmarking frameworks and clear documentation standards makes objective evaluation and comparison of models challenging. Efforts such as Molecule Net have been introduced to tackle these concerns; however, wider implementation and community acceptance are still necessary.[40]

4. Future Perspectives

The progress of artificial intelligence in chemistry will largely rely on better integration of technologies, greater transparency, and a focus on sustainable practices. Approaches based on explainable AI can connect highly accurate predictions with a clearer interpretation of the chemical mechanisms behind them. Furthermore, hybrid approaches such as physics-informed neural networks can embed established chemical principles within the learning models, which can enhance their reliability and enable them to make more accurate predictions even beyond the range of the training data. [41] As quantum computing technology advances, it may work together with machine learning to address complicated electronic structure calculations more effectively. At the same time, federated learning methods could allow different institutions to collaboratively train models while ensuring that sensitive data remains private and secure.[42]

Artificial intelligence can support the development of greener chemical processes by optimizing reactions to reduce waste generation, lower energy usage, and limit the use of

harmful reagents, thereby contributing to sustainable development objectives. The growth of autonomous laboratory systems is also expected to speed up research and discovery, allowing scientists to respond more quickly to global issues such as energy storage, climate change, and emerging health threats. Ultimately, AI is not intended to replace chemists; instead, it functions as a powerful tool that supports human expertise by encouraging innovative thinking, speeding up hypothesis evaluation, and allowing exploration of chemical spaces that were previously difficult to investigate through conventional experimental methods.

Conclusion:

Artificial intelligence has emerged as a powerful tool in chemical sciences, supporting advanced data-driven modeling, predictive analysis, and automated experimental workflows. Techniques such as machine learning and deep learning especially graph neural networks and transformer-based models have greatly enhanced the ability to predict molecular properties, anticipate chemical reactions, and accelerate the discovery of new materials. These technologies are now widely applied in areas including drug development, retrosynthetic planning, catalysis research, spectroscopy analysis, and autonomous laboratory systems, significantly improving research productivity and innovation. However, several challenges still limit widespread adoption, including issues related to data reliability, model interpretability, transferability across datasets, and the lack of standardized frameworks. Continued progress will likely depend on the development of explainable AI methods, physics guided modeling approaches, federated learning strategies, and the integration of emerging technologies such as quantum computing. In this context, AI acts not as a replacement for chemists but as an enabling tool that strengthens scientific expertise and supports sustainable and intelligent progress in modern chemical research.

Acknowledgement:

The authors are thankful to the Principal Dr. Shirish Landage and Head of the Department Dr. Ramesh Naval, Jijamata College of Science and Arts, Bhende (Bk), Tahshil - Newasa, Dist.- Ahilyanagar, Maharashtra, India, for providing infrastructure and research facilities.

References:

[1] Aspuru-Guzik, A.; Lindh, R.; Reiher, M. "The Matter Simulation (R)evolution." *ACS Central Science* **2018**, *4* (2), 144–152**.** https://doi.org/10.1021/acscentsci.7b00550

[2] Alexandre Varnek; Igor Baskin, Machine Learning Methods for Property Prediction in Chemoinformatics: An Overview. *Mol. Inform.* **2012**, *31*, 593–603. https://doi.org/10.1021/ci200409x

[3] Duvenaud, D. K.; Maclaurin, D.; Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. "Convolutional Networks on Graphs for Learning Molecular Fingerprints." *Advances in Neural Information Processing Systems* **2015**, *28*, 2224–2232. ISSN: 10495258

[4] Philippe Schwaller; Teodoro Laino; Thibault Gaudin; et al. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction, *ACS Cent. Sci.* **2019**, *5*, 1572–1583. https://doi.org/10.1021/acscentsci.9b00576

[5] Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery* **2019**, *18* (6), 463–477. https://doi.org/10.1038/s41573-019-0024-5

[6] Connor W. Coley; Marwin H. S. Segler; Klavs F. Jensen, Machine Learning in Computer-Aided Synthesis Planning, Accounts of Chemical Research, **2018**, *51*, 1281–1289. https://doi.org/10.1021/acs.accounts.8b00087

[7] Anubhav Jain et al., Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Materials, 2013, 1, 011002. https://doi.org/10.1063/1.4812323

[8] Leroy Cronin et. al., A universal System for Digitization and Automatic Execution of the Chemical Synthesis Literature Science **2020**, 370, 6512, 2631, 101-108 https://doi.org/10.1126/science.abc2986

[9] Cynthia Rudin, Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, Nature Machine Intelligence, **2019**, 1, 5, 206–215. https://doi.org/10.1038/s42256-019-0048-x

[10] Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A Machine Learning for Molecular and Materials Science, Nature, **2018**, 559, 7715, 547–555. https://doi.org/10.1038/s41586-018-0337-2

[11] von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Exploring Chemical Compound Space with Quantum-Based Machine Learning, Nature Reviews Chemistry, **2020**, 4, 7, 347–358. https://doi.org/10.1038/s41570-020-0189-9

[12] Mitchell, T. M., Machine Learning; McGraw-Hill: New York, **1997**.

[13] Breiman, Leo, Random Forests. Mach. Learn. **2001**, 45, 5–32. https://doi.org/10.1023/A:1010933404324

[14] Cherkasov A, Muratov EN, Fourches D, et. al. QSAR modeling: where have you been? Where are you going to? J Med Chem. 2014 Jun 26;57(12):4977-5010. doi: 10.1021/jm4004285. Epub 2014 Jan 6. PMID: 24351051; PMCID: PMC4074254.

[15] Jolliffe, I. T. Principal Component Analysis; 2nd edition Springer: New York, 2002. **ISBN** 978-0-387-95442-4.

[16] Settles, B., Active Learning Literature Survey, Computer Science Technical Report **2009**.

[17] Olivecrona, M., Blaschke, T., Engkvist, O. et al., Molecular de-novo design through deep reinforcement learning, J Cheminform **2017**, 9, 48, https://doi.org/10.1186/s13321-017-0235-x

[18] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature, **2015**, 521, 436–444. https://doi.org/10.1038/nature14539

[19] Gomes, J.; Ramsundar, B.; Feinberg, E.; Pande, V. Atomic Convolutional Networks for Predicting Protein–Ligand Binding Affinity. *arXiv* **2017**. https://arxiv.org/abs/1703.10603

[20] Segler, M. H. S., Kogej, T., Tyrchan, C., & Waller, M. P., Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Central Science, **2018** 4(1), 120–131. *https://doi.org/10.1021/acscentsci.7b00512*

[21] Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. "Neural Message Passing for Quantum Chemistry." In Proceedings of the 34th International Conference on Machine Learning (ICML), **2017** Volume 70, pages 1263–1272. https://alchemine.github.io/2023/09/12/MPNN.html?utm_source=chatgpt.com

[22] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I**.** "Attention Is All You Need." *Advances in Neural Information Processing Systems* (NeurIPS) **2017** 30**,** 5998–6008. This paper was presented at the *Advances in Neural Information Processing Systems 2017* conference. https://papers.nips.cc/paper/7181-attention-is-all-you-need?utm_source=chatgpt.com

[23] Schwaller, P.; et al. Molecular Transformer for Chemical Reaction Prediction. ACS Central Science **2019**, *5*, 1572–1583. https://doi.org/10.1021/acscentsci.9b00576

[24] Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. Journal of Chemical Information and Modeling. **2010**, *50*, 742–754. https://doi.org/10.1021/acs.jcim.3c00594

[25] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, Evan E Bolton, PubChem in 2021: new data content and improved web interfaces, Nucleic Acids Research, 8 January 2021, Volume 49, Issue D1, Pages D1388–D1395, https://doi.org/10.1093/nar/gkaa971

[26] Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. "How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge." Nature Reviews Drug Discovery **2010**, *9* (3), **203–214**. https://doi.org/10.1038/nrd3078

[27] Lavecchia, A. "Machine-Learning Approaches in Drug Discovery: Methods and Applications." *Drug Discovery Today* **2015**, *20* (3), 318–331**.** https://doi.org/10.1016/j.drudis.2014.10.012

[28] Xu, Y.; Dai, Z.; Chen, F.; Gao, S.; Pei, J.; Lai, L. "Deep Learning for Drug-Induced Liver Injury." *Journal of Chemical Information and Modeling* **2015**, *55* (10), 2085–2093**.** https://doi.org/10.1021/acs.jcim.5b00238

[29] Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. "Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules." *ACS Central Science* **2018**, *4* (2), 268–276**.** https://doi.org/10.1021/acscentsci.7b00572

[30] Corey, E. J.; Wipke, W. T**.** "Computer-Assisted Design of Complex Organic Syntheses." *Science* **1969**, *166* (3902), 178–192. https://doi.org/10.1126/science.166.3902.178

[31] Segler, M. H. S.; Preuss, M.; Waller, M. P. *Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI.* Nature **2018**, *555* (7698), 604–610**.** https://doi.org/10.1038/nature25978

[32] Ward, L.; Wolverton, C. *Atomistic Calculations and Materials Informatics: A Review. Current Opinion in Solid State & Materials Science* **2017,** *21* (3), 167–176**.** https://doi.org/10.1016/j.cossms.2016.07.002

[33] Jain, A.; et al. The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation." *APL Materials* **2013**, 1 (1), 011002. https://doi.org/10.1063/1.4812323

[34] Nørskov, J. K.; et al. Density Functional Theory in Surface Chemistry and Catalysis, *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108* (3), 937–943. https://doi.org/10.1073/pnas.1006652108

[35] Wishart, D. S**.** "NMR Spectroscopy for Metabolomics Research." *TrAC Trends in Analytical Chemistry* **2019**, *115* 129–137**.** https://doi.org/10.1021/jacsau.4c00220

[36] MacLeod, B. P.; et al. Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials." *Science Advances* **2020**, *6* (20), eaaz8867**.** https://doi.org/10.1126/sciadv.aaz8867

[37] Kapoor, S.; Narayanan, *Leakage and the Reproducibility Crisis in ML-based Science.* arXiv preprint *arXiv:2207.07048* (**2022**). https://doi.org/10.48550/arXiv.2207.07048

[38] Rudin, C. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence* **2019**, *1* (5), 206–215**.** https://doi.org/10.1038/s42256-019-0048-x

[39] Pan, S. J.; Yang, Q. "A Survey on Transfer Learning." *IEEE Transactions on Knowledge and Data Engineering* **2010**, *22* (10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

[40] Wu, Z.; et al. Molecule Net: A Benchmark for Molecular Machine Learning. *Chemical Science* **2018**, *9* (2), 513–530 https://doi.org/10.1039/C7SC02664A

[41] Karniadakis, G. E.; et al. Physics-Informed Machine Learning. *Nature Reviews Physics* **2021**, *3* (6), 422–440. https://doi.org/10.1038/s42254-021-00314-5

[42] Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. "Federated Machine Learning: Concept and Applications." ACM Transactions on Intelligent Systems and Technology **2019**, *10* (2), Article 12, 1–19. https://doi.org/10.1145/3298981