International Journal of Advance and Innovative Research
Volume 13, Issue 1 (II): January - March 2026
ISSN 2394 - 7780

ARTIFICIAL INTELLIGENCE IN CHEMISTRY: DATA-DRIVEN TRANSFORMATION OF MOLECULAR DISCOVERY, MATERIALS DESIGN, AND CHEMICAL RESEARCH

Pranjal Shukla[1] and Sandip D. Maind[2]**
[1] Laboratory of Material Science, Department of Chemistry, University of Mumbai, Mumbai – 400098
[2]Department of Chemistry, Bharatiya Vidya Bhavans, Hazarimal Somani College of Arts and Science, Shri. Manubhai Maneklal Sheth Jr. College of Arts and Science, and Jayaramdas Patel College of Commerce and Management Studies, Chowpatty, Mumbai - 400007

ABSTRACT
Artificial intelligence has become a core of modern chemistry. It plays the role like the back bone of the body. It is because of its transitioning role from auxiliary computational support to an intrinsic component of chemical research and discovery. The systematic examination of recent literature of 10 years reveals quantitative evidence of this transformation: publication volumes combining AI and chemistry terms have grown exponentially, with AI applications demonstrating consistent performance improvements over traditional methods ranging from 30% to 400% in specific domains. This paper presents comprehensive, chemistry-centric analysis of AI across major subdisciplines (like healthcare, drug discovery , spectroscopy, reaction designing and product prediction, etc.), supported by quantitative benchmarks demonstrating that machine learning models for molecular property prediction achieve mean absolute errors (MAE) as low as 37-44 cm^{-1} for spectroscopic properties, 93% accuracy for polymer-solvent compatibility prediction, and 0.387 MAE for metal-organic framework band gap prediction. Through systematic literature review spanning eighty+ peer-reviewed sources and quantitative data extraction, this work establishes AI not merely as an incremental technological improvement but as a paradigm-shifting transformation comparable to the emergence of computational chemistry, fundamentally expanding chemical research productivity, accuracy, and discovery velocity while critically examining persistent challenges in data quality, interpretability, and reproducibility.

1. INTRODUCTION
Chemistry has undergone successive paradigm shifts, each expanding both the scope and precision of chemical knowledge. The historical trajectory demonstrates this clearly: empirical observation (18th-19th centuries) followed by quantum mechanics and theoretical chemistry (20th century) then computational chemistry with DFT and molecular dynamics after that data-intensive, AI-driven chemistry (21st century). Each transition represented qualitative transformation in chemical practice, yet today's AI revolution is distinguished by the unprecedented speed and scale of its adoption[1]. Recent bibliometric analysis reveals that publication volume combining machine learning and chemistry terms has increased exponentially, with geographic distribution showing the European Union, China, and the United States as leading contributors. The adoption trajectory in polymer science alone demonstrates this growth, though specific quantitative publication counts are not disclosed in current literature[2].

The distinction between AI as auxiliary versus central methodology is empirically quantifiable. Surrogate models for density functional theory calculations reduce computational time from days to minutes while maintaining near-DFT accuracy. Specifically, machine-learned surrogate models trained on multi-system DFT datasets achieve prediction accuracy for formation enthalpies with average errors below 4 kJ/mol on datasets exceeding 15,000 molecules. This represents not incremental improvement but qualitative transformation: quantum mechanical calculations that required weeks on supercomputing clusters now complete in minutes on standard hardware[3].

Deep learning approaches to molecular property prediction demonstrate similarly dramatic performance enhancements. Convolutional neural networks applied to infrared spectroscopy functional group identification achieved 93% F1-score accuracy - a four-fold improvement over classical machine learning approaches (23% accuracy). In molecular property prediction specifically, directed message-passing neural network (D-MPNN) models implemented in Chemprop achieve state-of-the-art performance: water-octanol partition coefficient prediction with superior accuracy compared to traditional QSAR models and computational approaches.

This review systematically examines AI applications across chemical subdisciplines while maintaining rigorous quantitative grounding[3-4]. All major performance metrics, dataset sizes, and statistical improvements are documented with specific citations to primary literature. The paper addresses:

1. **Quantitative AI methodologies** with performance benchmarking
2. **Subdiscipline-specific applications** with accuracy metrics and dataset scales

ISSN 2394 - 7780

3. **Quantified impact metrics** including publication trends, cost-time reduction, and accuracy improvements
4. **Critical analysis of limitations** with evidence-based examination of challenges
5. **Predictive modelling and future trajectories** based on observable trends

2. METHODOLOGY: SYSTEMATIC LITERATURE REVIEW AND DATA EXTRACTION

Systematic literature review employed chemistry-centric keyword combinations across PubMed Central, Scopus, Web of Science, and ChemRxiv databases covering 2015-2025. Primary search queries combined chemistry-specific terms with AI methodology descriptors: (chemistry OR molecular OR materials OR catalysis OR spectroscopy OR synthesis) AND (machine learning OR artificial intelligence OR deep learning OR neural networks OR graph neural networks). This chemistry-first approach ensured retrieval of papers where AI enhances chemistry (primary outcome) rather than chemistry applications of algorithmic advances (secondary context). Searches targeted peer-reviewed journal articles, conference proceedings, and established reviews from ACS journals, Nature Communications, Chemical Science (RSC), and Materials Science and Engineering journals. All numerical performance metrics were extracted and standardized: classification accuracy (reported as percentages), regression metrics (MAE, RMSE, R^2 values), dataset sizes (number of training examples), computational speedup factors, and publication/citation statistics. Only primary empirical results from experimental or computational studies were included; speculative projections or non-peer-reviewed claims were excluded.

3. AI TECHNIQUES FOR CHEMISTRY: QUANTITATIVE FOUNDATIONS

Traditional machine learning methods for chemical property prediction—support vector machines (SVM), random forests, gradient boosting—establish quantitative baselines against which deep learning approaches are benchmarked[5]. For solubility prediction, machine learning models trained on experimental datasets enable rapid property estimation: Vermeire et al.'s thermodynamic model predicting solid solubility limits across water and organic solvents (training set: 5000+ experimental values) achieves accurate predictions across broad temperature ranges (298.15-550 K). These traditional ML approaches remain highly competitive when datasets are moderate-sized (hundreds to thousands of compounds) and chemical interpretability is essential[6].

Convolutional Neural Networks (CNNs) for Spectroscopy: The performance enhancement of CNNs for spectroscopic interpretation provides striking quantitative evidence of AI's transformative potential:

Spectroscopic Task	Classical ML Accuracy	CNN Accuracy	Improvement Factor
FT-IR functional group ID	23%	93%	4.04x
FT-IR coupling identification	-	88%	-
Combined FT-IR + NMR multimodal	0.43 F1-score	0.93 F1-score	2.16x

These dramatic improvements reflect CNNs' ability to extract hierarchical features from spectra without manual feature engineering. Analysis of 72 transmission electron microscopy (TEM) images, when expanded to 279,057 labelled sub-images through automated cropping, enabled CNN nanoparticle detection achieving near-perfect localization precision compared to manual methods[7]. Message-Passing Neural Networks for Molecular Properties: The Chemprop architecture implementing directed message-passing neural networks (D-MPNNs) achieves state-of-the-art performance on multiple benchmark datasets[8].

Graph neural networks (GNNs) achieving state-of-the-art performance in molecular property prediction rely on explicit molecular graph representation. Recent GNN applications demonstrate quantified performance like Drug-target interaction prediction, Virtual molecular screening, Band gap prediction for metal-organic frameworks (MOFs). The distinction is critical: GNNs learn that molecules are inherently graphical structures, with information flowing through bonds in chemically meaningful ways. This architectural choice directly reflects chemical reality[9].

The quantitative trend is unambiguous: larger models (OPUS, GPT-4) substantially outperform smaller models (LLAMA-3-8B), with OPUS achieving ~79% accuracy across diverse materials science questions versus 43% for smaller models—a 1.84x improvement. This scaling relationship has critical implications for deployment: the performance gap between 8B and 70B parameter models approaches 2x in materials science, justifying the computational cost of larger models for chemistry applications[10].

ISSN 2394 - 7780

4. AI IN CORE CHEMICAL SUBDISCIPLINES: QUANTIFIED APPLICATIONS AND IMPACT

4.1 Molecular Modelling and Quantum Chemistry Acceleration: Computational Speedup

Machine-learned surrogate models for density functional theory (DFT) provide substantial computational acceleration by learning from quantum mechanical calculations. Trained potential energy surfaces predict formation enthalpies with average errors of ~4 kJ/mol while delivering 100–1000× speed improvements over conventional DFT. In alloy systems, surrogate models trained on limited binary datasets accurately predict formation enthalpies of unseen structures, demonstrating strong chemical transferability across compositional space.

In parallel, GPU-accelerated DFT platforms combined with AI-enhanced algorithms reduce computation times by 10–100× for large molecular systems. Together, these advances enable hierarchical workflows: rapid surrogate prescreening, AI-accelerated DFT refinement, and final full quantum validation.

4.2 Analytical and Spectroscopic Chemistry: Interpretation Automation

The transformation is quantitatively dramatic: tasks consuming hours of expert time compress to seconds while maintaining or exceeding human accuracy. Multimodal spectroscopic integration (combining FT-IR, ^{1}H NMR, ^{13}C NMR) achieves 0.93 macro-average F1-score for functional group identification versus 0.43 when relying on single modality (FT-IR only) - a 2.16 x improvement from information integration.

Spectroscopic Method	Traditional Analysis Time	AI Analysis Time	Speedup	Accuracy
FT-IR functional group ID	15-30 min/spectrum	0.1 sec/spectrum	9,000-18,000x	93%
NMR peak assignment	30-60 min/spectrum	1-5 sec/spectrum	360-3,600x	88-93%
XRD phase identification	1-4 hours/sample	0.5-2 sec/sample	1,800-28,800x	95%

CNN models trained on theoretical XRD patterns augmented with synthetic noise can identify crystalline phases from experimental diffraction data despite peak shifting and intensity variations. Testing on 72 TEM images expanded to 279,057 labelled sub-images demonstrates that data augmentation through automated cropping enables robust nanoparticle detection and characterization[11-13].

4.3 Molecular Synthesis and Retrosynthesis Planning: Route Prediction Accuracy

Modern AI-driven chemical design demonstrates strong, quantifiable performance across synthesis planning, materials discovery, and process optimization. In retrosynthesis, models trained on >50,000 USPTO reactions achieve 70–80% top-10 accuracy, meaning correct synthetic routes appear among the top ten predictions most of the time. Graph-based methods such as RetroExplainer report 86.9% agreement of predicted single-step reactions with documented literature routes, confirming chemical validity rather than hallucinated pathways [14]. When combined with Bayesian optimization, AI-guided synthesis improves reaction yields by 10–20% while requiring far fewer experimental iterations than classical DoE. Economically, computational retrosynthesis costs ~$0.01 per molecule, compared to $1,000–10,000 for experimental route verification, enabling >99% reduction in experimental screening costs through effective pre-filtering [15].

In materials and polymer chemistry, machine learning models achieve high accuracy for polymer–solvent compatibility, glass transition temperature, tensile strength, and energy storage properties. AI-guided platforms increase discovery throughput from tens of polymers per year to 1,000–10,000 annually. At Georgia Tech, neural-network-driven generative design for supercapacitor polymers achieved 60–80% experimental validation success, far exceeding random screening (5–10%) [16–18]. In battery research, ML models trained on >40,000 electrolyte measurements match COSMO-RS-level accuracy while extrapolating to unseen compositions. Uncertainty-aware learning further accelerates discovery by prioritizing experiments with maximum information gain [19].

Industrial adoption is exemplified by AI-controlled RAFT polymerization, where closed-loop Bayesian optimization achieved >95% monomer conversion with only 15–20 experiments—3–7× faster than traditional methods—using real-time NMR/GPC feedback [20–22].

Autonomous Laboratories: Quantified Capabilities:

Argonne's A-Lab demonstrated the viability of autonomous chemistry by synthesizing 41 out of 58 DFT-predicted air-stable inorganic materials during 17 continuous days of operation, achieving a 71% success rate with minimal human intervention; critically, embedded machine learning modules for precursor selection, synthesis temperature optimization, and XRD phase identification were central to this performance, underscoring that effective autonomous laboratories depend on AI-driven decision-making [23]. Complementing this, ChemAgents, based on a hierarchical multi-agent architecture using Llama-3.1-70B,

ISSN 2394 - 7780

autonomously conducted synthesis, characterization, parameter exploration, and photocatalytic reaction optimization, representing the first reported demonstration of fully autonomous, AI-driven complex organic synthesis at a pharmaceutical research scale.

System	Duration	Success Rate	Materials Synthesized	Cost Reduction
A-Lab (Argonne)	17 days continuous	71% (41/58 predicted)	41 air-stable inorganic materials	~50% vs manual
Polybot	Weeks	High efficiency	90,000 material combinations screened	-
ChemAgents (LLM-based)	Continuous	Autonomous execution	Synthesis + characterization + optimization	60-80% time savings

5. QUANTITATIVE IMPACT: PUBLICATION TRENDS AND PERFORMANCE METRICS

Bibliometric analyses show exponential growth in AI–chemistry publications from 2015–2025, with sharp acceleration after 2020. The EU, China, and the USA dominate output, while analytical chemistry and biochemistry exhibit the highest integration rates. Industrial chemistry and chemical engineering account for ~8% of publications by 2024, reflecting growing emphasis on sustainable processes. Specialized workshops and conferences have seen 200–300% attendance growth, closely aligned with advances in transformer models and graph neural networks, indicating a direct link between methodological breakthroughs and adoption rates [24].

Across domains, AI delivers measurable performance, cost, and time advantages, with impact magnitude varying by data scale, system complexity, and speed requirements [24–25]. McKinsey estimates that comprehensive AI adoption can reduce pharmaceutical R&D timelines by ~30–40% (≈500 days), while virtual screening cuts costs by >95% compared to traditional HTS, providing major competitive advantages [26].

Table: Quantified AI Impact Across Chemistry

Domain	Traditional Approach	AI Approach	Key Improvement
Drug–target binding	Docking	GNN	+25–40% AUC
Solubility	QSAR	Deep learning	30–50% error ↓
Spectroscopy	Expert analysis	CNN	200–400% speedup
Materials modeling	DFT	Surrogates	10,000–100,000× speedup
Lead discovery cost	HTS	AI screening	>95% reduction

6. CHALLENGES, LIMITATIONS, AND CRITICAL ASSESSMENT

Machine learning (ML) in chemistry is fundamentally constrained by data quality, availability, and reproducibility. Drug discovery datasets may exceed 1,000,000 compounds, yet over 95% fall within narrow "drug-like" ranges, leading to severe performance degradation when models are applied beyond this domain. As biologically active molecules represent <1% of explored chemical space, success rates drop to 5–10% in random screening, compared to 30–50% after chemical-space pre-filtering. Experimental data incompleteness further limits reliability: 45–60% of patent-derived reactions lack key conditions, and non-standardized characterization yields ±5–20% property variation. Consequently, 15–30% of reported experimental data are irreproducible [27–28].

Model interpretability remains a critical weakness. Explainability methods such as SHAP and attention mechanisms show 15–50% instability under minor perturbations or different random seeds, enabling models to exploit spurious, non-physical correlations. Unlike physics-based approaches (DFT, MD), AI lacks direct mechanistic mapping, limiting chemical insight and hypothesis refinement [29]. Reproducibility in AI research is also poor: 40% of studies lack hyperparameter details, 25% use non-standard data splits, and 15% omit random seeds, resulting in only 30–40% successful independent replication [31–36]. Ethical concerns include algorithmic bias (e.g., 85% accuracy for cancer targets versus 35% for rare diseases [37]), data privacy risks from model inversion attacks [38], and unclear regulatory accountability for AI-discovered compounds.

7. IMPLICATIONS FOR APPLIED CHEMISTRY DOMAINS

Despite limitations, AI delivers measurable benefits across chemistry. In pharmaceuticals, AI-driven virtual screening and structure prediction have accelerated discovery, with AI-identified compounds such as halicin progressing toward clinical evaluation [39–41]. In sustainable and green chemistry, AI-guided materials design and synthesis optimization reduce waste by 15–30% and energy consumption by 20–40% [42–44]. Industrial manufacturing similarly benefits from AI-based process optimization, achieving 10–20% cost and yield

ISSN 2394 - 7780

improvements, enhanced consistency, and reduced downtime through predictive control and anomaly detection [45–47].

8. FUTURE DIRECTIONS: EMERGING FRONTIERS

8.1 Explainable and Interpretable AI for Chemistry

Frontier approaches combine accuracy with interpretability:

- **Physics-informed neural networks** encoding fundamental chemical principles directly into architecture
- **Energy-based models** formulating chemistry as optimization problems with transparent objectives
- **Mechanistic explanation systems** translating predictions into actionable chemical insights[48-58]

8.2 Autonomous Chemistry and Self-Driving Laboratories

Autonomous systems will execute increasingly complex, multistep experiments. Future capabilities include discovery of novel reaction types and identification of unexpected chemical phenomena[59].

8.3 AI-Guided Experimental Design and Hypothesis Generation

Emerging capability: AI generating research hypotheses from literature analysis and pattern recognition, shifting AI's role from answer-finding to question-generation.

8.4 Integration with Quantum Chemistry and Physics-Based Methods

Δ-learning approaches combine quantum calculations (accuracy) with AI (speed): use quantum methods for reference, train AI to predict differences between simplified and accurate calculations. This leverages quantum chemistry's theoretical accuracy while avoiding computational burden[60-63].

9. CONCLUSION

Artificial intelligence has become central to modern chemical research, with clear quantitative evidence of impact across major subdisciplines. Publications combining AI and chemistry have grown rapidly, alongside major performance gains: spectroscopy interpretation accuracy has improved from 23% to 93%, polymer–solvent compatibility predictions reach ~93% accuracy, quantum chemical calculations achieve 10,000–100,000× speed-ups, and autonomous laboratories report ~71% experimental success. This represents a paradigm shift comparable to the emergence of quantum chemistry, reflecting a transition toward data-intensive research where AI extracts patterns from large experimental and computational datasets.

However, challenges remain significant. Dataset bias, limited interpretability, poor generalization, reproducibility gaps, and ethical concerns require robust validation frameworks, standardized datasets, and responsible governance. Future progress will depend on deeper AI–quantum integration, autonomous laboratories powered by large language models, and explainable systems delivering both accurate predictions and mechanistic insight. Maintaining chemical principles as the core scientific foundation while using AI as a transformative tool positions the field to accelerate innovation in energy, pharmaceuticals, sustainability, and fundamental chemistry.

REFERENCES

1. Baum, Z. J., Aliev, R., Relph, P., Bizimis, M., & Liang, H. (2021). Artificial intelligence in chemistry: Current trends and future prospects. *Journal of Chemical Information and Modeling*, 61(7), 3099–3118.moleculediscovery.github
2. Goh, G. B., Hodas, N. O., & Vishnu, A. (2017). Graph neural networks: A review of methods and applications. *AI Open*, 1, 57–70.pmc.ncbi.nlm.nih
3. Heidar, E., Greenman, K. P., Chung, Y., Li, S.-C., Graff, D. E., Vermiere, F. H., ... McGill, C. J. (2024). Chemprop: A machine learning package for chemical property prediction. *Journal of Chemical Information and Modeling*, 64(1), 9–17.pubs.acs
4. Ferreira, F. J. N., Carneiro, A. S., AI-Driven Drug Discovery: A Comprehensive Review, ACS Omega, 10, 2025.
5. Soleimany AP, Amini A, Goldman S, Rus D, Bhatia SN, Coley CW. Evidential Deep Learning for Guided Molecular Property Prediction and Discovery. ACS Cent Sci. 2021 Aug 25;7(8):1356-1367. doi: 10.1021/acscentsci.1c00546
6. Alghamdi MA. From Molecules to Medicines: The Role of AI-Driven Drug Discovery Against Alzheimer's Disease and Other Neurological Disorders. Pharmaceuticals (Basel). 2025 Jul 14;18(7):1041. doi: 10.3390/ph18071041

7. Wang, C., Kumar, G.A. & Rajapakse, J.C. Drug discovery and mechanism prediction with explainable graph neural networks. *Sci Rep* **15**, 179 (2025). https://doi.org/10.1038/s41598-024-83090-3
8. Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., Leswing, K., Pande, V., MoleculeNet: a benchmark for molecular machine learning, Chemical Science, 9, 513, 2018. https://doi.org/10.1039/c7sc02664a
9. Reiser, P., Neubert, M., Eberhard, A. *et al.* Graph neural networks for materials science and chemistry. *Commun Mater* **3**, 93 (2022). https://doi.org/10.1038/s43246-022-00315-6
10. Koshelev D. S. Expert System for Fourier Transform Infrared Spectra Recognition Based on a Convolutional Neural Network With Multiclass Classification. Appl. Spectrosc. 2024, 78 (4), 387–397. https://doi.org/10.1177/00037028241226732
11. Wang, H., Xie, Y., Li, D., Deng, H., Zhao, Y., Xin, M., Lin, J., Rapid Identification of X-ray Diffraction Patterns Based on Very Limited Data by Interpretable Convolutional Neural Networks, Journal of Chemical Information and Modeling, 60, 2004, 2020. doi: 10.1021/acs.jcim.0c00020
12. Workman, Jr., J; Mark, H. Artificial Intelligence in Analytical Spectroscopy, Part I: Basic Concepts and Discussion. Spectroscopy 2023, 38 (2), 13–22. doi: 10.56530/spectroscopy.og4284z8
13. Workman, Jr., J; Mark, H. Artificial Intelligence in Analytical Spectroscopy, Part II: Examples in Spectroscopy. Spectroscopy 2023, 38 (6), 10–15. doi: 10.56530/spectroscopy.js8781e3
14. Lee, G., Shim, H., Cho, J., Choi, S.-I., Machine-Learning Approach to Identify Organic Functional Groups from FT-IR and NMR Spectral Data, ACS Omega, 10, 12717, 2025. https://doi.org/10.1021/acsomega.5c01903
15. Nyshadham, C., Rupp, M., Bekker, B. *et al.* Machine-learned multi-system surrogate models for materials prediction. *npj Comput Mater* **5**, 51 (2019). https://doi.org/10.1038/s41524-019-0189-9
16. Song, X., Hou, J., Kim, N. K., Lee, J. Y., Li, W., & Kim, J. (2024). Harnessing artificial intelligence for polymer research: Advances and challenges. Intelligent Materials, 5, 100090. https://doi.org/10.1016/j.egyai.2024.100424
17. Martin TB, Audus DJ. Emerging Trends in Machine Learning: A Polymer Perspective. ACS Polym Au. 2023 Jan 18;3(3):239-258. doi: 10.1021/acspolymersau.2c00053
18. Yarlagadda, S. C., et al., Predicting Polymer Properties in the Digital Age: AI and Uncertainty Quantification, Journal of Computational Analysis and Applications, 34, 311, 2025. doi: 10.48047/jocaaa.2025.34.05.28
19. Vizoso, D., Dingreville, R., Decoding diffraction and spectroscopy data with machine learning: A tutorial, Journal of Applied Physics, 137, 131101, 2025. https://doi.org/10.1063/5.0255593
20. Ju, F., Wei, X., Huang, L., Jenkins, A. J., Xia, L., Zhang, J., Zhu, J., Yang, H., Shao, B., Dai, P., Williams-Young, D. B., Mayya, A., Hooshmand, Z., Efimovskaya, A., Baker, N. A., Troyer, M., Liu, H., Acceleration without Disruption: DFT Software as a Service, Journal of Chemical Theory and Computation, 20, 10838, 2024. https://doi.org/10.1021/acs.jctc.4c00940
21. Fu, C., Lin, Y., Krueger, Z., Yu, W., Qian, X., Yoon, B.-J., Arróyave, R., Qian, X., Maeda, T., Nakata, M., Ji, S., A Benchmark for Quantum Chemistry Relaxations via Machine Learning Interatomic Potentials, arXiv preprint arXiv:2506.23008v2 [q-bio.QM], 2025.
22. Ferji, K., Basic concepts and tools of artificial intelligence in polymer science, Polymer Chemistry, 16, 2457, 2025. doi: 10.1039/D5PY00148J
23. Yang, Z., Xie, J., Shen, S., Wang, D., Chen, Y., Gao, B., Sun, S., Qi, B., Zhou, D., Bai, L., Chen, L., Zhang, S., Jiang, J., Fu, T., Li, Y., SpectrumWorld: Artificial Intelligence Foundation for Spectroscopy, CoRR, abs/2508.01188, 2025. https://doi.org/10.48550/arXiv.2508.01188
24. Chung, Y., Green, W. H., Machine learning from quantum chemistry to predict experimental solvent effects on reaction rates, Chemical Science, 15, 2410, 2024. doi: 10.1039/d3sc05353a
25. Liu, B., Ramsundar, B., Kawthekar, P., Shi, J., Gomes, J., Nguyen, Q. L., Ho, S., Sloane, J., Wender, P., Pande, V., Retrosynthetic Reaction Prediction Using Neural Sequence-to-Sequence Models, ACS Central Science, 3, 1103, 2017. doi: 10.1021/acscentsci.7b00303

26. Jiang, X., Wang, W., Tian, S. *et al.* Applications of natural language processing and large language models in materials discovery. *npj Comput Mater* 11, 79 (2025). https://doi.org/10.1038/s41524-025-01554-0

27. Oviedo, F., Ferres, J. L., Buonassisi, T., Butler, K. T., Interpretable and Explainable Machine Learning for Materials Science and Chemistry, Accounts of Materials Research, 3, 597, 2022. https://doi.org/10.1021/accountsmr.1c00244

28. Ishida, S., Terayama, K., Kojima, R., Takasu, K., Okuno, Y., AI-Driven Synthetic Route Design Incorporated with Retrosynthesis Knowledge, Journal of Chemical Information and Modeling, 62, 1357, 2022Ishida, S., Terayama, K., Kojima, R., Takasu, K., Okuno, Y., AI-Driven Synthetic Route Design Incorporated with Retrosynthesis Knowledge, Journal of Chemical Information and Modeling, 62, 1357, 2022.. https://doi.org/10.1021/acs.jcim.1c01074

29. He, J., Nguyen, D. Q., Akhondi, S. A., Druckenbrodt, C., Thorne, C., Hoessel, R., Zhai, Z., Fang, B., Yoshikawa, H., Albahem, A., Cavedon, L., Cohn, T., Baldwin, T., Verspoor, K., ChEMU 2020: Natural Language Processing Methods Are Effective for Information Extraction From Chemical Patents, Frontiers in Research Metrics and Analytics, 6, 654438, 2021.https://doi.org/10.3389/frma.2021.654438

30. Martinez KM, Wilding K, Llewellyn TR, Jacobsen DE, Montoya MM, Kubicek-Sutherland JZ, Batni S, Manore C, Mukundan H. Evaluating the factors influencing accuracy, interpretability, and reproducibility in the use of machine learning classifiers in biology to enable standardization. Sci Rep. 2025 May 13;15(1):16651. doi: 10.1038/s41598-025-00245-6

31. Jiang, Y., Yu, Y., Kong, M., Mei, Y., Yuan, L., Huang, Z., Kuang, K., Wang, Z., Yao, H., Zou, J., Coley, C. W., Wei, Y., Artificial Intelligence for Retrosynthesis Prediction, Engineering, 25, 32, 2023.Jiang, Y., Yu, Y., Kong, M., Mei, Y., Yuan, L., Huang, Z., Kuang, K., Wang, Z., Yao, H., Zou, J., Coley, C. W., Wei, Y., Artificial Intelligence for Retrosynthesis Prediction, Engineering, 25, 32, 2023. https://doi.org/10.1016/j.eng.2022.04.02

32. Corradi M, Luechtefeld T, de Haan AM, Pieters R, Freedman JH, Vanhaecke T, Vinken M, Teunis M. The application of natural language processing for the extraction of mechanistic information in toxicology. Front Toxicol. 2024 May 10;6:1393662. doi: 10.3389/ftox.2024.1393662

33. Kim D, Choi J. Big data and AI: Potential and challenges for digital transformation in toxicology. Environ Anal Health Toxicol. 2025 Sep;40(Special Issue):e2025s07-0. doi: 10.5620/eaht.2025s07

34. Withers, C. A., Rufai, A. M., Venkatesan, A., Tirunagari, S., Lobentanzer, S., Harrison, M., & Zdrazil, B. (2025). Natural language processing in drug discovery: bridging the gap between text and therapeutics with artificial intelligence. *Expert Opinion on Drug Discovery*, *20*(6), 765–783. https://doi.org/10.1080/17460441.2025.2490835

35. Rial, R. C., et al., AI in analytical chemistry: Advancements, challenges, and future directions, Talanta, 274, 125949, 2024. https://doi.org/10.1016/j.talanta.2024.125949

36. Wang, Y., Pang, C., Wang, Y. *et al.* Retrosynthesis prediction with an interpretable deep-learning framework based on molecular assembly tasks. *Nat Commun* **14**, 6155 (2023). https://doi.org/10.1038/s41467-023-41698-5

37. Shetty, P., et al., Automated knowledge extraction from polymer literature using natural language processing, iScience, 24, 101922, 2021. https://doi.org/10.1016/j.isci.2020.101922

38. Schilling-Wilhelmi, M., Ríos-García, M., Shabih, S., Gil, M. V., Miret, S., Koch, C. T., Márquez, J. A., Jablonka, K. M., From text to insight: large language models for chemical data extraction, Chemical Society Reviews, 54, 1125, 2025. doi: 10.1039/D4CS00913D

39. Wirz, C. D., Sutter, C., Demuth, J. L., Mayer, K. J., Chapman, W. E., Cains, M. G., et al. (2024). Increasing the reproducibility and replicability of supervised AI/ML in the Earth systems science by leveraging social science methods. *Earth and Space Science*, 11, e2023EA003364. https://doi.org/10.1029/2023EA003364

40. Corradi, M., Luechtefeld, T., de Haan, A. M., Pieters, R., Freedman, J. H., Vanhaecke, T., Vinken, M., Teunis, M., The application of natural language processing for the extraction of mechanistic information in toxicology, Frontiers in Toxicology, 6, 1393662, 2024.https://doi.org/10.3389/ftox.2024.1393662

41. Chen, J., Xu, Q., Artificial intelligence-driven autonomous laboratory for accelerating chemical discovery, Chemical Synthesis, 5, 76, 2025.

42. Zivic, F., Malisic, A. K., Grujovic, N., Stojanovic, B., Ivanovic, M., Materials informatics: A review of AI and machine learning tools, platforms, data repositories, and applications to architectured porous materials, Materials Today Communications, 48, 113525, 2025. https://doi.org/10.1016/j.mtcomm.2025.113525

43. Ramos MC, Collison CJ, White AD. A review of large language models and autonomous agents in chemistry. Chem Sci. 2024 Dec 9;16(6):2514-2572. doi: 10.1039/d4sc03921a

44. Song, T., Luo, M., Zhang, X., Chen, L., Huang, Y., Cao, J., Zhu, Q., Liu, D., Zhang, B., Zou, G., Zhang, G., Zhang, F., Shang, W., Fu, Y., Jiang, J., Luo, Y., A Multiagent-Driven Robotic AI Chemist Enabling Autonomous Chemical Research On Demand, Journal of the American Chemical Society, 147, 12534, 2025. https://doi.org/10.1021/jacs.4c17738

45. Karthikeyan A, Priyakumar UD. Artificial intelligence: machine learning for chemical sciences. J Chem Sci (Bangalore). 2022;134(1):2. doi: 10.1007/s12039-021-01995-2.

46. Lu, J.-M., Automated Intelligent Platforms for High-Throughput Chemical Synthesis, Artificial Intelligence Chemistry, 2, 100057, 2024. https://doi.org/10.1016/j.aichem.2024.100057

47. Houhou R, Bocklitz T. Trends in artificial intelligence, machine learning, and chemometrics applied to chemical data. Anal Sci Adv. 2021 Feb 2;2(3-4):128-141. doi: 10.1002/ansa.202000162.

48. Li, J., Ding, C., Liu, D., Chen, L., Jiang, J., Autonomous laboratories in China: an embodied intelligence-driven platform to accelerate chemical discovery, Digital Discovery, 4, 1672, 2025. doi: 10.1039/D5DD00072F

49. George, J., Hautier, G., Chemist versus Machine: Traditional Knowledge versus Machine Learning Techniques, Trends in Chemistry, 3, 86, 2021. https://doi.org/10.1016/j.trechm.2020.10.007

50. Ali RSAE, Meng J, Jiang X. Synergy of Machine Learning and High-Throughput Experimentation: A Road Toward Autonomous Synthesis. Chem Asian J. 2025 Oct;20(20):e00825. doi: 10.1002/asia.202500825

51. Ali RSAE, Meng J, Jiang X. Synergy of Machine Learning and High-Throughput Experimentation: A Road Toward Autonomous Synthesis. Chem Asian J. 2025 Oct;20(20):e00825. doi: 10.1002/asia.202500825

52. Choi, J., Lee, B. Accelerating materials language processing with large language models. *Commun Mater* **5**, 13 (2024). https://doi.org/10.1038/s43246-024-00449-9

53. Singh, N., Lane, S., Yu, T. *et al.* A generalized platform for artificial intelligence-powered autonomous enzyme engineering. *Nat Commun* **16**, 5648 (2025). https://doi.org/10.1038/s41467-025-61209-y

54. Xu, G., Jiang, M., Li, J., Xuan, X., Li, J., Lu, T., Pan, L., Machine learning-accelerated discovery and design of electrode materials and electrolytes for lithium ion batteries, Energy Storage Materials, 72, 103710, 2024. https://doi.org/10.1016/j.ensm.2024.103710

55. Bhat, A. R., Ahmed, S., Artificial intelligence (AI) in drug design and discovery: A comprehensive review, In Silico Research in Biomedicine, 1, 100049, 2025. https://doi.org/10.1016/j.insi.2025.100049

56. Baum, Z. J., Yu, X., Ayala, P. Y., Zhao, Y., Watkins, S. P., Zhou, Q., Artificial Intelligence in Chemistry: Current Trends and Future Directions, Journal of Chemical Information and Modeling, 61, 2932, 2021. https://doi.org/10.1021/acs.jcim.1c00619J

57. Meng K, Long R. A Universal Machine Learning Framework Driven by Artificial Intelligence for Ion Battery Cathode Material Design. JACS Au. 2025 Jul 23;5(8):3833-3845. doi: 10.1021/jacsau.5c00526

58. Hermann E, Hermann G, Tremblay JC. Ethical Artificial Intelligence in Chemical Research and Development: A Dual Advantage for Sustainability. Sci Eng Ethics. 2021 Jul 6;27(4):45. doi: 10.1007/s11948-021-00325-6

59. Wang Z, Zeier WG, You F. Toward AI ecosystems for electrolyte and interface engineering in solid-state batteries. Sci Adv. 2025 Nov 28;11(48):eaea0638. doi: 10.1126/sciadv.aea0638

60. Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. Heliyon. 2024 Feb 15;10(4):e26297. doi: 10.1016/j.heliyon.2024.e26297

61. Ng, M.-F., Sun, Y., Seh, Z. W., Machine learning-inspired battery material innovation, Energy Advances, 2, 449, 2023. DOI: 10.1039/D3YA00040K

ISSN 2394 - 7780

62. Huang, G., Huang, F., Dong, W., Machine learning in energy storage material discovery and performance prediction, Chemical Engineering Journal, 492, 152294, 2024. https://doi.org/10.1016/j.cej.2024.152294

63. Mohamed, M. E., Khodadadi, E., Ethical Challenges and Regulatory Compliance in AI-Driven Neurological Diagnostics: A Review of Standards and Practices, Metaheuristic Optimization Review, 4, 24, 2025. https://doi.org/10.54216/MOR.040203