International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 09 | Sept - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Green Chemistry and Artificial Intelligence: A Synergistic Approach for Sustainable Synthesis

[1] Usha Jaiswal,[2] Dr. Ambuj Pandey, [3]Anupama Pradhan,[4] Suresh Kumar Bharty

[1] Department of Chemistry,Chauksey Engineering College, Bilaspur, Chhattisgarh, India

[2] Department of Chemistry Govt. Bilasa Girls P.G. (Auto.) College, Bilaspur, Chhattisgarh, India

[3] Department of Chemistry Thakur Shobha Singh Govt. College Pathalgaon, Chhattisgarh, India

[4] Department of Chemistry Dr. Jwala Prasad Mishra Govt. Science College, Mungeli, Chhattisgarh, India

[3]pradhananupamatss@gmail.com

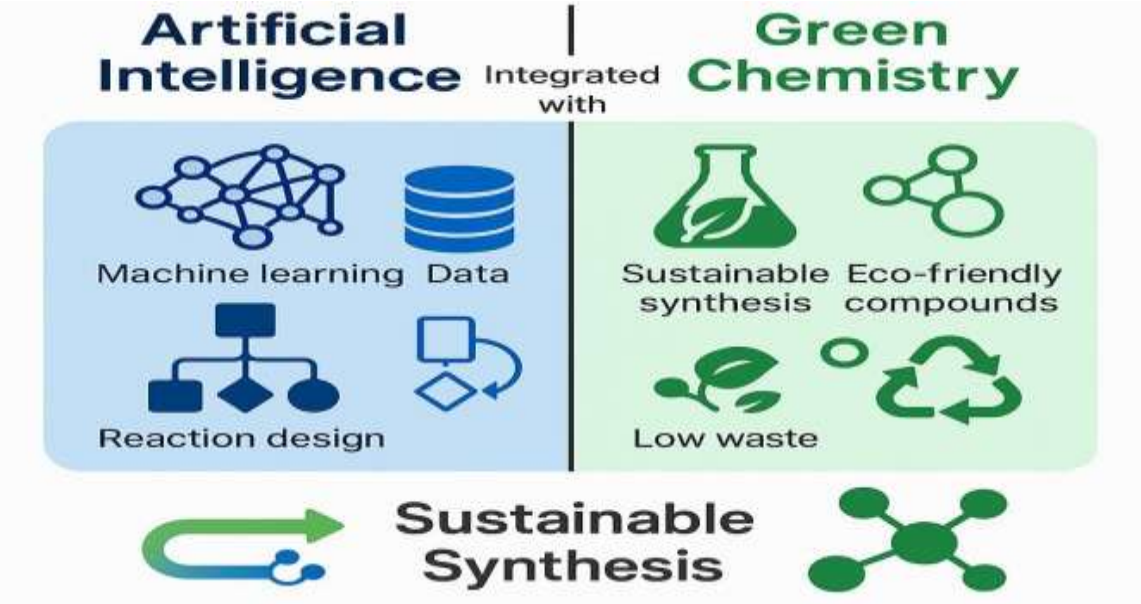
Abstract: The integration of Green Chemistry and Artificial Intelligence (AI) presents a transformative path toward sustainable chemical synthesis. While green chemistry emphasizes minimizing toxicity, waste, and environmental harm through safer and more efficient processes, AI provides data-driven capabilities such as predictive modeling, retrosynthetic planning, process optimization, and automation. This synergy enables intelligent systems that predict reaction outcomes, optimize solvent and catalyst selection, minimize hazardous by-products, and align closely with the Twelve Principles of Green Chemistry. Leveraging technologies like machine learning, deep neural networks, cheminformatics, and generative models, AI speeds the creation of eco-friendly chemicals and low-energy reaction pathways while decreasing the need for significant lab research. Tools that facilitate solvent-free synthesis, atom economy, and high-throughput testing, such Molecular Transformer, Chemistry42, AlphaFold, and Graph Neural Networks (GNNs), are prime examples of this progression. The expanding importance of robotic systems and autonomous labs that employ AI to dynamically optimize synthesis methods in real time is also examined in the article. AI's concrete contributions to safer, more economical, and environmentally friendly chemical innovation are demonstrated by case studies in the pharmaceutical, agrochemical, and fine chemicals industries. Based on a comprehensive literature analysis and expert discussions from the ASLLA Symposium, the report highlights the value of multidisciplinary education, interpretable AI models, and carefully selected green datasets. Even if there are still issues with model generalisation, ethical implementation, and data governance, integrating AI literacy into chemistry courses and encouraging cooperation between academia, industry, and politics will be crucial to expanding this integration. In the end, this multidisciplinary approach promotes global objectives for responsible manufacturing, sustainable innovation, and climate action in addition to chemical intelligence.

Keywords: Green Chemistry, Artificial Intelligence, Cheminformatics,, Environmental Impact, Smart Chemical Design.

1. Introduction

The demand for safer, cleaner, and more resource-efficient chemical processes has increased due to the pressing worldwide push for sustainable development. Green chemistry, as described by Anastas and Warner's Twelve Principles, has become a fundamental paradigm for creating less hazardous, waste-reducing, and energy-efficient chemical systems in response to growing concerns about climate change, ecological degradation, and resource depletion. Concurrently, artificial intelligence (AI) has developed as a potent catalyst for this change. AI enables high- throughput prediction, optimisation, and decision-making through machine learning, deep learning, natural language processing, and reinforcement learning. AI was first used in drug discovery, but it is currently used in environmental chemistry, materials science, sustainable manufacturing, and catalysis, among other areas. AI speeds up the transition to intelligent, sustainable, and life-cycle- oriented chemical innovation by reducing reliance on trial-and-error experimentation and increasing predictive efficiency. This helps to rethink synthesis and production paradigms for a more environmentally friendly future.

Figure 1: Conceptual Framework of AI Integration with Multiomics in Chemistry



The combination of green chemistry with artificial intelligence (AI) represents a paradigm shift in the way that chemical research is planned, carried out, and taught. By making predictive modelling for reaction outcomes, physicochemical attributes, retrosynthetic planning, solvent and catalyst selection, toxicity evaluation, and reaction condition optimisation possible, artificial intelligence (AI) improves green chemistry. Pre-synthesis screening for environmental effect, degradability, and synthetic feasibility is now possible because to sophisticated AI structures like graph neural networks (GNNs), transformer models, Bayesian frameworks, and generative models, which supplement traditional techniques like QSAR and QSPR.Using massive datasets like ChEMBL, ZINC, PubChem, and USPTO, generative AI models like VAEs, GANs, and reinforcement learning frameworks further facilitate the design of new, eco-optimized compounds. In the meanwhile, deep learning and search algorithms (e.g., MCTS, PNS) are used by AI-powered retrosynthesis platforms such as IBM RXN, SYNTHIA, and AiZynthFinder to suggest viable multistep synthesis paths. In accordance with the Twelve Principles of Green Chemistry, these instruments aid in minimising dangerous reagents, cutting waste, and improving conditions.

Through closed-loop systems like AlphaFlow, Coscientist, and Chemputer— which utilize real-time sensor data, robotic automation, and AI-driven feedback to dynamically optimize synthesis conditions—the integration of Artificial Intelligence (AI) into laboratory settings has enabled autonomous experimentation. These platforms exemplify AI's dual role in sustainable process development: acting both as an operational driver and as a predictive engine to enhance environmental performance, efficiency, and reproducibility.

This synergy is supported by data-rich ecosystems and cheminformatics infrastructures. Toolkits such as RDKit, Chemprop, and SMILES-based encoders, along with molecular fingerprinting techniques like ECFP4 and

MACCS, enable machine-readable molecular modeling and efficient chemical representation. Robust model training across diverse chemical domains is made possible by curated databases such as Reaxys, CAS Registry, ZINC, PubChem BioAssay, and the USPTO repository. These resources facilitate high-quality, domain-specific learning for AI systems applied in chemical synthesis and discovery.Beyond pharmaceuticals, AI is transforming green chemistry applications in fields like materials science, agrochemicals, and environmental chemistry. It aids in meeting regulatory requirements such as REACH, GHS, and GRAS, while accelerating the development of safer pesticides, biodegradable polymers, eco-friendly additives, and sustainable packaging solutions.

This report evaluated 98 peer-reviewed publications published between 2022 and 2023 in order to evaluate the present status of AI in sustainable chemical innovation. Prediction, design, pattern recognition, and workflow optimisation are the four areas into which the chosen papers were divided using the Web of Science, Scopus, and PubMed databases and the PRISMA methodology. According to the analysis, there is a growing body of research that supports the alignment of AI methods with green chemistry principles.

Universities and training institutions are increasingly incorporating AI into chemistry education through courses in cheminformatics, SMILES modeling, Python programming, and tools such as AlphaFold and Chemputer. This shift reflects the growing importance of digital competencies in sustainable research and prepares future chemists to navigate both experimental and computational domains effectively.
However, several challenges remain, including issues of model interpretability, algorithmic bias, data quality, the lack of negative reaction data, and ethical concerns surrounding the trade-offs between automation and sustainability. Addressing these concerns requires the adoption of Explainable AI (XAI), human- in-the-loop systems, and strong interdisciplinary collaboration among computer scientists, chemists, and ethicists. To ensure reproducibility and alignment with green chemistry metrics like life- cycle assessment (LCA), the development and use of open-access, sustainability- tagged datasets are critical. These resources will support transparent, data- driven research and responsible innovation.

In conclusion, the integration of AI with green chemistry signifies more than just technological progress; it represents a systemic transformation aligned with global sustainability targets such as the Paris Agreement and the UN Sustainable Development Goals. This study highlights the emerging advancements, opportunities, and responsibilities shaping the future of sustainable chemical innovation.

1.1 The Digital Wave in Chemistry: A Sustainable Catalyst

AI is driving the digital transformation of chemistry, which includes a wide range of instruments that support green chemistry:

Table 1**:** Summary of AI-Based Technologies in Chemistry and Their Impact

AI Technology	**Relevance to Green Chemistry and Sustainable Synthesis**
AI-Driven Drug Discovery	Helps design eco- friendly pharmaceuticals and reduce experimentation waste through virtual screening.
Big Data & Integrated Data	Consolidates heterogeneous data sets for environmentally conscious decision-making in synthesis design.
Automated Laboratory Platforms	Reduces chemical waste and enhances precision by minimizing manual errors in green synthetic experiments.
IoT-Integrated Lab Instruments	Enables real-time monitoring of green process metrics.
AI in Analytical Method	Enhances rapid, low- waste analytical procedures to

Development	monitor green chemistry compliance.
Digital Twins & Virtual Labs	Allows simulation of sustainable synthesis avoiding resource wastage.
Natural Language Processing (NLP)	Extracts environmentally safe synthesis routes from vast chemical literature databases.
Predictive Toxicology	Forecasts ecological and toxicity of compounds, essential for sustainable chemical product design.
AI in Environmental Chemistry	Models pollutant fate and optimizes waste reduction strategies in chemical manufacturing.
ML in Molecular Design	Identifies molecules with high efficacy and low environmental burden.
SAR via Deep Learning	Links chemical structure with biological activity for bio-safe, sustainable compound prediction.
High-Throughput Experimentation (HTE)	Enables green synthesis screening with minimal sample and energy usage.
Data-Driven Reaction Optimization	Reduces trial-and-error and chemical waste by predicting optimal reaction conditions.
Automated Synthesis Planning	Designs efficient, eco- friendly synthetic routes, saving time and raw materials.
Chemoinformatics	Enhances understanding of structure-property relationships critical for green chemical design.
AI in Quantum Chemistry	Reduces computational resources in modeling greener reaction pathways and catalysts

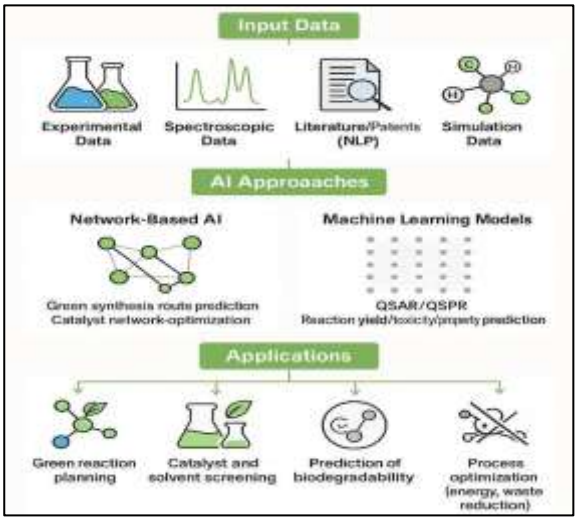


Figure 2

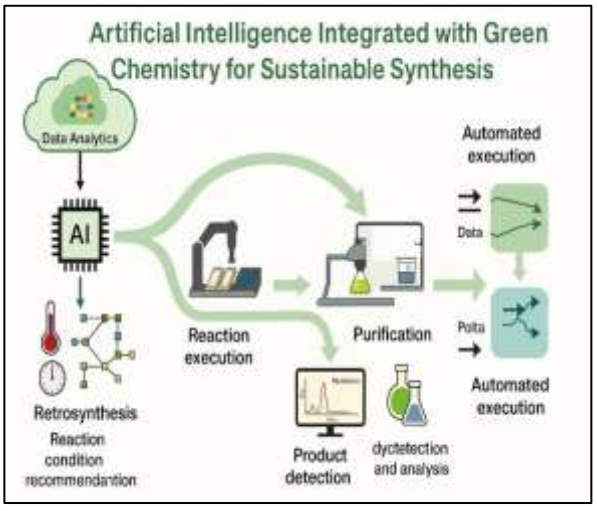


Figure 3

Figure 2:Integration of AI and Green Chemistry for Sustainable Synthesis
Figure 3: Workflow of AI-Driven Autonomous Organic Synthesis

By combining intelligence, automation, and sustainability, these developments have completely changed the field of chemistry research. The combination of AI and green chemistry principles enables chemists to develop new approaches that lead to safer products, cleaner technologies, and less environmental impact.

1.2 Toward Data-Driven Sustainable Synthesis

Conventional chemical synthesis requires a lot of resources and is frequently limited by trial-and-error techniques. AI provides instruments for environmentally sensitive chemical production through toxicity predictions, synthetic route design, and reaction prediction algorithms. For example, in order to prevent unnecessary bench tests, AI-powered software such as IBM's RXN, DeepChem, and Chemprop allows for the in-silico prediction of reaction outcomes and retrosynthesis processes. By connecting chemical descriptors with greenness criteria (such as atom economy, E-factor, and biodegradability), the combination of chemoinformatics and machine learning also makes it possible to screen for sustainable molecules. For businesses looking to adhere to green chemical regulations without sacrificing productivity or revenue, this is essential.

1.3 Implications for Innovation and Education

In addition to changing research procedures, these technologies are influencing the chemists of the future. These days, educational platforms use real- world datasets and coding tools like Python, RDKit, and deep learning frameworks to teach AI-powered cheminformatics, green synthesis modelling, and environmental impact prediction. This enables researchers and students to act responsibly, think computationally, and use science to create a more environmentally friendly future. Green chemistry principles may now be used at every stage of chemical synthesis thanks to the digital revolution in chemistry, which is led by artificial intelligence.

AI speeds up the process of sustainable innovation through automated optimisation, real-time monitoring, and predictive design. With an emphasis on real-world applications, case studies, and potential paths forward in sustainable molecular development, this study examines the relationship between artificial intelligence (AI) and green chemistry, offering a thorough analysis of how these game-changing technologies promote ecologically conscious synthesis.

AI-Based Technology / Digital Chemistry Trajectory	Key Application Area	Impact on Green & Sustainable Synthesis	Relative Publication Volume
AI- Driven Drug Discovery	Prediction of eco-friendly active compounds	Reduces experimentation, supports green pharmaceuticals	High
Big Data & Integrated Data	Harmonization of chemical data repositories	Enables holistic analysis for green chemical innovation	High
Automated Laboratory Platforms	Automated, minimal - waste synthesis setups	Enhances reproduci bility, minimizes chemical waste	High

Integration of Laboratory Instruments & IoT	Real- time monitoring in synthesis and purification	Improves reaction control and reduces resource consumption	High
AI in Spectroscopy & Analytical Chemistry	Real- time monitoring of green catalysts or pollutants	Accelerates eco- friendly process optimization	High
Predictive Toxicology	Prediction of harmful by- products or reagents	Enables early- stage elimination of toxic options	Medium- High
Machine Learning in Molecula r Design	Molecular property prediction, solvent- free synthesis design	Supports design of green molecule s and solvent alternatives	Medium
Deep Learning in Structure –Activity Relations hips (SAR)	QSAR/ QSPR modeling for biodegra dable and sustainable materials	Facilitates efficient design of safe and functional molecules	Medium
Data- Driven Chemical Reaction Optimization	AI Models for reaction condition selection	Increases reaction efficiency and eco-compatibility	Medium
Automated Synthesis Planning	Retrosynthetic route optimization	Reduces number of synthetic steps, saves energy and solvents	Medium
Digital Materials Design & Materials Informatics	Design of recyclable polymers, catalysts	Accelerates discovery of eco- friendly materials	Medium
Environmental Chemistry & AI for Sustainability	AI for pollution monitoring and lifecycle analysis	Promotes sustainable manufacturing and environmental protection	Medium
Smart Control Systems in Chemistry	Adaptive control in real- time synthesis systems	Enhances energy and material efficiency	Medium
Natural Language Processin g (NLP) in Chemical Space	Text mining from patents and journals for green synthesis data	Unlocks hidden sustainable synthesis pathways from literature	Medium- High
AI- Driven High-Throughput Experimentation (HTE	Rapid screening of reaction conditions and catalysts	Enables fast green reaction optimization with minimal waste	Medium
Digital Twins in Chemistry	Virtual modeling of chemical processes	Enables virtual validation of green process designs	Medium
Quantum Chemistry & AI- Enhanced Simulations	Simulating reaction mechanisms for catalyst and solvent- free reactions	Reduces experimental load, predicts reaction efficiency	Medium

Blockchain in Chemical Supply Chains	Verification and traceability of green chemical sources	Ensures responsible sourcing and transparency	Medium
Virtual Laboratories & Augmented Reality for Education	Simulation of green chemistry experiments	Enhances sustainab ility education in chemistry	Medium

Table 2: AI-Based Technologies and Their Influence on Green and Sustainable Chemistry

3.Background and Literature Review

This literature review is divided into two sections to give readers a structured understanding of how AI and Green Chemistry work together for sustainable synthesis. The first section is a Integrated Literature thorough thematic table (see Table 3), which is followed by a detailed narrative discussion of seminal contributions and new paradigms.

Table 3:Review on AI Applications in Green Chemistry and Sustainable Synthesis

Theme / Focus Area	Key Study / Authors	Key Insights / Contributions	Relevance to Green Chemistry
Foundations of AI in Chemistry	E. J. Corey & W. Todd Wipke, (1969)	Pioneered comput er-assisted synthesis planning (CASP) .	Established the comptational foundation for green synthesis route design .
Retrosynthesis &CASP Systems	Segler et al., (2018): Coley et al., (2019); Mikulak- Klucznik et al., (2020)	Symbol ic AI, neural-symbolic and Monte Carlo approaches for automated low-step retrosynthesis	Enhances atom econo my, avoids hazardous intermediates, and supports sustainable route planning.
Inverse Molecular Design & Generative Models	Sanchez- Lengeling & Aspuru- Guzik, (2018); Gómez- Bombarelli et al., (2018); Putin et al., (2018); Olivecrona et al., (2017)	Developed VAEs, GANs, RNNs to generate eco-optimized, biodegradable, and low- toxicity molecules.	Accelerates discovery of green alternatives with enhanced safety and sustain ability profiles.
Reaction Prediction & Optimization	Coley et al., (2019);Schwaller et al.,(2019); Jin et al., (2017); Ahneman et al., (2018)	GNNs, seq2seq models forecast yields, conditions, and selectivity.	Reduces failed trials and supports greener reagent/catalyst selection.

Reaction Condition Optimization	Gao et al., (2018); Zhou et al., (2017); Struebing et al., (2013); Seifrid et al., (2022)	Bayesian optimiz ation and ML fine-tune solvent, pressure, tempera ture.	Improves energy efficiency and minim izes toxic chemical usage.
Automated & Robotic Synthesis	Dragone et al., (2017);Granda et al., (2018);Roch et al., (2018);Boiko et al., (2023a); Mahjour et al., (2023)	Robotic labs and LLMs support closed-loop synthesis planning and execution.	Reduces waste and human error; improves scalability and reproducibility.
Catalysis & Interface Chemistry	Yang et al., (2019); Mai et al., (2022); Toyao et al., (2019); (Galushko et al., 2023);Schlexer Lamoureux et al., (2019)	ML and DFT models enhance catalyst performance prediction.	Enables selection of recyclable, energy - efficient, and non- toxic catalytic systems.
High- Through put Discovery & Automat ion	Bédard et al., (2018);Häse et al., ; Gómez-Bombarelli et al.,(2018)	Robotic and AI-guided platforms accelerate explorat ion of synthetic space.	Facilitates rapid identification of greener pathways.
MOF Synthesis & Text Mining	Zheng et al., (2023)	NLP and LLMs extract synthetic protocols from literature.	Supports efficient material discovery with lower energy and resource use.
Drug Discovery & Green Pharma	Vamatheva n et al., (2019);Zhavoronk ov et al., (2019); Stokes et al.,(2020); Griffen et al.(2020); Chan et al., (2019)	AI accelerates lead discovery, ADMET prediction, and safety profiling.	Reduces chemical waste and environmental toxicity in pharmaceutical pipelines.
Chemoinformatics & Descript or Learning	Varnek & Baskin, (2012);Rogers & Hahn, (2010); Duvenaud et al., (2015); Reker et al., (2016)	Developed molecular fingerpr ints, ECFPs, and pattern recognition tools.	Supports toxicity prediction and green compound screening.
Molecular Encoding & Representation	Weininger, (1988a); Heller et al., (2013a);Yang et al., (2019); Leach & Lewis, (1994)	Introduced SMILES, InChI, and GNN-friendly encodings.	Foundational for ML-driven sustain ability predict ion models.

QSAR/QSPR & Property Prediction	Hansch &Fujita, (1964); Dahl et al., (2014); Ma et al., (2015); Muratov et al., (2020)	ML models estimate toxicity, solubility, biodegradability, log P.	Supports eco- evalua tion of molecules prior to synthesis.
Material Discovery & Sustainable Design	Moosavi et al., (2020); Kim et al., (2019); Stach et al., (2021); Dimitrov et al., (2019)	Used HTVS and ML to develop green MOFs, COFs, and polymers.	Enables lifecycle-aligned material development for energy and filtration.
Quantum Chemistry & ML Integration	Smith et al., (2018); Grisafi et al., (2018); Aspuru- Guzik et al., (2018); Cao et al., (2012); Westermayr & Marquet and, (2020)	ML accelerates quantum property estimation with reduced comput ational load.	Optimizes energy -efficient reaction modeling and green catalyst screening.
Spectral Analysis & Quality Monitoring	Elyashberg & Williams, (2015); Schwaller et al., (2019); McCardle, (2022); Boiko et al., (2023); Ayres et al., (2021)	ML interprets NMR, MS, and IR data for real- time impurity profiling.	Supports green quality control by minimizing reprocessing.
Environmental Monitoring & IoT Integration	Capella et al., (2020); Nourani et al., (2018); Campelo et al.,(2022); Boger, (1992)	AI + IoT enables real-time pollutant/emission tracking	Enhances green compliance and closed-loop manufacturing.
AI Platforms & Synthesis Tools	Ivanenkov et al., (2023)	Provides end- to-end AI platforms for molecule generation and synthesis.	Democratizes access to green synthesis design capabilities.
Data Curation & FAIR Principles	Baum et al., (2021); Villalba et al., n.d.; Wilkinson et al., (2016)	Emphasized structured, interoperable, open-access datasets .	Enhances reproducibility, data reusability, and low-waste research planning.
Explainable & Ethical AI	Rudin, (2019); Azodi et al., (2020); Pavel et al., (2022); Bender & Cortés-Ciriano, (2021); Wallach & Heifets, (2018)	Called for transpar ent, auditable, and bias- mitigated AI models.	Supports responsible AI integration in high-stakes green applications.

Benchmarking & Model obustness	Maryasin et al., 2018); Walters & Murcko, (2020a)	Assessed generali zation,uncertainty, and reproducibility in chemical ML.	Ensures scienti fic credibility of AI-enable d green solutions
Integrate d AI-Green Chemistry Frameworks	Gómez- Bombarelli et al., (2018b); Schwaller et al., (2019); Walters & Murcko, (2020)	Theoretical and practical frameworks aligning AI tools with sustainability goals.	Frame s AI as a central enabler of sustain able molecular innovation.
Reaction Template Extraction & Rule Mining	Plehiers et al., (2018); Zhavoronk ov et al., (2019)	NLP and cheminf ormatics extract reusable, eco-optimized synthesis rules.	Promotes reaction reusability and avoids experimental redund ancy.
AI in Chemistry Education & Training	Gasteiger & Zupan, (1993); Kim et al., (2019); Wishart et al., (2018)	Developed ML education platforms and digital databases.	Prepares next- gen chemists for green, data-driven innova tion.

The field of sustainable chemical synthesis is undergoing a radical transformation due to the convergence of green chemistry and artificial intelligence (AI). Once limited to theoretical modeling, AI now enables intelligent automation and predictive, data- driven decision-making across the entire chemical workflow. This includes molecular design, retrosynthetic planning, reaction condition optimization, process control, and environmental monitoring. This multidisciplinary integration aligns closely with Anastas and Warner's Twelve Principles of Green Chemistry, which advocate for atom economy, waste minimization, energy efficiency, the use of safer solvents and reagents, and a cradle- to-grave approach to chemical design and innovation. The theoretical groundwork for this convergence dates back to the 1960s and 1970s, with the pioneering work of E. J. Corey & W. Todd Wipke, (1969)in computer-assisted synthesis planning (CASP), which laid the foundation for today's AI-driven retrosynthetic tools. Simultaneously, Hansch & Fujita, (1964) introduced Quantitative Structure– Activity Relationship (QSAR) models, which established correlations between molecular structures and their physicochemical or biological properties. These models continue to be fundamental in cheminformatics and the design of sustainable compounds.

The development of machine-readable molecular encodings such as SMILES Weininger, (1988), Heller et al., (2013) and molecular fingerprints Leach &Lewis, (1994) provided the digital infrastructure necessary for training AI algorithms. These tools allow modeling of structure– property relationships, toxicity predictions, and synthetic feasibility. These advancements have facilitated large- scale virtual screening and predictive modeling, enabling the creation of data- rich pipelines essential for driving green and sustainable chemical synthesis.The use of Artificial Intelligence (AI) in chemistry has surged in recent years, primarily due to the development of advanced deep learning architectures such as Graph Neural Networks (GNNs), Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and Reinforcement Learning (RL). These models now span the entire chemical lifecycle—from de novo molecule generation and retrosynthetic planning to reaction outcome prediction, green metric optimization, and end-of-life sustainability analysis. Seminal works by Segler et al., (2018), Coley et al., (2019), Gómez- Bombarelli et al., (2018)and Aspuru-Guzik et al., (2018); demonstrate these advances.

Platforms like IBM RXN, Chematica (now SYNTHIA), and the Molecular Transformer integrate neural-symbolic reasoning, rule-based logic, and probabilistic algorithms to offer efficient and environmentally

friendly synthesis pathways. These tools align synthesis planning with green chemistry principles by evaluating synthetic routes using green metrics such as atom economy, energy consumption, toxicity, and hazardous reagent use. Parallel advancements in sequence-to- sequence (seq2seq) models and GNNs, led by researchers like Schwaller et al., (2019)and Coley et al., (2019), have significantly improved reaction prediction and condition optimization. These models accurately forecast reaction yields, by- product profiles, and optimal parameters, enabling smarter experimental design and minimizing waste.

Complementary approaches such as Bayesian Optimization, Monte Carlo Tree Search (MCTS), and transfer learning— explored by Zhou, Bedard, and Gao— further enhance the precision and adaptability of reaction conditions. These methods allow real-time adjustments to solvents, catalysts, temperatures, and ressures, reducing environmental impact and experimental redundancy. In the realm of de novo molecular design, AI has delivered some of its most transformative contributions to green chemistry. Researchers like Gómez-Bombarelli et al., (2018), Putin et al., (2018) and Zhavoronkov et al., (2019) have employed generative models (VAEs, GANs, RNNs) to create novel molecules tailored for green criteria, such as low toxicity, high biodegradability, renewable functional groups, and synthetic accessibility.

These breakthroughs underscore AI's revolutionary potential in ecologically responsible molecular discovery and have played a vital role in developing next- generation green pharmaceuticals, agrochemicals, and sustainable functional materials. Real-time, closed-loop experimentation using robotic platforms like ChemOS and AlphaFlow has become possible due to the digitization and automation of synthetic processes. Pioneering efforts by researchers such as Matsubara, Mahjour et al., (2023) and Boiko et al., (2023a) have enabled autonomous reaction planning and condition optimization. These systems are increasingly integrated with large language models (LLMs) and AI-driven laboratory protocols.

These autonomous platforms exemplify the core goals of green chemistry by reducing reagent consumption, chemical waste, and energy use while enhancing safety, reproducibility, and process efficiency. Their application supports sustainable innovation across chemical domains.

Artificial Intelligence has also revolutionized catalysis—an essential pillar of green synthesis—by improving catalyst reactivity, selectivity, recyclability, and enabling the development of metal-free alternatives. Studies by Yang et al., (2019), Toyao et al., (2019), Galushko et al., (2023) and Mai et al., (2022) highlight how machine learning and quantum-enhanced models promote atom economy and reduce toxicity through energy-efficient, non- toxic catalytic systems.

In analytical chemistry, AI has enhanced the interpretation of spectroscopic data such as NMR, IR, and MS. Models developed by Schwaller et al.,(2019), McCardle, (2022) and Boiko et al., (2023a) allow real-time impurity detection, process control, and environmentally conscious quality assurance.

These capabilities are further expanded by integration with Internet of Things (IoT) networks. As shown by Capella et al., (2020)and Campelo et al., (2022), IoT- connected systems enable decentralized, cloud-based environmental monitoring, supporting proactive compliance with sustainability regulations.

Outside the lab, AI is increasingly used in environmental chemistry for tasks such as life-cycle assessment, wastewater analysis, and predictive pollution monitoring. Early work by Boger using neural networks, and more recent ensemble learning approaches by Nourani et al., (2018) demonstrate AI's ability to simulate pollutant behavior and emissions with accuracy.

Machine learning-augmented quantum simulations are also emerging as a low- energy alternative to traditional quantum chemical computations. Research by Westermayr & Marquetand, (2020) and Grisafi et al., (2018) explores their role in modeling photocatalytic and excited-state processes, helping to reduce computational and environmental costs. These advancements are supported by the availability of high-quality, FAIR- compliant datasets from resources such as PubChem, ChEMBL, Reaxys, and CAS Content. As noted by Villalba et al., and Wilkinson et al., (2016), such robust data infrastructure ensures cross-platform integration, model reliability, and reproducibility.

Finally, there is a growing emphasis on interpretable and ethical AI in chemistry. In high-stakes applications, researchers like Rudin, Azodi et al., (2020), and Pavel et al., (2022), advocate for explainable AI (XAI) and human-in-the-loop systems. These approaches ensure that AI-driven decisions are not only technically sound but also aligned with principles of safety, transparency, and environmental responsibility.Recent frameworks by Ivanenkov et al., (2023); and Walters & Murcko, (2020) support AI as a systemic catalyst for the concepts of the circular economy. Sustainability is now integrated into all phases of the chemical value chain, from molecule discovery to end-of-life product recovery, thanks to AI's influence on resource reuse, life-cycle optimisation, and design for degradation.

3. Methodology

This study uses a methodical, literature- based approach to examine how the Twelve Principles of Green Chemistry and Artificial Intelligence (AI) might work together to support sustainable chemical synthesis. As a conceptual overview, the study maps current trends, thematic intersections, and application areas at the junction of AI and green chemistry by synthesising information from peer- reviewed publications, technology reports, and open-access data platforms.

Using top scientific databases such as Scopus, Web of Science, PubMed, SpringerLink, IEEE Xplore, and ScienceDirect, a thorough literature review was carried out. Authoritative databases like ChEMBL, PubChem, USPTO, and ZINC provided more data and case studies, especially to put AI models (e.g., GNNs, VAEs, QSAR/QSPR, and retrosynthetic transformers) in perspective when working with chemical data. With a focus on contemporary work (2015–2024) pertinent to AI applications in chemical synthesis, sustainability, and environmental technologies, the inclusion criteria were centred on literature published between 2000 and 2024. The technical and ethical foundation for assessing sustainability relevance was the Twelve Principles of Green Chemistry, which were put out by Anastas and Warner.

A three-phase conceptual framework guided the analysis:

- **Technology Mapping:** AI tools and models were categorised based on their main chemical function, including environmental monitoring, catalyst design, reaction condition optimisation, property prediction, and retrosynthesis planning. The capabilities of well-known platforms in sustainable synthesis design, including IBM RXN, AiZynthFinder, Chemistry42, AlphaFold, SYNTHIA, and ChemOS, were profiled.

- **Theme Categorisation:** Using manual coding and co-occurrence analysis of important phrases, the literature was arranged into theme clusters. Significant intersections were identified, such as data-driven synthesis design, autonomous labs, AI in solvent selection and catalysis, and environmental lifecycle assessment.

- **Sustainability Alignment Assessment**: Each AI application was qualitatively assessed based on its contribution to green chemistry metrics such as atom economy, waste minimization, solvent sustainability, energy efficiency**,** and toxicity reduction. Consideration was also given to technology readiness levels and degree of AI integration (rule- based, ML, or hybrid).

This study critically interprets existing models, such as transformer-based retrosynthetic predictors, generative neural networks, and machine learning classifiers for eco-toxicity, to assess their applicability, scalability, and environmental alignment, even though no new models or experiments were carried out. ChemDraw, BioRender, Adobe Illustrator, and VOS viewer were among the visualisation technologies used to create:

- Conceptual diagrams illustrating pathways between AI and green chemistry; bibliometric maps of research trends; and infographics summarising the contributions of AI to certain green principles.

The article also summarises conceptual case findings from top published research. These consist of:

1. Tools for transformer-based retrosynthesis that reduce dangerous intermediates
2. Generative models that forecast low-toxicity, biodegradable compounds
3. Reaction parameter optimisation via reinforcement learning
4. Self-governing lab systems that facilitate instantaneous synthesis improvement

The importance of AI in sustainable design was further supported by a qualitative analysis of each instance using recognised green chemistry metrics (such as energy usage, E-factor, and EcoScale). All things considered; this literature-based technique offers a systematic, reproducible framework for examining the ways in which AI technologies are changing environmentally friendly chemical practices. It provides a technological and moral road map for upcoming studies at the nexus of green chemistry and artificial intelligence by bridging conceptual findings with applied significance.

5. Applications of AI in Green Chemistry

Artificial Intelligence (AI) is revolutionizing sustainable chemical synthesis by integrating machine learning (ML), cheminformatics, natural language processing (NLP), generative modeling, and robotics across all stages of the chemical value chain. This includes molecular design, reaction optimization, property prediction, solvent selection, autonomous experimentation, and green manufacturing.

In reaction optimization, AI models such as recurrent neural networks (RNNs) and Bayesian optimization accurately predict optimal parameters like temperature, pH, and catalyst selection. This improves yields while reducing energy use, waste, and hazardous by-products. Deep learning and search algorithms—like Monte Carlo Tree Search and Proof Number Search— underpin platforms such as IBM RXN, SYNTHIA, AiZynthFinder, and ASKCOS, which generate low-step, environmentally friendly synthetic routes that prioritize atom economy and minimize cost and reagent use.

Computer-Assisted Synthesis Design (CASD) tools are now sophisticated enough to challenge expert chemists in evaluating the feasibility and scalability of synthetic routes. Simultaneously, AI supports greener chemistry by predicting toxicity, volatility, and biodegradability to assist in solvent replacement and safer chemical selection. Autonomous laboratories like ChemOS and AlphaFlow, which combine robotic platforms with real-time self-optimization, allow for closed-loop, high-throughput experimentation. These systems significantly reduce trial-and-error cycles, reagent use, and operational inefficiencies—advancing the goals of green chemistry.

Beyond synthesis planning and execution, AI enhances the prediction of molecular properties such as HOMO-LUMO gaps, solubility, melting points, toxicity, and environmental persistence. This enables efficient pre-screening and helps avoid wasteful syntheses, particularly in data- scarce fields. Generative models—such as variational autoencoders (VAEs), generative adversarial networks (GANs), and graph neural networks (GNNs)—trained on databases like ZINC, QM9, and ChEMBL are used to design new compounds that meet green criteria, including low toxicity, synthetic accessibility, and biodegradability.

Although simulating complex phenomena like stereochemistry remains challenging, advanced GNNs and sequence-to-sequence models are advancing environmentally conscious reaction prediction. These innovations rest on a solid chemoinformatics foundation, which includes digital representations such as SMILES, InChI, molecular fingerprints, and graph-based encodings, supported by datasets like CAS, Reaxys, ChEMBL, and the Cambridge Structural Database (CSD).

AI also supports virtual screening, QSAR/QSPR modeling, and environmental impact prediction. Analytical tasks—including high- resolution mass spectrometry, GC, HPLC, NMR, and IR—are increasingly

automated using AI, enabling real-time impurity detection and product validation while minimizing chemical waste and reprocessing. AI's influence extends beyond the lab. Educational programs, such as the IIT Madras Pravartak initiative, train chemists in tools like RDKit, Chemprop, and Python to apply AI in green synthesis and property modeling. Workshops on technologies like AlphaFold2 and autonomous synthesis systems bridge the gap between traditional training and digital research.

In food chemistry, multimodal AI models enhance contaminant detection, traceability, and eco-compliance while reducing reliance on chemical reagents. Digital twins, IoT-enabled process monitoring, predictive toxicology, NLP tools for sustainability data extraction, and quantum simulations of reaction energetics are among over twenty AI-powered technologies advancing green innovation.

From a conceptual perspective, chemistry presents profound challenges to AI. Unlike bounded domains like chess or go, chemical space includes over 10^{60} possible drug-like molecules, along with NP- complete problems such as subgraph isomorphism, vague transformation rules, and poorly defined objectives—all constrained by sustainability metrics like atom economy and waste minimization. Tackling this complexity has led to breakthroughs in reinforcement learning, self-supervised learning, and transfer learning. AI systems now convert unstructured literature and chemical images into structured, machine-readable formats for downstream modeling and synthesis planning—building on foundational work like Lowe's USPTO dataset.

Despite these advances, challenges persist in data quality, regulatory compliance, and model interpretability. Researchers such as Rudin, Azodi, and Pavel emphasize the importance of explainable AI (XAI) and chemist-in-the-loop systems to ensure transparency, ethical implementation, and responsible decision-making in high-stakes chemical applications. In the end, the combination of AI with green chemistry is transformative rather than just additive. While chemistry poses practical problems that spur AI innovation, AI speeds up discovery, minimises experimentation, and unlocks sustainable solutions across disciplines. The future of chemical research is being redefined by this reciprocal interaction, which will make it safer, quicker, smarter, and more sustainable. The worldwide chemical industry is changing into a digital, ecologically conscious ecosystem that is suitable for 21st-century innovation as technologies become more automated, interpretable, and morally based.

Table 4: AI Capabilities Aligned with Green Chemistry Principles

Green Chemistry Principle	AI Contribution
Atom Economy	Reaction prediction, retrosynthetic route optimization
Less Hazardous Synthesis	Toxicity prediction using ML
Safer Solvents and Auxiliaries	Solvent selection via AI-based screening
Energy Efficiency	Process optimization algorithms
Renewable Feedstocks	Database mining for bio-based alternatives
Reduce Derivatives	Optimized synthetic planning
Catalysis	Catalyst design with generative models
Design for Degradation	Predictive modeling of environmental fate
Real-time Analysis for Pollution Prevention	AI-based sensor data analysis
Inherently Safer Chemistry	Risk assessment using AI classifiers

6. Challenges and Limitations

- Model accuracy is hampered by the availability and quality of data.
- In high-stakes chemical applications, interpretability of AI models is still an issue.
- Investment is needed to integrate AI with the current laboratory infrastructure.
- Due to a lack of training data, models for retrosynthesis perform poorly when using natural products.
- Stereochemistry and reaction kinetics are still hard to predict.

7. Future Directions and Conclusion

Looking ahead, there are a number of exciting opportunities to advance sustainable synthesis through the combination of Artificial Intelligence (AI) and Green Chemistry. The creation of high-quality, open-access green chemistry datasets that follow FAIR guidelines is one of the main future areas that will allow for greater involvement and more reliable model training. To provide aspiring researchers the tools they need to succeed in this changing environment, cross- disciplinary training programs that combine data science, artificial intelligence, and chemical sciences are crucial. Further advancements in explainable AI (XAI) will promote more use of AI tools in high-stakes chemical applications, increase model transparency, and build regulatory confidence.

Furthermore, developing AI algorithms that are specifically designed to tackle difficult problems—like stereochemistry, reaction kinetics, and the synthesis of natural products—will be essential to increasing the breadth and precision of AI-guided synthesis planning. Autonomous robotic execution systems combined with

AI-assisted retrosynthesis platforms will further facilitate closed-loop, self- optimizing experimentation that lowers material consumption, boosts repeatability, and quickens discovery cycles.

In summary, the combination of AI with green chemistry represents a revolutionary change in the design, synthesis, and application of molecules, giving efficiency, safety, and environmental sustainability first priority. Chemists may increase atom economy, speed up innovation in a more environmentally responsible way, and drastically cut waste by utilising AI's prediction and optimisation skills. The synergistic paradigm set forward in this study offers a guide for further research, highlighting AI's function as a strategic partner in the quest for sustainable chemical production rather than just as a computational tool. AI technologies are positioned to propel a new age of intelligent, environmentally friendly, and resilient chemical innovation as they develop further and become more interpretable, autonomous, and accessible.

References

- Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D., & Doyle, A. G. (2018). Predicting reaction performance in C–N cross- coupling using machine learning. *Science*, *360*(6385), 186–190.
- Aspuru-Guzik, A., Lindh, R., & Reiher, M. (2018). The matter simulation(r) evolution. *ACS Central Science*, *4*(2), 144–152.
- Ayres, L. B., Gomez, F. J., Linton, J. R., Silva, M. F., & Garcia, C. D. (2021). Taking the leap between analytical chemistry and artificial intelligence: A tutorial review. *Analytica Chimica Acta*, *1161*, 338403.
- Azodi, C. B., Tang, J., & Shiu, S.-H. (2020). Opening the black box:Interpretable machine learning for geneticists. *Trends in Genetics*, *36*(6), 442–455.
- Baum, Z. J., Yu, X., Ayala, P. Y., Zhao, Y., Watkins, S. P., & Zhou, Q. (2021). Artificial intelligence in chemistry: Current trends and future directions. *Journal of Chemical Information and Modeling*, *61*(7), 3197–3212.\
- Bédard, A.-C., Adamo, A., Aroh, K. C., Russell, M. G., Bedermann, A. A., Torosian, J., Yue, B., Jensen, K. F., & Jamison, T. F. (2018). Reconfigurable system for automated optimization of diverse chemical reactions. *Science*, *361*(6408), 1220–1225.
- Bender, A., & Cortés-Ciriano, I. (2021). Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. *Drug Discovery Today*, *26*(2), 511–524.
- Boger, Z. (1992). Application of neural networks to water and wastewater treatment plant operation. *ISA Transactions*, *31*(1), 25–33.
- Boiko, D. A., MacKnight, R., Kline, B., & Gomes, G. (2023a). Autonomous chemical research with large language models. *Nature*, *624*(7992), 570–578.
- Boiko, D. A., MacKnight, R., Kline, B., & Gomes, G. (2023b). Autonomous chemical research with large language models. *Nature*, *624*(7992), 570–578.
- Campelo, J. C., Capella, J. V., Ors, R., Peris, M., & Bonastre, A. (2022). IoT technologies in chemical analysis systems: Application to potassium monitoring in water. *Sensors*, *22*(3), 842.
- Cao, D.-S., Huang, J.-H., Liang, Y.-Z., Xu, Q.-S., & Zhang, L.-X. (2012). Tree-based ensemble methods and their applications in analytical chemistry. *TrAC Trends in Analytical Chemistry*, *40*, 158–167.
- Capella, J. V., Bonastre, A., Campelo, J.C., Ors, R., & Peris, M. (2020). IoT & environmental analytical chemistry: Towards a profitable symbiosis. *Trends in Environmental Analytical Chemistry*, *27*, e00095.
- Chan, H. S., Shan, H., Dahoun, T., Vogel, H., & Yuan, S. (2019). Advancing drug discovery via artificial intelligence. *Trends in Pharmacological Sciences*, *40*(8), 592–604.
- Coley, C. W., Thomas III, D. A., Lummiss, J. A., Jaworski, J. N., Breen, C. P., Schultz, V., Hart, T.,Fishman, J. S., Rogers, L., Gao, H., & others. (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science*, *365*(6453), eaax1566.
- Dahl, G. E., Jaitly, N., & Salakhutdinov, R. (2014). Multi-task neural networks for QSAR predictions. *arXiv Preprint arXiv:1406.1231*.
- Dimitrov, T., Kreisbeck, C., Becker, J. S., Aspuru-Guzik, A., & Saikin, S. K. (2019). Autonomous

molecular design: Then and now. *ACS Applied Materials & Interfaces*, *11*(28), 24825–24836.

- Dragone, V., Sans, V., Henson, A. B., Granda, J. M., & Cronin, L. (2017). An autonomous organic reaction search engine for chemical reactivity. *Nature Communications*, *8*(1), 15733.
- Duvenaud, D. K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., & Adams, R.P.(2015). Convolutional networks on graphs for learning molecular fingerprints. *Advances in Neural Information Processing Systems*, *28*.
- E. J. Corey & W. Todd Wipke. (1969). Computer-Assisted Design of Complex Organic Syntheses: Pathways for molecular synthesis can be devised with a computer and equipment for graphical communication. *Science*,*166*(3902),178192.https://doi.org/10.1126/science.166.3902.178.
- Elyashberg, M., & Williams, A. J. (2015). *Computer-based structure elucidation from spectral data* (Vol. 89). Springer.
- Galushko, A. S., Boiko, D. A., Pentsak, E. O., Eremin, D. B., & Ananikov, V.P. (2023). Time-resolved formation and operation maps of pd catalysts suggest a key role of single atom centers in cross-coupling. *Journal of the American Chemical Society*, *145*(16), 9092–9103.
- Gao, H., Struble, T. J., Coley, C. W., Wang, Y., Green, W. H., & Jensen, K. F. (2018). Using machine learning to predict suitable conditions for organic reactions. *ACS Central Science*, *4*(11), 1465–1476.
- Gasteiger, J., & Zupan, J. (1993). Neuronale netze in der Chemie. *Angewandte Chemie*, *105*(4), 510–536.
- 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. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, *4*(2), 268–276.
- Granda, J. M., Donina, L., Dragone, V., Long, D.-L., & Cronin, L. (2018). Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature*, *559*(7714), 377–381.
- Griffen, E. J., Dossetter, A. G., & Leach, G. (2020). Chemists: AI is here; unite to get the benefits. *Journal of Medicinal Chemistry*, *63*(16),8695–8704.
- Grisafi, A., Fabrizio, A., Meyer, B., Wilkins, D. M., Corminboeuf, C., & Ceriotti, M. (2018). Transferable machine-learning model of the electron density. *ACS Central Science*, *5*(1), 57–64.
- Hansch, C., & Fujita, T. (1964). p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal of the American Chemical Society*, *86*(8), 1616–1626.
- Häse, F., Roch, L. M., Kreisbeck, C., & Aspuru-Guzik, A. (n.d.). *Supplementary Information Phoenics: A Bayesian Optimizer for Chemistry*.
- Heller, S., McNaught, A., Stein, S., Tchekhovskoi, D., & Pletnev, I. (2013). InChI-the worldwide chemical structure identifier standard. *Journal of Cheminformatics*, *5*, 1–9.
- Ivanenkov, Y. A., Polykovskiy, D., Bezrukov, D., Zagribelnyy, B., Aladinskiy, V., Kamya, P., Aliper, A., Ren, F., & Zhavoronkov, A. (2023). Chemistry42: An AI-driven platform for molecular design and optimization. *Journal of Chemical Information and Modeling*, *63*(3), 695–701.
- Jin, W., Coley, C., Barzilay, R., & Jaakkola, T. (2017). Predicting organic reaction outcomes with weisfeiler-lehman network. *Advances in Neural Information Processing Systems*, *30*.
- Kim, S., Chen, J., Cheng, T., Gindulyte, A.,He, J.,He, S.,Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., & others. (2019). PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Research*, *47*(D1), D1102–D1109.
- Leach, A. R., & Lewis, R. A. (1994). A ring-bracing approach to computer- assisted ligand design. *Journal of Computational Chemistry*, *15*(2), 233–240.
- Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., & Svetnik, V. (2015). Deep neural nets as a method for quantitative structure–activity relationships. *Journal of Chemical Information and Modeling*, *55*(2), 263–274.
- Mahjour, B., Hoffstadt, J., & Cernak, T. (2023). Designing chemical reaction arrays using phactor and ChatGPT. *Organic Process Research & Development*, *27*(8), 1510–1516.
- Mai, H., Le, T. C., Chen, D., Winkler, D. A.,& Caruso, R.A.(2022).Machine learning for electrocatalyst and photocatalyst design and discovery. *Chemical Reviews*, *122*(16), 13478–13515.
- Maryasin, B., Marquetand, P., & Maulide, N. (2018). Machine learning for organic synthesis: Are robots replacing chemists? *Angewandte Chemie (International Ed. in English)*, *57*(24), 6978.

- McCardle, K. (2022). Accelerated mass spectra analysis. *Nature Computational Science*, *2*(9).
- Mikulak-Klucznik, B., Gołębiowska, P., Bayly, A. A., Popik, O., Klucznik, T., Szymkuć, S., Gajewska, E. P., Dittwald, P., Staszewska- Krajewska, O., Beker, W., & others. (2020). Computational planning of the synthesis of complex natural products. *Nature*, *588*(7836), 83–88.
- Moosavi, S. M., Jablonka, K. M., & Smit, B. (2020). The role of machine learning in the understanding and design of materials. *Journal of the American Chemical Society*, *142*(48), 20273–20287.
- Muratov, E. N., Bajorath, J., Sheridan, R. P., Tetko, I. V., Filimonov, D., Poroikov, V., Oprea, T. I., Baskin,I. I., Varnek, A., Roitberg, A., & others. (2020). QSAR without borders. *Chemical Society Reviews*, *49*(11), 3525–3564.
- Nourani, V., Elkiran, G., & Abba, S. (2018). Wastewater treatment plant performance analysis using artificial intelligence–an ensemble approach. *Water Science and Technology*, *78*(10), 2064–2076.
- Olivecrona, M., Blaschke, T., Engkvist, O., & Chen, H. (2017). Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, *9*, 1–14.
- Pavel, A., Saarimäki, L. A., Möbus, L., Federico, A., Serra, A., & Greco, D. (2022). The potential of a data centred approach & knowledge graph data representation in chemical safety and drug design. *Computational and Structural Biotechnology Journal*, *20*, 4837–
4849.
- Plehiers, P. P., Marin, G. B., Stevens, C. V., & Van Geem, K. M. (2018). Automated reaction database and reaction network analysis: Extraction of reaction templates using cheminformatics. *Journal of Cheminformatics*, *10*, 1–18.
- Putin, E., Asadulaev, A., Ivanenkov, Y., Aladinskiy, V., Sanchez- Lengeling, B., Aspuru-Guzik, A., & Zhavoronkov, A. (2018). Reinforced adversarial neural computer for de novo molecular design. *Journal of Chemical Information and Modeling*, *58*(6), 1194–1204.
- Reker, D., Schneider, P., & Schneider, G. (2016). Multi-objective active machine learning rapidly improves structure–activity models and reveals new protein–protein interaction inhibitors. *Chemical Science*, *7*(6), 3919–3927.
- Roch, L. M., Häse, F., Kreisbeck, C., Tamayo-Mendoza, T., Yunker, L. P., Hein, J. E., & Aspuru-Guzik, A. (2018). ChemOS: orchestrating autonomous experimentation.*Science Robotics*, *3*(19), eaat5559.
- Rogers, D., & Hahn, M. (2010). Extended- connectivity fingerprints. *Journal of Chemical Information and Modeling*, *50*(5), 742–754.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215.
- Sanchez-Lengeling, B., & Aspuru-Guzik, A. (2018). Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, *361*(6400), 360–365.
- Schlexer Lamoureux, P., Winther, K. T., Garrido Torres, J. A., Streibel, V., Zhao, M., Bajdich, M., Abild- Pedersen, F., & Bligaard, T. (2019). Machine learning for computational heterogeneous catalysis. *ChemCatChem*, *11*(16), 3581–3601.
- Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Hunter, C. A., Bekas, C., & Lee, A. A. (2019). Molecular transformer: A model for uncertainty-calibrated chemical reaction prediction. *ACS Central Science*, *5*(9), 1572–1583.
- Segler, M. H., Preuss, M., & Waller, M. P. (2018). Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*, *555*(7698), 604–610.
- Seifrid, M., Pollice, R., Aguilar-Granda, A., Morgan Chan, Z., Hotta, K.,Ser, C. T., Vestfrid, J., Wu, T. C., & Aspuru-Guzik, A. (2022). Autonomous chemical experiments: Challenges and perspectives on establishing a self- driving lab. *Accounts of Chemical Research*, *55*(17), 2454–2466.
- Smith, J. S., Roitberg, A. E., & Isayev, O. (2018). Transforming computational drug discovery with machine learning and AI. In *ACS medicinal chemistry letters* (Vol. 9, Issue 11, pp. 1065–1069). ACS Publications.
- Stach, E., DeCost, B., Kusne, A. G., Hattrick-Simpers, J., Brown, K. A., Reyes, K. G., Schrier, J., Billinge, S., Buonassisi, T., Foster, I., & others. (2021). Autonomous experimentation systems for materials development: A community perspective. *Matter*, *4*(9), 2702–2726.

- Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., & others. (2020). A deep learning approach to antibiotic discovery. *Cell*, *180*(4), 688–702.
- Struebing, H., Ganase, Z., Karamertzanis, P. G., Siougkrou, E., Haycock, P., Piccione, P. M., Armstrong, A., Galindo, A., & Adjiman, C. S. (2013). Computer-aided molecular design of solvents for accelerated reaction kinetics. *Nature Chemistry*, *5*(11), 952–957.
- Szymkuć, S., Gajewska, E. P., Klucznik, T., Molga, K., Dittwald, P., Startek, M., Bajczyk, M., & Grzybowski, B. A. (2016). Computer-assisted synthetic planning: The end of the beginning. *Angewandte Chemie International Edition*, *55*(20), 5904–5937.
- Toyao, T., Maeno, Z., Takakusagi, S., Kamachi, T., Takigawa, I., & Shimizu, K. (2019). Machine learning for catalysis informatics: Recent applications and prospects. *Acs Catalysis*, *10*(3), 2260–2297.
- Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li,
- B., Madabhushi, A., Shah, P., Spitzer, M., & others. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, *18*(6), 463–477.
- Varnek, A., & Baskin, I. (2012). Machine learning methods for property prediction in chemoinformatics: Quo Vadis? *Journal of Chemical Information and Modeling*, *52*(6), 1413–1437.
- Villalba, M., Wollenhaupt, M., & Ravitz, O. (n.d.). *Predicting New Chemistry: Impact of High-Quality Training Data on Prediction of Reaction Outcomes. CAS Whitepapers. 2022*.
- Wallach, I., & Heifets, A. (2018). Most ligand-based classification benchmarks reward memorization rather than generalization. *Journal of Chemical Information and Modeling*, *58*(5), 916–932.
- Walters, W. P., & Murcko, M. (2020). Assessing the impact of generative AI on medicinal chemistry. *Nature Biotechnology*, *38*(2), 143–145.
- Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, *28*(1), 31–36.
- Westermayr, J., & Marquetand, P. (2020). Machine learning for electronically excited states of molecules. *Chemical Reviews*, *121*(16), 9873–9926.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos,L. B., Bourne, P. E., & others. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, *3*(1), 1–9.
- Wishart, D. S., Feunang, Y. D., Guo, A.C., Lo, E. J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., & others. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, *46*(D1), D1074–D1082.
- Yang, K., Swanson, K., Jin, W., Coley, C.,Eiden, P., Gao, H., Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., & others. (2019). Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, *59*(8), 3370–3388.
- Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., Terentiev, V. A., Polykovskiy, D. A., Kuznetsov, M. D., Asadulaev, A., & others. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, *37*(9), 1038–1040.
- Zheng, Z., Zhang, O., Borgs, C., Chayes, J. T., & Yaghi, O. M. (2023).ChatGPT chemistry assistant for text mining and the prediction of MOF synthesis. *Journal of the American Chemical Society*, *145*(32), 18048–18062.