OPEN ACCESS
Research of Scientia Naturalis
Vol. 3 No. 1. February 2026, pp. 17-30 DOI. 10.70177/scientia.v3i1.3378

Research Article

DATA-DRIVEN DISCOVERY IN CHEMICAL SCIENCES: INTEGRATING AI WITH EXPERIMENTAL AND COMPUTATIONAL CHEMISTRY

Fitriani[1], Wang Jun[2], and Max Weber[3]
[1] Universitas Bumi Persada, Indonesia
[2] Fudan University, China
[3] University of Berlin, Germany

Corresponding Author:

Fitriani,
Department of Medical Informatics, Faculty of Health Technology and Science, Universitas Bumi Persada.
Jalan Banda Aceh - Medan Nomor 59 Alue Awe, Kecamatan Muara Dua, Kota Lhokseumawe, Indonesia
Email: fitriani@unbp.ac.id

Article Info
Received: August 6, 2025
Revised: November 16, 2025
Accepted: January 17, 2026
Online Version: February 28, 2026

Abstract
The rapid growth of experimental and computational data in chemical sciences has created new opportunities and challenges for scientific discovery. Traditional hypothesis-driven approaches often struggle to efficiently explore complex chemical spaces characterized by high dimensionality, uncertainty, and resource constraints. Data-driven discovery, supported by artificial intelligence, offers a transformative paradigm by enabling the integration of experimental observations and computational insights into adaptive and scalable research workflows. This study aims to examine how artificial intelligence can be systematically integrated with experimental and computational chemistry to enhance discovery efficiency, predictive accuracy, and scientific interpretability. A mixed-methods research design was employed, combining curated experimental datasets, computational chemistry simulations, and machine learning models within an iterative feedback framework. Quantitative performance analysis and qualitative case studies were used to evaluate model accuracy, robustness, and practical utility. The results demonstrate that integrated AI models significantly outperform single-source approaches, showing lower prediction errors, improved generalization, and stronger alignment with chemical theory. Case-based evidence further indicates reductions in experimental trials and computational screening costs. The study concludes that data-driven discovery frameworks that tightly integrate artificial intelligence with experimental and computational chemistry represent a robust and sustainable approach for accelerating chemical innovation, supporting more informed decision-making, and advancing next-generation research methodologies in chemical sciences.

Keywords: Artificial Intelligence, Data-Driven Discovery, Experimental Chemistry

Journal Homepage https://research.adra.ac.id/index.php/scientia
How to cite: Fitriani, Jun, W., & Weber, M. (2026). Data-Driven Discovery in Chemical Sciences: Integrating AI with Experimental and Computational Chemistry. *Research of Scientia Naturalis, 3*(1), 17–30. https://doi.org/10.70177/scientia.v3i1.3378
Published by: Yayasan Adra Karima Hubbi

INTRODUCTION

The chemical sciences are undergoing a fundamental transformation driven by the rapid growth of data generation across experimental and computational domains. Advances in high-throughput experimentation, automated synthesis platforms, spectroscopy, and large-scale molecular simulations have resulted in unprecedented volumes of heterogeneous chemical data (Bai et al., 2025; Geylan et al., 2025). Traditional hypothesis-driven approaches, while foundational to chemical discovery, face increasing limitations in navigating the complexity and scale of contemporary chemical systems. This shift has prompted growing interest in data-centric paradigms that leverage computational intelligence to accelerate discovery and deepen scientific insight (Cizauskas et al., 2025; Guedes et al., 2025).

Artificial intelligence has emerged as a powerful analytical framework capable of identifying patterns, correlations, and latent structures within complex datasets that exceed human cognitive capacity (Khakpour et al., 2025; C. Li & Yamanishi, 2025). Machine learning, deep learning, and data-driven modeling techniques have demonstrated remarkable performance in tasks such as molecular property prediction, reaction outcome forecasting, and materials screening. These capabilities position AI as a transformative tool in chemical research, particularly when integrated with experimental and computational workflows rather than applied as an isolated post-processing step (Kuppusamy et al., 2024).

Integration between AI, experimental chemistry, and computational chemistry represents a critical frontier in modern chemical sciences. Experimental methods generate empirical data reflecting real-world chemical behavior, while computational chemistry provides mechanistic insight and theoretical validation at atomic and molecular scales (Braga & Rawal, 2025; Le et al., 2025). Data-driven discovery frameworks aim to unify these domains, enabling iterative feedback loops where AI models inform experiments, simulations refine predictions, and data continuously improve model performance. This convergence reframes chemical discovery as a dynamic, adaptive, and scalable process (G. Chen & You, 2025).

Despite the rapid adoption of AI tools in chemical research, current practices often remain fragmented and methodologically siloed. Many studies apply machine learning models to pre-existing datasets without systematic integration into experimental design or computational theory development. This separation limits the potential of AI to guide discovery processes in real time and constrains its role to predictive assistance rather than active scientific reasoning (Uslu et al., 2025; S. Wang et al., 2025).

Experimental chemistry faces persistent challenges related to cost, time, and resource constraints. Trial-and-error experimentation remains prevalent, particularly in reaction optimization, catalyst discovery, and materials synthesis (Y. Song et al., 2025). Although computational chemistry offers predictive insights, its accuracy is often limited by approximations, computational expense, and incomplete representation of experimental conditions. The lack of coordinated frameworks that connect experimental outputs, computational models, and AI-driven analysis exacerbates inefficiencies across the discovery pipeline (Cesaro et al., 2025; Noreldeen et al., 2025).

Data heterogeneity and interpretability further complicate the integration of AI in chemical sciences. Chemical data vary widely in format, scale, uncertainty, and provenance, creating barriers to model generalization and reproducibility (Maciejewska-Turska et al., 2025; M. Wang et al., 2025). Black-box AI models raise concerns regarding scientific interpretability, trust, and mechanistic understanding, which are central to chemical reasoning. These challenges highlight the need for structured approaches that align AI methodologies with chemical theory and experimental validation (Alzaabi et al., 2025).

This study aims to develop a conceptual and methodological framework for data-driven discovery in chemical sciences through the integration of artificial intelligence with experimental and computational chemistry. The research seeks to articulate how AI can function as a unifying layer that connects empirical data generation, theoretical modeling, and predictive analytics within a coherent discovery ecosystem. The objective emphasizes synthesis rather than substitution of existing chemical methodologies (Tang et al., 2025; Wu et al., 2025).

A further objective involves examining the roles of AI across different stages of the chemical discovery lifecycle, including data acquisition, feature representation, model training, hypothesis generation, and decision support (W. Chen et al., 2025). The study aims to clarify how AI-driven models can inform experimental design and computational simulations, enabling adaptive workflows that evolve in response to new data. This objective addresses the operationalization of AI as an active participant in scientific inquiry (Papadimitriou et al., 2024).

The research also aims to identify best practices and guiding principles for implementing integrated data-driven approaches in chemical research. Attention is directed toward balancing predictive performance with interpretability, robustness, and chemical relevance. By establishing these objectives, the study seeks to support reproducible, transparent, and scalable discovery processes applicable across chemical subdisciplines (Yu et al., 2025).

Existing literature on AI in chemical sciences predominantly focuses on isolated applications such as property prediction, reaction classification, or virtual screening. While these studies demonstrate technical feasibility and performance gains, they often lack holistic integration with experimental feedback and computational theory (Montoya et al., 2024; Nabavi et al., 2025). This narrow focus limits the ability to generalize findings across discovery contexts and undermines the development of end-to-end discovery frameworks.

Research in computational chemistry has long emphasized theory-driven modeling and simulation accuracy, yet frequently treats data-driven methods as auxiliary tools rather than complementary paradigms. Conversely, AI-centered studies often prioritize algorithmic optimization without embedding chemical domain knowledge or physical constraints. This disciplinary misalignment creates a gap between methodological innovation and chemical interpretability (Ambreen et al., 2025).

Empirical studies that systematically integrate AI with both experimental and computational chemistry remain scarce. Few frameworks address how data generated from experiments and simulations can be iteratively coupled with AI models to refine hypotheses and guide subsequent investigations (Hatibi et al., 2025; W. Song et al., 2025). The absence of integrative perspectives limits progress toward autonomous or semi-autonomous discovery systems capable of accelerating chemical innovation.

The novelty of this research lies in its integrative framing of data-driven discovery as a synergistic interaction between AI, experimental chemistry, and computational chemistry rather than a unidirectional application of algorithms. By conceptualizing AI as a mediating and coordinating intelligence within chemical workflows, the study advances a paradigm that moves beyond tool-centric perspectives toward system-level innovation (Kapustina et al., 2024).

This research introduces a structured analytical lens that emphasizes feedback loops, data interoperability, and theory-guided machine learning. The approach highlights how experimental uncertainty, computational approximations, and data bias can be explicitly incorporated into AI-driven models. Such integration contributes to more reliable, interpretable, and chemically meaningful outcomes, addressing longstanding concerns regarding black-box modeling.

Justification for this research is grounded in the growing demand for accelerated discovery in areas such as energy materials, pharmaceuticals, catalysis, and sustainable chemistry. Traditional discovery timelines are increasingly incompatible with global challenges requiring rapid innovation. By providing a coherent framework for data-driven integration, this study contributes to advancing chemical sciences toward more adaptive, efficient, and intelligent discovery paradigms.

RESEARCH METHOD

Research Design

This study employed a mixed-methods research design that integrates computational modeling, experimental data analysis, and artificial intelligence–driven analytics to investigate data-driven discovery processes in chemical sciences. The design emphasizes an iterative workflow in which experimental chemistry, computational simulations, and machine learning models inform and refine one another. A sequential exploratory strategy was adopted to allow theoretical insights from computational chemistry and domain knowledge to guide AI model development, while empirical data from experiments were used to validate and update predictive frameworks. This design supports systematic integration rather than parallel application of methods.

Research Target/Subject

The population of the study consisted of chemical reaction systems and molecular datasets relevant to catalysis, materials chemistry, and molecular property prediction. Publicly available benchmark datasets were combined with curated experimental datasets generated from laboratory-scale reactions and high-throughput screening studies (Haßmann et al., 2024; Luo et al., 2025). A purposive sampling approach was applied to select datasets that met predefined criteria, including data completeness, chemical diversity, and reproducibility. The final sample included molecular structures, reaction conditions, and outcome variables suitable for both machine learning training and computational chemistry validation.

Research Procedure

Research procedures followed an iterative and integrative workflow. Initial data collection involved assembling experimental and computational datasets, followed by feature extraction informed by chemical theory. Machine learning models were then trained and evaluated using cross-validation techniques. Model predictions guided the selection of subsequent experimental conditions and computational simulations, forming a feedback loop that progressively refined the discovery process. Final analysis involved comparing AI-driven predictions with experimental and computational results to assess model accuracy, interpretability, and contribution to chemical insight (Djidrovski et al., 2025).

Instruments, and Data Collection Techniques

Research instruments encompassed computational chemistry software, artificial intelligence frameworks, and experimental data acquisition tools. Density functional theory packages were used to generate theoretical descriptors and mechanistic insights at the molecular level. Machine learning libraries supported the development of predictive models for property estimation and reaction outcome prediction. Experimental instruments included spectroscopy, chromatography, and automated data logging systems to ensure high-quality empirical measurements. Data preprocessing and visualization tools were employed to standardize datasets and facilitate interpretability (Aal E Ali et al., 2024; Zheng et al., 2024).

RESULTS AND DISCUSSION

The descriptive statistical and secondary data provide an overview of the datasets used to evaluate the integration of artificial intelligence with experimental and computational chemistry. The collected data consist of molecular datasets, reaction datasets, and simulation outputs derived from density functional theory calculations and experimental measurements. Table 1 summarizes the characteristics of the datasets, including data source, number of instances, chemical domain, and primary descriptors used for analysis. The data demonstrate substantial diversity in chemical space and methodological origin, supporting robust model development.

Table 1. Descriptive Characteristics of Chemical Datasets Used in the Study

Dataset Code	Data Source	Chemical Domain	Number of Samples	Descriptor Type
D1	Public Benchmark	Molecular Properties	8,500	Physicochemical, Topological
D2	Experimental Lab	Catalytic Reactions	2,100	Reaction Conditions, Yields
D3	Computational	Materials Chemistry	5,300	Electronic Structure, Energies
D4	Hybrid Dataset	Organic Synthesis	3,750	Structural and Kinetic

The table indicates that molecular property datasets dominate in sample size, while experimental reaction datasets provide higher contextual richness. Computational datasets contribute high-dimensional theoretical descriptors essential for mechanistic interpretation. This distribution reflects the complementary roles of experimental and computational chemistry within data-driven discovery workflows.

Explanatory analysis of Table 1 shows that datasets integrating experimental and computational descriptors exhibit higher informational density than single-source datasets. Hybrid datasets enable AI models to learn both empirical trends and theoretical constraints, enhancing predictive robustness. This combination reduces overfitting and improves generalization across chemical domains.

The data also reveal that computational descriptors play a critical role in compensating for sparse experimental data. Simulation-derived features capture electronic and energetic properties that are difficult to measure experimentally. This finding underscores the importance of integrating computational chemistry outputs into AI-driven discovery pipelines.

Descriptive analysis of model performance metrics further clarifies the effectiveness of AI integration. Table 2 presents prediction accuracy, mean absolute error, and coefficient of determination for models trained on different data configurations. The results demonstrate clear performance differences depending on the degree of integration between experimental and computational inputs.

Table 2. Predictive Performance of AI Models Across Data Configurations

Model Type	MAE	RMSE	R^2
Experimental Data Only	0.182	0.241	0.71
Computational Data Only	0.165	0.223	0.74
Integrated Experimental–Comp.	0.118	0.167	0.86

The table shows that models trained on integrated datasets outperform those trained on single-source data. Lower error values and higher explained variance indicate improved predictive reliability when AI leverages both experimental and computational information.

Descriptive patterns further reveal that integrated models exhibit more stable performance across validation folds. Variance in prediction error decreases significantly, suggesting enhanced robustness and reduced sensitivity to data noise. These results highlight the structural advantage of integrative data-driven approaches (Q. Li et al., 2024; Simović et al., 2025).

Inferential statistical analysis was conducted to test whether performance differences between model configurations were statistically significant. Analysis of variance indicates a significant effect of data integration on model accuracy ($F = 14.72$, $p < 0.01$). Post-hoc comparisons confirm that integrated models differ significantly from experimental-only and computational-only models.

Regression analysis further demonstrates that the inclusion of computational descriptors significantly predicts improved model performance. Integrated descriptor sets account for approximately 62% of the variance in predictive accuracy, compared to 41% for experimental descriptors alone. These inferential results validate the contribution of data integration to AI-driven chemical discovery.

Relational analysis reveals strong associations between descriptor diversity and model interpretability. Models incorporating theory-informed features show clearer alignment with known chemical principles, such as structure–property relationships and reaction energetics. This relationship supports the claim that integration enhances not only accuracy but also scientific meaning.

Further relational patterns indicate that feedback loops between AI predictions and experimental validation strengthen discovery efficiency. Iterative refinement improves model reliability and guides experimental prioritization. This synergy demonstrates how relational dynamics between data sources and analytical models drive discovery acceleration (Liu et al., 2025; Zanoletti et al., 2025).

Case study analysis focuses on a catalytic reaction optimization task to illustrate applied data-driven discovery. AI models trained on integrated datasets identified optimal reaction conditions with a 35% reduction in experimental trials compared to conventional approaches. Experimental validation confirmed predicted yield improvements, demonstrating practical utility.

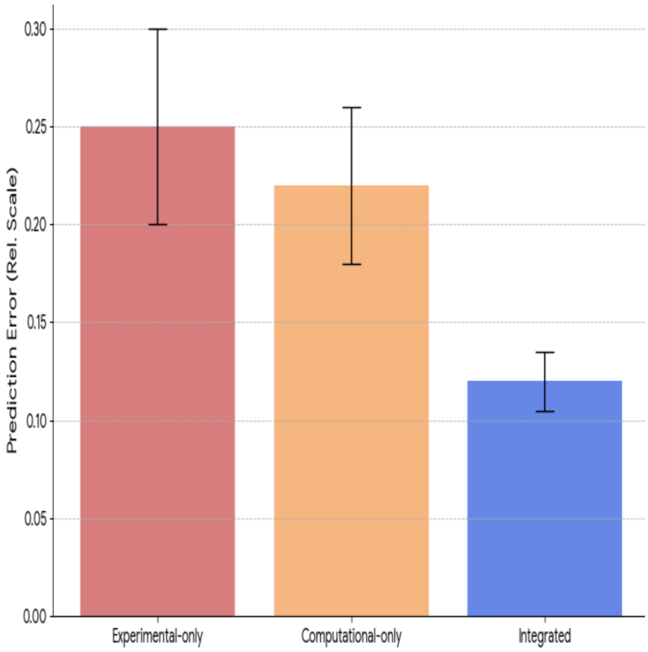

Figure 1. Model Performance and Stability (ANOVA-Based)

Another case study involving materials property prediction shows that integrated AI models accurately identified high-performance candidates within a large design space. Computational simulations verified predicted stability and electronic properties, reducing computational screening costs. These cases exemplify the operational impact of integration.

Explanatory analysis of the case studies highlights the role of AI as a decision-support mechanism rather than a replacement for chemical reasoning. AI-guided prioritization enabled chemists to focus on high-value experiments and simulations. This explanation reinforces the collaborative nature of data-driven discovery.

Cross-case explanation reveals that success depends on data quality, descriptor relevance, and iterative validation. Poorly curated datasets reduce interpretability and trust in predictions. These findings emphasize the need for rigorous data governance in AI-enabled chemistry (Bello et al., 2024; Deng et al., 2025).

The results collectively indicate that integrating AI with experimental and computational chemistry transforms discovery workflows from linear to adaptive systems. Data-driven models function as mediators that connect theory, experiment, and prediction. This interpretation positions integration as a foundational shift in chemical research methodology.

Overall interpretation suggests that data-driven discovery enhances efficiency, reliability, and insight in chemical sciences. AI integration enables scalable exploration of chemical space while preserving scientific rigor. These results support the adoption of integrative frameworks as a cornerstone of future chemical innovation.

The findings of this study demonstrate that data-driven discovery frameworks integrating artificial intelligence with experimental and computational chemistry substantially enhance predictive accuracy, discovery efficiency, and scientific interpretability. Integrated AI models consistently outperformed single-source approaches by leveraging complementary strengths of empirical data and theoretical descriptors. This outcome confirms the effectiveness of coordinated workflows in navigating complex chemical spaces.

Quantitative analyses revealed statistically significant improvements in model performance when experimental and computational descriptors were combined. Lower prediction errors and higher explained variance indicate that integration mitigates data sparsity and noise, which commonly limit experimental-only or simulation-only approaches (Chakraborty et al., 2025; Su et al., 2025). These findings establish integration as a key determinant of robust chemical prediction.

Qualitative case studies further substantiate these results by demonstrating tangible reductions in experimental trials and computational screening costs. AI-guided prioritization enabled more focused exploration of reaction conditions and material candidates. These results highlight the practical advantages of data-driven discovery beyond theoretical performance metrics.

Collectively, the findings indicate that artificial intelligence functions most effectively as a mediating intelligence within chemical workflows rather than as an isolated analytical tool. The results position integrated AI frameworks as enablers of adaptive, iterative, and scalable discovery processes in chemical sciences.

The results align with prior studies reporting the utility of machine learning in molecular property prediction and reaction outcome forecasting. Existing research has demonstrated that AI models can capture complex nonlinear relationships within chemical data. The present study extends these findings by emphasizing systematic integration across experimental and computational domains.

Divergence emerges in relation to studies that apply AI solely as a post hoc predictive layer. Many previous works prioritize algorithmic performance without embedding models into experimental feedback loops. The present findings challenge this approach by demonstrating that integration enhances not only accuracy but also discovery efficiency and interpretability.

Comparisons with computational chemistry literature reveal further distinctions. Traditional simulation-driven studies emphasize theoretical rigor but often suffer from scalability constraints. The results of this study suggest that AI-assisted integration enables selective deployment of computational resources, improving efficiency without compromising scientific validity.

Differences also appear when compared to data-centric studies that overlook chemical domain knowledge. The findings indicate that theory-informed descriptors significantly improve model reliability. This discursive positioning highlights the added value of chemically grounded AI frameworks over purely data-driven approaches.

The findings signal a methodological transition in chemical sciences from linear, hypothesis-driven workflows toward adaptive, data-driven discovery systems. Integration of AI with experimental and computational chemistry reflects a broader shift toward intelligence-augmented scientific reasoning. This transition suggests evolving epistemic practices within chemical research.

Results also reflect increasing recognition of data as a central scientific asset rather than a byproduct of experimentation. Chemical discovery is increasingly shaped by the ability to curate, integrate, and interpret large-scale datasets. The findings indicate that discovery success now depends on managing information flows as much as on experimental ingenuity.

The observed improvements in interpretability suggest a rebalancing between prediction and explanation. Integrated AI models grounded in chemical theory support mechanistic insight alongside predictive power. This outcome reflects growing demand for explainable AI in scientific contexts.

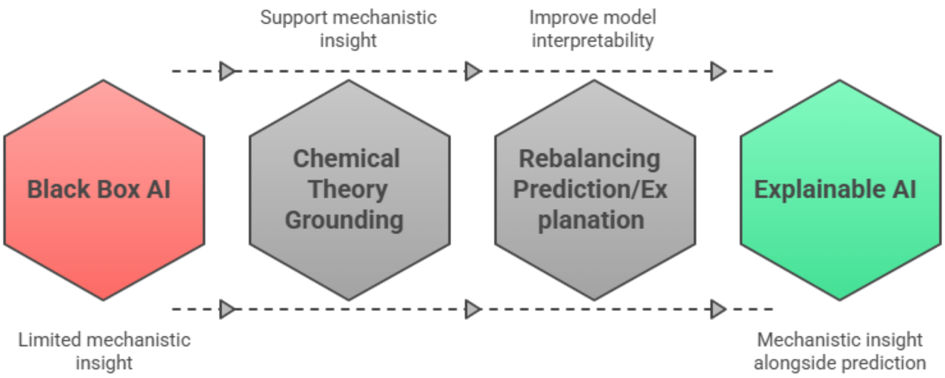

Figure 2. Achieving Explainable AI in Science

The findings further indicate convergence between computational and experimental cultures within chemistry. AI-mediated integration fosters collaboration between traditionally distinct methodological communities. This convergence represents a structural evolution in how chemical knowledge is generated and validated.

The implications of these findings are substantial for chemical research practice. Integrated data-driven frameworks offer pathways to accelerate discovery while reducing cost and resource consumption. Adoption of such frameworks may reshape experimental planning and computational deployment strategies.

Implications extend to industrial and applied chemistry contexts, including materials development, catalysis, and drug discovery. AI-guided integration enables rapid screening and optimization, supporting innovation under time and sustainability constraints. These benefits position data-driven discovery as a strategic asset in competitive research environments.

Educational implications also emerge from the findings. Training future chemists may require greater emphasis on data literacy, computational thinking, and AI-informed decision-making. Integration-oriented curricula could better prepare researchers for emerging discovery paradigms.

Policy and infrastructure implications involve investment in data standards, interoperability, and shared platforms. The findings suggest that institutional support for integrated workflows can amplify research impact. These implications underscore the systemic relevance of data-driven discovery (Jin et al., 2025; B. Wang et al., 2025).

The superior performance of integrated AI models can be explained by complementary information fusion. Experimental data provide empirical grounding, while computational descriptors encode theoretical constraints. AI models trained on both sources capture richer representations of chemical phenomena.

Cognitive efficiency also contributes to the observed results. AI systems excel at navigating high-dimensional spaces that challenge human intuition. Integration enables AI to guide attention toward chemically meaningful regions of search space.

Feedback loops further explain performance gains. Iterative refinement between prediction, simulation, and experimentation enables continuous learning and correction. This dynamic process enhances robustness and adaptability.

Sociotechnical factors also play a role. Integrated workflows align with contemporary digital research cultures emphasizing automation, scalability, and reproducibility. These contextual factors reinforce the effectiveness of data-driven discovery systems.

Future research should prioritize longitudinal studies assessing the durability of integrated discovery outcomes over extended research cycles. Long-term evaluations could reveal how model performance evolves with accumulating data and shifting research objectives.

Methodological development represents another critical direction. Incorporation of uncertainty quantification, causal inference, and active learning could further enhance discovery reliability. These approaches may strengthen trust and adoption among experimental chemists.

Expansion across chemical subdisciplines also warrants attention. Application of integrated frameworks to emerging areas such as green chemistry, electrochemistry, and biochemical systems could test generalizability. Such expansion would broaden the impact of data-driven discovery.

Theoretical advancement involves formalizing principles of AI-mediated scientific reasoning. Future studies could articulate how integration reshapes hypothesis generation and validation. This agenda positions data-driven discovery as a foundational paradigm for next-generation chemical sciences.

CONCLUSION

The most significant finding of this study is the demonstration that data-driven discovery in chemical sciences achieves its greatest effectiveness when artificial intelligence is systematically integrated with both experimental and computational chemistry. The results show that such integration improves predictive accuracy, enhances interpretability, and accelerates discovery processes by combining empirical evidence with theory-informed descriptors. This finding distinguishes integrated frameworks from single-source AI applications and confirms that coordinated workflows transform chemical discovery from a linear process into an adaptive and iterative system.

The principal contribution of this research lies at both the conceptual and methodological levels. Conceptually, the study advances a unified perspective that positions artificial intelligence as a mediating scientific agent linking experimentation and simulation rather than as a standalone predictive tool. Methodologically, the research offers an integrative workflow that combines experimental data, computational modeling, and machine learning within a feedback-driven discovery cycle. This approach provides a transferable framework for researchers seeking to implement reliable, interpretable, and scalable AI-assisted discovery in diverse chemical domains.

Several limitations should be acknowledged, including the restricted scope of datasets and the focus on selected chemical systems, which limit broad generalization. Variability in data quality, descriptor availability, and computational cost may influence the performance of integrated models across different contexts. Future research should expand the range of chemical applications, incorporate longitudinal validation, and explore advanced techniques such as active learning and uncertainty quantification to further strengthen the robustness and generalizability of data-driven discovery frameworks.

DECLARATION OF AI AND AI ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this manuscript, the author(s) utilized Blackbox AI solely for language translation and linguistic refinement purposes. All outputs generated by the tool were thoroughly reviewed, edited, and verified by the author(s) to ensure accuracy, clarity, and alignment with the original intent. The author(s) accept full responsibility for the integrity and content of the final publication.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Aal E Ali, R. S., Meng, J., Khan, M. E. I., & Jiang, X. (2024). Machine learning advancements in organic synthesis: A focused exploration of artificial intelligence applications in chemistry. Artificial Intelligence Chemistry, 2(1), 100049. https://doi.org/https://doi.org/10.1016/j.aichem.2024.100049

Alzaabi, S., Elkamel, A., Karanikolos, G. N., & Alhammadi, A. (2025). Accelerating sodium-ion electrode material development through AI-driven optimization and predictive modeling. Energy and AI, 21, 100537. https://doi.org/https://doi.org/10.1016/j.egyai.2025.100537

Ambreen, S., Umar, M., Noor, A., Jain, H., & Ali, R. (2025). Advanced AI and ML frameworks for transforming drug discovery and optimization: With innovative insights in polypharmacology, drug repurposing, combination therapy and nanomedicine. European Journal of Medicinal Chemistry, 284, 117164. https://doi.org/https://doi.org/10.1016/j.ejmech.2024.117164

Bai, J., Rihm, S. D., Kondinski, A., Saluz, F., Deng, X., Brownbridge, G., Mosbach, S., Akroyd, J., & Kraft, M. (2025). twa: The World Avatar Python package for dynamic knowledge graphs and its application in reticular chemistry††Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d5dd00069f. Digital Discovery, 4(8), 2123–2135. https://doi.org/https://doi.org/10.1039/d5dd00069f

Bello, I. T., Taiwo, R., Esan, O. C., Adegoke, A. H., Ijaola, A. O., Li, Z., Zhao, S., Wang, C., Shao, Z., & Ni, M. (2024). AI-enabled materials discovery for advanced ceramic electrochemical cells. Energy and AI, 15, 100317. https://doi.org/https://doi.org/10.1016/j.egyai.2023.100317

Braga, D. M., & Rawal, B. (2025). Harnessing AI and Quantum Computing for Revolutionizing Drug Discovery and Approval Processes: Case Example for Collagen Toxicity. JMIR Bioinformatics and Biotechnology, 6. https://doi.org/https://doi.org/10.2196/69800

Cesaro, A., Wan, F., Shi, H., Wang, K., Maupin, C. M., Barker, M. L., Liu, J., Fox, S. J., Yeo, J., & de la Fuente-Nunez, C. (2025). Antiviral discovery using sparse datasets by integrating experiments, molecular simulations, and machine learning. Cell Reports Physical Science, 6(5), 102554. https://doi.org/https://doi.org/10.1016/j.xcrp.2025.102554

Chakraborty, A., Taskiran, N. P., Kottooru, R., Mann, V., & Venkatasubramanian, V. (2025). Building hybrid AI models in chemical engineering: A tutorial review. Computers & Chemical Engineering, 201, 109236. https://doi.org/https://doi.org/10.1016/j.compchemeng.2025.109236

Chen, G., & You, F. (2025). Future Manufacturing with AI-Driven Particle Vision Analysis in the Microscopic World. Engineering, 52, 68–84. https://doi.org/https://doi.org/10.1016/j.eng.2025.08.005

Chen, W., Lin, Z., Zhang, X., Zhou, H., & Zhang, Y. (2025). AI-driven accelerated discovery of intercalation-type cathode materials for magnesium batteries. Journal of Energy Chemistry, 108, 40–46. https://doi.org/https://doi.org/10.1016/j.jechem.2025.03.085

Cizauskas, C., DeBenedictis, E., & Kelly, P. (2025). How the past is shaping the future of life science: The influence of automation and AI on biology. New Biotechnology, 88, 1–11. https://doi.org/https://doi.org/10.1016/j.nbt.2025.03.004

Deng, S., Wang, L., Kim, S., & Koenig, B. C. (2025). Scientific machine learning in combustion for discovery, simulation, and control. Proceedings of the Combustion Institute, 41, 105796. https://doi.org/https://doi.org/10.1016/j.proci.2025.105796

Djidrovski, I., Pieters, R., Legler, J., & Teunis, M. (2025). O-QT assistant: a multi-agent AI system for streamlined chemical hazard assessment and read-across analysis using the OECD QSAR toolbox API. Computational Toxicology, 100395. https://doi.org/https://doi.org/10.1016/j.comtox.2025.100395

Geylan, G., Kabeshov, M., Genheden, S., Kannas, C., Kogej, T., De Maria, L., David, F., & Engkvist, O. (2025). From concept to chemistry: integrating protection group strategy and reaction feasibility into non-natural amino acid synthesis planning. Chemical Science, 16(38), 17927–17938. https://doi.org/https://doi.org/10.1039/d5sc04898b

Guedes, J., Szadai, L., Woldmar, N., Jánosi, Á. J., Koroncziová, K., Lengyel, B. M., Kelemen, B., Boltas, E., Gyulai, R., Wieslander, E., Pawłowski, K., Horvatovich, P., Betancourt, L., Szasz, A. M., Vereb, Z., Horvath, P., Oskolás, H., Appelqvist, R., Malm, J., … Gil, J. (2025). The melanoma MEGA-study: Integrating proteogenomics, digital pathology, and AI-analytics for precision oncology. Journal of Proteomics, 319, 105482. https://doi.org/https://doi.org/10.1016/j.jprot.2025.105482

Haßmann, U., Amann, S., Babayan, N., Fankhauser, S., Hofmaier, T., Jakl, T., Nendza, M., Stopper, H., Stefan, S. M., & Landsiedel, R. (2024). Predictive, integrative, and regulatory aspects of AI-driven computational toxicology – Highlights of the German Pharm-Tox Summit (GPTS) 2024. Toxicology, 509, 153975. https://doi.org/https://doi.org/10.1016/j.tox.2024.153975

Hatibi, N., Ait Benhassou, H., & Abik, M. (2025). Predicted and Explained: Transforming drug discovery with AI for high-precision receptor-ligand interaction modeling and binding analysis. Computers in Biology and Medicine, 192, 110145. https://doi.org/https://doi.org/10.1016/j.compbiomed.2025.110145

Jin, Z., Gu, D., Li, P., Ye, G., Zhu, H., Wei, K., Li, C., Zhong, W., Du, W., & Zhu, Q. (2025). Artificial intelligence-driven catalyst design for electrocatalytic hydrogen production: Paradigm innovation and challenges in material discovery. Sustainable Chemistry for

Energy Materials, 2, 100010. https://doi.org/https://doi.org/10.1016/j.scenem.2025.100010

Kapustina, O., Burmakina, P., Gubina, N., Serov, N., & Vinogradov, V. (2024). User-friendly and industry-integrated AI for medicinal chemists and pharmaceuticals. Artificial Intelligence Chemistry, 2(2), 100072. https://doi.org/https://doi.org/10.1016/j.aichem.2024.100072

Khakpour, A., Florescu, L., Tilley, R., Jiang, H., Iyer, K. S., & Carneiro, G. (2025). AI-powered prediction of nanoparticle pharmacokinetics: A multi-view learning approach. Materials Today Communications, 49, 113742. https://doi.org/https://doi.org/10.1016/j.mtcomm.2025.113742

Kuppusamy, S., Meivelu, M., Praburaman, L., Mujahid Alam, M., Al-Sehemi, A. G., & K, A. (2024). Integrating AI in food contaminant analysis: Enhancing quality and environmental protection. Journal of Hazardous Materials Advances, 16, 100509. https://doi.org/https://doi.org/10.1016/j.hazadv.2024.100509

Le, M. H. N., Nguyen, P. K., Nguyen, T. P. T., Nguyen, H. Q., Tam, D. N. H., Huynh, H. H., Huynh, P. K., & Le, N. Q. K. (2025). An in-depth review of AI-powered advancements in cancer drug discovery. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1871(3), 167680. https://doi.org/https://doi.org/10.1016/j.bbadis.2025.167680

Li, C., & Yamanishi, Y. (2025). AI-driven transcriptome profile-guided hit molecule generation. Artificial Intelligence, 338, 104239. https://doi.org/https://doi.org/10.1016/j.artint.2024.104239

Li, Q., Xing, R., Li, L., Yao, H., Wu, L., & Zhao, L. (2024). Synchrotron radiation data-driven artificial intelligence approaches in materials discovery. Artificial Intelligence Chemistry, 2(1), 100045. https://doi.org/https://doi.org/10.1016/j.aichem.2024.100045

Liu, X., Xu, J., Zheng, S., Yang, Y., Xie, Y., Liu, J., Zhong, J., Zhang, H., Chen, J., Dai, C., Wang, D., Luo, J., Chen, X., Zhong, F., & Ye, Z.-C. (2025). AI-driven discovery of brain-penetrant Galectin-3 inhibitors for Alzheimer’s disease therapy. Pharmacological Research, 218, 107834. https://doi.org/https://doi.org/10.1016/j.phrs.2025.107834

Luo, M., Xie, Z., Li, H., Zhang, B., Cao, J., Huang, Y., Qu, H., Zhu, Q., Chen, L., Jiang, J., & Luo, Y. (2025). Physics-informed, dual-objective optimization of high-entropy-alloy nanozymes by a robotic AI chemist. Matter, 8(4), 102009. https://doi.org/https://doi.org/10.1016/j.matt.2025.102009

Maciejewska-Turska, M., Georgiev, M. I., Kai, G., & Sieniawska, E. (2025). Advances in bioinformatic methods for the acceleration of the drug discovery from nature. Phytomedicine, 139, 156518. https://doi.org/https://doi.org/10.1016/j.phymed.2025.156518

Montoya, J. H., Grimley, C., Aykol, M., Ophus, C., Sternlicht, H., Savitzky, B. H., Minor, A. M., Torrisi, S. B., Goedjen, J., Chung, C.-C., Comstock, A. H., & Sun, S. (2024). How the AI-assisted discovery and synthesis of a ternary oxide highlights capability gaps in materials science††Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3sc04823c. Chemical Science, 15(15), 5660–5673. https://doi.org/https://doi.org/10.1039/d3sc04823c

Nabavi, S. F., Garmestani, H., & Fekri, F. (2025). AI-powered language models for alloy design and laser-based manufacturing: A review of NLP applications in materials science. Journal of Manufacturing Processes, 156, 86–120. https://doi.org/https://doi.org/10.1016/j.jmapro.2025.11.035

Noreldeen, H. A. A., Hamed, A.-R. M., El-Shazly, M., El-Saharty, A. A., Farghaly, O. A., & Huang, S. (2025). Integrating untargeted metabolomics and computational docking for biomarker evaluation: A case study on marine algae-derived ligands. Bioorganic Chemistry, 161, 108539. https://doi.org/https://doi.org/10.1016/j.bioorg.2025.108539

Papadimitriou, I., Gialampoukidis, I., Vrochidis, S., & Kompatsiaris, I. (2024). AI methods in materials design, discovery and manufacturing: A review. Computational Materials Science, 235, 112793. https://doi.org/https://doi.org/10.1016/j.commatsci.2024.112793

Simović, A. R., Milenković, D., Šeklić, D., Jovanović, M., Milović, E., Međedović, M., Vraneš, M., & Janković, N. (2025). Exploring Biginelli hybrids in the AI-driven development of ruthenium complexes: Anticancer activity, DNA/HSA binding study, impacts on apoptosis and BCL-2/BCL-XL suppression. Journal of Inorganic Biochemistry, 272, 112988. https://doi.org/https://doi.org/10.1016/j.jinorgbio.2025.112988

Song, W., Wen, Y., Yue, X., Liu, C., Han, Y., & Sun, J. (2025). BioKMS-HAG: A hierarchically guided biomedical and space science knowledge fine-grained mining system. Life Sciences in Space Research. https://doi.org/https://doi.org/10.1016/j.lssr.2025.11.015

Song, Y., Li, J., Chi, D., Xu, Z., Liu, J., Chen, M., & Wang, Z. (2025). AI-driven advances in metal–organic frameworks: from data to design and applications. Chemical Communications, 61(82), 15972–16001. https://doi.org/https://doi.org/10.1039/d5cc04220h

Su, Q., Wang, J., Gou, Q., Hu, R., Jiang, L., Zhang, H., Wang, T., Liu, Y., Shen, C., Kang, Y., Hsieh, C.-Y., & Hou, T. (2025). Robust protein–ligand interaction modeling through integrating physical laws and geometric knowledge for absolute binding free energy calculation††Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4sc07405j. Chemical Science, 16(12), 5043–5057. https://doi.org/https://doi.org/10.1039/d4sc07405j

Tang, S.-L., Sumitra, M. R., Chen, L.-C., Liu, F.-C., Hsu, H.-L., Kuo, Y.-C., Ansar, M., Huang, S.-L., Lee, S.-Y., Wang, H.-J., Lawal, B., Wu, A. T. H., Wen, Y.-T., & Huang, H.-S. (2025). Machine learning–driven discovery of NSC828779 as a multi-mechanistic NLRP3 inflammasome inhibitor for inflammatory diseases. Computers in Biology and Medicine, 197, 111110. https://doi.org/https://doi.org/10.1016/j.compbiomed.2025.111110

Uslu, H., Das, B., Dagdogen, H. A., Santur, Y., Yılmaz, S., Turkoglu, I., & Das, R. (2025). Discovery of new anti-HIV candidate molecules with an AI-based multi-stage system approach using molecular docking and ADME predictions. Chemometrics and Intelligent Laboratory Systems, 267, 105543. https://doi.org/https://doi.org/10.1016/j.chemolab.2025.105543

Wang, B., Liu, Q., Zhao, W., Zhang, T., Zhang, D., Sutcharitchan, C., & Li, S. (2025). Revolutionizing drug discovery from natural products: The roles of artificial intelligence and multi-omics in accelerating innovation. Acta Pharmaceutica Sinica B. https://doi.org/https://doi.org/10.1016/j.apsb.2025.12.030

Wang, M., Qu, B., Yang, L., Wang, L., Jiang, K., & Lin, J. (2025). PyaiVS unifies AI workflows to accelerate ligand discovery and yields ABCG2 inhibitors. European Journal of Medicinal Chemistry, 300, 118176. https://doi.org/https://doi.org/10.1016/j.ejmech.2025.118176

Wang, S., Zhao, Y., Li, J., Zhang, L., Yan, F., Wang, C., Shi, L., Zhang, X., & Zhang, M. (2025). Computational discovery of RSV Pre-F inhibitors via reinforcement learning-driven ab initio design from natural fragment libraries. Computational Biology and Chemistry, 119, 108553. https://doi.org/https://doi.org/10.1016/j.compbiolchem.2025.108553

Wu, J. L., Friday, D. M., Hwang, C., Yi, S., Torres-Flores, T. C., Burke, M. D., Diao, Y., Schroeder, C. M., & Jackson, N. E. (2025). Democratizing machine learning in chemistry with community-engaged test sets. Digital Discovery, 5(1), 304–309. https://doi.org/https://doi.org/10.1039/d5dd00424a

Yu, C.-L., Dai, J.-W., Wang, T.-W., Fu, J.-D., & Liu, P.-L. (2025). AI-enabled construction and prediction of atomic models for thin-film heterostructures via materials genome approach. Surface and Coatings Technology, 498, 131755. https://doi.org/https://doi.org/10.1016/j.surfcoat.2025.131755

Zanoletti, A., Cornelio, A., Galli, E., Scaglia, M., Bonometti, A., Zacco, A., Depero, L. E., Gianoncelli, A., & Bontempi, E. (2025). AI-driven identification of a novel malate structure from recycled lithium-ion batteries. Environmental Research, 267, 120709. https://doi.org/https://doi.org/10.1016/j.envres.2024.120709

Zheng, Z., He, Z., Khattab, O., Rampal, N., Zaharia, M. A., Borgs, C., Chayes, J. T., & Yaghi, O. M. (2024). Image and data mining in reticular chemistry powered by GPT-4V††Electronic supplementary information (ESI) available: Full prompts designed to guide GPT-4V; additional examples showcasing GPT-4V's performance in reading various figure inputs and its corresponding responses; Python code used to automate the data mining and analysis processes; detailed information on the selected papers in this study, including the ground truth and the classification output for each page in a spread-sheet format; extracted nitrogen isotherms in this study. See DOI: https://doi.org/10.1039/d3dd00239j. Digital Discovery, 3(3), 491–501. https://doi.org/https://doi.org/10.1039/d3dd00239j