

Mykola Golovenko

A. V. Bogatsky Physico-Chemical Institute of the National Academy of Sciences of Ukraine, Odesa, Ukraine
ORCID: 0000-0003-1485-128X

Viktor Kuzmin

A. V. Bogatsky Physico-Chemical Institute of the National Academy of Sciences of Ukraine, Odesa, Ukraine
ORCID: 0000-0002-2753-0453

Vitalii Larionov

A. V. Bogatsky Physico-Chemical Institute of the National Academy of Sciences of Ukraine, Odesa, Ukraine
ORCID: 0000-0003-2678-4264

ARTIFICIAL INTELLIGENCE AND CHEMOINFORMATICS: COMPLEMENTARITY OF TWO FIELDS OF MODERN CHEMISTRY

Abstract. The theses examine the complementary roles of artificial intelligence and cheminformatics in modern chemistry. The applications of machine learning in predicting chemical properties, planning organic synthesis, forecasting chemical reactions, drug discovery, and experimental optimization are discussed. The advantages and limitations of AI-based approaches are analyzed, including data quality, model interpretability, prediction uncertainty, and the need for experimental validation. Future perspectives involving autonomous laboratories, materials discovery, quantum chemistry, and the development of digitally skilled chemists are considered.

Keywords: Artificial intelligence; cheminformatics; machine learning; prediction of chemical properties; organic synthesis; chemical reaction prediction; pharmaceutical chemistry.

Микола Головенко

Фізико-хімічний інститут ім. О. В. Богатського НАН України, Одеса, Україна
ORCID: 0000-0003-1485-128X

Віктор Кузьмін

Фізико-хімічний інститут ім. О. В. Богатського НАН України, Одеса, Україна
ORCID: 0000-0002-2753-0453

Віталій Ларіонов

Фізико-хімічний інститут ім. О. В. Богатського НАН України, Одеса, Україна
ORCID: 0000-0003-2678-4264

ШТУЧНИЙ ІНТЕЛЕКТ І ХЕМОІНФОРМАТИКА: ВЗАЄМОДОПОВНЮВАНІСТЬ ДВОХ ГАЛУЗЕЙ СУЧАСНОЇ ХІМІЇ

Анотація. У тезах розглянуто взаємодоповнювальну роль штучного інтелекту та хемоінформатики в сучасній хімії. Проаналізовано застосування машинного навчання для прогнозування хімічних властивостей, планування органічного синтезу, прогнозування хімічних реакцій, пошуку лікарських засобів та оптимізації експериментів. Розглянуто переваги й обмеження підходів на основі ШІ, зокрема якості даних, інтерпретованість моделей, невизначеність прогнозів і необхідність експериментальної перевірки. Окреслено перспективи, пов'язані з автономними лабораторіями, пошуком нових матеріалів, квантовою хімією та підготовкою хіміків із розвиненими цифровими компетентностями.

Ключові слова: штучний інтелект; хемоінформатика; машинне навчання; прогнозування хімічних властивостей; органічний синтез; прогнозування хімічних реакцій; фармацевтична хімія

Introduction

Contemporary chemistry is undergoing a period of profound transformation driven by the active development of digital technologies, computational methods, and artificial intelligence (AI). A prominent place in this process is occupied by chemoinformatics, which utilizes mathematical, statistical, and information technology methods for the storage, processing, and analysis of data regarding chemical compounds and reactions. AI opens new horizons for chemoinformatics due to its capacity to identify patterns within large-scale chemical datasets, predict substance properties, evaluate reaction probability, and facilitate the discovery of prospective drug molecules. The complementarity of artificial intelligence and chemoinformatics lies not in the replacement of a chemist by a computer, but rather in the formulation of a novel paradigm for scientific endeavor. Within this approach, the human scientist defines the research problem, the computational model assists in information processing, and the experiment serves to validate and refine the prediction. It is precisely this synergy that establishes the framework for a more rapid, rational, and efficient advancement of modern chemistry.

The Synergy of Chemical Knowledge and Artificial Intelligence Algorithms

One of the most critical avenues for applying AI in chemoinformatics is the prediction of chemical substance properties. When analyzing a molecule, a chemist considers its structure, functional groups, electronic characteristics, spatial configuration, and potential interactions with other substances. This knowledge enables the formulation of hypotheses regarding the compound's solubility, melting point, stability, toxicity, or biological activity. Machine learning algorithms significantly expand these capabilities. They can be trained on datasets of previously investigated substances to establish statistical relationships between molecular structure and specific properties. To achieve this, molecular descriptors, structural fingerprints, molecular graph representations, and other methods of encoding chemical information are employed [1]. However, predictions do not emerge in a vacuum. The quality of the outcome depends on how accurately the chemist has described the molecule, what data were utilized for training, and whether the predictive conditions correspond to actual experimental settings. Consequently, chemical intuition and theoretical knowledge remain an indispensable complement to mathematical algorithms (Table).

**Table 1. Examples of Interaction
Between Chemical Knowledge and AI**

Chemical Task	AI and Chemoinformatics Tool	Practical Outcome
Solubility prediction	Machine learning based on molecular descriptors	Preliminary evaluation of compound solubility
Toxicity assessment	QSAR models and classification algorithms	Identification of potentially hazardous substances

Continuation of Table 1

Chemical Task	AI and Chemoinformatics Tool	Practical Outcome
Prediction of material properties	Neural networks, structure and composition analysis	Screening for materials with target characteristics
Chemical reaction prediction	Models based on reaction data	Assessment of probable products and reaction conditions
Drug molecule discovery	Virtual screening and generative models	Selection of prospective candidates for further research

Consequently, the application of AI in chemistry facilitates a transition from the analysis of individual compounds to the systematic investigation of vast chemical spaces.

Planning of organic synthesis and prediction of chemical reactions.

Organic synthesis often requires the development of a reaction sequence, selection of reagents, catalysts, solvents, temperature regimes, and product purification methods. Chemoinformatics and AI assist in automating individual stages of this process. Retrosynthetic analysis involves the conceptual disassembly of a target molecule into simpler structural fragments from which it can be synthesized. Algorithms can suggest potential chemical bond cleavages, analyze reactions from databases, and identify prospective synthetic routes.

Reaction prediction models, specifically transformer and graph neural networks, are trained on examples of known chemical transformations [2]. They can evaluate probable reaction products, predict the reactivity of substrates, and guide the selection of appropriate conditions. However, a model may fail to account for intermediate instability, the complexity of purification, reagent availability, or the hazards associated with certain conditions. Consequently, the final decision must be made by the chemist, relying on theoretical knowledge and experimental experience.

Applications in pharmaceutical chemistry and the discovery of new medicinal agents

The discovery and development of a new medicinal agent is a prolonged and complex process encompassing the identification of biologically active compounds, as well as the evaluation of their efficacy, safety, pharmacokinetic properties, and potential for subsequent manufacturing. AI can assist at various stages of this pathway. One of the most prevalent approaches is virtual screening—the computational evaluation of a vast number of molecules to identify compounds capable of interacting with a specific biological target. This method significantly reduces the number of substances requiring high-priority experimental testing [3].

Another prominent avenue is the prediction of molecular properties associated with therapeutic effects. Computational models can evaluate potential biological activity, permeability through biological membranes, metabolic stability, and specific toxicity risks. At the same time, pharmaceutical predictions face substantial limitations. Biological systems are extraordinarily complex, and the activity of a molecule within a computer model does not guarantee its efficacy within a living organism. A drug candidate must undergo multi-stage experimental and clinical validation. Consequently, AI serves as a tool to accelerate the discovery process rather than a replacement for pharmacological and medical research.

Optimization of experimental studies.

Traditional chemistry relies heavily on conducting numerous laboratory trials. This approach is often resource-intensive, and a significant portion of experiments may ultimately prove non-viable. AI facilitates a transition from random or sequential testing of conditions to a more targeted planning strategy. Active learning and Bayesian optimization methods enable the model to determine which subsequent experiment can yield the most valuable information. Following each trial, the model updates its predictions and proposes new conditions.

The integration of AI with automated laboratory systems is particularly promising. Robotic equipment can perform reaction series, measure outcomes, and transmit data back to the computational model. This 'prediction–experiment–analysis–new prediction' cycle establishes the foundation for self-driving research processes [4]. It enables the investigator to gain a deeper insight into process regularities, as experiments are selected in accordance with a specific informational objective.

The chemist, computer models, and AI as a new model of scientific activity.

The application of AI alters not only individual methods of chemical research but also the very organization of scientific labor. In the traditional model, the chemist formulates a hypothesis, conducts an experiment, and analyzes the outcome. In the digital model, a computational system is integrated into this process, capable of processing large-scale datasets, identifying patterns, and generating new hypotheses. However, a computational model is not an autonomous bearer of chemical understanding; it operates solely with the data and mathematical representations embedded during the training process. The novel paradigm of scientific endeavor is grounded in the synergy of three core components: fundamental chemical knowledge, computational methods, and experimental validation. Within this framework, AI serves as an intellectual assistant, capable of rapidly comparing alternatives and suggesting new avenues of research. Concurrently, the responsibility for problem definition, outcome evaluation, and scientific conclusions remains strictly with the human researcher.

Advantages and challenges of integrating artificial intelligence into chemoinformatics

One of the primary advantages of integrating AI into chemistry is the enhancement of research speed and efficiency. Computational models are capable of analyzing thousands and millions of molecular structures in a significantly shorter timeframe than is possible through an exclusively manual approach. This is particularly crucial in medicinal chemistry, materials science, and the discovery of novel catalysts. Another benefit is the potential to minimize non-viable experiments. If a model accurately evaluates compound properties or reaction conditions, the researcher can concentrate attention on the most well-founded alternatives. However, the efficacy of such an approach directly depends on the quality of the chemical data. Data quality, reliability, and representativeness constitute one of the core challenges in chemoinformatics. Datasets may contain errors, incomplete descriptions of

substances, varied measurement conditions, or biases toward specific types of molecules. For instance, a model trained predominantly on thoroughly investigated organic compounds may inaccurately predict the properties of rare or fundamentally novel materials.

The interpretability of models remains a critical issue. Certain algorithms, particularly deep neural networks, are capable of generating precise predictions, yet the underlying reasons for achieving a specific result are not always transparent. This is of fundamental importance for chemistry, as it is vital for a scientist not only to predict a substance's property but also to understand its molecular rationale.

The necessity for experimental validation represents a fundamental limitation of any predictions. While a model may prove useful for screening candidates, its outcomes cannot be automatically equated with proven facts. The experiment remains the definitive method for testing hypotheses, refining models, and discovering new regularities. Furthermore, the digitalization of chemistry introduces ethical, educational, and professional challenges. Concurrently, questions arise regarding the responsibility for errors in automated systems, data protection, academic integrity, and the accessibility of digital tools.

Prospects for the development of the interaction between artificial intelligence and chemoinformatics

One of the most significant prospects is the creation of autonomous laboratories, wherein robotic systems will be capable of independently executing a portion of the research cycle. Such laboratories will integrate experiment planning, substance synthesis, property measurement, and machine-driven analysis of the outcomes [5]. The development of novel materials, catalysts, and medicinal agents represents another critical avenue. Generative models can assist researchers in screening for molecular structures with pre-defined target characteristics. In materials science, this unlocks opportunities to discover compounds possessing required electrical, mechanical, optical, or catalytic properties.

The integration of chemoinformatics, quantum chemistry, and AI will be of particular importance. Quantum chemical calculations provide insight into electronic structure, energetic characteristics, and reaction mechanisms. Machine learning is capable of utilizing the results of these calculations to build faster, approximate models, thereby opening up prospects for investigating more complex chemical systems. As a result, a novel archetype of the chemist is emerging—one who possesses fundamental chemical knowledge and understands the principles of digital technologies. Such a specialist must be able to critically evaluate data, operate computational models, comprehend the limitations of algorithms, and concurrently retain the capacity for creative chemical thinking.

Consequently, the developmental prospects of autonomous laboratories, generative molecular design, quantum chemical modeling, and intelligent data analysis signify the formation of a new stage in chemical science. The future of chemistry lies not in the opposition of human and machine, but rather in their productive

collaboration, wherein chemical thinking defines the research direction, artificial intelligence expands the horizon of possibilities, and the experiment ensures the scientific validity of the outcomes.

REFERENCES / СПИСОК ДЖЕРЕЛ

1. Tetko, I. V., & Tanchuk, V. Y. (2002). Associative neural network as a powerful tool for QSAR analysis and database mining. *Journal of Chemical Information and Computer Sciences*, 42(5), 1136–1145. <https://doi.org/10.1021/ci010132r>.
2. Coley, C. W., Rogers, L., Jamison, T. F., & Jensen, K. F. (2017). A graph-convolutional neural network model for the prediction of chemical reactivity. *Chemical Science*, 8, 533–542. <https://doi.org/10.1039/C6SC02721J>.
3. Vamathevan, J., Clark, D., Czodrowski, P., et al. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18, 463–477. <https://doi.org/10.1038/s41573-019-0024-5>.
4. Häse, F., Roch, L. M., & Aspuru-Guzik, A. (2019). Next-generation experimentation with self-driving laboratories. *Trends in Chemistry*, 1(3), 282–291. <https://doi.org/10.1016/j.trechm.2019.02.007>.
5. Stach, E., DeCost, B., Kusne, A. G., et al. (2021). Autonomous experimentation systems for materials development: A community perspective. *Matter*, 4(9), 2702–2726. <https://doi.org/10.1016/j.matt.2021.06.007>