

Qsaria: Agentic AI for End-to-End QSAR Modeling

Abraham Brisset, Alexey Orlov, Dragos Horvath, Alexandre Varnek

Laboratoire de Chémoinformatique, UMR7140 CNRS/UniStra, University of Strasbourg, 4 rue Blaise Pascal, 67081, Strasbourg, France

Recent progress in agentic AI is opening new possibilities for automating scientific machine-learning workflows. General purpose systems such as, for example, AIDE^[1] and the AI Scientist^[2], have shown that large language model (LLM)-based agents can assist with machine learning experimentation, including dataset preparation, code execution, model training, and iterative performance improvement. In chemistry, domain-specific systems such as MolAgent^[3] and ChemLint^[4] demonstrate the growing relevance of agentic workflows for molecular property modeling.

Here, we present Qsaria, a multi-agent system for automated, end-to-end quantitative structure-activity relationship (QSAR) modeling. The system is designed to translate a natural-language user query into a complete QSAR workflow, spanning data preprocessing, feature generation, model training, ensemble modeling, outlier analysis, performance evaluation, and report generation. Qsaria supports multiple families of modeling algorithms, including message-passing neural networks, tabular foundation models, and classical machine learning methods. It also enables the use of diverse molecular representations, including molecular descriptors, fingerprints, and embeddings derived from foundation models. For model evaluation, Qsaria combines classical cross-validation strategies with random, scaffold-based, and cluster-based data splits, enabling a thorough assessment of model generalization across chemical space. Finally, ensemble modeling is applied to improve predictive robustness and integrate complementary modeling approaches.

Experiments on benchmark QSAR datasets, including Therapeutics Data Commons, MoleculeNet, and OpenADMET datasets, demonstrate performance comparable to established machine-learning baselines and state-of-the-art agentic QSAR approaches. While Qsaria is available as a standalone agentic system, it is also integrated into the broader ChemSpace Copilot platform^[5], where it operates as a QSAR-focused team alongside other cheminformatics agents, supporting the broader goal of making cheminformatics research increasingly autonomous.

Bibliography:

[1] [2502.13138] AIDE: AI-Driven Exploration in the Space of Code. <https://arxiv.org/abs/2502.13138>.

[2] Lu, C. et al. Towards end-to-end automation of AI research. *Nature* 651, 914–919 (2026).

[3] Gómez-Tamayo, J. C. et al. MolAgent: Biomolecular Property Estimation in the Agentic Era. *J. Chem. Inf. Model.* 65, 10808–10818 (2025).

[4] Van Tilborg, D. & Grisoni, F. ChemLint: Conversational Cheminformatics with Large Language Models. Preprint at <https://doi.org/10.26434/chemrxiv.15000386/v1> (2026).

[5] Orlov, A. A., Volkov, M., Milova, E. S., Horvath, D. & Varnek, A. ChemSpace Copilot: Agentic AI for Interactive Visualization and Exploration of Chemical Space. *ChemRxiv* 2026.